

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 874 230**

51 Int. Cl.:

C07K 19/00 (2006.01)

C07K 14/415 (2006.01)

C07K 14/435 (2006.01)

C12N 15/09 (2006.01)

C12N 15/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **30.05.2017 PCT/JP2017/020076**

87 Fecha y número de publicación internacional: **07.12.2017 WO17209122**

96 Fecha de presentación y número de la solicitud europea: **30.05.2017 E 17806675 (9)**

97 Fecha y número de publicación de la concesión europea: **07.04.2021 EP 3466978**

54 Título: **Proteína de fusión para mejorar la expresión de proteínas a partir de ARNm diana**

30 Prioridad:

03.06.2016 US 201662345252 P

17.06.2016 JP 2016120524

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

04.11.2021

73 Titular/es:

KYUSHU UNIVERSITY, NATIONAL UNIVERSITY CORPORATION (100.0%)

**744, Motooka, Nishi-ku, Fukuoka-shi
Fukuoka 819-0395, JP**

72 Inventor/es:

**NAKAMURA, TAKAHIRO y
YAGI, YUSUKE**

74 Agente/Representante:

ARIAS SANZ, Juan

ES 2 874 230 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

Proteína de fusión para mejorar la expresión de proteínas a partir de ARNm diana

5 **[Campo técnico]**

La presente invención se refiere a proteínas de fusión para mejorar los niveles de expresión de proteínas a partir de los ARNm diana.

10 **[Antecedentes de la técnica]**

En los últimos años se han establecido y utilizado técnicas de unión de factores proteicos de unión a ácidos nucleicos reveladas por una variedad de análisis a secuencias de interés. El uso de esta unión específica de secuencia permite la eliminación de una secuencia de ADN diana o la regulación (activación o inactivación) de la expresión de un gen que codifica una proteína presente en dirección 3' de la secuencia de ADN diana en cierta medida.

Mientras que la nucleasa de dedos de zinc (ZFN, por sus siglas en inglés *zinc finger nuclease*), la nucleasa efectora TAL (TALEN, *TAL effector nuclease*), la Crispr-cas9 y similares, se conocen como técnicas que utilizan factores proteicos que actúan sobre el ADN, el desarrollo de técnicas que utilizan factores proteicos que actúan específicamente sobre el ARN es todavía limitado.

Los presentes inventores han propuesto un método para diseñar una proteína que se puede unir específicamente a una secuencia de ARN diana usando las propiedades de las proteínas PPR (proteína que tiene uno o más motivos de repetición pentatricopeptídica (PPR, *pentatricopeptide repeat*)), que son proteínas que se encuentran principalmente en plantas (Publicación de patente 1).

[Listado de citas]

[Publicación de patentes]

[Publicación de patente 1]

(1) WO2013/058404

(2) EP 2 784 157

(3) US 2016/075744

[Sumario de la invención]

[Problema técnico]

En la divulgación de acuerdo con la Publicación de patente 1, se identificaron los aminoácidos que funcionan cuando un motivo PPR demuestra propiedades de unión al ARN, y se reveló la relación entre la estructura del motivo PPR y la base diana, permitiendo así la construcción de proteínas que tienen uno o más motivos PPR y que se pueden unir a ARN que tienen cualquier secuencia y longitud. Sin embargo, nunca se ha descubierto ningún método que regule realmente los ARN diana utilizando las técnicas de acuerdo con la Publicación de patente 1.

La Publicación de patente 2 se refiere a proteínas de unión a ADN que utilizan un motivo PPR. El documento enseña un complejo que comprende una región funcional unida a una proteína que contiene el motivo PPR. Sin embargo, no hay datos experimentales sobre un complejo que comprenda una región de motivo PPR ligada a una región funcional. Además, la Publicación de patente 2 solo sugiere la "RNAsa" (ribonucleasa) como una región funcional proteica, pero no enseña ni sugiere en absoluto un dominio que guía al ribosoma hacia el ARNm, un dominio que inicia o promueve la traducción del ARNm, un dominio asociado con la exportación nuclear del ARNm, un dominio que se une a una membrana del retículo endoplasmático, un dominio que contiene una secuencia señal de retención del retículo endoplasmático (señal de retención del RE) o un dominio que contiene una secuencia señal del retículo endoplasmático.

La Publicación de patente 3 se refiere a proteínas de unión a ADN que utilizan el motivo PPR, pero no enseña ni sugiere la mejora de la expresión de proteínas a partir de un ARN diana mediante el uso de la proteína PPR

[Solución al problema]

Como resultado de una extensa investigación sobre un método para mejorar el nivel de expresión de una proteína a partir de un ARNm diana utilizando una proteína PPR, los presentes inventores han descubierto que una proteína de fusión de un dominio funcional predeterminado y una proteína PPR mejora el nivel de expresión de la proteína a partir del ARNm diana y han completado la presente invención.

Específicamente, una realización de la presente invención se refiere a una proteína de fusión para mejorar el nivel de expresión de una proteína a partir de un ARNm diana, comprendiendo la proteína de fusión:

- 5 (A) uno o más dominios funcionales que mejoran el nivel de expresión de una proteína a partir de un ARNm; y
 (B) un resto polipeptídico que se puede unir a un ARNm diana de una manera selectiva de bases de ARN o específica de secuencia de bases de ARN, en donde el resto polipeptídico (B) es un resto polipeptídico que comprende uno o más motivos PPR, comprendiendo cada motivo PPR un polipéptido que consiste en de 30 a 38 aminoácidos de longitud y está representado por la Fórmula 1:

10 [Fórmula 1]

(Hélice A)-X-(Hélice B)-L

(Fórmula 1)

donde

- 15 La hélice A es un resto que consiste en 12 aminoácidos de longitud y puede formar una estructura de α -hélice, y está representada por la Fórmula 2:

[Fórmula 2]

A₁-A₂-A₃-A₄-A₅-A₆-A₇-A₈-A₉-A₁₁-A₁₁-A₁₂

(Fórmula 2)

20

donde A₁ a A₁₂ representa cada uno independientemente un aminoácido;
 X no está presente, o es un resto que consiste en de 1 a 9 aminoácidos de longitud;
 La hélice B es un resto que consiste en de 11 a 13 aminoácidos de longitud y puede formar una estructura;
 L es un resto que consiste en de 2 a 7 aminoácidos de longitud y está representado por la Fórmula 3:

25

[Fórmula 3]

L_{vii}-L_{vi}-L_v-L_{iv}-L_{iii}-L_{ii}-L_i

(Fórmula 3)

30

donde los aminoácidos se numeran desde el C-terminal como "i" (-1), "ii" (-2), ... y L_{iii} a L_{vii} pueden no estar presentes, y

una combinación de tres aminoácidos A₁, A₄, y L_{ii} o una combinación de dos aminoácidos A₄ y L_{ii} se corresponde con una base o secuencia de bases del ARNm diana,

35

en donde uno o más dominios funcionales (A) se seleccionan del grupo que consiste en un dominio que guía al ribosoma hacia el ARNm, un dominio que inicia o promueve la traducción del ARNm, un dominio asociado con la exportación nuclear del ARNm, un dominio que se une a una membrana del retículo endoplasmático, un dominio que contiene una secuencia señal de retención del retículo endoplasmático (señal de retención del RE) y un dominio que contiene una secuencia señal del retículo endoplasmático.

40

En una realización de acuerdo con la presente invención, el resto polipeptídico (B) comprende de 2 a 30 motivos PPR, y la pluralidad de motivos PPR se dispone para unirse específicamente a la secuencia de bases del ARNm diana.

Además, en una realización de acuerdo con la presente invención, el resto polipeptídico (B) comprende de 5 a 25 motivos PPR.

45

Además, en una realización de acuerdo con la presente invención, uno o más dominios funcionales (A) se unen cada uno a un lado de N-terminal y/o a un lado C-terminal del resto polipeptídico (B).

50

Además, en una realización de acuerdo con la presente invención, el dominio que guía al ribosoma hacia el ARNm es un dominio que contiene todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en DENR (proteína regulada por densidad), MCT-1 (secuencia 1 amplificada de linfocitos T malignos), TPT1 (proteína tumoral controlada traduccionalmente) y Lerepo4 (dominio CCCH de dedo de zinc),

el dominio que inicia o promueve la traducción del ARNm es un dominio que contiene todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en eIF4E y eIF4G,

55

el dominio asociado con la exportación nuclear del ARNm es un dominio que contiene todo o la parte funcional de SLBP (proteína de unión al bucle del tallo),

el dominio que se une a una membrana del retículo endoplasmático es un dominio que contiene todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en SEC61B, TRAP-alfa (proteína alfa asociada a Translocón), SR-alfa, Dial (citocromo b5 reductasa 3) y p180,

60

la secuencia señal de retención del retículo endoplasmático (señal de retención del RE) es una secuencia señal que contiene una secuencia KDEL (KEEL), o

la secuencia señal del retículo endoplasmático es una secuencia señal que contiene MGWSCII LFLVATATGAHS (SEQ ID NO: 22).

65

Además, en una realización de acuerdo con la presente invención, la combinación de los tres aminoácidos A₁, A₄, y L_{ii} en cada uno de los motivos PPR es:

(valina, treonina, asparagina), (fenilalanina, serina, asparagina), (fenilalanina, treonina, asparagina), (isoleucina, asparagina, ácido aspártico) o (treonina, treonina, asparagina) en orden de (A_1 , A_4 , L_{ii}) si una base diana para el motivo PPR es A (adenina);

- 5 (ácido glutámico, glicina, ácido aspártico), (valina, treonina, ácido aspártico), (lisina, treonina, ácido aspártico) o (leucina, treonina, ácido aspártico) en el orden de (A_1 , A_4 , L_{ii}) si la base diana para el motivo PPR es G (guanina); (valina, asparagina, ácido aspártico), (isoleucina, asparagina, asparagina), (isoleucina, asparagina, ácido aspártico), (isoleucina, metionina, ácido aspártico), (fenilalanina, prolina, ácido aspártico) o (tirosina, prolina, ácido aspártico) en orden de (A_1 , A_4 , L_{ii}) si la base diana para el motivo PPR es U (uracilo); o
- 10 (valina, asparagina, asparagina), (isoleucina, asparagina, asparagina), (valina, asparagina, serina) o (isoleucina, metionina, ácido aspártico) en el orden de (A_1 , A_4 , L_{ii}) si la base diana para el motivo PPR es C (citosina).

Además, en una realización de acuerdo con la presente invención, la combinación de los dos aminoácidos A_4 y L_{ii} en cada uno de los motivos PPR es:

- 15 (treonina, asparagina), (serina, asparagina) o (glicina, asparagina) en orden de (A_4 , L_{ii}) si una base diana para el motivo PPR es A (adenina); (treonina, ácido aspártico) o (glicina, ácido aspártico) en el orden de (A_4 , L_{ii}) si la base diana para el motivo PPR es G (guanina); (asparagina, ácido aspártico), (prolina, ácido aspártico), (metionina, ácido aspártico) o (valina, treonina) en orden de (A_4 , L_{ii}) si la base diana para el motivo PPR es U (uracilo); o
- 20 (asparagina, asparagina), (asparagina, serina) o (leucina, ácido aspártico) en el orden de (A_4 , L_{ii}) si la base diana para el motivo PPR es C (citosina).

- 25 Otra realización de acuerdo con la presente invención, se refiere a un ácido nucleico que codifica la proteína de fusión de acuerdo con la presente invención.

Otra realización más de acuerdo con la presente invención, se refiere a un vector (preferentemente un vector de expresión) que comprende el ácido nucleico de acuerdo con la presente invención.

- 30 Además, otra realización más de acuerdo con la presente invención, se refiere a un método *in vitro* para mejorar el nivel de expresión de una proteína a partir de un ARNm diana dentro de una célula, comprendiendo el método:

- una etapa de proporcionar la proteína de fusión de acuerdo con la presente invención o el vector de acuerdo con la presente invención; y
- 35 una etapa de introducir la proteína de fusión o el vector en la célula.

Además, en una realización de acuerdo con la presente invención, la célula es una célula eucariota.

- 40 Además, en una realización de acuerdo con la presente invención, la célula es una célula animal.

Además, en una realización de acuerdo con la presente invención, la célula animal es una célula humana.

- 45 Las invenciones que tienen cualquier combinación de una o más características de la presente invención descritas anteriormente también se incluyen en el alcance de la presente invención.

[Breve descripción de las figuras]

- [Figura 1] La Figura 1 ilustra una vista esquemática de un plásmido efector y de un plásmido indicador usados en los Ejemplos, y una vista esquemática de un esquema experimental. La Figura 1A ilustra una vista esquemática del plásmido efector y del plásmido indicador usados en los Ejemplos. Una proteína de fusión de motivos PPR y eIF4G se expresa a partir del plásmido efector. En los ejemplos, se utilizó una proteína CRR4, cuya secuencia diana está bien investigada. A partir del plásmido indicador, la luciferasa de renilla (RLuc) y la luciferasa de luciémaga (FLuc), se transcriben en forma de un ARNm dicistrónico. Se insertó una secuencia de unión a PPR (en el presente documento, secuencia de unión a CRR4) en un sitio en el extremo 5' de FLuc. La Figura 1B ilustra una vista esquemática de un esquema experimental de Ejemplos. Independientemente de la presencia/ausencia de la secuencia de unión a PPR, RLuc se traduce a un nivel similar. Por este motivo, el valor de actividad de RLuc se puede tratar como un control en la transfección en este sistema indicador. La traducción de FLuc se inicia solo cuando PPR-eIF4G se une a la secuencia de unión a PPR y los factores de traducción pueden ser atraídos por los efectos de eIF4G. En cambio, la traducción de FLuc permanece a un nivel bajo si la secuencia de unión a PPR no está presente y, por lo tanto, PPR-eIF4G no se puede unir a la secuencia de unión a PPR.
- 50
- 55
- 60

[Figura 2] La Figura 2 ilustra un procedimiento experimental de un ensayo indicador utilizando células HEK293T.

- [Figura 3] La Figura 3 muestra los resultados experimentales del Ejemplo 1. La activación de la traducción específica de secuencia depende de CRR4-eIF4G y de la secuencia de unión a PPR. Este experimento se realizó utilizando un plásmido efector, en el que se insertó CRR4-Flag (sin factor de activación de la traducción, en blanco) o CRR4-eIF4G (con factor de activación de la traducción, en gris), y un vector indicador con o sin una secuencia de unión a PPR insertada. A partir de los resultados, se verificó que la actividad de traducción específica aumentaba
- 65

2,75 veces en presencia tanto de PPR-eIF4G como de la secuencia de unión a PPR. El valor representa el promedio y la desviación estándar (N = 3).

[Figura 4] La Figura 4 ilustra un esquema del experimento en el Ejemplo 2.

[Figura 5] La Figura 5 ilustra los resultados experimentales del Ejemplo 2 y las funciones de los dominios.

5 [Figura 6] La Figura 6 ilustra los resultados experimentales del Ejemplo 2 y las funciones de los dominios.

[Descripción de la realización]

[Motivos PPR y proteínas PPR]

10 A menos que se especifique de otro modo, la expresión "motivo PPR", tal como se usa en el presente documento, indica un polipéptido que está compuesto por 30 a 38 aminoácidos y tiene una secuencia de aminoácidos que tiene un valor E igual o menor que un valor predeterminado (deseablemente E-03), obteniéndose el valor E en PF01535 en Pfam y PS51375 en Prosite durante el análisis de la secuencia de aminoácidos con un programa de búsqueda de dominios de proteínas en la Web. El número de posición de un aminoácido que forma el motivo PPR definido en la presente invención, es sustancialmente como se define como PF01535 mientras que corresponde al número obtenido al restar 2 de la ubicación del aminoácido en PS51375 (por ejemplo, la posición 1 en la presente invención se corresponde con la posición 3 en PS51375). Téngase en cuenta que el término "ii" (-2)-ésimo aminoácido, se refiere al segundo aminoácido del extremo de la cola (lado C-terminal) de los aminoácidos que forman un motivo PPR o el aminoácido cercano al N-terminal por dos aminoácidos del primer aminoácido del siguiente motivo PPR (es decir, -2 aminoácidos). Si el siguiente motivo PPR no está claramente identificado, el aminoácido directo por dos aminoácidos del primer aminoácido de la siguiente estructura de hélice se define como "ii". Véase <http://pfam.sanger.ac.uk/> para Pfam y <http://www.expasy.org/prosite/> para Prosite.

25 Aunque la secuencia de aminoácidos conservada del motivo PPR tiene propiedades de conservación bajas a nivel de aminoácidos, dos α -hélices están bien conservadas en la estructura secundaria. Aunque un motivo típico de PPR se compone de 35 aminoácidos, su longitud es variable de 30 a 38 aminoácidos.

30 De acuerdo con la invención, la expresión motivo PPR, usada en la presente invención, está compuesta por un polipéptido que tiene de 30 a 38 aminoácidos de longitud y está representado por la Fórmula 1:

[Fórmula 4]

$$\text{(Hélice A)-X-(Hélice B)-L} \qquad \text{(Fórmula 1)}$$

35 donde
La hélice A es un resto que consiste en 12 aminoácidos de longitud y puede formar una estructura, y está representada por la Fórmula 2: [Fórmula 5]

$$\text{A}_1\text{-A}_2\text{-A}_3\text{-A}_4\text{-A}_5\text{-A}_6\text{-A}_7\text{-A}_8\text{-A}_9\text{-A}_{10}\text{-A}_{11}\text{-A}_{12} \qquad \text{(Fórmula 2)}$$

40 donde A_1 a A_{12} representa cada uno independientemente un aminoácido;
X no está presente, o es un resto que consiste en de 1 a 9 aminoácidos de longitud;
La hélice B es un resto que consiste en de 11 a 13 aminoácidos de longitud y puede formar la estructura; y L es un resto que consiste en de 2 a 7 aminoácidos de longitud y está representado por la Fórmula 3:

[Fórmula 6]

$$\text{L}_{vii}\text{-L}_{vi}\text{-L}_v\text{-L}_{iv}\text{-L}_{iii}\text{-L}_{ii}\text{-L}_i \qquad \text{(Fórmula 3)}$$

50 donde los aminoácidos se numeran desde el lado de C-terminal como "i" (-1), "ii" (-2), ... y L_{iii} a L_{vii} pueden no estar presentes, y una combinación de tres aminoácidos A_1 , A_4 , y L_{ii} o una combinación de dos aminoácidos A_4 y L_{ii} se corresponde con una base o secuencia de bases del ARNm diana,
55 en donde uno o más dominios funcionales (A) se seleccionan del grupo que consiste en un dominio que guía al ribosoma hacia el ARNm, un dominio que inicia o promueve la traducción del ARNm, un dominio asociado con la exportación nuclear del ARNm, un dominio que se une a una membrana del retículo endoplasmático, un dominio que contiene una secuencia señal de retención del retículo endoplasmático (señal de retención del RE) y un dominio que contiene una secuencia señal del retículo endoplasmático.

60 A menos que se especifique de otro modo, la expresión "proteína PPR" utilizada en la presente invención, indica una proteína PPR que comprende uno o más motivos PPR descritos anteriormente, preferentemente dos o más motivos PPR descritos anteriormente. A menos que se especifique de otro modo, el término "proteína" utilizado en el presente documento, generalmente indica sustancias que consisten en polipéptidos (cadenas de varios aminoácidos unidas a través de un enlace peptídico), incluyendo también las que consisten en polipéptidos de peso molecular relativamente bajo. El término "aminoácido" utilizado en la presente invención, puede indicar una molécula de aminoácido habitual,

o de otro modo, en algunos casos, puede indicar un residuo de aminoácido que forma una cadena peptídica. Los expertos en la técnica entienden claramente a partir de los contextos qué caso indica el término.

5 A menos que se especifique de otro modo, el término "selectivo" utilizado en la presente invención, sobre las propiedades de unión del motivo PPR a las bases de ARN, indica que la actividad de unión de un motivo PPR a una de las bases de ARN es mayor que la actividad de unión del mismo a otras bases. Los expertos en la técnica pueden planificar el experimento para esta selectividad y verificarlo, y también pueden determinarlo mediante cálculo.

10 A menos que se especifique de otro modo, la expresión "base de ARN" utilizada en la presente invención, indica una base de un ribonucleótido que forma un ARN, específicamente adenina (A), guanina (G), citosina (C) o uracilo (U). Téngase en cuenta que, aunque la proteína PPR puede tener selectividad a la base en el ARN, no se une a un monómero de ácido nucleico.

15 La proteína PPR está presente en muchas plantas y en *Arabidopsis thaliana* se pueden encontrar 500 proteínas y aproximadamente 5000 motivos. Los motivos PPR y las proteínas PPR que tienen una variedad de secuencias de aminoácidos también están presentes en muchas plantas terrestres como *Oryza*, *Populus* y *Selaginella tamariscina*. En la presente invención, se pueden usar, por ejemplo, motivos PPR y proteínas PPR presentes en la naturaleza, o motivos PPR y proteínas PPR diseñados basándose en el método desvelado en el documento WO2013/058404. Específicamente, los motivos PPR y las proteínas PPR deseados se pueden diseñar basándose en la siguiente información desvelada en el documento WO2013/058404.

(I) Información sobre la posición del aminoácido esencial para la unión selectiva

25 La combinación (A1, A4, Lii) de tres, es decir, 1º, 4º y "ii" (-1)-ésimo aminoácidos de un motivo PPR o la combinación (A4, Lii) de dos, es decir, 4º y "ii" (-1)-ésimo aminoácidos, son esenciales para la unión selectiva a la base de ARN, y la base de ARN diana para la unión se puede determinar mediante estas combinaciones.

La presente invención puede utilizar los hallazgos sobre la combinación de tres aminoácidos A1, A4 y Lii, y/o la combinación de dos aminoácidos A4 y Lii desvelados en el documento WO2013/058404.

30

(II) Información sobre la correspondencia de la combinación de tres aminoácidos A1, A4 y Lii a bases de ARN

(3-1) Si la combinación de tres aminoácidos A1, A4 y Lii es valina, asparagina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a C es la segunda más fuerte, seguida de unión a A o G.

(3-2) Si la combinación de tres aminoácidos A1, A4 y Lii es valina, treonina y asparagina en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a A es la más fuerte y la unión a G es la segunda más fuerte, seguida de unión a C sin unión a U.

40 (3-3) Si la combinación de tres aminoácidos A1, A4 y Lii es valina, asparagina y asparagina en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a C es la más fuerte y la unión a A o U es la segunda más fuerte, sin unión a G.

(3-4) Si la combinación de tres aminoácidos A1, A4 y Lii es ácido glutámico, glicina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a G es fuerte, sin unión a A, U o C.

45 (3-5) Si la combinación de tres aminoácidos A1, A4 y Lii es isoleucina, asparagina y asparagina en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a C es la más fuerte y la unión a U es la segunda más fuerte, seguida de unión a A, sin unión a G.

(3-6) Si la combinación de tres aminoácidos A1, A4 y Lii es valina, treonina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a G es la más fuerte y la unión a U es la segunda más fuerte, sin unión a A o C.

(3-7) Si la combinación de tres aminoácidos A1, A4 y Lii es lisina, treonina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a G es la más fuerte y la unión a A es la segunda más fuerte, sin unión a U o C.

50 (3-8) Si la combinación de tres aminoácidos A1, A4 y Lii es fenilalanina, serina y asparagina en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a A es la más fuerte y la unión a C es la segunda más fuerte, seguida de unión a G y U.

(3-9) Si la combinación de tres aminoácidos A1, A4 y Lii es valina, asparagina y serina en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a C es la más fuerte y la unión a U es la segunda más fuerte, sin unión a A o G.

60 (3-10) Si la combinación de tres aminoácidos A1, A4 y Lii es fenilalanina, treonina y asparagina en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a A es fuerte, sin unión a G, U o C.

(3-11) Si la combinación de tres aminoácidos A1, A4 y Lii es isoleucina, asparagina, ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a A es la segunda más fuerte, sin unión a G o C.

65 (3-12) Si la combinación de tres aminoácidos A1, A4, y Lii es treonina, treonina y asparagina en este orden, el

motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a A es fuerte, sin unión a G, U o C.

(3-13) Si la combinación de tres aminoácidos A1, A4 y Lii es isoleucina, metionina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a C es la segunda más fuerte, sin unión a A o G.

(3-14) Si la combinación de tres aminoácidos A1, A4 y Lii es fenilalanina, prolina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a C es la segunda más fuerte, sin unión a A o G.

(3-15) Si la combinación de tres aminoácidos A1, A4 y Lii es tirosina, prolina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es fuerte, sin unión a A, G o C.

(3-16) Si la combinación de tres aminoácidos A1, A4 y Lii es leucina, treonina y ácido aspártico en este orden, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a G es fuerte, sin unión a A, U o C.

(II) Información sobre la correspondencia de la combinación de dos aminoácidos A4 y Lii a las bases de ARN

(2-1) Si A4 y Lii en este orden son asparagina y ácido aspártico, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a C es la segunda más fuerte, seguida de unión a A y G.

(2-2) Si A4 y Lii en este orden son asparagina y asparagina, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a C es la más fuerte, la unión a U es la segunda más fuerte, seguida de unión a A y G.

(2-3) Si A4 y Lii en este orden son treonina y asparagina, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva con unión fuerte a A y unión débil a G, U y C.

(2-4) Si A4 y Lii en este orden son treonina y ácido aspártico, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva con unión fuerte a G y unión débil a A, U y C.

(2-5) Si A4 y Lii en este orden son serina y asparagina, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a A es la más fuerte y la unión a G, U y C es la segunda más fuerte.

(2-6) Si A4 y Lii en este orden son glicina y ácido aspártico, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a G es la más fuerte y la unión a U es la segunda más fuerte, seguida de unión a A, sin unión a C.

(2-7) Si A4 y Lii en este orden son asparagina y serina, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a C es la más fuerte y la unión a U es la segunda más fuerte, seguida de unión a A y G.

(2-8) Si A4 y Lii en este orden son prolina y ácido aspártico, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a G, C, y C es la segunda más fuerte, sin unión a A.

(2-9) Si A4 y Lii en este orden son glicina y asparagina, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a A es la más fuerte y la unión a G es la segunda más fuerte, sin unión a C o U.

(2-10) Si A4 y Lii en este orden son metionina y ácido aspártico, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva con unión fuerte a U y unión débil a A, G y C.

(2-11) Si A4 y Lii en este orden son leucina y ácido aspártico, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a C es la más fuerte y la unión a U es la segunda más fuerte, sin unión a A o G.

(2-12) Si A4 y Lii en este orden son valina y treonina, el motivo PPR tiene una capacidad de unión a bases de ARN selectiva de la siguiente manera: la unión a U es la más fuerte y la unión a A es la segunda más fuerte, sin unión a G o C.

[Uso de motivos PPR y de proteínas PPR]

Identificación y diseño:

Un motivo PPR puede reconocer una base específica de un ARN. De acuerdo con la presente invención, los motivos PPR selectivos a A, U, G o C, se pueden seleccionar o diseñar disponiendo los aminoácidos apropiados en posiciones específicas de un motivo PPR. Por otro lado, una proteína que contiene una serie apropiada de tales motivos PPR puede reconocer su secuencia específica correspondiente. Además, de acuerdo con los hallazgos descritos anteriormente, se puede diseñar un motivo PPR que se puede unir selectivamente a una base de ARN deseada y una proteína que tiene una pluralidad de motivos PPR que se pueden unir de forma específica en secuencia a un ARN deseado. En diseño, la información de secuencia de un motivo PPR de origen natural puede referirse con respecto a restos distintos de los aminoácidos dispuestos en las posiciones importantes del motivo PPR. Como alternativa, se puede diseñar un motivo PPR usando un motivo PPR de origen natural como un todo y reemplazando sólo los aminoácidos en las posiciones importantes por otros aminoácidos. El número de repeticiones del motivo PPR se puede determinar apropiadamente de acuerdo con la secuencia diana; por ejemplo, el número de repeticiones puede ser de 2 o más, o de 2 a 30.

El motivo PPR o la proteína PPR así diseñados se pueden preparar mediante un método bien conocido por los expertos en la técnica. Por ejemplo, se puede determinar una secuencia de ácido nucleico que codifica una secuencia de aminoácidos del motivo PPR diseñado o de la proteína PPR a partir de la secuencia de aminoácidos, y se puede clonar para preparar un transformante (tal como un vector de expresión) que produzca un motivo PPR o una proteína PPR deseados.

Preparación y uso de la proteína de fusión:

La presente invención se refiere a una proteína de fusión del motivo PPR o de la proteína PPR descritos anteriormente, tal como se define en las reivindicaciones (es decir, un polipéptido que puede unirse a la base del ARN selectivamente o a la secuencia de la base del ARN específicamente al ARNm diana) y uno o más dominios funcionales que mejoran el nivel de expresión de una proteína a partir de un ARNm.

El "dominio funcional que mejora el nivel de expresión de una proteína a partir de un ARNm" que se puede usar en la presente invención, puede ser todo o parte funcional de un dominio funcional de una proteína conocida que promueve directa o indirectamente la traducción del ARNm, por ejemplo. De acuerdo con la invención, el dominio funcional que se puede usar en la presente invención, se selecciona de un dominio que guía al ribosoma hacia el ARNm, un dominio que inicia o promueve la traducción del ARNm, un dominio asociado con la exportación nuclear del ARNm, un dominio que se une a una membrana del retículo endoplasmático, un dominio que contiene una secuencia señal de retención en retículo endoplasmático (señal de retención en RE) y un dominio que contiene una secuencia señal del retículo endoplasmático, por ejemplo.

De manera más específica, el dominio que guía al ribosoma hacia el ARNm puede ser un dominio que contenga todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en DENR (proteína regulada por densidad), MCT-1 (secuencia 1 amplificada de linfocitos T malignos), TPT1 (proteína tumoral controlada traduccionalmente) y Lerepo4 (dominio CCCH de dedo de zinc). El dominio que inicia o promueve la traducción del ARNm puede ser un dominio que contenga todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en eIF4E y eIF4G. El dominio asociado con la exportación nuclear del ARNm puede ser un dominio que contenga todo o la parte funcional de SLBP (proteína de unión al bucle del tallo). El dominio que se une a una membrana del retículo endoplasmático puede ser un dominio que contenga todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en SEC61B, TRAP-alfa (proteína alfa asociada a Translocón), SR-alfa, Dial (citocromo b5 reductasa 3) y p180. La secuencia señal de retención del retículo endoplasmático (señal de retención del RE) puede ser una secuencia señal que contiene una secuencia KDEL (KEEL). La secuencia señal del retículo endoplasmático puede ser una secuencia señal que contiene MGWSCIIILFLVATATGAHS (SEQ ID NO: 22).

En la proteína de fusión de acuerdo con la presente invención, el dominio funcional se puede fusionar con el lado N-terminal de la proteína PPR, se puede fusionar con el lado C-terminal de la proteína PPR, o se puede fusionar tanto con el lado N-terminal como con el lado C-terminal del mismo. Además, la proteína de fusión de acuerdo con la presente invención puede incluir varios dominios funcionales (por ejemplo, de 2 a 5 dominios funcionales). Por otro lado, en la proteína de fusión de acuerdo con la presente invención, el dominio funcional y la proteína PPR, se pueden fusionar indirectamente a través de un enlazador, por ejemplo.

La presente invención también se refiere a un ácido nucleico que codifica la proteína de fusión descrita anteriormente y a un vector (tal como un vector de expresión) que comprende el ácido nucleico. El vector de expresión en el presente documento se refiere a, por ejemplo, un vector que comprende un ADN que tiene una secuencia promotora, un ADN que codifica una proteína deseada y un ADN que tiene una secuencia de terminación, en este orden desde la dirección 5'. Es posible que el vector de expresión no tenga estos ADN en este orden siempre que demuestre las funciones deseadas. En la presente invención se puede utilizar una variedad de vectores de expresión que normalmente pueden utilizar los expertos en la técnica.

Dado que la proteína de fusión de acuerdo con la presente invención usa el mecanismo de traducción de ARN de eucariotas, ésta puede funcionar en células de eucariotas (como animales, plantas, microorganismos (por ejemplo, levaduras) y protistas). La proteína de fusión de acuerdo con la presente invención, puede funcionar dentro de células animales (*in vitro* o *in vivo*) en particular. Ejemplos de células animales en las que se puede introducir la proteína de fusión de acuerdo con la presente invención o un vector que expresa la proteína de fusión de acuerdo con la presente invención, pueden incluir células que provienen de ser humano, mono, cerdo, vaca, caballo, perro, gato, ratón y rata. Ejemplos de células cultivadas en las que se puede introducir la proteína de fusión de acuerdo con la presente invención o un vector que expresa la proteína de fusión de acuerdo con la presente invención pueden incluir, pero no debe limitarse a, células de ovario de hámster chino (CHO), células COS-1, células COS-7, células VERO (ATCC CCL-81), células BHK, células MDCK que provienen de riñón de perro, células de hámster AV-12-664, células HeLa, células WI38, células 293, células 293T y células PER.C6.

Los términos utilizados en el presente documento, excluyendo los particularmente definidos, se utilizan para ilustrar las realizaciones específicas y no se pretende que sean limitativos de la invención.

El término "comprender" usado en el presente documento, a menos que los contextos requieran claramente diferentes entendimientos, pretende expresar que una entrada descrita (tal como un miembro, una etapa, un componente o un número) está presente, y no pretende excluir la presencia de otras entradas (como un miembro, una etapa, un componente o un número).

5 A menos que se definan de otro modo, todos los términos utilizados en el presente documento (incluidos los términos técnicos y científicos) tienen el mismo significado que los entendidos ampliamente por los expertos en la técnica a la que pertenece la presente invención. A menos que se defina claramente lo contrario, los términos utilizados en el presente documento deben interpretarse con el significado que corresponde a los mismos y a su campo técnico relacionado, y no deben interpretarse como significados idealizados o excesivamente formales.

10 En lo sucesivo, en el presente documento, la presente invención se describirá más en detalle con referencia a los Ejemplos. Sin embargo, la presente invención se puede implementar con una variedad de aspectos y no debe interpretarse como limitativa de los Ejemplos descritos a continuación.

15 [Ejemplos]

Ejemplo 1: Mejora en el nivel de expresión de proteínas del ARNm diana mediante la proteína de fusión del motivo PPR y eIF4G

20 Materiales

(Equipo)

- 25 - Instalación básica para experimentos de biología molecular (para la construcción de plásmidos, por ejemplo)
- Microscopio invertido (DM IL S40, Leica Microsystems, Wetzlar, Alemania)
- Incubadora de CO₂ (KM-CC17RH2, Panasonic Healthcare, Tokio, Japón)
- Mesa de laboratorio aséptica (MHE-S1300A2, Panasonic Healthcare, Tokio, Japón)
- Aspirador (SP-30, Air Liquide Medical Systems, Bovezzo BS, Italia)
- 30 - Centrífuga (rotor oscilante) (LC-200, Tomy Seiko, Tokio, Japón)
- Congelador de temperatura ultrabaja (-80 °C) (MDF-C8V, Panasonic Healthcare, Tokio, Japón)
- Lector de placas (EnSight Kaleido, PerkinElmer, Waltham, MA, EE.UU.)

(Cultivo celular)

- 35 - Línea celular HEK293T (véase la nota 1)
- Medio de cultivo de Eagle modificado por Dulbecco (DMEM, rico en glucosa) (véase la nota 2)
- Solución de penicilina-estreptomina a 100x
- Suero fetal bovino (FBS) (véase la nota 3)
- 40 - Solución EDTA-NaCl: EDTA 10 mM y NaCl al 0,85 % (p/v), pH ajustado de 7,2 a 7,4, esterilizada en autoclave, almacenada a temperatura ambiente
- Placa de Petri de cultivo celular de 100 x 20 mm (Greiner bio one, Frickenhausen, Alemania)
- Pipeta esterilizada desechable de 10 ml
- Tubos de centrifuga de plástico de 15 ml y 50 ml
- 45 - Criotubo de 1,8 ml (Nunc; Thermo Fisher Scientific, Waltham, MA, EE.UU.)
- Recipiente de congelación (Nalgene; Thermo Fisher Scientific, Waltham, MA, EE.UU.)
- Bambanker (Lymphotec, Tokio, Japón)

(Transfección)

- 50 - Plásmido efector: pcDNA3.1 (Thermo Fisher Scientific, Waltham, MA, EE.UU.) se utilizó como vector básico. Se inserta un gen de fusión de PPR y eIF4G en un casete de expresión (100 ng/μl) (véase la nota 4).
- Plásmido indicador: pcDNA3.1 (Thermo Fisher Scientific, Waltham, MA, EE.UU.) se utilizó como vector básico. Los genes de luciferasa se insertan en un casete de expresión y una secuencia de unión a PPR se inserta en su 5'-UTR (100 ng/μl).
- 55 - Placa de 96 pocillos recubierta con poli-L-lisina (AGC Techno glass, Shizuoka, Japón)
- Solución salina tamponada con fosfato a 1x, PBS(-): KH₂PO₄ 1,47 mM, Na₂HPO₄ 8,1 mM, NaCl 137 mM y KCl 2,7 mM. pH ajustado a 7,4, esterilizada en autoclave, almacenada a temperatura ambiente
- Hemocitómetro (para contar el número de células) (placa de recuento de células tipo Neubauer mejorada, Watson, Hyogo, Japón)
- 60 - Reactivo de transfección (HilyMax, Dojindo Molecular Technologies, Kumamoto, Japón)

(Ensayo de luciferasa)

- 65 - Sistema de ensayo de luciferasa Dual-Glo (Promega, Madison, WI, EE.UU.)
- Placa de luminómetro de 96 pocillos (PerkinElmer, Waltham, MA, EE.UU.).

Método experimental

(Construcción del vector)

5 El ensayo indicador requiere un plásmido efector y un plásmido indicador. Ambos plásmidos se construyen basándose en pcDNA3.1. El plásmido efector incluye un gen de fusión que codifica una proteína PPR y un dominio parcial de eIF4G humano (SEQ ID NO: 1) (Figura 1A). El resto de proteína PPR utilizado fue CRR4 (SEQ ID NO: 2). El plásmido indicador incluye dos marcos de lectura abiertos (ORF, siglas del inglés *open reading frames*), específicamente, luciferasa de renilla (RLuc) y luciferasa de luciérnaga (FLuc), que se transcriben dicistricamente (Figura 1A). El gen de RLuc se encuentra en el lado del extremo 5' del gen de FLuc y se utilizó como control de la expresión génica. La región de unión a PPR se inserta en la 5'-UTR del ORF de FLuc y consiste en tres repeticiones de una secuencia de reconocimiento de CRR4 (5'-UAUCUUGUCUUUA-3') (SEQ ID NO: 3) interrumpida con secuencias de cuatro bases (ATCG y GATC). Para expresar tanto el gen efector fusionado como el gen indicador, se utilizaron un promotor de citomegalovirus (CMV) y una señal de poliadenilación que proviene del gen de la hormona del crecimiento bovino. Para un experimento de control, se construyó un plásmido efector sin eIF4G, fusionando una etiqueta de epítipo FLAG a la PPR. También se construyó un plásmido indicador de control sin una región de unión a PPR.

20 En la Figura 2 se muestra el esquema de los procedimientos desde el cultivo celular hasta el ensayo indicador en los Ejemplos.

(Cultivo de células de solución madre congelada)

25 Esta etapa se realiza asépticamente. Todas las herramientas se someten previamente a tratamiento antiséptico con etanol al 70 %.

1. Se coloca un medio de cultivo DMEM de 9 ml en un tubo de centrifuga de 15 ml (esterilizado).
2. Se incuba 1 ml de células HEK293T congeladas en un criotubo en un baño de agua a 37 °C para fundir rápidamente las células.
- 30 3. Las células se colocan en el tubo de centrifuga de 15 ml que contiene 9 ml de DMEM.
4. El tubo de centrifuga se centrifuga a temperatura ambiente y 1100 x g durante dos minutos y se retira el sobrenadante.
5. Las células se resuspenden en 10 ml de DMEM (se añade FBS de manera que la concentración final sea del 10 %).
- 35 6. Las células suspendidas se transfieren a una placa de Petri de 100 mm. La placa de Petri se dejó reposar en una incubadora a 37 °C y con CO₂ al 5 %. Si el cultivo se inició a partir de la solución madre congelada, las células cultivadas se subcultivaron después de 24 horas.

40 Para mantener las células sanas (véase la [nota 5](#)), la densidad celular en la superficie de la placa de Petri se mantiene entre el 10 % y el 80 %. El pase se realiza básicamente cada tres días (dos veces por semana), o se realiza de acuerdo con la tasa de crecimiento de las células. Por otro lado, para mantener un número reducido de pases, las células se cultivan recientemente a partir de la reserva congelada una vez al mes. Para que la transfección de ADN sea eficaz, es importante que el número de pases sea reducido y que las células se mantengan sanas.

45 (Pases para mantener las células)

1. Se proporcionan nuevas placas de Petri de 100 mm según sea necesario. Se colocan preliminarmente 8 ml de DMEM y 1 ml de FBS en cada una de las placas de Petri.
2. El medio de cultivo de una placa de Petri que contiene las células cultivadas, se retira con un aspirador (véase la [nota 6](#)).
- 50 3. Se añaden cuidadosamente 2 ml de solución de EDTA-NaCl a las células adheridas en la superficie de la placa de Petri para que éstas no se desprendan. La placa de Petri se gira para distribuir uniformemente la solución por toda la superficie de la placa de Petri. La solución de EDTA-NaCl se retira con un aspirador. Para que las células se desprendan, la placa de Petri se golpetea cuidadosamente.
- 55 4. Se añaden 10 ml de DMEM a las células en la placa de Petri y éstas se suspenden pipeteando cuidadosamente.
5. A las placas de Petri anteriormente proporcionadas, se añade 1 ml de células en suspensión (10 % de células cultivadas) y cada una de ellas contiene 9 ml de medio de cultivo. Cada una de las placas de Petri se gira para distribuir las células por toda su superficie.

60 (Almacenamiento de células por congelación)

Se prepara una reserva congelada con reactivo Bambanker y las células se cultivan en una fase de crecimiento logarítmico a una densidad celular de hasta el 50 %. El uso de Bambanker proporciona una alta tasa de recuperación y facilita el almacenamiento a largo plazo.

- 65 1. Las células del segundo día desde el pase, se desprenden de acuerdo con el procedimiento del pase. Se añaden

de 5 a 10 ml de DMEM y las células se recuperan en un tubo de centrifuga de 50 ml.

2. El tubo de centrifuga se centrifuga a temperatura ambiente y 1100 x g durante dos minutos y se retira el sobrenadante.

3. Se añade 1 ml de Bambanker por placa de Petri para suspender las células.

5 4. Las células suspendidas se dispensan rápidamente en criotubos y los criotubos se cubren con sus tapas.

5. Los criotubos se colocan en un recipiente de congelación específico y se dejan reposar a -80 °C durante 12 horas (véase la nota 7).

6. Los criotubos se transfieren a una caja de muestra estándar y se almacenan a -80 °C o en nitrógeno líquido.

10 (Introducción transitoria de genes (transfección))

1. Antes de comenzar la transfección, según sea necesario, se proporcionan placas de Petri, conteniendo cada una de ellas, las células del segundo día desde el pase y se comprueba si las células están sanas (normales) o no (véase la nota 8). Se pueden realizar aproximadamente 96 ensayos con una placa de Petri como estimación.

15 2. Las células del segundo día desde el pase, se desprenden de acuerdo con el procedimiento del pase, y las células suspendidas se transfieren a un tubo de centrifuga de 50 ml.

3. El tubo de centrifuga se centrifuga a temperatura ambiente y 1100 x g durante dos minutos y se retira el sobrenadante.

20 4. Los grupos de células se dispersan completamente en 10 ml de DMEM (se añade FBS de manera que la concentración final sea del 10 %).

5. El número de células se cuenta con un hemocitómetro y un microscopio invertido. Las células se suspenden en una cantidad apropiada de DMEM (se añade FBS de manera que la concentración final sea del 10 %) de modo que el número de células sea de 1 a 2 x 10⁵ células/ml.

25 6. Se proporciona una placa de 96 pocillos. En cada pocillo, se colocan 200 µl (de 2 a 4 x 10⁴ células/ml) de células cultivadas suspendidas, y la placa se deja reposar durante la noche en una incubadora a 37 °C con CO₂ al 5 %. Se utiliza un pocillo para un ensayo.

7. Al día siguiente, el medio de cultivo se retira cuidadosamente de cada pocillo y se reemplaza por 100 µl de DMEM nuevo (se añade FBS de manera que la concentración final sea del 10 %).

30 8. Se colocan 400 ng de plásmido efector (4 µl de 100 ng/µl) y 100 ng de plásmido indicador (1 µl de 100 ng/µl) en un solo pocillo de una nueva placa de PCR de 96 pocillos (o en un tubo de 0,2 ml).

9. Para un ensayo, se diluye 1 µl de HilyMAX con 10 µl de DMEM sin suero.

35 10. Se colocan 11 µl de solución diluida en cada uno de los pocillos que contienen los plásmidos. La solución se mezcla bien con los plásmidos mediante pipeteo.

11. La solución se deja reposar a temperatura ambiente durante 15 minutos. La cantidad total de la mezcla se coloca en los pocillos que contienen las células cultivadas. La placa se deja reposar en una incubadora a 37 °C con CO₂ al 5 % durante 24 horas.

(Ensayo de luciferasa)

40 El ensayo de luciferasa dual se realiza utilizando el sistema de ensayo de luciferasa Dual-Glo de acuerdo con las instrucciones de uso del fabricante, excepto por algunas modificaciones.

1. Después de 24 horas desde la transfección, el medio de cultivo de cada pocillo se reemplaza por 40 µl 1 x PBS(-).

45 2. Se colocan 40 µl de reactivo de luciferasa Dual-Glo en cada pocillo y se mezcla bien con el medio de cultivo mediante pipeteo.

3. La mezcla se deja reposar a temperatura ambiente durante 10 minutos y la cantidad total de la misma se transfiere a una placa luminométrica de 96 pocillos.

50 4. La emisión de luz de luciferasa de luciérnaga relacionada con la expresión del gen de FLuc se mide con un lector de placas.

5. Un sustrato Stop & Glo se diluye 100 veces con un tampón Dual-Glo Stop & Glo. Se añaden 40 µl de la solución diluida a cada pocillo.

6. La placa se deja reposar al menos a temperatura ambiente durante 10 minutos, y después se mide la emisión de luz de la luciferasa de renilla relacionada con la expresión del gen de RLuc.

55 (Análisis de los datos)

1. El valor de FLuc/RLuc se calcula para corregir una diferencia en la eficacia de transfección entre los ensayos o errores experimentales.

60 2. Se determina un aumento en la actividad de la expresión del gen indicador en presencia de la región de unión a PPR y en ausencia de la misma, dividiendo un valor experimental obtenido usando el plásmido de acuerdo con la presente invención (plásmido que codifica una proteína de fusión de CRR4 y un dominio de activación de la traducción eIF4G) entre un valor experimental obtenido usando un plásmido de control (plásmido que codifica una proteína de fusión de CRR4 y etiqueta FLAG).

65

Resultados experimentales

Los resultados del ensayo de luciferasa se muestran en la Figura 3. Tal como se muestra en la Figura 3, se verificó específicamente una actividad de traducción 2,75 veces mayor en presencia tanto de PPR-eIF4G como de la secuencia de unión a PPR. Es decir, se demuestra que la proteína de fusión de la proteína PPR y el dominio funcional, que mejora un nivel de expresión de la proteína a partir de un ARNm, mejoran el nivel de expresión de la proteína a partir del ARNm diana.

Notas

(Nota 1) HEK293T es una línea celular de riñón que proviene de feto humano que expresa un antígeno T grande del SV40. La línea celular se cultiva fácilmente y se puede transfectar con alta eficacia mediante una variedad de métodos. Las células HEK293T están disponibles en RIKEN BRC (ja.brc.riken.jp) o en la ATCC (www.atcc.org).

(Nota 2) Se añade una solución de penicilina-estreptomina a 1x al DMEM para evitar la contaminación con microorganismos.

(Nota 3) Antes de usar, el FBS se inactiva a 56 °C durante 30 minutos y se almacena a 4 °C.

(Nota 4) La pureza del plásmido es significativamente importante para la eficacia de la transfección. El plásmido se debe aislar utilizando un kit de calidad transfección.

(Nota 5) Una tasa de crecimiento diaria es un índice que indica que las células están sanas. Para evitar la supresión del crecimiento celular, las células se deben cultivar siempre en un espacio suficiente y en condiciones nutricionales suficientes.

(Nota 6) Las células HEK293T se deben tratar cuidadosamente cuando se reemplaza el medio de cultivo, porque las células se desprenden fácilmente de la placa de Petri de cultivo.

(Nota 7) El recipiente de congelación específico es una caja cuya velocidad de congelación se puede ajustar (aproximadamente de -1 °C por minuto a -80 °C) y permite congelar las células almacenadas en un congelador no programable a -80 °C.

(Nota 8) En la transfección, las células se utilizan a una densidad de cultivo del 50 al 80 %. Sin embargo, una densidad celular apropiada depende del reactivo de transfección. Adicionalmente, la proporción del reactivo de transfección (µl) frente al ADN plasmídico (µg) también debe optimizarse de acuerdo con las instrucciones de uso del fabricante. El procedimiento descrito en el presente documento está optimizado para una condición en la que se usa una placa de 96 pocillos, células HEK293T y HilyMAX como reactivo de transfección.

Ejemplo 2: Mejora en el nivel de expresión de la proteína del ARNm diana por la proteína de fusión de PPR y otro dominio funcional

En el caso de que se produzcan sustancias útiles utilizando células, las cantidades de proteína sintetizadas por genes endógenos y genes exógenos deben controlarse con precisión. La cantidad final de proteína sintetizada está determinada por las posiciones de inserción de los genes, la cantidad de transcripción de ARNm, la regulación postranscripcional (regulación a nivel de ARN), la modificación postraduccional y similares. Por estas razones, los presentes inventores han ideado un método para mejorar la traducción de los ARNm aprovechando el hecho de que una secuencia de proteína PPR se une específicamente a una molécula de ARN diana (Figura 4). En la traducción de los ARNm en eucariotas, un ARNm se somete a la acción de un factor de iniciación de la traducción (factor de iniciación eucariota; eIF, siglas del inglés *eukaryotic initiation factor*). Como resultado, el ribosoma se recluta cerca del punto de inicio de la traducción y después se inicia la traducción del ARNm. En otras palabras, los presentes inventores han considerado que la traducción del ARNm se puede mejorar artificialmente si el ribosoma solo puede reclutarse en el ARNm. Además, la traducción de un ARNm en una proteína generalmente se realiza en el RE. Por este motivo, los presentes inventores han considerado que la traducción del ARNm se puede mejorar localizando intencionalmente el ARNm diana en el RE.

Verificación por experimento

Para verificar la idea anterior, se preparó un sistema de ensayo indicador usando células animales (HEK293T) cultivadas (el experimento se realizó mediante el mismo método que en el Ejemplo 1, excepto que se usaron dominios funcionales diferentes). El sistema se construyó utilizando la proteína CRR4 (una de las proteínas PPR de *Arabidopsis thaliana*), que se sabe que se une a una secuencia de ARN específica (UAUCUUGUCUUUA) (SEQ ID NO: 3). En primer lugar, se preparó un vector de expresión de la proteína de fusión (plásmido efector) de CRR4 y un dominio candidato funcional de la proteína. Los dominios candidatos seleccionados fueron (a) proteínas eIF (eIF4E y eIF4G), (b) proteínas unidas a ribosoma (DENR, MCT-1, TPT1 y Lerepo4), (c) factores de regulación de la traducción (SLBP) de histona que promueven el transporte del ARNm transcrito desde el núcleo al citoplasma, (d) proteínas de anclaje al RE (SEC61B, TRAP-alfa, SR-alfa, Dial y p180), (e) señal de retención del RE (KDEL) y (f) péptido señal del RE. Las proteínas de fusión se clonaron para expresarse en forma de HA-CRR4-XX o XX-CRR4-HA (HA: etiqueta de epítipo

(SEQ ID NO: 4); XX: dominio candidato).

El plásmido indicador incluía un casete de expresión en el que se transcriben la luciferasa de renilla (RLuc) y la luciferasa de luciérnaga (FLuc) en forma de un ARNm dicistrónico bajo el control de un promotor del CMV. Se insertan tres secuencias de unión a PPR (UAUCUUGUCUUUA) (SEQ ID NO: 3) en un sitio en el extremo 5' de Fluc.

El plásmido efector y el plásmido indicador se transfectaron en células HEK293T y se midieron las intensidades de emisión de luz de RLUC y FLUC. La intensidad de la emisión de luz de RLUC se trató como un control de transfección, y el valor de la intensidad de la emisión de luz de FLUC/la intensidad de la emisión de luz de RLUC se trató como una cantidad de actividad de traducción.

Resultados

Los resultados que se muestran en las Figuras 5 y 6 se examinaron utilizando los siguientes índices (A) y (B).

(A) Comparación entre la ausencia y la presencia de la diana

La comparación muestra una cantidad de cambio específico de secuencia en la traducción.

(B) Comparación con la presencia de la diana y la ausencia del efector (vacío) (línea discontinua negra)

La comparación muestra una cantidad de cambio en la traducción causada por la adición del dominio.

1. Se fusionó eIF4E con el lado C-terminal de CRR4.

- (A) 2,7 veces
- (B) 1,6 veces

2. Se fusionó eIF4G con el lado C-terminal de CRR4.

- (A) 4,5 veces
- (B) 3,3 veces

3. Se fusionó DENR con el lado N-terminal de CRR4.

- (A) 1,7 veces
- (B) 1,3 veces

4. Se fusionó DENR con el lado C-terminal de CRR4.

- (A) 2,4 veces
- (B) 1,7 veces

5. Se fusionó MCT-1 con el lado N-terminal de CRR4.

- (A) 1,3 veces
- (B) 1,0 vez

6. Se fusionó MCT-1 con el lado C-terminal de CRR4.

- (A) 2,0 veces
- (B) 1,2 veces

7. Se fusionó TPT-1 con el lado N-terminal de CRR4.

- (A) 1,4 veces
- (B) 1,0 vez

8. Se fusionó TPT-1 con el lado C-terminal de CRR4.

- (A) 2,4 veces
- (B) 1,9 veces

9. Se fusionó Lerepo4 con el lado N-terminal de CRR4.

- (A) 3,0 veces

- (B) 1,8 veces
10. Se fusionó Lerepo4 con el lado C-terminal de CRR4.
- 5 (A) 3,3 veces
(B) 2,6 veces
11. Se fusionó SLBP con el lado C-terminal de CRR4.
- 10 (A) 4,1 veces
(B) 3,3 veces
12. Se fusionó Sec61B con el lado C-terminal de CRR4.
- 15 (A) 1,6 veces
(B) 1,6 veces
13. Se fusionó Sec61BTM con el lado C-terminal de CRR4.
- 20 (A) 2,4 veces
(B) 1,9 veces
14. Se fusionó TRAP-alfa con el lado C-terminal de CRR4.
- 25 (A) 3,5 veces
(B) 4,5 veces
15. Se fusionó TRAPTM con el lado C-terminal de CRR4.
- 30 (A) 2,3 veces
(B) 1,6 veces
16. Se fusionó SR-alfa con el lado N-terminal de CRR4.
- 35 (A) 1,7 veces
(B) 1,5 veces
17. Se fusionó DialTM con el lado N-terminal de CRR4.
- 40 (A) 1,8 veces
(B) 1,2 veces
18. Se fusionó P180TM2R con el lado N-terminal de CRR4.
- 45 (A) 2,1 veces
(B) 1,5 veces
19. Se fusionó P180TMH con el lado N-terminal de CRR4.
- 50 (A) 2,3 veces
(B) 2,5 veces
20. Se fusionó P180TM2 con el lado N-terminal de CRR4.
- 55 (A) 3,0 veces
(B) 2,1 veces
21. Se fusionó KDEL con el lado C-terminal de CRR4.
- 60 (A) 1,8 veces
(B) 1,4 veces
22. Se fusionó KEEL con el lado C-terminal de CRR4.
- 65 (A) 2,3 veces
(B) 2,1 veces

23. El péptido señal (SP, por sus siglas en inglés, *Signal Peptide*) se fusionó con el lado N-terminal de CRR4.

- 5 (A) 1,4 veces
- (B) 2,0 veces

Como se ha mostrado anteriormente, se descubrió un aumento en la traducción en todos los dominios funcionales tanto en los índices (A) como en las dianas (B). En concreto, se demostró claramente que la proteína de fusión de acuerdo con la presente invención puede potenciar la traducción del ARNm diana.

10 Las secuencias de aminoácidos de los dominios funcionales utilizados en los ejemplos se enumeran a continuación:

[Tabla 1-1]

Dominio	Secuencia
eIF4E	MATVEPETTPTPNPPTTEEEKTESNQEVANPEHYIKHPLQNRWALW FFKNDKSKTWQANLRLISKFDTVEDFWALYNHIQLSSNLMPGCDYS LFDKDGIEPMLEDEKKNKRGGRWLITLNKQQRSDLDLRFWLETLLCLI GESFDDYSDDVCGAVVNVRAKGDKIAIWTTECENREAVTHIGRVYK ERLGLPPKIVIGYQSHADTATKSGSTTKNRFVVGRY (SEQ ID NO: 5)
eIF4G	EEKKRYDREFLGQFIFASMQKPEGLPHISDVVLDKANKTPLRPL DPTRLQGINCGPDFTPSFANLGRITLSTRGPPRGGPGGELPRGPQA GLGPRRSQQGPRKEPRKIIATVLMTEDIKLNKAEKAWKPSSKRTAA DKDRGEEDADGSKTQDLFRRVRSILNKLTPQMFQQLMKQVTQLAID TEERLKGVIDLIFEKAISEPNFSVAYANMCRCLMALKVPTTEKPTV TVNFRKLLLNRCQKEFEKDKDDDEVFEKKQKEMDEAATAEERGRK EELEEARDIARRRSLGNIKFIGELFKLKMTEAIMHDCVVKLLKNH DEESLECLCRLTTIGKDLDFEKAKPRMDQYFNQMEKIKEKKTSS RIRFMLQDVLDLRGSNWVPRRGDQGPKTIDQIHKEAEMEHEHREHIK VQQLMAKGS DKRRGGPPGPPISRGLPLVDDGGWNTVPI SKGSRPID TSRLTKITKPGSIDSNNQLFAPGGRLSWGKGS SGGSGAKPSDAASE AARPATSTLNRFSALQQAVPTESDNRRVVQRSSLSRERGEKAGDR GDRLERSERGGDRGLDRARTPATKRSFSKEVEERSRERPSQPEG LRKAASLTEDRDRGRDAVKREAAALPPVSPLKAALSEEELEKSKAI IEEYLHLNDMKEAVQCVQELASPSLLFIFVRHGVESTLERSAIARE HMGQLLHQLLCAGHLSTAQYYQGLYEILELAEDMEIDI PHVWLYLA ELVTPILQEGGVPMGELFREITKPLRPLGKAASLLEILGLLCKSM GPKKVGT LWREAGLSWKEFLPEGQDIGAFVAEQKVEYTLGEESEAP GQRALPSEELNRQLEKLLKEGSSNQRVFDWIEANLSEQQIVSNTLV RALMTAVCYSAIIFETPLRVDVAVLKARAKLLQKYL CDEQKELQAL YALQALVVTLEQPPNLLRMFPDALYDEDVVKEDAFYSWESSKDPAE QQGKGVALKSVTAFFKWLREABEESDH (SEQ ID NO: 1)
DENR	MAADISESSGADCKGDPNRS AKLDADYPLRVLYCGVCSLPTEYCEY MPDVAKCRQWLEKNFPNEFAKLTVENSPKQEAGISEGQGTAGEEEE KKKQKRGGRGQIKQKKKTVPOKVTIAKIPRAKKKYVTRVCGLATFE IDLKEAQRFFAQKFSFGASVTGEDEIIIQGDFTDDIIDVIQEKWPE VDDDSIEDLGEVKK (SEQ ID NO: 6)

(continuación)

Dominio	Secuencia
MCT-1	MFKKFDEKENVSNCIQLKTSVIKGIKNQLIEQFPGIEPWLNQIMPK KDPVKIVRCHEHIEILTVNGELLFFRQREGPFYPTLRLLLHKYPFIL PHQQVDKGAIKFVLSGANIMCPGLTSPGAKLYPAAVDTIVAIMAEG

[Tabla 1-2]

	KQHALCVGVMKMSAEDIEKVNKGIENIHYLNDGLWHMKTYP (SEQ IND NO: 7)
TPT-1	MI IYRDLISHDEMFSDIYKIREIADGLCLEVEGKMVSRTEGNIDDS LIGNASAEGPEGEGTESTVITGVDIVMNHHLQETSFTKEAYKKYI KDYMKSIGKLEEQRPERVKPFMTGAAEQIKHILANFKNYOFFIGE NMNPDGMVALLDYREDGVTPYMIFFKDGLEMEKC (SEQ ID NO: 8)
Leropo4	P PKKQAQAGGSKKAEQKKKEKI IEDKTFGLKNKKGAKQQKFIKAVT HQVKFGQQNPRQVAQSEAEKKLKKDDKKELQELNELFKPVVAAQK ISKGADPKSVVCAFFKQGQCTKGDCKFSDHDLTLERKCEKRSVYID ARDEELEKDTMDNWDEKKLEEVVNKKHGAEKKKPKTQIVCKHFLE AIENNKYGFVWVCPGGGDI CMYRHALPPGFVLKKDKKKEEKEDEIS LEDLIERERSALGPNVTKITLESFLAWKKRKRQEKIDKLEQDMERR KADFKAGKALVISGREVFEPPELVNDDDEEADDTRYTQGTGGDEV DDSVSVNDIDLSLYIPRDVDETGITVASLERFSTYTSKDKENKLSE ASGGRAENGERSDLEEDNEREGTENGAI DAVPVDEKSFHWRGFG (SEQ ID NO: 9)
SLBP	ACRPRSPRRHQSRCGDASPPSPARWSLGRKRRADGRRWRPEDAEE AEHRGAERRPESFTTPEGPKPRSRCSDWASAVEEDEMTRVNKEMA RYKRKLLINDFGRERKSSSGSSDSKESMSTVPADFETDESVLNRRQ KQINYGKNTIAYDRYI KEVPRHLRQPGIHPKTPNKFKKYSRRSWDQ QIKLWKVALHFWDPPAEEGCDLQEIHPVDLESAESSSEPQTSSQDD FDVYSGTPTKVRHMDSQVEDEFDLEACLEPLRDFSAMS (SEQ ID NO: 10)
Sec61B	PGPTPSGTNVGSSGRSPSKAVARAAGSTVRQRKNASCGRSAGRT TSAGTGGMWRFYTEDSPGLKVGPPVPLVMSLLFIASVFMLHIWVKY TRS (SEQ ID NO: 11)
sec61B-TM	VGPVPLVMSLLFIASVFMLHIW (SEQ IND NO: 12)
TRAP-alfa	RLLPRLLLLLLVFPATVLFRRGGPRGLLAVAQDLTEDEETVEDSI I EDEDDEAEVEEPTDLVEDKKEEDVSGEPEASPSADTTILFVKGE DFPANNIVKFLVGFTNKGTEDFIVESLDASFRYPQDYQFYIQNFTA LPLNTVPPQRQATFEYSFIPAEPMGGRPFGLVINLNYKDLGNVVF QDAVFNQTVTVIEREDGLDGETIFMYMFLAGLGLLVIVGLHQLES RKRKRPIQKVEMGTSSQNDVMSWI PQETLNQINKASPRRLPRKRA QKR SVGSDE (SEQ ID NO: 13)
TRAP-TM	TIFMYMFLAGLGLLVIVGLHQLL (SEQ IND NO: 14)

(continuación)

SR-alfa	LDFFTIFSKGGLVWLWCFQGVSDSCTGPVNALIRSVLLQVGFQKILT
---------	--

[Tabla 1-3]

	LT YVDKLI DDVHRLFRDKYRTEIQQQSALSLLNGTFDFQNDFLRL REAEESSKIRAPTTMKKFEDSEKAKKPVRSMIETRGEKPKKAKNS KKKGAKKEGSDGPLATSKPVPAEKSGLPVGPENGVELSKEELIRK REEFIQKHGRGMEKSNKSTKSDAPKEKGGKAPRVWELGGCANKEVL DYSTPTTNGTPEAALSEDINLIRGTGSGGQLQDLDCSSSDDEGAAQ NSTKPSATKGTGGMFGMLKGLVGSKSLSRDMESVLDKMRDHLIA KNVAADIAVQLCESVANKLEGKVMGTFSTVTSTVKQALQESLVQIL QPQRRVDMLRDIMDAQRRQRPYVVTFCGVNGVGKSTNLAKISFWLL ENGFVSLIAACDTFRAGAVEQLRTHTRRLSALHPPEKHGGRTMVQL FEKGYGKDAAGIAMEAIAFARNQGFVVLVDTAGRMQDNAPLMTAL AKLITVNTPDLVLFVGEALVGNEAVDQLVKFNALADHSMAQT IDGIVLTKFDTIDDKVGAASMTYITSKPIVFGVTGQTYCDLRS AKAVVAALMKA (SEQ ID NO: 15)
DiaTM	STLGHMVLFPVWFLYSLL (SEQ ID NO: 16)
P180TMR2	DIYDTQTLGVVVFGGFMVSAIGIFLVSTFPMKETS YBEALANQRK EMAKTHHQKVEKKKKEKTVEKKGKTKKKBEKPNGKIPDHDPA VNTVLLREPVRAPAVAVAPTVPVQPIIVAPVATVPAMPQEK LASSPKDKKKKEKKVAKVEPAVSSVNSIQVLTSKAAI LETAPKEGRNTDVAQSPKQEPAPAKKSGSKKKGPPDADG PLYLPYKTLVSTVGSVMVFNEGQAQRLIEILSEKAGI IQDTWHKATQKGDV (SEQ ID NO: 17)
P180TMH	LGVVVFGGMVSAIGIFLVSTF (SEQ ID NO: 18)
P180TM2	DIYDTQTLGVVVFGGFMVSAIGIFLVSTF (SEQ ID NO: 19)
KDEL	KDEL (SEQ ID NO: 20)
KEEL	KEEL (SEQ ID NO: 21)
Péptido señal del RE	MGWSCIIFLVATATGAHS (SEQ ID NO: 22)

LISTADO DE SECUENCIAS

- 5 <110> KYUSHU UNIVERSITY, NATIONAL UNIVERSITY CORPORATION
- <120> Proteína de fusión para aumentar la expresión de proteínas a partir de un ARNm diana
- 10 <130> EDFP1601F
- <150> US62/345252
- <151> 03/06/2016
- 15 <150> JP2016-120524
- <151> 17/06/2016
- <160> 22

<170> PatentIn versión 3.5

<210> 1

<211> 993

5 <212> PRT

<213> *Homo sapiens*

<400> 1

Glu Glu Lys Lys Arg Tyr Asp Arg Glu Phe Leu Leu Gly Phe Gln Phe
 1 5 10 15

Ile Phe Ala Ser Met Gln Lys Pro Glu Gly Leu Pro His Ile Ser Asp
 20 25 30

Val Val Leu Asp Lys Ala Asn Lys Thr Pro Leu Arg Pro Leu Asp Pro
 35 40 45

Thr Arg Leu Gln Gly Ile Asn Cys Gly Pro Asp Phe Thr Pro Ser Phe
 50 55 60

Ala Asn Leu Gly Arg Thr Thr Leu Ser Thr Arg Gly Pro Pro Arg Gly
 65 70 75 80

Gly Pro Gly Gly Glu Leu Pro Arg Gly Pro Gln Ala Gly Leu Gly Pro
 85 90 95

Arg Arg Ser Gln Gln Gly Pro Arg Lys Glu Pro Arg Lys Ile Ile Ala
 100 105 110

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

Trp Lys Pro Ser Ser Lys Arg Thr Ala Ala Asp Lys Asp Arg Gly Glu
 130 135 140

Glu Asp Ala Asp Gly Ser Lys Thr Gln Asp Leu Phe Arg Arg Val Arg

10

ES 2 874 230 T3

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

ES 2 874 230 T3

His Lys Glu Ala Glu Met Glu Glu His Arg Glu His Ile Lys Val Gln
 405 410 415
 Gln Leu Met Ala Lys Gly Ser Asp Lys Arg Arg Gly Gly Pro Pro Gly
 420 425 430
 Pro Pro Ile Ser Arg Gly Leu Pro Leu Val Asp Asp Gly Gly Trp Asn
 435 440 445
 Thr Val Pro Ile Ser Lys Gly Ser Arg Pro Ile Asp Thr Ser Arg Leu
 450 455 460
 Thr Lys Ile Thr Lys Pro Gly Ser Ile Asp Ser Asn Asn Gln Leu Phe
 465 470 475 480
 Ala Pro Gly Gly Arg Leu Ser Trp Gly Lys Gly Ser Ser Gly Gly Ser
 485 490 495
 Gly Ala Lys Pro Ser Asp Ala Ala Ser Glu Ala Ala Arg Pro Ala Thr
 500 505 510
 Ser Thr Leu Asn Arg Phe Ser Ala Leu Gln Gln Ala Val Pro Thr Glu
 515 520 525
 Ser Thr Asp Asn Arg Arg Val Val Gln Arg Ser Ser Leu Ser Arg Glu
 530 535 540
 Arg Gly Glu Lys Ala Gly Asp Arg Gly Asp Arg Leu Glu Arg Ser Glu
 545 550 555 560
 Arg Gly Gly Asp Arg Gly Asp Arg Leu Asp Arg Ala Arg Thr Pro Ala
 565 570 575
 Thr Lys Arg Ser Phe Ser Lys Glu Val Glu Glu Arg Ser Arg Glu Arg
 580 585 590
 Pro Ser Gln Pro Glu Gly Leu Arg Lys Ala Ala Ser Leu Thr Glu Asp
 595 600 605
 Arg Asp Arg Gly Arg Asp Ala Val Lys Arg Glu Ala Ala Leu Pro Pro
 610 615 620
 Val Ser Pro Leu Lys Ala Ala Leu Ser Glu Glu Glu Leu Glu Lys Lys
 625 630 635 640
 Ser Lys Ala Ile Ile Glu Glu Tyr Leu His Leu Asn Asp Met Lys Glu
 645 650 655

ES 2 874 230 T3

Ala Val Gln Cys Val Gln Glu Leu Ala Ser Pro Ser Leu Leu Phe Ile
660 665 670

Phe Val Arg His Gly Val Glu Ser Thr Leu Glu Arg Ser Ala Ile Ala
675 680 685

Arg Glu His Met Gly Gln Leu Leu His Gln Leu Leu Cys Ala Gly His
690 695 700

Leu Ser Thr Ala Gln Tyr Tyr Gln Gly Leu Tyr Glu Ile Leu Glu Leu
705 710 715 720

Ala Glu Asp Met Glu Ile Asp Ile Pro His Val Trp Leu Tyr Leu Ala
725 730 735

Glu Leu Val Thr Pro Ile Leu Gln Glu Gly Gly Val Pro Met Gly Glu
740 745 750

Leu Phe Arg Glu Ile Thr Lys Pro Leu Arg Pro Leu Gly Lys Ala Ala
755 760 765

Ser Leu Leu Leu Glu Ile Leu Gly Leu Leu Cys Lys Ser Met Gly Pro
770 775 780

Lys Lys Val Gly Thr Leu Trp Arg Glu Ala Gly Leu Ser Trp Lys Glu
785 790 795 800

Phe Leu Pro Glu Gly Gln Asp Ile Gly Ala Phe Val Ala Glu Gln Lys
805 810 815

Val Glu Tyr Thr Leu Gly Glu Glu Ser Glu Ala Pro Gly Gln Arg Ala
820 825 830

Leu Pro Ser Glu Glu Leu Asn Arg Gln Leu Glu Lys Leu Leu Lys Glu
835 840 845

Gly Ser Ser Asn Gln Arg Val Phe Asp Trp Ile Glu Ala Asn Leu Ser
850 855 860

Glu Gln Gln Ile Val Ser Asn Thr Leu Val Arg Ala Leu Met Thr Ala
865 870 875 880

Val Cys Tyr Ser Ala Ile Ile Phe Glu Thr Pro Leu Arg Val Asp Val
885 890 895

Ala Val Leu Lys Ala Arg Ala Lys Leu Leu Gln Lys Tyr Leu Cys Asp
900 905 910

ES 2 874 230 T3

Glu Gln Lys Glu Leu Gln Ala Leu Tyr Ala Leu Gln Ala Leu Val Val
 915 920 925

Thr Leu Glu Gln Pro Pro Asn Leu Leu Arg Met Phe Phe Asp Ala Leu
 930 935 940

Tyr Asp Glu Asp Val Val Lys Glu Asp Ala Phe Tyr Ser Trp Glu Ser
 945 950 955 960

Ser Lys Asp Pro Ala Glu Gln Gln Gly Lys Gly Val Ala Leu Lys Ser
 965 970 975

Val Thr Ala Phe Phe Lys Trp Leu Arg Glu Ala Glu Glu Glu Ser Asp
 980 985 990

His

<210> 2

<211> 561

<212> PRT

5 <213> *Arabidopsis thaliana*

<400> 2

Ala Phe Ala Ser Ser Arg Arg Pro Tyr Leu Ala Asp Phe Ala Arg Cys
 1 5 10 15

Val Phe His Glu Tyr His Val Cys Ser Phe Ser Phe Gly Glu Val Glu
 20 25 30

Asp Pro Phe Leu Trp Asn Ala Val Ile Lys Ser His Ser His Gly Lys
 35 40 45

Asp Pro Arg Gln Ala Leu Leu Leu Leu Cys Leu Met Leu Glu Asn Gly
 50 55 60

Val Ser Val Asp Lys Phe Ser Leu Ser Leu Val Leu Lys Ala Cys Ser
 65 70 75 80

Arg Leu Gly Phe Val Lys Gly Gly Met Gln Ile His Gly Phe Leu Lys
 85 90 95

Lys Thr Gly Leu Trp Ser Asp Leu Phe Leu Gln Asn Cys Leu Ile Gly
 100 105 110

Leu Tyr Leu Lys Cys Gly Cys Leu Gly Leu Ser Arg Gln Met Phe Asp
 115 120 125

ES 2 874 230 T3

Arg Met Pro Lys Arg Asp Ser Val Ser Tyr Asn Ser Met Ile Asp Gly
 130 135 140

Tyr Val Lys Cys Gly Leu Ile Val Ser Ala Arg Glu Leu Phe Asp Leu
 145 150 155 160

Met Pro Met Glu Met Lys Asn Leu Ile Ser Trp Asn Ser Met Ile Ser
 165 170 175

Gly Tyr Ala Gln Thr Ser Asp Gly Val Asp Ile Ala Ser Lys Leu Phe
 180 185 190

Ala Asp Met Pro Glu Lys Asp Leu Ile Ser Trp Asn Ser Met Ile Asp
 195 200 205

Gly Tyr Val Lys His Gly Arg Ile Glu Asp Ala Lys Gly Leu Phe Asp
 210 215 220

Val Met Pro Arg Arg Asp Val Val Thr Trp Ala Thr Met Ile Asp Gly
 225 230 235 240

Tyr Ala Lys Leu Gly Phe Val His His Ala Lys Thr Leu Phe Asp Gln
 245 250 255

Met Pro His Arg Asp Val Val Ala Tyr Asn Ser Met Met Ala Gly Tyr
 260 265 270

Val Gln Asn Lys Tyr His Met Glu Ala Leu Glu Ile Phe Ser Asp Met
 275 280 285

Glu Lys Glu Ser His Leu Leu Pro Asp Asp Thr Thr Leu Val Ile Val
 290 295 300

Leu Pro Ala Ile Ala Gln Leu Gly Arg Leu Ser Lys Ala Ile Asp Met
 305 310 315 320

His Leu Tyr Ile Val Glu Lys Gln Phe Tyr Leu Gly Gly Lys Leu Gly
 325 330 335

Val Ala Leu Ile Asp Met Tyr Ser Lys Cys Gly Ser Ile Gln His Ala
 340 345 350

Met Leu Val Phe Glu Gly Ile Glu Asn Lys Ser Ile Asp His Trp Asn
 355 360 365

Ala Met Ile Gly Gly Leu Ala Ile His Gly Leu Gly Glu Ser Ala Phe

ES 2 874 230 T3

370	375	380																	
Asp	Met	Leu	Leu	Gln	Ile	Glu	Arg	Leu	Ser	Leu	Lys	Pro	Asp	Asp	Ile				
385					390					395					400				
Thr	Phe	Val	Gly	Val	Leu	Asn	Ala	Cys	Ser	His	Ser	Gly	Leu	Val	Lys				
				405					410					415					
Glu	Gly	Leu	Leu	Cys	Phe	Glu	Leu	Met	Arg	Arg	Lys	His	Lys	Ile	Glu				
			420					425					430						
Pro	Arg	Leu	Gln	His	Tyr	Gly	Cys	Met	Val	Asp	Ile	Leu	Ser	Arg	Ser				
		435					440					445							
Gly	Ser	Ile	Glu	Leu	Ala	Lys	Asn	Leu	Ile	Glu	Glu	Met	Pro	Val	Glu				
	450					455						460							
Pro	Asn	Asp	Val	Ile	Trp	Arg	Thr	Phe	Leu	Thr	Ala	Cys	Ser	His	His				
465					470					475					480				
Lys	Glu	Phe	Glu	Thr	Gly	Glu	Leu	Val	Ala	Lys	His	Leu	Ile	Leu	Gln				
				485					490						495				
Ala	Gly	Tyr	Asn	Pro	Ser	Ser	Tyr	Val	Leu	Leu	Ser	Asn	Met	Tyr	Ala				
			500					505					510						
Ser	Phe	Gly	Met	Trp	Lys	Asp	Val	Arg	Arg	Val	Arg	Thr	Met	Met	Lys				
		515					520					525							
Glu	Arg	Lys	Ile	Glu	Lys	Ile	Pro	Gly	Cys	Ser	Trp	Ile	Glu	Leu	Asp				
	530					535					540								
Gly	Arg	Val	His	Glu	Phe	Phe	Val	Asp	Ser	Ile	Glu	Val	Ser	Ser	Thr				
545					550					555					560				

Leu

- <210> 3
- <211> 13
- 5 <212> ARN
- <213> Secuencia artificial
- <220>
- <223> ARN sintético
- 10 <400> 3 uaucuugucu uua 13
- <210> 4
- <211> 12
- 15 <212> PRT

ES 2 874 230 T3

<213> Secuencia artificial

<220>

<223> Péptido sintético

5

<400> 4

Met	Ala	Gly	Tyr	Pro	Tyr	Asp	Val	Pro	Asp	Tyr	Ala
1				5					10		

10

<210> 5

<211> 220

<212> PRT

<213> *Homo sapiens*

15

<400> 5

ES 2 874 230 T3

Met Ala Thr Val Glu Pro Glu Thr Thr Pro Thr Pro Asn Pro Pro Thr
 1 5 10 15

Thr Glu Glu Glu Lys Thr Glu Ser Asn Gln Glu Val Ala Asn Pro Glu
 20 25 30

His Tyr Ile Lys His Pro Leu Gln Asn Arg Trp Ala Leu Trp Phe Phe
 35 40 45

Lys Asn Asp Lys Ser Lys Thr Trp Gln Ala Asn Leu Arg Leu Ile Ser
 50 55 60

Lys Phe Asp Thr Val Glu Asp Phe Trp Ala Leu Tyr Asn His Ile Gln
 65 70 75 80

Leu Ser Ser Asn Leu Met Pro Gly Cys Asp Tyr Ser Leu Phe Lys Asp
 85 90 95

Gly Ile Glu Pro Met Leu Glu Asp Glu Lys Asn Lys Arg Gly Gly Arg
 100 105 110

Trp Leu Ile Thr Leu Asn Lys Gln Gln Arg Arg Ser Asp Leu Asp Arg
 115 120 125

Phe Trp Leu Glu Thr Leu Leu Cys Leu Ile Gly Glu Ser Phe Asp Asp
 130 135 140

Tyr Ser Asp Asp Val Cys Gly Ala Val Val Asn Val Arg Ala Lys Gly
 145 150 155 160

Asp Lys Ile Ala Ile Trp Thr Thr Glu Cys Glu Asn Arg Glu Ala Val
 165 170 175

Thr His Ile Gly Arg Val Tyr Lys Glu Arg Leu Gly Leu Pro Pro Lys
 180 185 190

Ile Val Ile Gly Tyr Gln Ser His Ala Asp Thr Ala Thr Lys Ser Gly
 195 200 205

Ser Thr Thr Lys Asn Arg Phe Val Val Gly Arg Tyr
 210 215 220

<210> 6
 <211> 198
 <212> PRT
 <213> *Homo sapiens*

5

<400> 6

ES 2 874 230 T3

Met Ala Ala Asp Ile Ser Glu Ser Ser Gly Ala Asp Cys Lys Gly Asp
 1 5 10 15

Pro Arg Asn Ser Ala Lys Leu Asp Ala Asp Tyr Pro Leu Arg Val Leu
 20 25 30

Tyr Cys Gly Val Cys Ser Leu Pro Thr Glu Tyr Cys Glu Tyr Met Pro
 35 40 45

Asp Val Ala Lys Cys Arg Gln Trp Leu Glu Lys Asn Phe Pro Asn Glu
 50 55 60

Phe Ala Lys Leu Thr Val Glu Asn Ser Pro Lys Gln Glu Ala Gly Ile
 65 70 75 80

Ser Glu Gly Gln Gly Thr Ala Gly Glu Glu Glu Glu Lys Lys Lys Gln
 85 90 95

Lys Arg Gly Gly Arg Gly Gln Ile Lys Gln Lys Lys Lys Thr Val Pro
 100 105 110

Gln Lys Val Thr Ile Ala Lys Ile Pro Arg Ala Lys Lys Lys Tyr Val
 115 120 125

Thr Arg Val Cys Gly Leu Ala Thr Phe Glu Ile Asp Leu Lys Glu Ala
 130 135 140

Gln Arg Phe Phe Ala Gln Lys Phe Ser Cys Gly Ala Ser Val Thr Gly
 145 150 155 160

Glu Asp Glu Ile Ile Ile Gln Gly Asp Phe Thr Asp Asp Ile Ile Asp
 165 170 175

Val Ile Gln Glu Lys Trp Pro Glu Val Asp Asp Asp Ser Ile Glu Asp
 180 185 190

Leu Gly Glu Val Lys Lys
 195

<210> 7
 <211> 181
 <212> PRT
 <213> *Homo sapiens*
 <400> 7

5

ES 2 874 230 T3

Met Phe Lys Lys Phe Asp Glu Lys Glu Asn Val Ser Asn Cys Ile Gln
 1 5 10 15

Leu Lys Thr Ser Val Ile Lys Gly Ile Lys Asn Gln Leu Ile Glu Gln
 20 25 30

Phe Pro Gly Ile Glu Pro Trp Leu Asn Gln Ile Met Pro Lys Lys Asp
 35 40 45

Pro Val Lys Ile Val Arg Cys His Glu His Ile Glu Ile Leu Thr Val
 50 55 60

Asn Gly Glu Leu Leu Phe Phe Arg Gln Arg Glu Gly Pro Phe Tyr Pro
 65 70 75 80

Thr Leu Arg Leu Leu His Lys Tyr Pro Phe Ile Leu Pro His Gln Gln
 85 90 95

Val Asp Lys Gly Ala Ile Lys Phe Val Leu Ser Gly Ala Asn Ile Met
 100 105 110

Cys Pro Gly Leu Thr Ser Pro Gly Ala Lys Leu Tyr Pro Ala Ala Val
 115 120 125

Asp Thr Ile Val Ala Ile Met Ala Glu Gly Lys Gln His Ala Leu Cys
 130 135 140

Val Gly Val Met Lys Met Ser Ala Glu Asp Ile Glu Lys Val Asn Lys
 145 150 155 160

Gly Ile Gly Ile Glu Asn Ile His Tyr Leu Asn Asp Gly Leu Trp His
 165 170 175

Met Lys Thr Tyr Lys
 180

<210> 8
 <211> 172
 <212> PRT
 <213> *Homo sapiens*

5

<400> 8

ES 2 874 230 T3

Met Ile Ile Tyr Arg Asp Leu Ile Ser His Asp Glu Met Phe Ser Asp
 1 5 10 15

Ile Tyr Lys Ile Arg Glu Ile Ala Asp Gly Leu Cys Leu Glu Val Glu
 20 25 30

Gly Lys Met Val Ser Arg Thr Glu Gly Asn Ile Asp Asp Ser Leu Ile
 35 40 45

Gly Gly Asn Ala Ser Ala Glu Gly Pro Glu Gly Glu Gly Thr Glu Ser
 50 55 60

Thr Val Ile Thr Gly Val Asp Ile Val Met Asn His His Leu Gln Glu
 65 70 75 80

Thr Ser Phe Thr Lys Glu Ala Tyr Lys Lys Tyr Ile Lys Asp Tyr Met
 85 90 95

Lys Ser Ile Lys Gly Lys Leu Glu Glu Gln Arg Pro Glu Arg Val Lys
 100 105 110

Pro Phe Met Thr Gly Ala Ala Glu Gln Ile Lys His Ile Leu Ala Asn
 115 120 125

Phe Lys Asn Tyr Gln Phe Phe Ile Gly Glu Asn Met Asn Pro Asp Gly
 130 135 140

Met Val Ala Leu Leu Asp Tyr Arg Glu Asp Gly Val Thr Pro Tyr Met
 145 150 155 160

Ile Phe Phe Lys Asp Gly Leu Glu Met Glu Lys Cys
 165 170

<210> 9
 <211> 412
 <212> PRT
 <213> *Homo sapiens*

<400> 9

Pro Pro Lys Lys Gln Ala Gln Ala Gly Gly Ser Lys Lys Ala Glu Gln
 1 5 10 15

5

10

ES 2 874 230 T3

Lys Lys Lys Glu Lys Ile Ile Glu Asp Lys Thr Phe Gly Leu Lys Asn
 20 25 30

Lys Lys Gly Ala Lys Gln Gln Lys Phe Ile Lys Ala Val Thr His Gln
 35 40 45

Val Lys Phe Gly Gln Gln Asn Pro Arg Gln Val Ala Gln Ser Glu Ala
 50 55 60

Glu Lys Lys Leu Lys Lys Asp Asp Lys Lys Lys Glu Leu Gln Glu Leu
 65 70 75 80

Asn Glu Leu Phe Lys Pro Val Val Ala Ala Gln Lys Ile Ser Lys Gly
 85 90 95

Ala Asp Pro Lys Ser Val Val Cys Ala Phe Phe Lys Gln Gly Gln Cys
 100 105 110

Thr Lys Gly Asp Lys Cys Lys Phe Ser His Asp Leu Thr Leu Glu Arg
 115 120 125

Lys Cys Glu Lys Arg Ser Val Tyr Ile Asp Ala Arg Asp Glu Glu Leu
 130 135 140

Glu Lys Asp Thr Met Asp Asn Trp Asp Glu Lys Lys Leu Glu Glu Val
 145 150 155 160

Val Asn Lys Lys His Gly Glu Ala Glu Lys Lys Lys Pro Lys Thr Gln
 165 170 175

Ile Val Cys Lys His Phe Leu Glu Ala Ile Glu Asn Asn Lys Tyr Gly
 180 185 190

Trp Phe Trp Val Cys Pro Gly Gly Gly Asp Ile Cys Met Tyr Arg His
 195 200 205

Ala Leu Pro Pro Gly Phe Val Leu Lys Lys Asp Lys Lys Lys Glu Glu
 210 215 220

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

Ala Leu Gly Pro Asn Val Thr Lys Ile Thr Leu Glu Ser Phe Leu Ala
 245 250 255

Trp Lys Lys Arg Lys Arg Gln Glu Lys Ile Asp Lys Leu Glu Gln Asp
 260 265 270

ES 2 874 230 T3

Met Glu Arg Arg Lys Ala Asp Phe Lys Ala Gly Lys Ala Leu Val Ile
 275 280 285

Ser Gly Arg Glu Val Phe Glu Phe Arg Pro Glu Leu Val Asn Asp Asp
 290 295 300

Asp Glu Glu Ala Asp Asp Thr Arg Tyr Thr Gln Gly Thr Gly Gly Asp
 305 310 315 320

Glu Val Asp Asp Ser Val Ser Val Asn Asp Ile Asp Leu Ser Leu Tyr
 325 330 335

Ile Pro Arg Asp Val Asp Glu Thr Gly Ile Thr Val Ala Ser Leu Glu
 340 345 350

Arg Phe Ser Thr Tyr Thr Ser Asp Lys Asp Glu Asn Lys Leu Ser Glu
 355 360 365

Ala Ser Gly Gly Arg Ala Glu Asn Gly Glu Arg Ser Asp Leu Glu Glu
 370 375 380

Asp Asn Glu Arg Glu Gly Thr Glu Asn Gly Ala Ile Asp Ala Val Pro
 385 390 395 400

Val Asp Glu Lys Ser Phe His Trp Arg Gly Phe Gly
 405 410

<210> 10
 <211> 269
 <212> PRT
 <213> *Homo sapiens*

5

<400> 10

Ala Cys Arg Pro Arg Ser Pro Pro Arg His Gln Ser Arg Cys Asp Gly
 1 5 10 15

Asp Ala Ser Pro Pro Ser Pro Ala Arg Trp Ser Leu Gly Arg Lys Arg
 20 25 30

Arg Ala Asp Gly Arg Arg Trp Arg Pro Glu Asp Ala Glu Glu Ala Glu
 35 40 45

His Arg Gly Ala Glu Arg Arg Pro Glu Ser Phe Thr Thr Pro Glu Gly
 50 55 60

Pro Lys Pro Arg Ser Arg Cys Ser Asp Trp Ala Ser Ala Val Glu Glu
 65 70 75 80

10

ES 2 874 230 T3

Asp Glu Met Arg Thr Arg Val Asn Lys Glu Met Ala Arg Tyr Lys Arg
85 90 95

Lys Leu Leu Ile Asn Asp Phe Gly Arg Glu Arg Lys Ser Ser Ser Gly
100 105 110

Ser Ser Asp Ser Lys Glu Ser Met Ser Thr Val Pro Ala Asp Phe Glu
115 120 125

Thr Asp Glu Ser Val Leu Met Arg Arg Gln Lys Gln Ile Asn Tyr Gly
130 135 140

Lys Asn Thr Ile Ala Tyr Asp Arg Tyr Ile Lys Glu Val Pro Arg His
145 150 155 160

Leu Arg Gln Pro Gly Ile His Pro Lys Thr Pro Asn Lys Phe Lys Lys
165 170 175

Tyr Ser Arg Arg Ser Trp Asp Gln Gln Ile Lys Leu Trp Lys Val Ala
180 185 190

Leu His Phe Trp Asp Pro Pro Ala Glu Glu Gly Cys Asp Leu Gln Glu
195 200 205

Ile His Pro Val Asp Leu Glu Ser Ala Glu Ser Ser Ser Glu Pro Gln
210 215 220

Thr Ser Ser Gln Asp Asp Phe Asp Val Tyr Ser Gly Thr Pro Thr Lys
225 230 235 240

Val Arg His Met Asp Ser Gln Val Glu Asp Glu Phe Asp Leu Glu Ala
245 250 255

Cys Leu Thr Glu Pro Leu Arg Asp Phe Ser Ala Met Ser
260 265

<210> 11
<211> 95
5 <212> PRT
<213> *Homo sapiens*
<400> 11

Pro Gly Pro Thr Pro Ser Gly Thr Asn Val Gly Ser Ser Gly Arg Ser
1 5 10 15

Pro Ser Lys Ala Val Ala Ala Arg Ala Ala Gly Ser Thr Val Arg Gln
20 25 30

10

ES 2 874 230 T3

Arg Lys Asn Ala Ser Cys Gly Thr Arg Ser Ala Gly Arg Thr Thr Ser
 35 40 45

Ala Gly Thr Gly Gly Met Trp Arg Phe Tyr Thr Glu Asp Ser Pro Gly
 50 55 60

Leu Lys Val Gly Pro Val Pro Val Leu Val Met Ser Leu Leu Phe Ile
 65 70 75 80

Ala Ser Val Phe Met Leu His Ile Trp Gly Lys Tyr Thr Arg Ser
 85 90 95

<210> 12
 <211> 23
 <212> PRT
 <213> *Homo sapiens*

5

<400> 12

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

Val Phe Met Leu His Ile Trp
 20

10

<210> 13
 <211> 285
 <212> PRT
 <213> *Homo sapiens*

15

<400> 13

Arg Leu Leu Pro Arg Leu Leu Leu Leu Leu Leu Leu Val Phe Pro Ala
 1 5 10 15

Thr Val Leu Phe Arg Gly Gly Pro Arg Gly Leu Leu Ala Val Ala Gln
 20 25 30

Asp Leu Thr Glu Asp Glu Glu Thr Val Glu Asp Ser Ile Ile Glu Asp
 35 40 45

Glu Asp Asp Glu Ala Glu Val Glu Glu Asp Glu Pro Thr Asp Leu Val
 50 55 60

Glu Asp Lys Glu Glu Glu Asp Val Ser Gly Glu Pro Glu Ala Ser Pro
 65 70 75 80

Ser Ala Asp Thr Thr Ile Leu Phe Val Lys Gly Glu Asp Phe Pro Ala
 85 90 95

ES 2 874 230 T3

Asn Asn Ile Val Lys Phe Leu Val Gly Phe Thr Asn Lys Gly Thr Glu
 100 105 110

Asp Phe Ile Val Glu Ser Leu Asp Ala Ser Phe Arg Tyr Pro Gln Asp
 115 120 125

Tyr Gln Phe Tyr Ile Gln Asn Phe Thr Ala Leu Pro Leu Asn Thr Val
 130 135 140

Val Pro Pro Gln Arg Gln Ala Thr Phe Glu Tyr Ser Phe Ile Pro Ala
 145 150 155 160

Glu Pro Met Gly Gly Arg Pro Phe Gly Leu Val Ile Asn Leu Asn Tyr
 165 170 175

Lys Asp Leu Asn Gly Asn Val Phe Gln Asp Ala Val Phe Asn Gln Thr
 180 185 190

Val Thr Val Ile Glu Arg Glu Asp Gly Leu Asp Gly Glu Thr Ile Phe
 195 200 205

Met Tyr Met Phe Leu Ala Gly Leu Gly Leu Leu Val Ile Val Gly Leu
 210 215 220

His Gln Leu Leu Glu Ser Arg Lys Arg Lys Arg Pro Ile Gln Lys Val
 225 230 235 240

Glu Met Gly Thr Ser Ser Gln Asn Asp Val Asp Met Ser Trp Ile Pro
 245 250 255

Gln Glu Thr Leu Asn Gln Ile Asn Lys Ala Ser Pro Arg Arg Leu Pro
 260 265 270

Arg Lys Arg Ala Gln Lys Arg Ser Val Gly Ser Asp Glu
 275 280 285

<210> 14
 <211> 23
 <212> PRT
 <213> *Homo sapiens*

5

<400> 14

Thr Ile Phe Met Tyr Met Phe Leu Ala Gly Leu Gly Leu Leu Val Ile
 1 5 10 15

Val Gly Leu His Gln Leu Leu
 20

10

ES 2 874 230 T3

<210> 15
 <211> 609
 <212> PRT
 <213> *Homo sapiens*

5

<400> 15

Leu Asp Phe Phe Thr Ile Phe Ser Lys Gly Gly Leu Val Leu Trp Cys
 1 5 10 15

Phe Gln Gly Val Ser Asp Ser Cys Thr Gly Pro Val Asn Ala Leu Ile
 20 25 30

Arg Ser Val Leu Leu Gln Val Gly Phe Gln Lys Ile Leu Thr Leu Thr
 35 40 45

Tyr Val Asp Lys Leu Ile Asp Asp Val His Arg Leu Phe Arg Asp Lys
 50 55 60

Tyr Arg Thr Glu Ile Gln Gln Gln Ser Ala Leu Ser Leu Leu Asn Gly
 65 70 75 80

Thr Phe Asp Phe Gln Asn Asp Phe Leu Arg Leu Leu Arg Glu Ala Glu
 85 90 95

Glu Ser Ser Lys Ile Arg Ala Pro Thr Thr Met Lys Lys Phe Glu Asp
 100 105 110

Ser Glu Lys Ala Lys Lys Pro Val Arg Ser Met Ile Glu Thr Arg Gly
 115 120 125

Glu Lys Pro Lys Glu Lys Ala Lys Asn Ser Lys Lys Lys Gly Ala Lys
 130 135 140

Lys Glu Gly Ser Asp Gly Pro Leu Ala Thr Ser Lys Pro Val Pro Ala
 145 150 155 160

Glu Lys Ser Gly Leu Pro Val Gly Pro Glu Asn Gly Val Glu Leu Ser
 165 170 175

Lys Glu Glu Leu Ile Arg Arg Lys Arg Glu Glu Phe Ile Gln Lys His
 180 185 190

Gly Arg Gly Met Glu Lys Ser Asn Lys Ser Thr Lys Ser Asp Ala Pro
 195 200 205

Lys Glu Lys Gly Lys Lys Ala Pro Arg Val Trp Glu Leu Gly Gly Cys
 210 215 220

ES 2 874 230 T3

Ala Asn Lys Glu Val Leu Asp Tyr Ser Thr Pro Thr Thr Asn Gly Thr
 225 230 235 240

Pro Glu Ala Ala Leu Ser Glu Asp Ile Asn Leu Ile Arg Gly Thr Gly
 245 250 255

Ser Gly Gly Gln Leu Gln Asp Leu Asp Cys Ser Ser Ser Asp Asp Glu
 260 265 270

Gly Ala Ala Gln Asn Ser Thr Lys Pro Ser Ala Thr Lys Gly Thr Leu
 275 280 285

Gly Gly Met Phe Gly Met Leu Lys Gly Leu Val Gly Ser Lys Ser Leu
 290 295 300

Ser Arg Glu Asp Met Glu Ser Val Leu Asp Lys Met Arg Asp His Leu
 305 310 315 320

Ile Ala Lys Asn Val Ala Ala Asp Ile Ala Val Gln Leu Cys Glu Ser
 325 330 335

Val Ala Asn Lys Leu Glu Gly Lys Val Met Gly Thr Phe Ser Thr Val
 340 345 350

Thr Ser Thr Val Lys Gln Ala Leu Gln Glu Ser Leu Val Gln Ile Leu
 355 360 365

Gln Pro Gln Arg Arg Val Asp Met Leu Arg Asp Ile Met Asp Ala Gln
 370 375 380

Arg Arg Gln Arg Pro Tyr Val Val Thr Phe Cys Gly Val Asn Gly Val
 385 390 395 400

Gly Lys Ser Thr Asn Leu Ala Lys Ile Ser Phe Trp Leu Leu Glu Asn
 405 410 415

Gly Phe Ser Val Leu Ile Ala Ala Cys Asp Thr Phe Arg Ala Gly Ala
 420 425 430

Val Glu Gln Leu Arg Thr His Thr Arg Arg Leu Ser Ala Leu His Pro
 435 440 445

Pro Glu Lys His Gly Gly Arg Thr Met Val Gln Leu Phe Glu Lys Gly
 450 455 460

Tyr Gly Lys Asp Ala Ala Gly Ile Ala Met Glu Ala Ile Ala Phe Ala
 465 470 475 480

ES 2 874 230 T3

Arg Asn Gln Gly Phe Asp Val Val Leu Val Asp Thr Ala Gly Arg Met
 485 490 495

Gln Asp Asn Ala Pro Leu Met Thr Ala Leu Ala Lys Leu Ile Thr Val
 500 505 510

Asn Thr Pro Asp Leu Val Leu Phe Val Gly Glu Ala Leu Val Gly Asn
 515 520 525

Glu Ala Val Asp Gln Leu Val Lys Phe Asn Arg Ala Leu Ala Asp His
 530 535 540

Ser Met Ala Gln Thr Pro Arg Leu Ile Asp Gly Ile Val Leu Thr Lys
 545 550 555 560

Phe Asp Thr Ile Asp Asp Lys Val Gly Ala Ala Ile Ser Met Thr Tyr
 565 570 575

Ile Thr Ser Lys Pro Ile Val Phe Val Gly Thr Gly Gln Thr Tyr Cys
 580 585 590

Asp Leu Arg Ser Leu Asn Ala Lys Ala Val Val Ala Ala Leu Met Lys
 595 600 605

Ala

<210> 16
 <211> 18
 <212> PRT
 <213> *Homo sapiens*
 <400> 16

5

Ser Thr Leu Gly His Met Val Leu Phe Pro Val Trp Phe Leu Tyr Ser
 1 5 10 15

10

Leu Leu

<210> 17
 <211> 261
 <212> PRT
 <213> *Homo sapiens*
 <400> 17

15

20

Asp Ile Tyr Asp Thr Gln Thr Leu Gly Val Val Val Phe Gly Gly Phe
 1 5 10 15

ES 2 874 230 T3

Met Val Val Ser Ala Ile Gly Ile Phe Leu Val Ser Thr Phe Ser Met
 20 25 30

Lys Glu Thr Ser Tyr Glu Glu Ala Leu Ala Asn Gln Arg Lys Glu Met
 35 40 45

Ala Lys Thr His His Gln Lys Val Glu Lys Lys Lys Lys Glu Lys Thr
 50 55 60

Val Glu Lys Lys Gly Lys Thr Lys Lys Lys Glu Glu Lys Pro Asn Gly
 65 70 75 80

Lys Ile Pro Asp His Asp Pro Ala Pro Asn Val Thr Val Leu Leu Arg
 85 90 95

Glu Pro Val Arg Ala Pro Ala Val Ala Val Ala Pro Thr Pro Val Gln
 100 105 110

Pro Pro Ile Ile Val Ala Pro Val Ala Thr Val Pro Ala Met Pro Gln
 115 120 125

Glu Lys Leu Ala Ser Ser Pro Lys Asp Lys Lys Lys Lys Glu Lys Lys
 130 135 140

Val Ala Lys Val Glu Pro Ala Val Ser Ser Val Val Asn Ser Ile Gln
 145 150 155 160

Val Leu Thr Ser Lys Ala Ala Ile Leu Glu Thr Ala Pro Lys Glu Gly
 165 170 175

Arg Asn Thr Asp Val Ala Gln Ser Pro Glu Ala Pro Lys Gln Glu Ala
 180 185 190

Pro Ala Lys Lys Lys Ser Gly Ser Lys Lys Lys Gly Pro Pro Asp Ala
 195 200 205

Asp Gly Pro Leu Tyr Leu Pro Tyr Lys Thr Leu Val Ser Thr Val Gly
 210 215 220

Ser Met Val Phe Asn Glu Gly Glu Ala Gln Arg Leu Ile Glu Ile Leu
 225 230 235 240

Ser Glu Lys Ala Gly Ile Ile Gln Asp Thr Trp His Lys Ala Thr Gln
 245 250 255

Lys Gly Asp Pro Val
 260

ES 2 874 230 T3

<210> 18
<211> 23
<212> PRT
<213> *Homo sapiens*

5
<400> 18

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

Ile Phe Leu Val Ser Thr Phe
20

10
<210> 19
<211> 30
<212> PRT
<213> *Homo sapiens*

15
<400> 19

Asp Ile Tyr Asp Thr Gln Thr Leu Gly Val Val Val Phe Gly Gly Phe
1 5 10 15

Met Val Val Ser Ala Ile Gly Ile Phe Leu Val Ser Thr Phe
20 25 30

20
<210> 20
<211> 4
<212> PRT
<213> *Homo sapiens*

25
<400> 20

Lys Asp Glu Leu
1

30
<210> 21
<211> 4
<212> PRT
<213> *Homo sapiens*

<400> 21

Lys Glu Glu Leu
1

35
<210> 22
<211> 19
<212> PRT
<213> *Homo sapiens*

<400> 22

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

45
Ala His Ser

REIVINDICACIONES

1. Una proteína de fusión para mejorar el nivel de expresión de una proteína a partir de un ARNm diana, comprendiendo la proteína de fusión:

- 5 (A) uno o más dominios funcionales que mejoran el nivel de expresión de una proteína a partir de un ARNm; y
(B) un resto polipeptídico que se puede unir a un ARNm diana de una manera selectiva de bases de ARN o específica de secuencia de bases de ARN,

10 en donde el resto polipeptídico (B) es un resto polipeptídico que comprende uno o más motivos de repetición pentatricopeptídica (PPR), comprendiendo cada motivo PPR un polipéptido que consiste en de 30 a 38 aminoácidos de longitud y está representado por la Fórmula 1:
[Fórmula 1]

15 **(Hélice A)-X-(Hélice B)-L** **(Fórmula 1)**

donde

La hélice A es un resto que consiste en 12 aminoácidos de longitud y puede formar una estructura de α -hélice, y está representada por la Fórmula 2:

20 **A₁-A₂-A₃-A₄-A₅-A₆-A₇-A₈-A₉-A₁₁-A₁₁-A₁₂** **(Fórmula 2)**

donde A₁ a A₁₂ representa cada uno independientemente un aminoácido;

25 X no está presente, o es un resto que consiste en de 1 a 9 aminoácidos de longitud;

La hélice B es un resto que consiste en de 11 a 13 aminoácidos de longitud y puede formar una estructura de α -hélice;

L es un resto que consiste en de 2 a 7 aminoácidos de longitud y está representado por la Fórmula 3:

30 [Fórmula 3]

L_{vii}-L_{vi}-L_v-L_{iv}-L_{iii}-L_{ii}-L_i **(Fórmula 3)**

donde los aminoácidos se numeran desde el C-terminal como "i" (-1), "ii" (-2), ... y

35 L_{iii} a L_{vii} pueden no estar presentes, y

una combinación de tres aminoácidos A₁, A₄, y L_{ii} o una combinación de dos aminoácidos A₄ y L_{ii} se corresponde con una base o secuencia de bases del ARNm diana,

40 en donde uno o más dominios funcionales (A) se seleccionan del grupo que consiste en un dominio que guía al ribosoma hacia el ARNm, un dominio que inicia o promueve la traducción del ARNm, un dominio asociado con la exportación nuclear del ARNm, un dominio que se une a una membrana del retículo endoplasmático, un dominio que contiene una secuencia señal de retención del retículo endoplasmático (señal de retención del RE) y un dominio que contiene una secuencia señal del retículo endoplasmático.

45 2. La proteína de fusión de acuerdo con la reivindicación 1, en donde el resto polipeptídico (B) comprende de 2 a 30 motivos PPR, y la pluralidad de motivos PPR se dispone para unirse específicamente a la secuencia de bases del ARNm diana.

50 3. La proteína de fusión de acuerdo con la reivindicación 2, el resto polipeptídico (B) comprende de 5 a 25 motivos PPR.

4. La proteína de fusión de acuerdo con una cualquiera de las reivindicaciones 1 a 3, en donde uno o más dominios funcionales (A) se unen, cada uno, a un N-terminal y/o a un C-terminal del resto polipeptídico (B).

55 5. La proteína de fusión de acuerdo con una cualquiera de las reivindicaciones anteriores, en donde el dominio que guía al ribosoma hacia el ARNm, es un dominio que contiene todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en DENR (proteína regulada por densidad), MCT-1 (secuencia 1 amplificada de linfocitos T malignos), TPT1 (proteína tumoral controlada traduccionalmente) y Lerepo4 (dominio CCCH de dedo de zinc),

60 el dominio que inicia o promueve la traducción del ARNm es un dominio que contiene todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en eIF4E y eIF4G,

el dominio asociado con la exportación nuclear del ARNm es un dominio que contiene todo o la parte funcional de SLBP (proteína de unión al bucle del tallo),

65 el dominio que se une a una membrana del retículo endoplasmático es un dominio que contiene todo o la parte funcional de un polipéptido seleccionado del grupo que consiste en SEC61B, TRAP-alfa (proteína alfa asociada a Translocón), SR-alfa, Dial (citocromo b5 reductasa 3) y p180,

la secuencia señal de retención del retículo endoplasmático (señal de retención del RE) es una secuencia señal que contiene una secuencia KDEL (KEEL), o
la secuencia señal del retículo endoplasmático es una secuencia señal que contiene MGWSCILFLVATATGAHS.

5 6. La proteína de fusión de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en donde la combinación de los tres aminoácidos A₁, A₄, y L_{ii} en cada uno de los motivos PPR es:

10 (valina, treonina, asparagina), (fenilalanina, serina, asparagina), (fenilalanina, treonina, asparagina), (isoleucina, asparagina, ácido aspártico) o (treonina, treonina, asparagina) en orden de (A₁, A₄, L_{ii}) si una base diana para el motivo PPR es A (adenina);

(ácido glutámico, glicina, ácido aspártico), (valina, treonina, ácido aspártico), (lisina, treonina, ácido aspártico) o (leucina, treonina, ácido aspártico) en el orden de (A₁, A₄, L_{ii}) si la base diana para el motivo PPR es G (guanina);

15 (valina, asparagina, ácido aspártico), (isoleucina, asparagina, asparagina), (isoleucina, asparagina, ácido aspártico), (isoleucina, metionina, ácido aspártico), (fenilalanina, prolina, ácido aspártico) o (tirosina, prolina, ácido aspártico) en orden de (A₁, A₄, L_{ii}) si la base diana para el motivo PPR es U (uracilo); o

(valina, asparagina, asparagina), (isoleucina, asparagina, asparagina), (valina, asparagina, serina) o (isoleucina, metionina, ácido aspártico) en el orden de (A₁, A₄, L_{ii}) si la base diana para el motivo PPR es C (citosina).

20 7. La proteína de fusión de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en donde la combinación de los dos aminoácidos A₄ y L_{ii} en cada uno de los motivos PPR es:

(treonina, asparagina), (serina, asparagina) o (glicina, asparagina) en orden de (A₄, L_{ii}) si una base diana para el motivo PPR es A (adenina);

25 (treonina, ácido aspártico) o (glicina, ácido aspártico) en el orden de (A₄, L_{ii}) si la base diana para el motivo PPR es G (guanina);

(asparagina, ácido aspártico), (prolina, ácido aspártico), (metionina, ácido aspártico) o (valina, treonina) en orden de (A₄, L_{ii}) si la base diana para el motivo PPR es U (uracilo); o

30 (asparagina, asparagina), (asparagina, serina) o (leucina, ácido aspártico) en el orden de (A₄, L_{ii}) si la base diana para el motivo PPR es C (citosina).

8. Un ácido nucleico que codifica la proteína de fusión de acuerdo con una cualquiera de las reivindicaciones 1 a 7.

9. Un vector que comprende el ácido nucleico de acuerdo con la reivindicación 8.

35 10. El vector de acuerdo con la reivindicación 9, en donde el vector es un vector de expresión.

11. Un método *in vitro* para mejorar el nivel de expresión de una proteína a partir de un ARNm diana dentro de una célula, comprendiendo el método:

40 una etapa de proporcionar la proteína de fusión de acuerdo con una cualquiera de las reivindicaciones 1 a 8 o el vector de acuerdo con la reivindicación 9 o 10; y
una etapa de introducir la proteína de fusión o el vector en una célula.

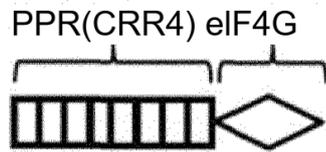
45 12. El método de acuerdo con la reivindicación 11, en donde la célula es una célula eucariota.

13. El método de acuerdo con la reivindicación 12, en donde la célula es una célula animal.

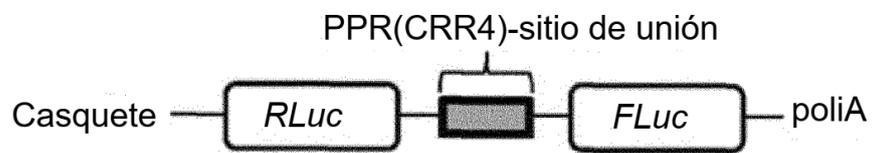
50 14. El método de acuerdo con la reivindicación 13, en donde la célula animal es una célula humana.

[Figura 1]

A Plásmido efector

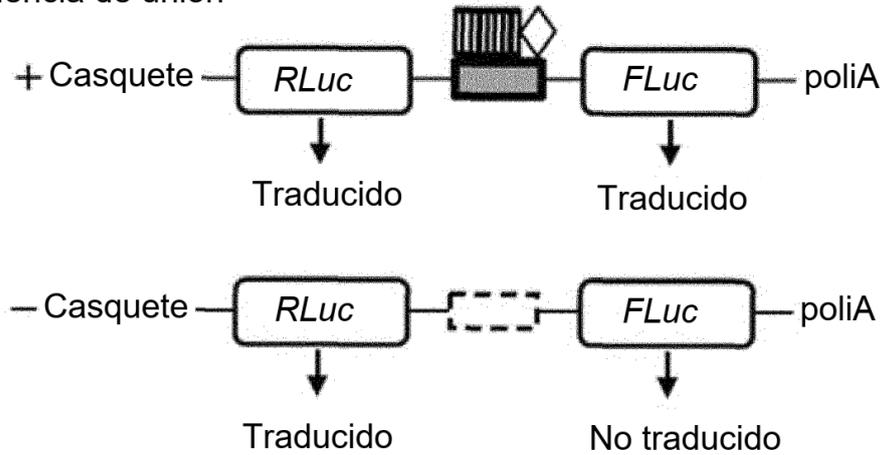


Plásmido indicador

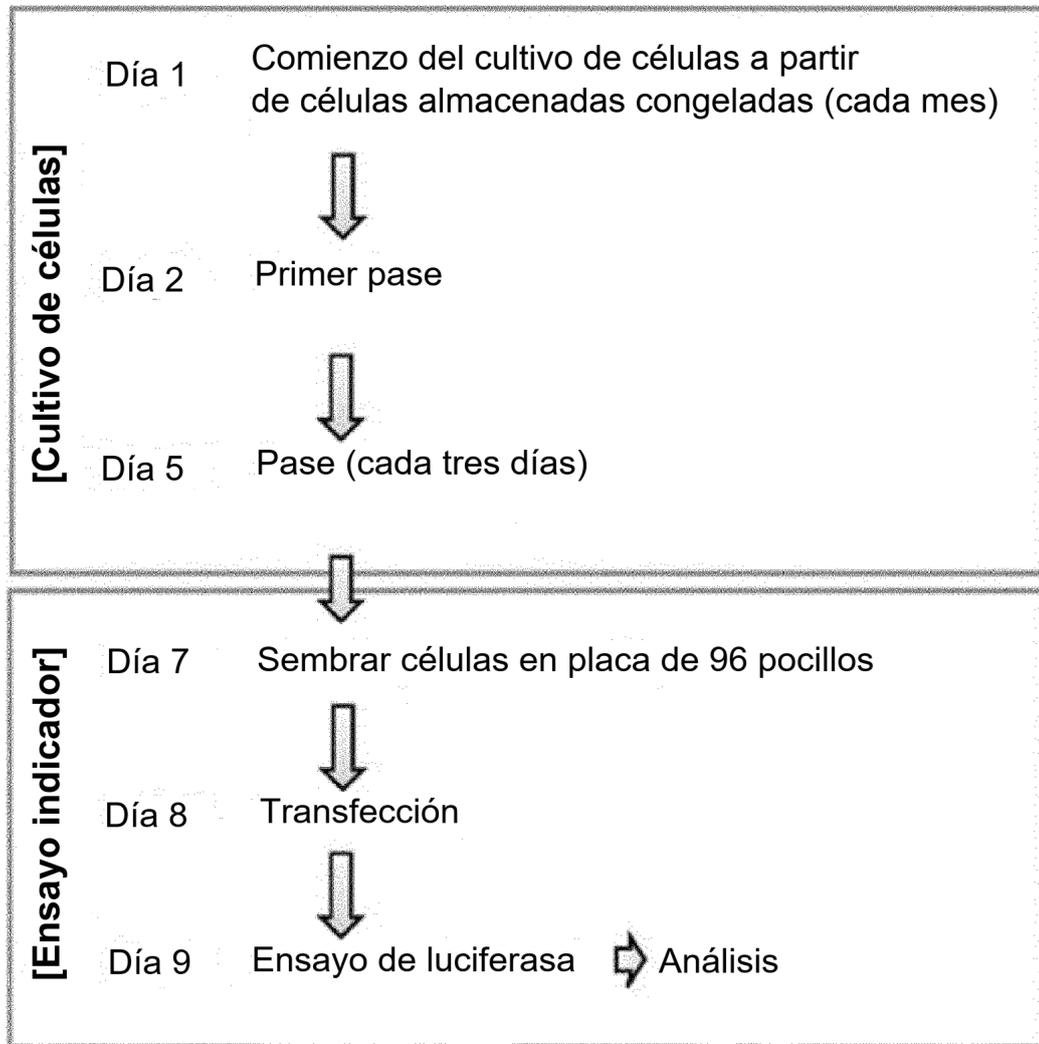


B

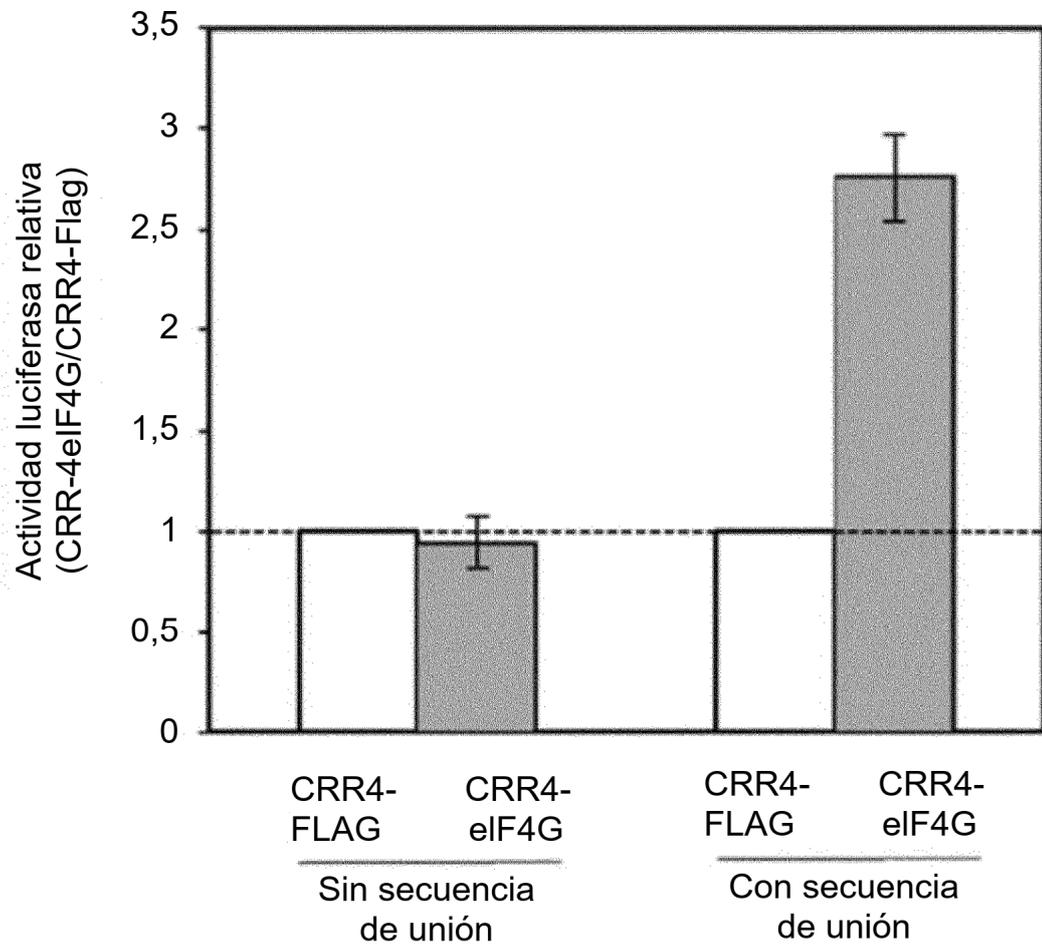
Secuencia de unión



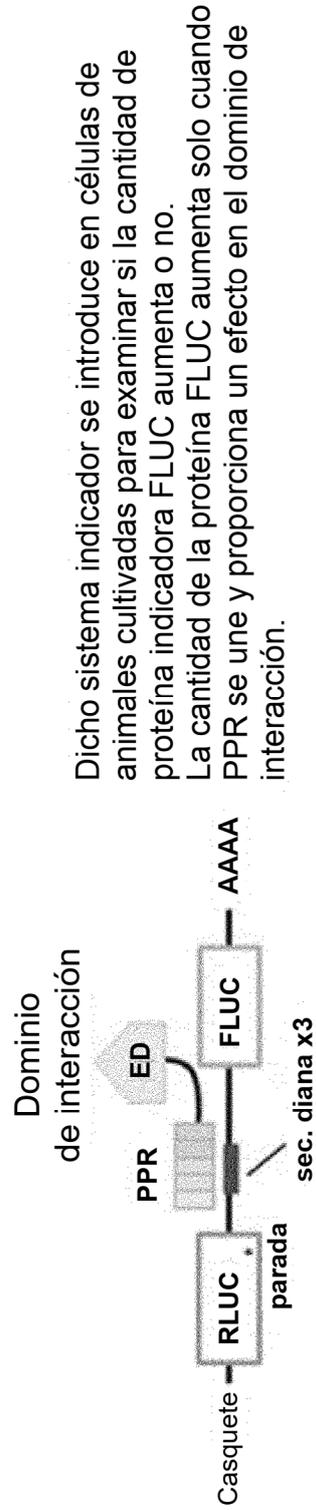
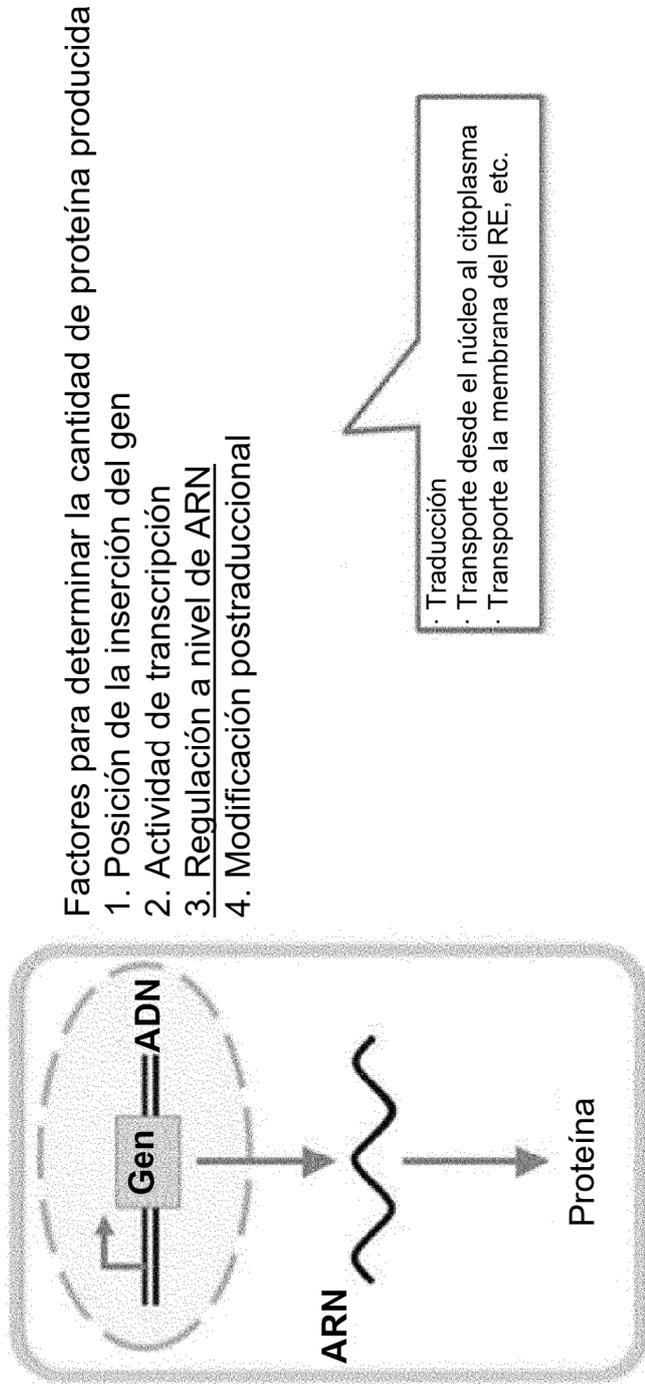
[Figura 2]



[Figura 3]

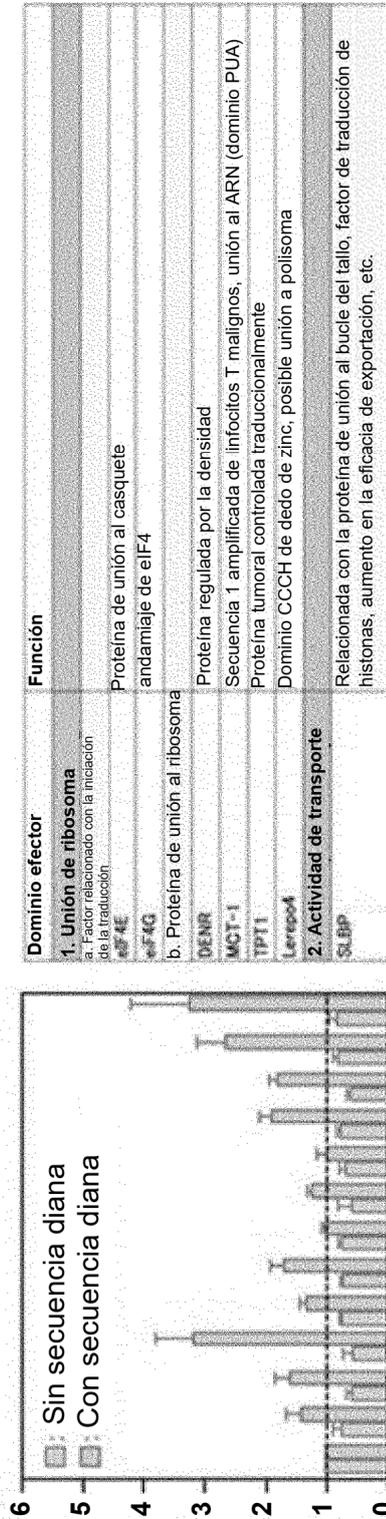


[Figura 4]



[Figura 5]

1. Activación de la traducción por reclutamiento de los ribosomas

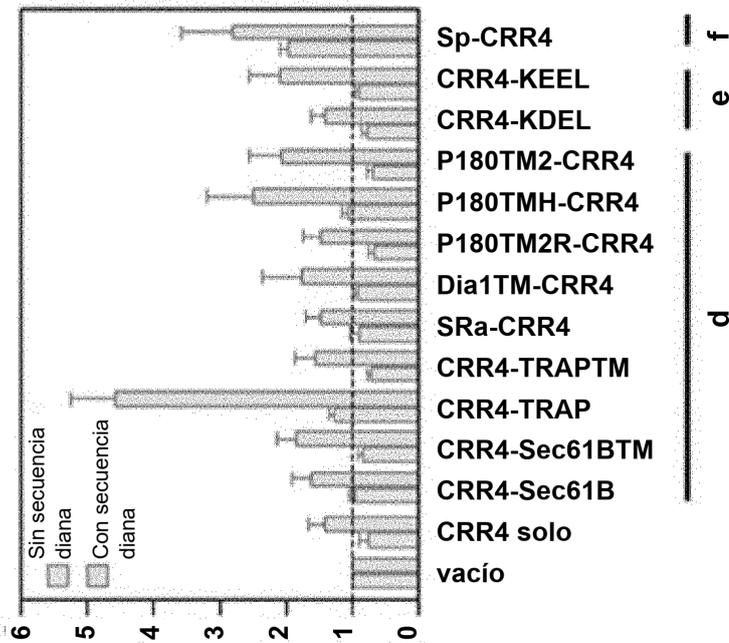


Dominio efector	Función
1. Unión de ribosoma	
a. Factor relacionado con la iniciación de la traducción	
eIF4E	Proteína de unión al casquete andamiaje de eIF4
b. Proteína de unión al ribosoma	
DENR	Proteína regulada por la densidad
MCT-1	Secuencia 1 amplificada de linfocitos T malignos, unión al ARN (dominio PUA)
TPT1	Proteína tumoral controlada traduccionalmente
Lerepo4	Dominio CCH de dedo de zinc, posible unión a polisoma
2. Actividad de transporte	
SLBP	Relacionada con la proteína de unión al bucle del tallo, factor de traducción de histonas, aumento en la eficacia de exportación, etc.

- a. Factor de iniciación de la traducción
- b. Proteína de unión al ribosoma
- c. Transporte de núcleo-citoplasma

[Figura 6]

2. Activación de la traducción por fijación al RE



Dominio efector	Función
3. Unión al RE	
a. Anclaje de la membrana del RE	
SECY1B	Subunidad del compuesto para insertar la proteína en la membrana del RE
TRAP-alfa	Subunidad del compuesto para insertar la proteína en la membrana del RE
SRa-alfa	Subunidad del compuesto para insertar la proteína en la membrana del RE
Dia1	Citocromo b5 reductasa 3
P180	¿Sirve para pasar el ARNm al ribosoma en el RE?
b. Señal de retención en RE	
KDEL/KEEL	
c. Péptido señal del RE	
MGWSCLFLVATATGAS	Añadido a N-terminal, reconocido por SRP, y localizado en el RE

- d. Factor de iniciación de la traducción de anclaje al RE
- e. Retención del RE
- f. Localizado en el RE