

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 432 391**

51 Int. Cl.:

C12N 15/74 (2006.01)

C07K 14/245 (2006.01)

C12N 15/70 (2006.01)

C12P 21/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.03.2007 E 07752601 (0)**

97 Fecha y número de publicación de la concesión europea: **21.08.2013 EP 1991680**

54 Título: **Sistema para la expresión de componentes de traducción ortogonales en células huésped eubacterianas**

30 Prioridad:

09.03.2006 US 780973 P

17.03.2006 US 783497 P

29.10.2006 US 855336 P

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

03.12.2013

73 Titular/es:

**THE SCRIPPS RESEARCH INSTITUTE (100.0%)
10550 NORTH TORREY PINES ROAD
LA JOLLA, CA 92037, US**

72 Inventor/es:

**RYU, YOUNGHA y
SCHULTZ, PETER G.**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 432 391 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

Sistema para la expresión de componentes de traducción ortogonales en células huésped eubacterianas

5 **Campo de la invención**

La invención se refiere al campo de la química de las proteínas, por ejemplo la bioquímica de la traducción. La invención se refiere a composiciones y procedimientos para la producción *in vivo* de polipéptidos que comprende uno o más aminoácidos no naturales.

10

Antecedentes de la invención

El estudio de la estructura y función de las proteínas ha dependido históricamente de las propiedades químicas que están disponibles usando los grupos R de los aminoácidos naturales. Por desgracia, todos los organismos conocidos, desde bacterias a seres humanos, codifica los mismos veinte aminoácidos comunes (con las raras excepciones de la selenocisteína (véase, por ejemplo, Bock y col.,(1991), *Molecular Microbiology* 5:515-20) y la pirrolisina. Esta limitada selección de grupos R ha restringido el estudio de la estructura y función de las proteínas, en el que los estudios están limitados por las propiedades químicas de los aminoácidos naturales.

15

20

Se ha desarrollado una metodología general para la incorporación específica de sitio *in vivo* de aminoácidos no naturales químicamente diversos con nuevas propiedades fisicoquímicas y biológicas en proteínas de organismos procariontes y eucariontes (Wang y col., *Science* 292, 498-500 (2001); Chin, y col., *Science* 301, 964-967 (2003); Wang y Schultz, *Angew. Chem. Int. Ed.* 44, 34-66 (2005)). Este procedimiento depende de un único par de codón-ARNt y la correspondiente aminoacil-ARNt sintetasa (aaRS o simplemente RS) para cada aminoácido no natural que funciona con eficiencia en la traducción de proteínas pero que no reacciona de forma cruzada con cualquiera de los ARNt endógenos, RS, aminoácidos o codones en el organismo del huésped (es decir, debe ser ortogonal). El uso de dichos pares ARNt-RS ortogonales ha posibilitado codificar genéticamente un gran número de aminoácidos de estructura diversa que incluyen aquéllos con reactividad química (Wang et al., *Proc. Natl. Sci. Acad. USA.* 100, 56-61 (2003)) y fotoquímica (Chin et al., *Proc. Natl. Acad. Sci. USA* 99, 11020-11024 (2002); Wu et al., *J. Am. Chem. Soc.*, 126, 14306-14307 (2004)), así como glicosilada (Zhang et al., *Science* 303, 371-373 (2004)) fluorescente (Wang y Schultz, *Angew. Chem. Int. Ed.* 44:34-66 (2005)), unión a metales (Wang y Schultz, *Angew. Chem. Int. Ed.* 44:34-66 (2005)) y aminoácidos activos en redox (Alfonta et al., *J. Am. Chem. Soc.* 125:14662-14663 (2003)). Un par M^tRNA-Tyr(CUA)-M^tTyrRS mutante concreto ha sido particularmente útil para codificar nuevos aminoácidos en *E. coli* (Wang y Schultz, *Chem. Biol.* 8:883-890 (2001)).

25

30

35

El documento WO 2005/007870 describe composiciones de pares de leucil-ARNt y aminoacil-ARNt sintetasa y usos de las mismas.

40

El documento WO 2004/094593 describe composiciones de pares de ortogonal-ARNt y aminoacil ARNt sintetasa ortogonal y usos de los mismos para incorporar aminoácidos no naturales en una proteína de interés.

El documento US 2005/272121 describe la incorporación específica de sitio de aminoácidos no naturales que contienen átomos pesados en proteínas para determinar la estructura.

45

Wang y Schultz, 2001, *Chemistry & Biology*, vol. 8, pp. 883-890, describen un abordaje general para la generación de ARNt ortogonales.

50

No obstante, a pesar del éxito de esta técnica en la incorporación de una matriz diversa de aminoácidos no naturales *in vivo*, la eficiencia del sistema de expresión para la producción de proteínas mutantes que contienen aminoácidos no naturales no se ha optimizado y la eficacia de la supresión del sistema ortogonal para superar el codón selector puede ser mala. En la técnica existe la necesidad de desarrollar reactivos para mejorar la eficiencia en la supresión de los sistemas de traducción. La invención descrita en el presente documento cumple estas y otras necesidades y será evidente a partir de la revisión de la divulgación siguiente.

55

Sumario de la invención

La presente invención proporciona una composición que comprende un constructo de ácido nucleico, comprendiendo dicho constructo: Secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una primera secuencia de nucleótidos expresable que codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en el que dicho O-ARNt de *Archaea* comprende un par C1-G72 y en el que dichas secuencias del promotor y del terminador están unidas operativamente a dicha primera secuencia de nucleótidos expresable y en el que dicha primera secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, y en el que dicha secuencia de nucleótidos del promotor tiene la secuencia de nucleótidos SEC ID N°: 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36.

65

En ciertos casos de la composición de la presente invención, el constructo de ácido nucleico comprende además una secuencia de nucleótidos del promotor de *E. coli glns* que tiene una secuencia de nucleótidos de la SEC ID N° 13 y una segunda secuencia de nucleótidos expresable, en el que dicha secuencia de nucleótidos del promotor de *E. coli glns* está unida operativamente a dicha segunda secuencia de nucleótidos expresable, de modo que la segunda secuencia de nucleótidos expresable codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila, preferentemente, dicha O-ARNt con un aminoácido no natural.

En ciertos casos de la composición de la presente invención, dicho O-ARNt está codificado por una secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha secuencia de nucleótidos expresable es un operón policistrónico que comprende una pluralidad de O-ARNt cada uno de los cuales tiene la secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*).

En algunos casos, dicha secuencia de nucleótidos expresable comprende una pluralidad de dicho operón policistrónico.

En algunos casos, dicha O-RS es una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o (ii) dicha O-RS es una tirosil-ARNt sintetasa de *Methanococcus jannaschii*; o (iii) dicha O-RS tiene una sustitución de ácido aspártico en arginina en la posición de aminoácidos 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* proporcionada en la SEC ID N° 2 (*Mj-tRNA-Tyr RS* de tipo silvestre).

En un aspecto adicional, la presente invención proporciona una célula huésped que comprende la composición de la presente invención.

En algunos casos, dicha célula huésped es una célula huésped eubacteriana o dicha célula huésped es una célula de *E. coli*.

En un aspecto adicional, la presente invención proporciona un sistema de traducción para la expresión de un polipéptido de interés que comprende al menos un aminoácido no natural en una posición especificada; comprendiendo el sistema:

- (a) un aminoácido no natural;
- (b) un constructo de ácido nucleico, comprendiendo dicho constructo:

- (i) las secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una secuencia de nucleótidos expresable, en la que dicha secuencia de nucleótidos expresable codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en el que dicho O-ARNt de *Archaea* comprende un par C1-G72 y en el que dichas secuencias del promotor y del terminador están unidas operativamente a dicha secuencia de nucleótidos expresable y en la que dicha secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, y en la que dicha secuencia de nucleótidos del promotor tiene la secuencia de nucleótidos SEC ID N° 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36; y
- (ii) una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural; y

- (c) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt, en el que la posición del codón selector en el polinucleótido controla la posición especificada del aminoácido no natural en el polipéptido de interés tras la expresión del polinucleótido para producir el polipéptido.

En algunos casos, dicho constructo de ácido nucleico comprende además al menos uno de:

- (i) una secuencia de nucleótidos que tiene un promotor de *glnS* de *E. coli* que tiene una secuencia de nucleótidos de SEC ID N° 13, en la que dicha secuencia de nucleótidos de *glnS* de *E. coli* está unida operativamente a dicha secuencia de nucleótidos que codifica dicha O-RS; y
- (ii) un operón policistrónico que comprende una pluralidad de dichas secuencias de nucleótidos del gen de O-ARNt, en el que al menos un gen de O-ARNt se separa de al menos un gen de O-ARNt adyacente por un adaptador polinucleotídico heterólogo de un adaptador polinucleotídico de un operón de ARNt.

En algunos casos, dichas secuencias del promotor y del terminador del ARNt de prolina de *E. coli* se proporcionan en las SEC ID N° 32 (promotor) y 33 (terminador), respectivamente; o en el que dicho operón policistrónico comprende una pluralidad de adaptadores polinucleotídicos heterólogos idénticos.

o en el que dicho operón policistrónico comprende una pluralidad de adaptadores polinucleotídicos heterólogos, en el que al menos dos de los adaptadores polinucleotídicos heterólogos son diferentes;

O en el que dicho adaptador polinucleotídico heterólogo comprende un nucleótido de timidina en el extremo 5' o un nucleótido adenosina en el extremo 3' o ambos, un nucleótido de timidina en el extremo 5' y un nucleótido adenosina

en el extremo 3'.

o en el que dicho adaptador polinucleotídico heterólogo es el adaptador polinucleotídico localizado entre los genes de ARNt endógenos de *Escherichia coli* seleccionados de: *valU* y *valX*, *ileT* y *alaT*; *serV* y *argV*; *valV* y *valW*; *glyT* y *thrT*; *metT* y *leuW*; *glnW* y *metU*; *hisR* y *leuT*; *glnU* y *glnW*; *leuP* y *leuV*; *glnV* y *glnX*; *alaW* y *alaX*; *ileU* y *alaU*; *ileV* y *alaV*; *metU* y *glnV*; *glyW* y *cysT*; *argX* y *hisR*; y *argY* y *argZ*;

o en el que dicho adaptador polinucleotídico heterólogo comprende la secuencia de nucleótidos de SEC ID N°: 14 (adaptador *valU/valX*) o 15 (adaptador *ileT/alaT*).

En algunos casos, dicha secuencia de nucleótidos que codifica un O-ARNt es un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos que codifican un O-ARNt.

o en el que dicha secuencia de nucleótidos que codifica un O-ARNt comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*);

o en el que dicha secuencia de nucleótidos que codifica un O-ARNt es un operón policistrónico que comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*);

o en el que dicha O-RS es una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*;

o en el que dicha O-RS es una tirosil-ARNt sintetasa de *Methanococcus jannaschii*;

o en el que dicha O-RS tiene una sustitución de ácido aspártico en arginina en la posición del aminoácido 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* que se proporciona en la SEC ID N° 2 (*Mj-tRNATyr RS* de tipo silvestre).

En algunos casos, el sistema de traducción comprende una célula huésped que comprende (a), (b) y (c).

En algunos casos, la célula huésped es una célula huésped eubacteriana.

En un aspecto adicional, la presente invención proporciona un procedimiento para producir, en una célula huésped, un polipéptido de interés, que comprende un aminoácido no natural en una posición especificada; comprendiendo el procedimiento:

(a) proporcionar:

(i) un aminoácido no natural;

(ii) un constructo de ácido nucleico que comprende:

secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una secuencia de nucleótidos expresable, en la que dicha secuencia de nucleótidos expresable codifica un ARNt ortogonal de *Archaea* (O-ARNt).

en el que dicho O-ARNt de *Archaea* comprende un par C1-G72 y en el que dichas secuencias del promotor y del terminador está unidos operativamente a dicha secuencia de nucleótidos expresable y en el que dicha secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, en el que dicha secuencia de nucleótidos del promotor tiene una secuencia de nucleótidos de las SEC ID N°: 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36; y

una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural; y

(iii) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt y en el que la posición del codón selector se correlaciona con la posición especificada del aminoácido no natural en el polipéptido de interés.

(iv) una célula huésped que comprende (i), (ii) y (iii); y

(b) cultivar dicha célula huésped; y

(c) incorporar dicho aminoácido no natural en dicha posición especificada en dicho polipéptido durante la traducción de dicho polipéptido en dicha célula huésped, de modo que se produce dicho polipéptido de interés que comprende dicho aminoácido no natural en la posición especificada.

En algunos casos, proporcionar un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos que codifican una o más especies de O-ARNt: o en los que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifican un O-ARNt que comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de la secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico proporcionar una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o en el que dicha provisión de un constructo de ácido nucleico proporcionar una tirosil-ARNt sintetasa de *Methanococcus jannaschii*; o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifica una O-RS que tiene una sustitución de ácido aspártico en arginina en la posición de aminoácidos 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-

ARNt sintetasa de *Methanococcus jannaschii* proporcionada en la SEC ID N° 2 (Mj-tRNATyrRS de tipo silvestre); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar al menos uno de:

- 5 (I) secuencias del promotor y del terminador que comprende las secuencias del promotor y del terminador de proK de *E. coli* proporcionadas en las SEC ID N°: 32 (promotor) y 33 (terminador), respectivamente;
- (II) una secuencia de nucleótidos del promotor que tiene un promotor de glnS modificado de *E. coli* que tiene una secuencia de nucleótidos de SEC ID N° 13, en la que dicha secuencia de nucleótidos de glnS de *E. coli* está unida operativamente a dicha secuencia de nucleótidos que codifica dicha O-RS; y
- 10 (III) un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos del gen de O-ARNt, en el que al menos un gen de O-ARNt se separa de al menos un gen de O-ARNt adyacente por un adaptador polinucleotídico heterólogo de un adaptador polinucleotídico de un operón de ARNt.

En algunos casos, proporcionar un operón policistrónico de (III) comprende proporcionar una pluralidad de adaptadores polinucleotídicos heterólogos idénticos; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar una pluralidad de adaptadores polinucleotídicos heterólogos en los que al menos dos de los adaptadores polinucleotídicos heterólogos son diferentes; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar un adaptador polinucleotídico heterólogo que comprende un nucleótido timidina en el extremo 5', un nucleótido adenosina en el extremo 3' o ambos un nucleótido timidina en el extremo 5' y un nucleótido adenosina en el extremo 3'; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar un adaptador polinucleotídico heterólogo del adaptador polinucleotídico localizado entre los genes de ARNt de *Escherichia coli* seleccionados de: *valU* y *valX*; *ileT* y *alaT*; *serV* y *argV*; *valV* y *valW*; *glyT* y *thrT*; *metT* y *leuW*; *glnW* y *metU*; *hisR* y *leuT*; *glnU* y *glnW*; *leuP* y *leuV*; *glnV* y *glnX*; *alaW* y *alaX*; *ileU* y *alaU*; *ileV* y *alaV*; *metU* y *glnV*; *glyW* y *cysT*; *argX* y *hisR*; y *argY* y *argZ*; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar un adaptador polinucleotídico heterólogo que tiene la secuencia de nucleótidos de SEC ID N°: 14 (adaptador *valU/valX*) o 15 (adaptador *ileT/alaT*).

En algunos casos, proporcionar una célula huésped comprende proporcionar una célula huésped eubacteriana o una célula huésped de *Escherichia coli*.

30 En un aspecto adicional, la presente invención proporciona un procedimiento para producir, en una célula huésped, un polipéptido de interés, que comprende un aminoácido no natural en una posición especificada; comprendiendo el procedimiento:

(a) proporcionar:

- 35 (i) un aminoácido no natural;
- (ii) un constructo de ácido nucleico que comprende una secuencia de nucleótidos que codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en el que el O-ARNt de *Archaea* comprende un par C1-G72.
- 40 (iii) un constructo de ácido nucleico que comprende una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural;
- (iv) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt y en el que la posición del codón selector se correlaciona con la posición especificada del aminoácido no natural en el polipéptido de interés; y
- 45 (v) una célula huésped que comprende (i), (ii), (iii) y (iv);

en el que dicho constructo de ácido nucleico de (ii) comprende secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli*, en el que dichas secuencias de nucleótidos del promotor y del terminador están unidas operativamente a dicha secuencia de nucleótidos que codifica dicho O-ARNt, y en el que dicha secuencia de nucleótidos que codifica dicho O-ARNt es heteróloga de dichas secuencias del promotor y del terminador, en el que dicha secuencia de nucleótidos del promotor tiene una secuencia de nucleótidos de SEC ID N°: 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36;

(b) cultivar dicha célula huésped; y

55 (c) incorporar dicho aminoácido no natural en dicha posición especificada en dicho polipéptido durante la traducción de dicho polipéptido en dicha célula huésped, de modo que se produce dicho polipéptido de interés que comprende dicho aminoácido no natural en la posición especificada.

En algunos casos, proporcionar un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos que codifican una o más especies de O-ARNt: o en los que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifican un O-ARNt que comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de la secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una aminoacil-ARNt sintetasa de *Methanococcus jannaschii* o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una tirosil-ARNt sintetasa de

Methanococcus jannaschii;

o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifica una O-RS que tiene una sustitución de ácido aspártico en arginina en la posición del aminoácido 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* que se proporciona en la SEC ID N°: 2 (*Mj*-tRNATyrRS de tipo silvestre); o en el que dicha provisión de un constructo de ácido nucleico de (I) comprende secuencias del promotor y del terminador que comprenden las secuencias del promotor y del terminador de proK de *E. coli* proporcionadas en las SEC ID N° 32 (promotor) y 33 (terminador), respectivamente; o en el que dicha provisión de una célula huésped comprende proporcionar una célula huésped eubacteriana; o en el que dicha provisión de una célula huésped comprende proporcionar una célula huésped de *Escherichia coli*.

En un aspecto adicional, la presente invención proporciona un procedimiento para producir un sistema de traducción para la expresión de un polipéptido de interés que comprende al menos un aminoácido no natural en una posición especificada, comprendiendo el procedimiento proporcionar:

(a) un aminoácido no natural;

(b) un constructo de ácido nucleico, comprendiendo dicho constructo:

(i) las secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una secuencia de nucleótidos expresable, en la que dicha secuencia de nucleótidos expresable codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en el que dicho O-ARNt de *Archaea* comprende un par C1-G72 y en el que dichas secuencias del promotor y del terminador están unidas operativamente a dicha secuencia de nucleótidos expresable y en la que dicha secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, y en la que dicha secuencia de nucleótidos del promotor tiene la secuencia de nucleótidos SEC ID N° 32 o 34; y en el que dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos seleccionada de SEC ID N° 33, 35 y 36; y

(ii) una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural; y

(c) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt, en el que la posición del codón selector en el polinucleótido controla la posición especificada del aminoácido no natural en el polipéptido de interés tras la expresión del polinucleótido para producir el polipéptido.

En algunos casos, (a), (b) y (c) se proporcionan en una célula huésped.

En otras realizaciones del procedimiento, el constructo de ácido nucleico de (II) usa una pluralidad de adaptadores polinucleotídicos heterólogos idénticos. Opcionalmente, al menos dos de los adaptadores polinucleotídicos heterólogos son diferentes.

Opcionalmente, el operón policistrónico de (III) en estos procedimientos usa un adaptador polinucleotídico heterólogo que comprende un nucleótido de timidina en el extremo 5', un nucleótido adenosina en el extremo 3' o ambos, un nucleótido de timidina en el extremo 5' y un nucleótido adenosina en el extremo 3'. El adaptador polinucleotídico heterólogo usado en el presente documento puede derivar de un adaptador polinucleotídico natural localizado entre los genes de ARNt endógenos de *Escherichia coli* seleccionados de: *valU* y *valX*; *ileT* y *alaT*; *serV* y *argV*; *valV* y *valW*; *glyT* y *thrT*; *metT* y *leuW*; *glnW* y *metU*; *hisR* y *leuT*; *glnU* y *glnW*; *leuP* y *leuV*; *glnV* y *glnX*; *alaW* y *alaX*; *ileU* y *alaU*; *ileV* y *alaV*; *metU* y *glnV*; *glyW* y *cysT*; *argX* y *hisR*; y *argY* y *argZ*. Por ejemplo, el adaptador polinucleotídico heterólogo puede derivar de la secuencia de nucleótidos de SEC ID N°: 14 (adaptador *valU/valX*) o 15 (adaptador *ileT/alaT*). Opcionalmente, estos procedimientos se realizan en una célula huésped eubacteriana tal como *Escherichia coli*.

En otros aspectos, la invención también proporciona sistemas de traducción para la expresión de un polipéptido de interés que tiene al menos un aminoácido no natural en una posición especificada. Esencialmente, estos sistemas incluyen:

(a) un aminoácido no natural;

(b) un constructo de ácido nucleico, comprendiendo el constructo una secuencia de nucleótidos que codifica un ARNt ortogonal (O-ARNt) y una secuencia de nucleótidos que codifica un aminoacil-ARNt sintetasa ortogonal (O-RS), en la que la O-RS aminoacila preferentemente el O-ARNt con el aminoácido no natural; y

(c) un polinucleótido que codifica el polipéptido de interés, comprendiendo el polinucleótido al menos un codón selector que es reconocido por el O-ARNt, en el que la posición del codón selector en el polinucleótido controla la posición especificada del aminoácido no natural en el polipéptido de interés tras la expresión del

polinucleótido para producir el polipéptido.

Opcionalmente, estos componentes del sistema se integran en una célula huésped.

5 Definiciones

Antes de describir con detalle la invención debe entenderse que la presente invención no se limita a sistemas biológicos concretos que, por supuesto, pueden variar. Se entenderá también que la terminología usada en el presente documento es para el propósito de describir solo realizaciones particulares y no se desea que sea limitante.

10 Como se usa en la presente memoria descriptiva y las reivindicaciones adjuntas, las formas en singular "un", "una" y "el/la" incluyen las referencias en plural a menos que el contenido indique claramente lo contrario. Por tanto, por ejemplo, la referencia a "una célula" incluye combinaciones de dos o más células; la referencia a un "polinucleótido" incluye, como materia práctica, muchas copias de dicho polinucleótido.

15 A menos que se defina otra cosa en el presente documento y más adelante en el resumen de la memoria descriptiva, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que un experto en la técnica a la que la presente divulgación pertenece entiende habitualmente. Al describir y reivindicar la presente invención se usará la siguiente terminología de acuerdo con las definiciones expuestas a continuación.

20 Ortogonal: Como se usa en el presente documento, el término "ortogonal" hace referencia a una molécula (p. ej., un ARNt ortogonal (O-ARNt) y/o una aminoacil-ARNt sintetasa ortogonal (O-RSA)) que funciona con componentes endógenos de una célula con menor eficiencia en comparación con una correspondiente molécula que es endógena a la célula o sistema de traducción o que no funciona con los componentes endógenos de la célula. En el contexto de los ARNt y las aminoacil-ARNt sintetetasas, ortogonal se refiere a una incapacidad o eficiencia reducida, por ejemplo una eficiencia menor del 20 %, una eficiencia menor del 10 %, una eficiencia menor del 5 % o una eficiencia menor del 1 %, de un ARNt ortogonal para funcionar como una ARNt sintetasa endógena en comparación con un ARNt endógeno para funcionar como la ARNt sintetasa endógena, o de una aminoacil-ARNt sintetasa ortogonal para funcionar con un ARNt endógeno en comparación con una ARNt sintetasa endógena para funciona con el ARNt endógeno. La molécula ortogonal carece de una molécula complementaria endógena funcionalmente normal en la célula. Por ejemplo, un ARNt ortogonal en una célula se aminoacila por la acción de cualquier RS endógena de la célula con una eficiencia reducida, o incluso de cero, en comparación con la aminoacilación de un ARNt endógeno mediante la RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno en una célula de interés con una eficiencia reducida, o incluso de cero, en comparación con la aminoacilación del ARNt endógeno mediante una RS endógena. Una segunda molécula ortogonal se puede introducir en la célula que funciona con la primera molécula ortogonal. Por ejemplo, un par ARNt/RS ortogonal incluye componentes complementarios introducidos que funcionan juntos en la célula con una eficiencia (p. ej., una eficiencia del 45 %, una eficiencia del 50 %, una eficiencia del 60 %, una eficiencia del 70 %, una eficiencia del 75 %, una eficiencia del 80 %, una eficiencia del 90 %, una eficiencia del 95 % o una eficiencia del 99 % o mayor) en comparación con la de un control, por ejemplo un par ARNt/RS ortogonal endógeno correspondiente o un par ortogonal activo (p. ej., un par tirosil ARNt/RS ortogonal).

45 Tirosil-ARNt ortogonal: Como se usa en el presente documento, un tirosil-ARNt ortogonal (tirosil-O-ARNt) es un ARNt que es ortogonal a un sistema de traducción de interés, en el que el ARNt es: (1) idéntico o sustancialmente similar a un tirosil-ARNt de origen natural, (2) procedente de un tirosil-ARNt de origen natural mediante mutagénesis natural o artificial, (3) procedente de cualquier proceso que tiene en cuenta una secuencia de un tirosil-ARNt de tipo silvestre o mutante de (1) o (2), (4) homólogo a un tirosil-ARNt de tipo silvestre o mutante; (5) homólogo a cualquier ARNt ilustrativo que se ha diseñado como sustrato para una tirosil-ARNt sintetasa, por ejemplo sintetetasas de las SEC ID N°: 2, 4, 6, 8 o 10, (6) una variante conservadora de cualquier ARNt ilustrativo que se ha diseñado como sustrato de la tirosil-ARNt sintetasa en, por ejemplo, el O-ARNt de la SEC ID N°: 1. El tirosil-ARNt puede existir cargado con un aminoácido o en un estado sin carga. También se entiende que un "tirosil-O-ARNt" opcionalmente está cargado (aminoacilado) mediante una sintetasa conocida con un aminoácido que no es tirosina ni leucina, respectivamente, por ejemplo con un aminoácido no natural. De hecho, se apreciará que un tirosil-O-ARNt de la invención se usa de forma ventajosa para insertar esencialmente cualquier aminoácido, sea natural o artificial, en un polipéptido en crecimiento, durante la traducción, en respuesta a un codón selector.

60 Tirosil aminoácido sintetasa ortogonal: Como se usa en el presente documento, una tirosil aminoácido sintetasa ortogonal (tirosil-O-RS) es una enzima que aminoacila, preferentemente, la tirosil-O-ARNt con un aminoácido en un sistema de traducción de interés. El aminoácido en el la tirosil-O-RS se carga sobre el tirosil-O-ARNt puede ser cualquier aminoácido, natural, no natural o artificial, y no está limitado en el presente documento. La sintetasa es opcionalmente la misma u homóloga a una tirosil aminoácido sintetasa natural o la misma u homóloga de una sintetasa diseñada como una O-RS en, por ejemplo, las SEC ID N°: 4, 6, 8 o 10. Por ejemplo, la O-RS puede ser una variante conservadora de una tirosil-O-R de, por ejemplo, las SEC ID N°: 4, 6, 8 o 10, y/o puede tener una identidad de secuencia de al menos un 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o mayor con una O-RS de, por ejemplo, las SEC ID N°: 4, 6, 8 o 10.

Afín: El término “afín” se refiere a componentes que funcionan juntos, por ejemplo un ARNt ortogonal y una aminoacil-ARNt sintetasa ortogonal. Los componentes también se pueden denominar complementarios.

Aminoacila de forma preferente: Como se usa en el presente documento en referencia a los sistemas de traducción ortogonal, una O-RS “aminoacila de forma preferente” un O-ARNt afín cuando la O-RS carga el O-ARNt con un aminoácido de un modo más eficiente que carga cualquier ARNt endógeno en un sistema de expresión. Es decir, cuando la O-ARNt y cualquier ARNt endógeno están presentes en un sistema de traducción a proporciones molares aproximadamente iguales, la O-RS cargará el O-ARNt con más frecuencia que cargará el ARNt endógeno. Preferentemente, la proporción relativa entre el O-ARNt cargado por la O-RS y el ARNt endógeno cargado por la O-RS es elevada y, preferentemente, tiene como resultado la carga por la O-RS del O-ARNt exclusivamente, o casi exclusivamente, cuando el O-ARNt y el ARNt endógeno están presentes en concentraciones molares iguales en el sistema de traducción. La proporción relativa entre el O-ARNt y el ARNt endógeno cargado por la O-RS cuando el O-ARNt y la O-RS están presentes en concentraciones molares iguales, es superior a 1:1, preferentemente de al menos aproximadamente 2:1, más preferentemente de 5:1, todavía más preferentemente de 10:1, todavía más preferentemente 20:1, todavía más preferentemente 50:1, todavía más preferentemente 75:1, todavía más preferentemente 95:1, 98:1, 99:1, 100:1, 500:1, 1,000:1, 5,000:1 o superior.

La O-RS “aminoacila de forma preferente un O-ARNt con un aminoácido no natural” cuando (a) la O-RS aminoacila de forma preferente el O-ARNt en comparación con un ARNt endógeno, y (b) cuando la aminoacilación es específica del aminoácido no natural en comparación con la aminoacilación del O-ARNt mediante la O-RSA con un aminoácido natural. Es decir, cuando los aminoácidos no naturales y naturales es la están presentes en cantidades molares iguales en un sistema de traducción que comprende la O-RS y el O-ARNt, la O-RS cargará el O-ARNt con el aminoácido no natural con más frecuencia que con el aminoácido natural. Preferentemente, la proporción relativa entre el O-ARNt cargado con el aminoácido no natural y el O-ARNt cargado con el aminoácido natural es alta. Más preferentemente, la O-RS carga el O-ARNt exclusivamente, o casi exclusivamente, con el aminoácido no natural. La proporción relativa entre la carga del O-ARNt con el aminoácido no natural y la carga del O-ARNt con el aminoácido natural, cuando los aminoácidos natural y no natural están presentes en el sistema de traducción en concentraciones molares iguales, es superior a 1:1, preferentemente de al menos aproximadamente 2:1, más preferentemente de 5:1, todavía más preferentemente de 10:1, todavía más preferentemente 20:1, todavía más preferentemente 50:1, todavía más preferentemente 75:1, todavía más preferentemente 95:1, 98:1, 99:1, 100:1, 500:1, 1.000:1, 5.000:1 o superior.

Codón selector: La expresión “codón selector” se refiere a codones reconocidos por el O-ARNt en el proceso de traducción que no es reconocido por un ARNt endógeno. El bucle del anticodón del O-ARNt reconoce el codón selector sobre el ARNm e incorpora su aminoácido, por ejemplo un aminoácido no natural, y este sitio en el polipéptido. Los codones selectores pueden incluir, por ejemplo, codones sin sentido, tales como codones de terminación, por ejemplo los codones ámbar, ocre y ópalo; cuatro o más codones de bases; codones raros; codones derivados de pares de bases naturales o no naturales y/o similares.

ARNt supresor: Un ARNt supresor es un ARNt que altera la lectura de un ARN mensajero (ARNm) en un sistema de traducción determinado, por ejemplo, proporcionando un mecanismo para incorporar un aminoácido en una cadena polipeptídica en respuesta a un codón selector. Por ejemplo, un ARNt supresor puede leer, por ejemplo, un codón de terminación (p. ej., un codón ámbar, ocre u ópalo), un codón de cuatro bases, un codón raro etc.

Actividad de supresión: Como se usa en el presente documento, la expresión “actividad de supresión” hace referencia, en general, a la capacidad de un ARNt (p. ej., un ARNt supresor) para permitir la lectura traduccional completa de un codón (p. ej., un codón selector que es un codón ámbar o un codón de 4 o más bases) que, de otro modo, tendría como resultado la terminación de la traducción o una traducción errónea (p. ej., desplazamiento del marco). La actividad de supresión de un ARNt supresor se puede expresar como un porcentaje de actividad de lectura traduccional completa observada en comparación con un segundo ARNt supresor o en comparación con un sistema control, por ejemplo un sistema control que carece de una O-RS.

La presente invención proporciona varios procedimientos por los cuales se puede cuantificar la actividad de supresión. El porcentaje de supresión de un O-ARNt u O-RS concreto contra un codón selector (p. ej., un codón ámbar) de interés se refiere al porcentaje de actividad de un marcador de prueba expresado dado (p. ej., LacZ), que incluye un codón selector, en un ácido nucleico que codifica el marcador de prueba expresado, en un sistema de traducción de interés, cuando el sistema de traducción de interés incluye una O-RS y un O-ARNt en comparación con un constructo control positivo, en el que el control positivo carece del O-ARNt, la O-RS y el codón selector. Por tanto, por ejemplo, si un constructo marcador control positivo activo que carece de un codón selector tiene una actividad observada de X en un sistema de traducción dado, en unidades relevantes para el ensayo marcador en cuestión, el porcentaje de supresión de un constructo de prueba que comprende el codón selector es el porcentaje de X que el constructo marcador de prueba muestra en esencialmente las mismas condiciones ambientales en las que se expresó el marcador control positivo, a excepción de que el constructo marcador de prueba se expresa en un sistema de traducción que también incluye el O-ARNt y la O-RS. Normalmente, el sistema de traducción que expresa el marcador de prueba también incluye un aminoácido que es reconocido por la O-RS y el O-ARNt. Opcionalmente, la medición del porcentaje de supresión se puede perfeccionar por comparación del marcador de prueba con un constructo marcador control “de fondo” o “negativo” que incluye el mismo codón selector que el

marcador de prueba, pero en un sistema que no incluye el O-ARNt y/o el aminoácido relevante reconocido por el O-ARNt y/o la O-RS. Este control negativo es útil en la normalización de las mediciones del porcentaje de supresión para representar los efectos de la señal de fondo desde el marcador en el sistema de traducción de interés.

- 5 La eficiencia de supresión se puede determinar con cualquiera de una serie de ensayos conocidos en la técnica. Por ejemplo, se puede usar un ensayo indicador de β -galactosidasa, por ejemplo un plásmido *lacZ* derivado (cuando el constructo tiene un codón selector en la secuencia de ácido nucleico de *lacZ*) se introduce en las células de un organismo adecuado (p. ej., un organismo en el que se pueden usar los componentes ortogonales) junto con el plásmido que comprende un O-ARNt de la invención. También se puede introducir una sintetasa afín (bien como polipéptido o como polinucleótido que codifica la sintetasa afín cuando se expresa). Las células se cultivan en medio hasta una densidad deseada, por ejemplo hasta una DO_{600} de aproximadamente 0,5 y se realizan ensayos de β -galactosidasa, por ejemplo usando el kit de ensayo con β -galactosidasa BetaFluor™ (Novagen). El porcentaje de supresión se puede calcular como el porcentaje de actividad para una muestra relativa a un control comparable, por ejemplo el valor observado del constructo *lacZ* derivado, donde el constructo tiene un codón sentido correspondiente a una posición deseada en lugar de un codón selector.

15 Sistema de traducción: La expresión "sistema de traducción" se refiere a componentes que incorporan un aminoácido en una cadena polipeptídica en crecimiento (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetasa, ARNm y similares. El O-ARNt y/o las O-RS de la invención se pueden añadir o formar parte de un sistema de traducción *in vitro* o *in vivo*, por ejemplo, en una célula no eucariota, por ejemplo, una eubacteria (tal como *E. coli*), o en una célula eucariota, por ejemplo, una célula de levadura, una célula de mamífero, una célula vegetal, una célula de alga, una célula de hongo, una célula de insecto y/o similares.

25 Aminoácido no natural: Como se usa en el presente documento "aminoácido no natural" hace referencia a cualquier aminoácido, aminoácido modificado y/o análogo de aminoácido, que no es uno de los 20 aminoácidos de origen natural habituales o los aminoácidos naturales raros selenocisteína o pirrolisina. Por ejemplo, los aminoácidos no naturales *p*-benzoil-L-fenilalanina (Bpa), *para*-acetil-L-fenilalanina (*pAcPhe*), *para*-azido-L-fenilalanina (*pAzPhe*) y *para*-yodo-L-fenilalanina (*pIPhe*) son útiles en la invención.

30 En respuesta a: Como se usó en la presente invención, la expresión "en respuesta a" hace referencia al proceso en el cual un O-ARNt de la invención reconoce un codón selector y participa en la incorporación del aminoácido no natural, que se acopla al ARNt, en la cadena polipeptídica en crecimiento.

35 Polipéptido: Un polipéptido es cualquier oligómero de aminoácidos (naturales o no naturales, o una combinación de los mismos) de cualquier longitud, normalmente, pero no exclusivamente, unidos por enlaces peptídicos covalentes. Un polipéptido puede proceder de cualquier fuente, por ejemplo un polipéptido natural, un polipéptido producido por técnicas de genética molecular recombinante, un polipéptido de una célula o sistema de traducción o un polipéptido producido por medios sintéticos sin células. Un polipéptido se caracteriza por su secuencia de aminoácidos, por ejemplo la estructura primaria de sus componentes aminoácidos. Como se usa en el presente documento, la secuencia de aminoácidos de un polipéptido no está limitada a las secuencias de longitud completa pero pueden ser secuencias parciales o completas. Adicionalmente, no se pretende que un polipéptido se limite a poseer o no poseer ninguna actividad biológica concreta. Como se usa en el presente documento, el término "proteína" es sinónimo de polipéptido. El término "péptido" se refiere a un polipéptido pequeño, por ejemplo, de una longitud de, entre otras, 2-45 aminoácidos.

50 Variante conservadora: Como se usa en el presente documento, la expresión "variante conservadora en el contexto de un componente de traducción, se refiere a un componente de traducción, por ejemplo, un O-ARN y variante conservador o una O-RS variante conservadora, que funcionalmente se comporta de un modo similar a un componente base al cual es similar la variante conservadora, por ejemplo un O-ARNt o O-RS, que tienen variaciones en la secuencia en comparación con un O-ARNt u O-RS de referencia. Por ejemplo, una O-RS, o una variante conservadora de dicha O-RS, aminoacilará un O-ARNt afín con un aminoácido no natural, por ejemplo un aminoácido que comprende un resto de N-acetilgalactosamina. En este ejemplo, la O-RS y la O-RS variante conservadora no tienen las mismas secuencias de aminoácidos. La variante conservadora puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones o cinco o más variaciones en la secuencia, siempre que la variante conservadora siga siendo complementaria al correspondiente O-ARNt u O-RS.

60 En algunas realizaciones, una variante conservadora de O-RS comprende una o más sustituciones conservadoras de aminoácidos en comparación con la O-RS de la que derivó. En algunas realizaciones, una variante conservadora de la O-RS comprende una o más sustituciones conservadoras de aminoácidos en comparación con la O-RS de la que derivó y, adicionalmente, conserva la actividad biológica de la -RS; por ejemplo, una variante conservadora de la O-RS que conserva al menos un 10 % de la actividad biológica de la molécula de O-RS parental de la que derivó o, como alternativa, al menos un 20 %, al menos un 30 % o al menos un 40 %. En algunas realizaciones preferidas, la variante conservadora de O-RS conserva al menos un 50 % de la actividad biológica de la molécula de O-RS parental de la que derivó. Las sustituciones conservadoras de aminoácidos de una variante conservadora de O-RS se pueden producir en cualquier dominio de la O-RS, incluido el punto de unión del aminoácido.

Polinucleótido o ácido nucleico: Las expresiones “ácido nucleico”, “secuencia de ácido nucleico”, “secuencia de nucleótidos”, “oligonucleótido”, “polinucleótido” o “molécula de ácido nucleico” o expresiones similares como se usa en el presente documento se refieren a oligómeros de bases normalmente unidas por una estructura de azúcar-fosfato, tales como oligonucleótidos o polinucleótidos y fragmentos o partes de los mismos y a ADN o ARN de origen genómico o sintético que pueden ser mono o bicatenarios y representa la cadena sentido o antisentido. Los términos ácido nucleico, polinucleótido y nucleótido también incluyen, específicamente, ácidos nucleicos compuestos por bases distintas a las cinco bases de origen biológico (es decir, adenina, guanina, timina, citosina y uracilo) y también incluyen ácidos nucleicos que tienen estructuras de base no naturales, tales como moléculas de PNA.

Se dice que las moléculas de ácido nucleico (p. ej., ADN o ARN) que tienen extremos “5'” y extremos “3'” porque los mononucleótidos reaccionan para fabricar oligonucleótidos o polinucleótidos de un modo tal que el 5' fosfato de un anillo de pentosa mononucleótido está unido al oxígeno en 3' de su vecino en una dirección mediante un enlace fosfodiéster. Por tanto, un polinucleótido normalmente tendrá un “extremo 5'” que comprende un fosfato en 5' y un “extremo 3'” que comprende un oxígeno en 3'. También se dice que una secuencia de polinucleótidos, aunque esté interna en un ácido nucleico más grande, que tiene una direccionalidad 5' y 3'. En una molécula de ADN lineal o circular, se hace referencia a los elementos pequeños estando “cadena arriba” o en 5' de los elementos “cadena abajo” o en 3'. Esta terminología refleja el hecho de que la transcripción progresa en dirección 5' a 3' a lo largo de la hebra de ADN.

Gen: Como se usa en el presente documento, el término “gen” hace referencia, más generalmente, a una combinación de elementos polinucleotídicos, que cuando están unidos operativamente de un modo tanto nativo como recombinante proporcionan algún producto o función. El término “gen” se debe interpretar de forma amplia en el presente documento, y abarca las formas ARNm, ADNc, ARNc y ADN genómico de un gen. En algunos casos, un gen es heredable. En algunos aspectos, los genes comprenden secuencias de codificación (p. ej., un “marco de lectura abierto” o “región de codificación”) necesarias para la producción de un polipéptido, mientras que en otros aspectos, los genes no codifican un polipéptido. Ejemplos de genes que no codifican polipéptidos incluyen genes de ARN ribosómico (ARNr) y ARN de transferencia (ARNt).

El término “gen” puede abarcar, opcionalmente, secuencias reguladoras no codificantes que residen en un locus genético. Por ejemplo, además de una región de codificación de un ácido nucleico, el término “gen” también abarca las secuencias de nucleótidos transcritas del ARNm de longitud completa adyacentes a los extremos 5' y 3' de la región de codificación. Estas regiones no codificantes son de tamaño variable y normalmente se extienden en ambos extremos, 5' y 3', de la región de codificación. Las secuencias que se localizan en 5' y 3' de la región de codificación y están contenidas en el ARNm se denominan secuencias no traducidas en 5' y 3' (5' UT y 3' UT). Las UR en 5' y 3' pueden desempeñar papeles reguladores, incluyendo iniciación de la traducción, escisión postranscripcional y poliadenilación. El término “gen” abarca tanto las formas ARNm, ADNc y genómicas de un gen.

En algunos aspectos, la forma genómica o clon genómico de un gen incluye las secuencias del ARNm transcrito, así como otras secuencias no transcritas que se encuentran fuera del transcrito. Las regiones reguladoras que están fuera de la unidad de transcripción del ARNm se denominan en ocasiones secuencias flanqueantes en “5'” o “3'”. Una forma genómica funcional del un gen normalmente contiene elementos reguladores necesarios para la regulación de la transcripción. Por ejemplo, el término “promotor” normalmente se usa para describir una región de ADN, normalmente pero no exclusivamente en 5' del sitio de iniciación de la transcripción, suficiente para conferir un inicio de la transcripción preciso. En algunas realizaciones, un promotor es activo constitutivamente, aunque en realizaciones alternativas, el promotor es condicionalmente activo (p. ej., cuando la transcripción se inicia únicamente en determinadas condiciones fisiológicas). En procariotas, la actividad de un promotor se puede modular mediante una secuencia “operadora” adyacente. En algunas realizaciones, la región flanqueante en 3' contiene secuencias adicionales que regulan la terminación de la transcripción, en ocasiones denominadas “secuencias de terminación”. En general, la expresión “elemento regulador” se refiere a un elemento genético que controla algún aspecto de la expresión de secuencias de ácido nucleico.

Secuencia de nucleótidos expresable: Como se usa en el presente documento, la expresión “secuencia de nucleótidos expresable” se refiere a cualquier secuencia de nucleótidos que se puede transcribir (p. ej., transcribir mediante una ARN polimerasa dependiente de ADN) para generar un transcrito. La secuencia del transcrito no es limitada y puede ser una secuencia de codificación de proteínas (p. ej., puede codificar una aminoacil-ARNt sintetasa) o no codificante (p. ej., puede codificar una molécula de ARNt).

Unido operativamente: Como se usa en el presente documento, las expresiones “en combinación operable”, “en orden operable”, “unido operativamente” y frases similares, cuando se usan en referencia a los ácidos nucleicos hacen referencia a la unión de secuencias de ácido nucleico que están en relación funcional entre sí. Por ejemplo, una secuencia promotora unida operativamente, el marco de lectura abierto y la secuencia de terminación tiene como resultado la producción precisa de una molécula de ARN. En algunos aspectos, elementos de ácido nucleico unidos operativamente tienen como resultado la transcripción de un marco de lectura abierto y, en última instancia, la producción de un polipéptido (es decir, la expresión del marco de lectura abierto).

Operón: Como se usa en el presente documento, el término “operón” se refiere a una unidad genética (p. ej., una región cromosómica) que controla la expresión génica en procariontes. Un operón comprende normalmente uno o más genes que codifican uno o más polipéptido(s) o ARN y la región (o regiones) reguladora adyacente que controla la transcripción de los genes. La región reguladora normalmente comprende un promotor y un operador. La región de codificación de un gen procarionte históricamente se denomina “cistrón”. Los operones que contienen múltiples cistrones se denominan “policistrónicos”. Los genes en un operón policistrónico normalmente están relacionados en términos de función y normalmente se co-transcriben como una única unidad y se expresa de un modo coordinado.

Constructo: Como se usa en el presente documento, el término “constructo” se usa con referencia a cualquier polinucleótido u otra molécula que puede transferir segmento(s) de ácido nucleico a una célula. El término “vector” o “vehículo” se usa, en ocasiones, de forma intercambiable con “vector”. Un vector comprende, opcionalmente, partes que participan en la propagación y manipulación del vector (p. ej., secuencias necesarias para la replicación, genes que imparte resistencia a fármacos o a antibióticos, un sitio de clonación múltiple, elementos promotores/potenciadores unidos operablemente que permiten la expresión de un gen clonado, etc.). A menudo los vectores derivan de plásmidos, bacteriófagos o virus de plantas o de animales. Un “vector de clonación” o “vector lanzadora” o “vector de subclonación” contiene partes unidas operablemente que facilitan las etapas de subclonación (p. ej., sitio de clonación múltiple que contiene múltiples sitios para las endonucleasas de restricción).

Vector de expresión: La expresión “vector de expresión” como se usa en el presente documento se refiere a un vector recombinante que comprende secuencias polinucleotídicas unidas operablemente que facilitan la expresión de una secuencia de codificación en un organismo huésped concreto (p. ej., un vector de expresión bacteriana). Las secuencias polinucleotídicas que facilitan la expresión en procariontes normalmente incluyen, por ejemplo, un promotor, secuencias de terminación de la transcripción, es decir secuencias terminadoras, un operador (opcional) y un sitio de unión al ribosoma, a menudo junto con otras secuencias.

Codificar: Como se usa en el presente documento, el término “codificar” se refiere a cualquier procedimiento a través del cual la información en una macromolécula o tira de secuencia polimérica se usa para dirigir la producción de una segunda molécula o tira de secuencia que es diferente de la primera molécula o tira de secuencia. Como se usa en el presente documento, el término se usa ampliamente y puede tener varias aplicaciones. En algunos aspectos, el término “codificar” describe el procedimiento de replicación semiconservadora del ADN, en el que una hebra de una molécula de ADN bicatenaria se usa como molde para codificar una hebra hermana complementaria recién sintetizada mediante una ADN polimerasa dependiente de ADN.

En otro aspecto, el término “codificar” se refiere a cualquier procedimiento a través del cual la información en una macromolécula se usa para dirigir la producción de una segunda molécula que tiene una naturaleza química diferente de la de la primera molécula. Por ejemplo, una molécula de ADN puede codificar una molécula de ARN (p. ej., mediante el proceso de transcripción que incorpora una enzima ARN polimerasa dependiente de ADN). Asimismo, una molécula de ARN puede codificar un polipéptido, como en el proceso de traducción. Cuando se usa para describir el proceso de traducción, el término “codificar” también se extiende al codón triplete que codifica un aminoácido. En algunos aspectos, una molécula de ARN puede codificar una molécula de AD, por ejemplo mediante el proceso de transcripción inversa que incorpora una ADN polimerasa dependiente de ARN. En otro aspecto, una molécula de ADN puede codificar un polipéptido, en la que se entiende que “codifica” como se usa en ese caso, incorpora ambos procesos de transcripción y de traducción.

Heterólogo: Como se usa en el presente documento, los términos “heterólogo” o “exógeno” como se aplica a polinucleótidos o polipéptidos se refiere a moléculas que se han reorganizado o suministrado artificialmente a un sistema biológico y no están en una configuración nativa (p. ej., con respecto a la secuencia, la posición genómica o la reorganización de partes) o no son nativos de dicho sistema biológico concreto. Los términos indican que el material relevante originado de una fuente distinta a la fuente natural o se refiere a moléculas que tienen una configuración, localización genética o reorganización de partes no natural. Los términos “exógeno” y “heterólogo” se usan, en ocasiones, de forma intercambiable con “recombinante”.

Recombinante: El término “recombinante” en referencia a un ácido nucleico o polipéptido indica que el material (p. ej., un ácido nucleico recombinante, gen, polinucleótido, polipéptido etc.) ha sido alterado mediante intervención humana. En general, la reorganización de partes de una molécula recombinante no es una configuración nativa o la secuencia primaria del polinucleótido o polipéptido recombinante se ha manipulado de algún modo. La alteración para dar el material recombinante se puede realizar sobre el material en el interior o eliminar de su ambiente o estado natural. Por ejemplo, un ácido nucleico natural se convierte en un ácido nucleico recombinante y, si se altera o si se transcribe a partir de ADN que se ha alterado, por medio de la intervención humana dentro de la célula de la cual procede. Una marco de lectura abierto de la secuencia génica es recombinante si dicha secuencia de nucleótidos se ha eliminado de su contexto natural y se ha clonado en cualquier tipo de vector de ácido nucleico artificial. El término recombinante también puede hacer referencia a un organismo que aloja material recombinante. Los protocolos y reactivos para producir moléculas recombinantes, especialmente ácidos nucleicos recombinantes, son habituales y rutinarios en la técnica (véase, por ejemplo, Maniatis et al. (eds.), Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, NY, [1982]; Sambrook et al. (eds.), Molecular Cloning: A Laboratory Manual, Second Edition, Volumes 1-3, Cold Spring Harbor Laboratory Press, NY, [1989]; y Ausubel et al. (eds.),

Current Protocols in Molecular Biology, Vol. 1-4, John Wiley & Sons, Inc., New York [1994]).

Nativo o endógeno: En contraste con una molécula heteróloga o exógena, una molécula "nativa" o "endógena" es nativa del sistema biológico, especie o cromosoma en estudio. Un gen "nativo" o "endógeno" es un gen que no contiene elementos de ácido nucleico codificados por fuentes distintas al cromosoma en el que normalmente se encuentra en la naturaleza. Un gen, transcrito o polipéptido endógeno está codificado por su locus cromosómico natural y no se suministra artificialmente a la célula.

Célula huésped: La expresión "célula huésped" se refiere a una célula que contiene un ácido nucleico heterólogo, tal como un vector, y soporta la replicación y/o la expresión del ácido nucleico. Las células huésped pueden ser células procariontas, tales como *E. coli*, o células eucariotas tales como de levadura, de insecto, de anfibio o de mamífero. Preferentemente, las células huésped son células vegetales. En el contexto de la invención, una célula huésped particularmente preferida es una célula huésped de soja.

Eucariota: Como se usa en el presente documento, el término "eucariota" se refiere a organismos pertenecientes al reino Eucarya. Generalmente, los eucariotas se pueden distinguir de los procariontas por su organización normalmente pluricelular (pero exclusivamente pluricelular, por ejemplo levaduras), la presencia de un núcleo unido a la membrana y otros orgánulos unidos a la membrana, material genético lineal (es decir, cromosomas lineales), la ausencia de operones, la presencia de intrones, ARN de mensaje protegido y poli-A, y otras características bioquímicas, tales como una estructura ribosómica distinguible. Los organismos eucariotas incluyen, por ejemplo, animales (mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (p. ej., monocotiledóneas, dicotiledóneas, algas etc.), hongos, levaduras, flagelados, microsporidios, protistas etc.

Procarionta: Como se usa en el presente documento, el término "procarionta" se refiere a organismos pertenecientes al reino Monera (también denominado Procarya). Generalmente, los organismos procariontas se pueden distinguir de los eucariotas por su organización unicelular, reproducción asexual mediante gemación o fisión, la ausencia de un núcleo unido a la membrana u otros orgánulos unidos a la membrana, un cromosoma circular, la presencia de operones, la ausencia de intrones, ARNm de mensaje protegido y poli-A, y otras características bioquímicas, tales como una estructura ribosómica distinguible. El reino Prokarya incluye los subreinos Eubacteria y Archaea (en ocasiones denominados "Archaeobacteria"). En ocasiones, se da a las cianobacterias (algas marrón azuladas) y micoplasma clasificaciones distintas en el reino Monera.

Bacteria: Como se usa en el presente documento, los términos "bacteria" y "eubacteria" se refieren a organismos procariontas que se pueden distinguir de Archaea. De un modo similar, Archaea hace referencia a procariontas que se pueden distinguir de eubacterias. Eubacteria y Archaea se pueden distinguir por una serie de criterios morfológicos y bioquímicos. Por ejemplo, las diferencias en las secuencias de ARN ribosómico, la estructura de la ARN polimerasa, la presencia o la ausencia de intrones, la sensibilidad a antibióticos, la presencia o ausencia de peptidoglicanos en la pared celular y otros componentes de la pared celular, las estructuras ramificada frente ano ramificada de los lípidos de la membrana y la presencia / ausencia de histonas y proteínas similares a las histonas y se usan para asignar un organismo a Eubacterias o Archaea.

Ejemplos de Eubacterias incluyen *Escherichia coli*, *Thermus thermophilus* y *Bacillus stearothermophilus*. Ejemplos de Archaea incluyen *Methanococcus jannaschii* (Mj), *Methanosarcina mazei* (Mm), *Methanobacterium thermoautotrophicum* (Mt), *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Halobacterium* tales como *Haloferax volcanii* y *Halobacterium* especies NRC-1, *Archaeoglobus fulgidus* (Af), *Pyrococcus furiosus* (Pf), *Pyrococcus horikoshii* (Ph), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Aeuropyrum pernix* (Ap), *Thermoplasma acidophilum* y *Thermoplasma volcanium*

Procedente de: Como se usa en el presente documento, la expresión de "procedente de" se refiere a un componente que se aísla o se prepara usando una molécula u organismo especificado, o información de la molécula o el organismo especificado. Por ejemplo, un polipéptido que procede de un segundo polipéptido puede incluir una secuencia de aminoácidos que es idéntica o sustancialmente similar a la secuencia de aminoácidos del segundo polipéptido. En el caso de los polipéptidos, las especies derivadas se pueden obtener mediante, por ejemplo, mutagénesis natural, mutagénesis directa artificial o mutagénesis aleatoria artificial. La mutagénesis usada para obtener polipéptidos puede estar dirigida intencionadamente o ser aleatoria intencionadamente, o una mezcla de las dos. La mutagénesis de un polipéptido para crear un polipéptido diferente derivado del primero puede ser un acontecimiento aleatorio (p. ej., causado por infidelidad de la polimerasa) y la identificación del polipéptido derivado puede realizarse mediante procedimientos de detección selectiva adecuados, por ejemplo como los que se tratan en el presente documento. La mutagénesis de un polipéptido normalmente implica manipulación del polinucleótido que codifica el polipéptido.

De un modo similar, la expresión "procedente de" se puede aplicar a los polinucleótidos. Un polipéptido procedente de una fuente de polinucleótidos puede incluir una secuencia de nucleótidos que es idéntica o sustancialmente similar a la secuencia de nucleótidos fuente. En el caso de los polinucleótidos, las especies derivadas se pueden obtener mediante, por ejemplo, mutagénesis natural, mutagénesis directa artificial o mutagénesis aleatoria artificial. La mutagénesis usada para obtener polinucleótidos puede estar dirigida intencionadamente o ser aleatoria

intencionadamente, o una mezcla de las dos. En algunos aspectos, un polinucleótido derivado se genera colocando un polinucleótido fuente en un contexto heterólogo, es decir en un contexto que es diferente de su contexto nativo o endógeno. Por ejemplo, un promotor génico se puede obtener de un promotor génico endógeno eliminando dicho dominio del promotor endógeno y colocándolo en una combinación operable con diferentes secuencias de nucleótidos con las que normalmente no está asociado.

Marcador de selección o identificación positiva: Como se usa en el presente documento, la expresión “marcador de selección o identificación positiva” hace referencia a un marcador que, cuando está presente, por ejemplo está expresado, activado o similares, tiene como resultado la identificación de una célula, que comprende el rasgo, por ejemplo una célula con el marcador de selección positiva, de aquéllos sin el rasgo.

Marcador de selección o identificación negativa: Como se usa en el presente documento, la expresión “marcador de selección o identificación negativa” se refiere a un marcador que, cuando está presente, por ejemplo está expresado, activado o similares, permite la identificación de una célula, que no comprende una propiedad o rasgo seleccionado (p. ej., en comparación con una célula que posee la propiedad o rasgo).

Agente de selección o identificación: Como se usa en el presente documento, la expresión “agente de selección o identificación” hace referencia a un agente que, cuando está presente, permite la selección/identificación de determinados componentes de una población. Por ejemplo, un agente de selección o identificación puede ser, entre otros, por ejemplo, un nutriente, un antibiótico, una longitud de onda de la luz, un anticuerpo, un polinucleótido expresado o similares. El agente de selección se puede variar en, por ejemplo, concentración, intensidad etc.

Indicador: Como se usa en el presente documento, el término "indicador" o términos equivalentes se refiere en sentido general a cualquier componente que se puede detectar fácilmente en un sistema en estudio, en el que la detección del indicador se correlaciona con la presencia o ausencia de algunas otras moléculas o propiedades, o se puede usar para identificar, seleccionar y/o detectar dianas en un sistema de interés. La elección del indicador más adecuado para usar para una aplicación concreta depende del uso al que está destinado y otras variables conocidas para una persona familiarizada en la técnica. En algunos aspectos, un indicador es un gen indicador.

En la técnica se conocen una amplia variedad de moléculas y genes indicadores, cada indicador tiene un ensayo concreto para la detección de dicho indicador. Algunos ensayos de detección de indicador pueden ser ensayos enzimáticos, mientras que otros ensayos pueden tener una naturaleza inmunológica (p. ej., ELISA o análisis inmunohistoquímico) o, por ejemplo, colorimétrico. Todavía más, un indicador puede incluir una proteína (p. ej., una enzima que confiere resistencia o sensibilidad a antibiótico (p. ej., lactamasa, cloranfenicol acetiltransferasa (CAT), y similares), un marcador fluorescente (p. ej., una proteína fluorescentes verde tales como GFP, YFP, EGFP, RFP, etc.), un marcador luminiscente (p. ej., una proteína luciferasa de luciérnaga), un marcador de identificar según la afinidad, una actividad enzimática tal como *lacZ* (β -galactosidasa) u otros genes marcadores seleccionables positivos o negativos tales como ADH (alcohol deshidrogenasa), *his3*, *ura3*, *leu2*, *lys2* o similares.

Breve descripción de las figuras

La **FIG. 1** proporciona las estructuras y las correspondientes denominaciones de cuatro aminoácidos no naturales, que son *p*-benzoil-L-fenilalanina (Bpa), *para*-acetil-L-fenilalanina (*pAcPhe*), *para*-azido-L-fenilalanina (*pAzPhe*) y *para*-yodo-L-fenilalanina (*pIPhe*).

La **FIG. 2** proporciona un histograma que muestra las eficiencias de supresión de plásmidos (respecto a la β -galactosidasa de tipo silvestre) con el promotor y el terminador *proK* para el gen *MjtRNA-Tyr*(CUA), la mutación D286R en el gen *BpaRS*, una forma mutada del promotor *glsS* para el gen *BpaRS* y/o múltiples copias del gen del ARNt. Las barras de error indican la desviación estándar y $n = 3$.

La **FIG. 3** proporciona una imagen de quimioluminiscencia de un análisis de transferencia de tipo Northern del supresor ámbar *MjtRNA-Tyr*(CUA) expresado de los plásmidos de supresión indicados.

La **FIG. 4** proporciona una imagen de quimioluminiscencia tras un análisis de transferencia de tipo western de *BpaRS* expresado bajo el control del promotor de *glnS* de tipo silvestre y la forma mutada del promotor *glnS* La transferencia usó un conjugado anticuerpo-HRP anti-His (extremo C) (Invitrogen).

La **FIG. 5** muestra UN mapa plasmídico de *pSup-BpaRS-6TRN*. Otros genes de sintetasa se subclonaron a partir de sus correspondientes plásmidos de *pBK* en los sitios *NdeI/PstI* de este plásmido.

La **FIG. 6** muestra las eficiencias de supresión del sistema nuevo para la incorporación de *Bpa*, *pAcPhe*, *pAzPhe* e *pIPhe*. Las barras de error indican la desviación estándar y $n = 3$.

La **FIG. 7** proporciona una imagen de quimioluminiscencia tras transferencia de tipo western de una mioglobina mutantes que contiene BPa en lugar de Ser-4 expresada en ausencia o presencia de *Bpa*. La transferencia usó un conjugado anticuerpo-HRP anti-His (extremo C) (Invitrogen).

La FIG. 8 proporciona varias secuencias de polipéptidos y polinucleótidos que encuentra uso en la invención.

Descripción detallada de la invención

5 La invención proporciona composiciones, células huésped, sistemas de traducción y procedimientos como se define en las reivindicaciones.

10 La invención proporciona nuevas características del vector de expresión que tienen como resultado una mejora significativa de la eficiencia en la incorporación de aminoácidos no naturales en proteínas en eubacterias (p. ej., *E. coli*), y tiene como resultado una expresión de alto rendimiento de las proteínas mutantes que contienen los aminoácidos no naturales en sitios específicos diseñados genéticamente mediante codones selectores. La mejor eficiencia en la incorporación de aminoácidos no naturales en una proteína de interés se debe, posiblemente (al menos en parte) a la mejora de la expresión de la aminoacil-ARNt sintetasa ortogonal y del ARNt supresor, aunque no es necesario entender el mecanismo de la mejora de la eficiencia para fabricar o usar la invención.

15 Las características del nuevo vector de expresión de la invención son ampliamente compatibles con diversas estructuras del vector de expresión de *E. coli* y cepas de *E. coli*, y también se adaptan fácilmente para la expresión de otras proteínas o ARNt de interés, además de la expresión de aminoacil-ARNt sintetasas ortogonales o ARNt supresores ortogonales.

20 En algunos aspectos, las características del nuevo vector de expresión proporcionado por la invención se usan de forma independiente en diferentes plásmidos. En otros aspectos, una única característica nueva o combinación de características se usan en una pluralidad de plásmidos. En otras realizaciones más se usa una pluralidad de estas características en combinación en el mismo plásmido. La invención proporciona una serie de mejoras en los sistemas del vector de expresión bacteriana que se pueden usar para mejorar la expresión de proteínas mutantes que comprenden uno o más aminoácidos no naturales y, en algunos casos, se pueden usar más ampliamente para mejorar la expresión de cualquier polipéptido concreto o ARNt de interés.

30 La invención proporciona, por ejemplo, las mejoras siguientes en los sistemas de vectores de expresión bacteriana:

(A) En el presente documento se describen vectores de expresión en los que el gen de la aminoacil-ARNt sintetasa ortogonal y el gen del ARNt supresor ortogonal son portados en el mismo plásmido. Eso simplifica la expresión de estos componentes ortogonales, como se ha citado anteriormente, estos dos componentes eran portados cada uno de ellos en diferentes plásmidos.

35 (B) La invención proporciona mejores secuencias promotoras y terminadoras como se define en las reivindicaciones.

(C) También se describe en el presente documento mejores operones policistrónicos recombinantes para la expresión de genes de ARNt, donde dos cualquiera de los genes de ARNt en el operón están separados por una secuencia adaptadora heteróloga derivada de un adaptador de un operón policistrónico de ARNt natural, por ejemplo el adaptador que se produce de origen natural entre los genes *valU* y *valX* de *E. coli* o, como alternativa, por ejemplo entre los genes *ileT* y *alaT*.

40 (D) En el presente documento también se describe una nueva secuencia promotora derivada del promotor *glnS* de *E. coli* para la menor expresión de un marco de lectura abierto, por ejemplo un marco de lectura abierto que codifica una aminoacil-ARNt sintetasa ortogonal.

Sistemas de vectores para la coexpresión de genes O-ARNt y O-RS

50 En el presente documento se describen vectores de expresión en los que el gen de la aminoacil-ARNt sintetasa ortogonal y el gen del ARNt supresor ortogonal son portados en el mismo plásmido. Esta característica es una mejora sobre la técnica, en la que anteriormente era necesario co-transformar una célula huésped con dos vectores de expresión distintos portadores de forma independiente los genes de O-ARNt y O-RS.

55 Como se describe en los ejemplos se construye una serie de vectores de expresión relacionados que son adecuados para la co-expresión de especies de O-ARNt y O-RS. Estos plásmidos incluyen:

- 60 pYR-BpaRS1
- pYR-BpaRS5
- pYR-BpaRS5(D286R)
- pYR-BpaRS-TRN
- pYR-BpaRS-TRN(D286R)
- pYR-BpaRS-3TRN(D286R)
- pYR-BpaRS-6TRN(D286R)
- 65 pSup-BpaRS-6TRN(D286R)
- pSup-pAcPheRS-6TRN

pSup-pAzPheRS-6TRN
pSup-pIPheRS-6TRN

5 Cada uno de estos plásmidos es como se describe en el presente documento. No obstante, no se pretende que la invención se limite a estos plásmidos, como un experto reconocerá, la construcción de variantes de estos plásmidos entra dentro de la divulgación.

10 Por ejemplo, no se pretende limitar ningún plásmido de la invención a la expresión de cualquier especie de O-ARNt u O-RS para producir una proteína que comprende cualquier aminoácido no natural. Los ejemplos proporcionados en el presente documento describen el uso con éxito de un *Mjt*RNA-Tyr(CUA) (SEC ID N°: 1) y cuatro especies de O-RS diferentes que tienen una especificidad de la carga del ARNt por *p*-benzoil-L-fenilalanina (Bpa), *para*-acetil-L-fenilalanina (*p*AcPhe), *para*-azido-L-fenilalanina (*p*AzPhe) y *para*-yodo-L-fenilalanina (*p*IPhe) (véanse las **FIG. 1**, y **FIG. 8**, SEC ID N°: 4, 6, 8 y 10).

15 Estos ejemplos de trabajo anteriores sirven para ilustrar la aplicabilidad más amplia de la invención que se va a usar con otras especies de O-ARNt y O-RS. De hecho, la invención encuentra uso en la expresión de cualquier O-ARNt o cualquier O-RS de interés y, en particular, componentes ortogonales de la traducción que operan óptimamente en células eubacterianas. En algunos aspectos, la invención encuentra un uso concreto con especies de O-RS que proceden de aminoacil-ATNt sintetasas de Archaea (*ep. ej.*, *Methanococcus jannaschii*) de origen natural o especies de O-ARNt derivadas del ARNt de Archaea. La amplia variedad de especies de O-ARNt y O-RS que encuentran uso con la invención se conocen en la técnica y se describen en numerosas fuentes. Véase, por ejemplo, los números de publicación internacional WO 2002/086075, titulada "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS;" el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACID;" el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE;" el documento WO 2005/019415, presentado el 7 de julio de 2004; el documento WO 2005/007870, presentado el 7 de julio de 2004; el documento WO 2005/00762A., presentado el 7 de julio de 2004 y el documento WO 2006/110182, presentado el 27 de octubre de 2005, titulado "ORTHOGONAL TRANSLATION COMPONENTS FOR THE IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS. Para una discusión adicional de sistemas de traducción ortogonales que incorporan aminoácidos no naturales y procedimientos para su producción y uso, véase también Wang y Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34,66 (2005), Xie y Schultz, "An Expanding Genetic Code," *Methods* 36(3):227-238 (2005); Xie y Schultz, "Adding Amino Acids to the Genetic Repertoire," *Curr. Opinion in Chemical Biology* 9(6):548-554; y Wang et al., "Expanding the Genetic Code," *Annu. Rev. Biophys. Biomol. Struct.*, epub Jan 13, 2006.

35 La técnica anterior (p. ej., la técnica citada en el presente documento) también proporciona guías para la construcción y uso de numerosas variantes (p. ej., variantes conservadoras) y fragmentos de especies de O-RS y O-ARNt conocidas. Estas variantes y fragmentos también encuentran uso con los vectores de expresión de la invención. La técnica también proporciona la identificación y construcción de nuevas especies de O-RS y O-ARNt, que también encuentran uso en la invención.

40 Construcciones de plásmidos y células huésped eubacterianas

Como se describe y usa en el Ejemplo 6, los plásmidos proporcionados en la memoria descriptiva se basan en una estructura del vector pACYC184. No obstante, no se pretende que los plásmidos de la invención estén limitados al uso de dicho vector estructural concreto. Un experto en la técnica reconoce que uno cualquiera de diversos plásmidos (incluyendo otros plásmidos pública o comercialmente disponibles) se puede usar para construir los plásmidos de la invención. Por ejemplo, los vectores plasmídicos pACYC177 y pRARE2 (Novagen; véase en *Novations*, N° 12, junio de 2001) también se pueden usar junto con la invención. En algunos aspectos, cualquier plásmido portador de un origen de replicación compatible (p. ej., el origen de replicación p15A) y al menos un marcador de selección se pueden usar junto con la invención. Los derivados del plásmidos pSC101 también se pueden usar de acuerdo con la invención.

También como se describe en el ejemplo 6, los plásmidos proporcionados por la invención se usaron para transformar *E. coli* electrocompetente One Shot[®] TOP-10 Invitrogen[™]). No obstante, no se pretende que la cepa de eubacteria usada como célula huésped para producir proteínas que comprenden uno o más aminoácidos no naturales esté limitada al uso de dicha célula huésped concreta. Un experto en la técnica reconoce que una cualquiera de diversas células huésped (incluyendo otros plásmidos pública o comercialmente disponibles así como cepas producidas por el usuario) se puede usar fácilmente para producir proteínas que comprenden aminoácidos no naturales. Por ejemplo, otras células huésped, tal como la cepa de *E. coli* DH10B[™] (Invitrogen[™]), Electrocomp[™] GeneHogs[®] (Invitrogen[™]), BL21 One Sbot[®] (Invitrogen[™]), y BL21(DE3) One Shot[®] (Invitrogen[™]). De hecho, cualquier cepa de *E. coli* sin ningún gen supresor de ARNt endógeno es una célula huésped adecuada. Otras especies de eubacteria además de *E. coli* también encuentran uso en la invención. Por ejemplo se contempla que las cepas de *Bacillus subtilis* también se pueden usar como células huésped para los vectores de la invención.

65

Secuencias promotoras y terminadoras mejoradas para la expresión de O-ARNt

Con el fin de mejorar la eficacia de la supresión del sistema de traducción ortogonal se construyó un nuevo operón del ARNt supresor ámbar con un promotor y terminador del ARNt de *E. coli* natural. Una visión de los genes de ARNt de *E. coli* reveló que los ARNt de prolina de *E. coli* tienen el mismo par C1-G72 que los ARNt de Archaea; este par de bases es un determinante principal de la identidad para el reconocimiento selectivo de *Mj*tRNA-Tyr(CUA) por *Mj*TyrRS en *E. coli* (Wang y Schultz, Chem. Biol., 8:883-890 (2001)).

A la luz de esta observación se construyó un gen del ARNt supresor ámbar sintético de un modo tal que un gen de O-ARNt heterólogo sustituya al gen de *proK* de *E. coli* de la misma longitud (77 nucleótidos) en el operón *proK* monocistrónico. Un vector de expresión mejorado (pYR-BpaRS5) se generó sustituyendo el operón del ARNt supresor original en pYR-BpaRS 1 con el gen *Mj*tRNA-Tyr(CUA) bajo el control del promotor *proK* (SEC ID N°: 32) y el terminador *proK* (SEC ID N°: 33). Este constructo de expresión mostró un incremento por dos en la expresión del O-ARNt (véase la FIG. 3), y tuvo como resultado una mejora significativa de la eficacia de la supresión (véase la FIG. 2), ambos en relación con las actividades del vector pYR-BpaRS1.

Por tanto, la invención proporciona mejores vectores de expresión para la expresión de un ARNt de interés, en el que la expresión de un ARNt policistrónico está dirigida por secuencias de nucleótidos promotoras y terminadoras derivadas del gen *proK* del ARNt de prolina de *E. coli*.

Como se describe en el Ejemplo 1, el ARNt concreto usado para demostrar este vector de expresión mejorado fue un ARNt ortogonal (O-ARNt), más específicamente *Mj*tRNA-Tyr(CUA). No obstante, no se pretende que la mejor eficiencia de la expresión del ARNt esté limitada a *Mj*tRNA-Tyr(CUA), no limitada a un O-ARNt. De hecho, esta característica de la invención se puede usar para mejorar la expresión de cualquier ARNt de interés.

En *E. coli*, tres especies de ARNt se cargan con prolina durante la traducción. Además del gen de ARNt *proK*, *E. coli* también usa dos genes de prolil-ARNt adicionales. Estos son *proL* y *proM*. A la luz de su estructura similar con la del locus de *proK*, se contempla que la secuencia promotora de *proL* (SEC ID N°: 34) y las secuencias terminadoras de *proL* y *proM* (SEC ID N°: 35 y 36, respectivamente) también encuentran uso en la construcción de mejores vectores de expresión de la invención. También es una característica de la invención las combinaciones de promotores y terminadores de diferentes genes de prolil-ARNt de *E. coli* también se pueden usar para conseguir una expresión mejorada. Por ejemplo, la secuencia promotora de *proK* (SEC ID N°: 32) se puede usar junto con la secuencia terminadora de *proL* (SEC ID N°: 35).

Estructuras de operón policistrónico mejoradas

La invención proporciona mejores operones policistrónicos recombinantes para la expresión de genes de ARNt. Estos operones policistrónicos comprenden múltiples copias (p. ej., tres copias) de genes de ARNt de interés, donde las secuencias de ARNt están separadas por una secuencia adaptadora heteróloga derivada de un adaptador de un operón policistrónico de ARNt natural, por ejemplo el adaptador que se produce de forma natural entre los genes *valU* y *valX* de *E. coli* (SEC ID N°: 14) o, como alternativa, por ejemplo entre los genes de ARNt *ileT* y *alaT* de *E. coli* (SEC ID N°: 15).

Por tanto, la invención proporciona mejores vectores de expresión para la expresión de operones de ARNt policistrónicos, en los que el operón comprende al menos un adaptador de ARNt heterólogo que separa al menos dos secuencias de ARNt expresadas. En algunas realizaciones, como se describe en el Ejemplo 3, se usan múltiples adaptadores de ARNt para separar tres o más secuencias de ARNt expresadas en el operón. En este caso, el adaptador de ARNt usado entre cada par de ARNt expresado puede ser diferente (como en el Ejemplo 3) o puede ser el mismo adaptador entre cada gen de ARNt.

No se pretende que la invención se limite al uso del adaptador de los genes *valU* y *valX* de *E. coli* (SEC ID N°: 14) o, e adaptador del gen del ARNt *ileT* y *alaT* de *E. coli* (SEC ID N°: 15). De hecho, los adaptadores de ARNt naturales adicionales también encuentran uso en la invención. Por ejemplo, cada uno de los siguientes adaptadores localizados entre los genes nativos de ARNt de *E. coli* enumerados más adelante encuentra uso con la invención, de modo que el adaptador que se usa en el sistema recombinante es heterólogo de cualquiera de las secuencias de ARNt expresadas en el operón recombinante. Estos adaptadores de ARNt útiles incluyen:

Adaptador del ARNt de <i>E. coli</i> nativo	Secuencia	SEC ID N°:
<i>valU</i> y <i>valX</i>	ACTACTTTTATGTAGTCTCCGCCGTGTAGCAAGAAATTGAGAAGT	14
<i>ileT</i> y <i>alaT</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT	15
<i>serV</i> y <i>argV</i>	TTT	16
<i>valV</i> y <i>valW</i>	TCCT	17
<i>glyT</i> y <i>thrT</i>	AGATGT .	18

<i>metT</i> y <i>leuW</i>	TCTTTTTT	19
<i>glnW</i> y <i>metU</i>	TCGAAGAAACAATCT	20
<i>hisR</i> y <i>leuT</i>	TTATTAGAAGTTGTGACAAT	21
<i>glnU</i> y <i>glnW</i>	TCTTCTTCGAGTAAGCGGTTACCCGCCGGTTAT	22
<i>leuP</i> y <i>leuV</i>	AACGAGGCGATATCAAAAAAGTAAGATGACTGT	23
<i>glnV</i> y <i>glnX</i>	ATTTATTCAAGACGCTTACCTTGTAAAGTGCACCCAGT	24
<i>alaW</i> y <i>alaX</i>	AATTTTGCACCCAGCAAACCTTGGTACGTAAACGCATCGT	25
<i>ileU</i> y <i>alaU</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT	26
<i>ileV</i> y <i>alaV</i>	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT	27
<i>metU</i> y <i>glnV</i>	AATTCTGAATGTATCGAATATGTTCCGGCAAATTCAAAACCA ATTTGT	28
<i>glyW</i> y <i>cysT</i>	GTTTAAAAGACATCGGCGTCAAGCGGATGTCTGGCTGAAA GGCCTGAAGAATTT	29
<i>argX</i> y <i>hisR</i>	TTTAGTCCCGGCGCTTGAGCTGCGGTGGTAGTAATACCGC GTAACAAGATTTGTAGT	30
<i>argY</i> y <i>argZ</i>	TCTCTTACTTGATATGGCTTTAGTAGCGGTATCAATATCAGCAG TAAAATAAATTTCCCGAT	31

En algunas realizaciones, los adaptadores de ARNt preferidos que encuentran uso con la invención contienen uno los dos nucleótidos T(-1) y A(77). Se ha demostrado que estas dos posiciones de nucleótidos en los adaptadores de ARNt son óptimas para el procesamiento 5' y 3' eficiente de los precursores de ARNt cuando están en su contexto nativo (es decir, endógeno). Véase, por ejemplo, Li y Deutscher, "Maturation pathways for E. coli tRNA precursors: A random multienzyme process in vivo," *Cell* 86:503-512 (1996); y Zahler et al., "Recognition of the 5' leader of pre-tRNA substrates by the active site of ribonuclease P," *RNA* 9:734-745 (2003). En otras realizaciones, los adaptadores del ARNt que encuentran uso con la invención comprenden sitios de restricción (naturales o modificados genéticamente).

En algunas realizaciones, la invención proporciona construcciones que comprenden una pluralidad del mismo operón policistrónico, opcionalmente en tándem. Por tanto, si un único operón policistrónico comprende tres copias de una secuencia de nucleótidos expresable (tal como un gen de ARNt), dos de los operones tendrán como resultado la expresión de un total de seis secuencias génicas de ARNt. Este tipo de configuración en clúster génico se demuestra en el Ejemplo 3 y la **FIG. 5**.

El operón policistrónico recombinante mejorado descrito en el Ejemplo 3 expresa el ARNt ortogonal *M*tRNA-Tyr(CUA). No obstante, no se pretende que la invención esté limitada a la expresión de *M*tRNA-Tyr(CUA). Tampoco está la invención limitada a la expresión de especies de ARNt ortogonal. De hecho, los operones policistrónicos mejorados de la invención se pueden usar para expresar cualquier especie de ARNt deseada.

Promotor de *glnS* de E. coli mejorado para la expresión de polipéptidos

La divulgación proporciona una nueva secuencia promotora derivada del promotor *glnS* de E. coli para una expresión mejorada de un marco de lectura abierto. Como se describe en el ejemplo 2, el promotor de *glnS* mutante (SEC ID N°: 12) descrito en Plumbridge y Söll (*Biochimie* 69:539-541 (1987)) se subclonó en un vector de expresión de la invención. La secuenciación de la región promotora *glnS* subclonada reveló la introducción inadvertida de una deleción en la secuencia del promotor (además de la sustitución descrita en Plumbridge y Söll). Esta nueva variante del promotor *glnS* modificado se denominó *glnS*-TNR (proporcionada en la SEC ID N°: 13). Sorprendentemente, esta mutación tuvo como resultado una mejora de la eficiencia de la traducción del sistema en comparación con la actividad promotora de *glnS* de tipo silvestre determinada mediante transferencia de tipo western (véase la **FIG. 4**).

Como se describe en los ejemplos 2 y 5, el promotor *glnS*-TNR se usó para expresar las sintetasas ortogonales BpaRS, pAcPheRS, pAzPheRS y pIPheRS. No obstante, no se pretende que la invención esté limitada a la expresión de ninguna aminoacil-ARNt sintetasa ortogonal concreta o esté limitada a ninguna aminoacil-ARNt sintetasa en general. Este mejor promotor encuentra un uso más amplio en la expresión bacteriana de cualquier marco de lectura abierto del polipéptido deseado.

Tecnología de ARNt/aminoacil-ARNt sintetasa ortogonal

Una comprensión de las nuevas composiciones y procedimientos de la presente invención se facilita mediante una comprensión de las actividades asociadas con los pares de ARNt ortogonal y aminoacil-ARNt sintetasa ortogonal. Con el fin de añadir aminoácidos no naturales adicionales al código genético, se necesitan nuevos pares ortogonales

que comprenden una aminoacil-ARNt sintetasa y un ARNt adecuado que puedan funcionar con eficiencia en la maquinaria traduccional del huésped, pero que son "ortogonales" al sistema de traducción en cuestión, lo que significa que funciona de forma independiente de las sintetasas y ARNt endógenos al sistema de traducción. Las características deseadas del par ortólogo incluyen ARNt que descodifican o reconocen únicamente un codón específico, por ejemplo un codón selector, que no se descodifica mediante ningún ARNt endógeno y aminoacil-ARNt sintetatas que aminoacilan, preferentemente, (o "cargan") su ARNt afín con solo un aminoácido no natural específico. El O-ARNt tampoco se aminoacila normalmente mediante sintetasas endógenas. Por ejemplo, en *E. coli*, un par ortogonal incluirá una aminoacil-ARNt sintetasa que no sufre reacción cruzada con ninguno de los ARNt endógenos, por ejemplo que hay 40 en *E. coli*, y un ARNt ortogonal que no se aminoacila por ninguna de las sintetasas endógenas, por ejemplo de las cuales hay 21 en *E. coli*. Hasta la fecha. Una amplia variedad de aminoácidos no naturales estructuralmente diversos se ha incorporado en las proteínas que usan tecnología de traducción ortogonal, como se sabe en la técnica.

La capacidad para incorporar un aminoácido no natural específico de sitio en un polipéptido puede facilitar el estudio de proteínas permitiendo la modificación postraduccional altamente selectiva de dichas proteínas, además de permitir la ingeniería de las proteínas con propiedades nuevas. Por ejemplo, la expresión de proteínas que contienen uno o más aminoácidos no naturales puede facilitar el estudio de proteínas mediante marcaje específico, alterar la función catalítica de las enzimas, mejorar la actividad biológica o reducir la reactividad cruzada con un sustrato, reticular una proteína con otras proteínas, moléculas pequeñas o biomoléculas, reducir o eliminar la degradación de proteínas, mejorar la semivida de las proteínas in vivo (p. ej., mediante pegilación u otras modificaciones de sitios reactivos introducidos) etc.

Los sistemas de traducción ortogonales que son adecuadas para fabricar proteínas que comprenden uno o más aminoácidos no naturales se conocen en la técnica, como son los procedimientos generales para producir sistemas de traducción ortogonales. Por ejemplo, véase, los números de publicación internacional WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS;" el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACID;" el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE;" el documento WO 2005/019415, presentado el 7 de julio de 2004; el documento WO 2005/007870, presentado el 7 de julio de 2004; el documento WO 2005/00762A., presentado el 7 de julio de 2004 y el documento WO 2006/110182, presentado el 27 de octubre de 2005, titulado "ORTHOGONAL TRANSLATION COMPONENTS FOR THE IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS. Para una discusión de sistemas de traducción ortogonales que incorporan aminoácidos no naturales y procedimientos para su producción y uso, véase también Wang y Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005), Xie y Schultz, "An Expanding Genetic Code," *Methods* 36(3):227-238 (2005); Xie y Schultz, "Adding Amino Acids to the Genetic Repertoire," *Curr. Opin. in Chemical Biology* 9(6):548-554; y Wang et al., "Expanding the Genetic Code," *Annu. Rev. Biophys. Biomol. Struct.*, 35:225-249 (2006); y Xie y Schultz, "A chemical toolkit for proteins - an expanded genetic code," *Nat. Rev. Mol. Cell Biol.*, 7(10):775-782 (2006; epub Aug 23, 2006)..

Dichos sistemas de traducción comprenden, generalmente, células (que pueden ser células procariotas tales como *E. coli* o células eucariotas como las levaduras) que incluyen un ARNt ortogonal (O-ARNt), una aminoacil ARNt sintetasa ortogonal (O-RS) y un aminoácido no natural, de modo que la O-RS aminoacila el O-ARNt con el aminoácido natural. Un par ortogonal de la invención puede incluir un O-ARNt y, por ejemplo un ARNt supresor, un ARNt de desplazamiento de marco o similar, y una O-RS afín.

En general, cuando un par ortogonal reconoce un codón selector y carga un aminoácido en respuesta al codón selector, se dice que el par ortogonal "suprime" el codón selector. Es decir, un codón selector que no es reconocido por la maquinaria endógena del sistema de traducción (p. ej., de las células) no se carga de forma habitual, que tiene como resultado el bloqueo de la producción de un polipéptido que, de otro modo, se traduciría del ácido nucleico. En un sistema de par ortogonal, la O-RS aminoacila el O-ARNt con un aminoácido no natural específico. El O-ARNt cargado reconoce el codón selector y suprime el bloque traduccional causado por el codón selector. La célula usa el par O-ARNt/O-RS para incorporar el aminoácido no natural en una cadena en crecimiento, por ejemplo mediante un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, de modo que el polinucleótido comprende un codón selector que es reconocido por el O-ARNt. En determinados aspectos deseables, la célula puede incluir un par adicional de O-ARNt/O-RS, donde el O-ARNt adicional es cargado por la O-RS adicional con un aminoácido no natural diferente. Por ejemplo, uno de los O-ARNt puede reconocer un codón de cuatro bases y el otro puede reconocer un codón de terminación. De forma alternativa, múltiples codones de terminación diferentes o múltiples codones de cuatro bases diferentes pueden reconocer de forma específica diferentes codones selectores

En determinadas realizaciones, los sistemas comprenden una célula tal como una célula *E. coli* o una célula de levadura que incluye un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), un aminoácido no natural y un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende el codón selector que es reconocido por el O-ARNt. El sistema de traducción también puede ser un sistema sin células, por ejemplo, cualquiera de diversos sistemas de transcripción/traducción "in vitro" disponibles en el mercado en combinación con un par de O-ARNt/O-RS y un aminoácido no natural como se describe

en el presente documento.

Como se ha indicado, en algunas realizaciones existen múltiples pares de O-ARNt/O-RS en una célula u otro sistema de traducción, que permite la incorporación de más de un aminoácido no natural en un polipéptido. Por ejemplo, la célula puede incluir adicionalmente un par adicional diferente de O-ARNt/O-RS y un segundo aminoácido no natural, donde este O-ARNt adicional reconoce un segundo codón selector y esta O-RS adicional aminoacila de forma preferente el O-ARNt con el segundo aminoácido no natural. Por ejemplo, una célula que incluye un par de O-ARNt/O-RS (en el que el O-ARNt reconoce, por ejemplo, un codón selector ámbar) puede comprender adicionalmente un segundo par ortogonal, en el que el segundo O-ARNt reconoce un codón selector diferente, por ejemplo, un codón ópalo, de cuatro bases o similares. De forma deseable, los diferentes pares ortogonales proceden de diferentes fuentes, lo cual puede facilitar el reconocimiento de diferentes codones selectores.

El O-ARNt y/o la O-RS pueden ser de origen natural, o pueden obtenerse, por ejemplo, mediante mutación de un ARNt y/o RS de origen natural, por ejemplo, generando bibliotecas de ARNt y/o bibliotecas de RS, a partir de cualquiera de diversos organismos y/o usando cualquiera de las diversas estrategias de mutación disponibles. Por ejemplo, una estrategia para producir un par ortogonal de ARNt/aminoacil-ARNt sintetasa implica importar un par heterólogo (a la célula huésped) de ARNt/sintetasa de, por ejemplo, una fuente diferente de la célula huésped, o fuentes múltiples, en la célula huésped. Las propiedades del candidato de sintetasa heteróloga incluyen, por ejemplo, que no carga ningún ARNt de célula huésped, y las propiedades del candidato de ARNt heterólogo incluyen, por ejemplo, que no se aminoacila por ninguna sintetasa de célula huésped. Además, el ARNt heterólogo es ortogonal a todas las sintetasas de la célula huésped.

Una segunda estrategia para generar un par ortogonal implica generar bibliotecas mutantes a partir de las cuales se identifica y/o selecciona un O-ARNt o una O-RS. Estas estrategias también se pueden combinar.

ARNt ortogonal (O-ARNt)

Un ARNt ortogonal (O-ARNt) de la invención de forma deseable media en la incorporación de un aminoácido no natural, en una proteína que está codificada por un polinucleótido que comprende un codón selector que se reconoce por el O-ARNt, por ejemplo, *in vivo* o *in vitro*, con una eficiencia de supresión elevada. La eficiencia de supresión se puede determinar con cualquiera de una serie de ensayos conocidos en la técnica. Por ejemplo, se puede usar un ensayo indicador de β -galactosidasa, por ejemplo un plásmido *lacZ* derivado (cuando el constructo tiene un codón selector en la secuencia de ácido nucleico de *lacZ*) se introduce en las células de un organismo adecuado (p. ej., un organismo en el que se pueden usar los componentes ortogonales) junto con el plásmido que comprende un O-ARNt de la invención. También se puede introducir una sintetasa afín (bien como polipéptido o como polinucleótido que codifica la sintetasa afín cuando se expresa). Las células se cultivan en medio hasta una densidad deseada, por ejemplo hasta una DO_{600} de aproximadamente 0,5 y se realizan ensayos de β -galactosidasa, por ejemplo usando el kit de ensayo con β -galactosidasa BetaFluor™ (Novagen). El porcentaje de supresión se puede calcular como el porcentaje de actividad para una muestra relativa a un control comparable, por ejemplo el valor observado del constructo *lacZ* derivado, donde el constructo tiene un codón sentido correspondiente a una posición deseada en lugar de un codón selector.

Los O-ARNt también pueden proceder de variaciones conservadoras de O-ARNt conocidos. Por ejemplo, las variaciones conservadoras de O-ARNt incluyen las moléculas que funcionan como los O-ARNt particulares, por ejemplo, como en la lista de secuencias del presente documento y que conservan la estructura en forma del L del ARNt mediante auto-complementariedad apropiada, pero que no tienen una secuencia idéntica a los mismos, por ejemplo, en la lista de secuencias, las figuras o los ejemplos en este documento (y, de forma deseable, son diferentes de las moléculas de ARNt de tipo silvestre).

La composición que comprende un O-ARNt puede incluir adicionalmente una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que la O-RS aminoacila de forma preferente el O-ARNt con un aminoácido no natural. En determinadas realizaciones, una composición que incluye un O-ARNt puede incluir adicionalmente un sistema de traducción (por ejemplo, *in vitro* o *in vivo*). También puede estar presente en la célula un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende un codón selector que es reconocido por el O-ARNt, o una combinación de uno o más de los mismos.

Se conocen los procedimientos para producir un ARNt ortogonal (O-ARNt). En determinadas realizaciones de la invención, los O-ARNt se pueden producir generando una biblioteca de mutantes. La biblioteca de ARNt mutantes se puede generar usando diversas técnicas de mutagénesis conocidas en la materia. Por ejemplo, los ARNt mutantes se pueden generar mediante mutaciones con especificidad de sitio, mutaciones puntuales aleatorias, recombinación homóloga, barajado de ADN u otros procedimientos de mutagénesis recursivos, construcción química o cualquier combinación de los mismos.

Se pueden introducir mutaciones adicionales en una(s) posición (posiciones) específica(s), por ejemplo, en una(s) posición (posiciones) no conservadora(s), o en una posición conservadora, en una(s) posición (posiciones) aleatoria(s) o una combinación de ambas en un bucle deseado o región de un ARNt, por ejemplo, un bucle

anticodón, el tallo aceptor, el brazo o bucle D, el bucle variable, el brazo o bucle TPC, otras regiones de la molécula de ARNt o una combinación de los mismos. Normalmente, las mutaciones en un ARNt incluyen mutar el bucle anticodón de cada miembro de la biblioteca de ARNt mutantes para permitir el reconocimiento de un codón selector. El procedimiento puede incluir adicionalmente añadir secuencias adicionales al O-ARNt. Normalmente, un OARNt

5 posee una mejora de la ortogonalidad para un organismo deseado en comparación con el material de partida, por ejemplo, en la pluralidad de secuencias de ARNt, mientras que conserva su afinidad hacia una RS deseada.

Los procedimientos incluyen opcionalmente analizar la similitud (y/u homología deducida) de secuencias de ARNt y/o aminoacil-ARNt sintetizadas para determinar candidatos potenciales para un O-ARNt, una O-RS y/o pares de los mismos que parecen ser ortogonales para un organismo específico. Se pueden usar programas informáticos conocidos en la técnica y descritos en el presente documento para el análisis, por ejemplo, se pueden usar programas BLAST y pileup. En un ejemplo, para elegir componentes de traducción ortogonales potenciales para usar en *E. coli*, se elige una sintetasa y/o un ARNt que no muestra similitud de secuencia cercana a organismos eubacterianos.

10

15

Normalmente se obtiene un O-ARNt sometiendo, por ejemplo, a selección negativa a una población de células de una primera especie, en la que las células comprenden un miembro de la pluralidad de O-ARNt potenciales. La selección negativa elimina células que comprenden un miembro de la biblioteca de O-ARNt potenciales que se aminoacilan mediante una aminoacil-ARNt sintetasa (RS) que es endógena a la célula. Esto proporciona un conjunto de ARNt que son ortogonales a la célula de la primera especie.

20

En determinadas realizaciones, en la selección negativa se introduce un(os) codón (codones) selector(es) en un polinucleótido que codifica un marcador de selección negativa, por ejemplo, una enzima que confiere resistencia a antibióticos, por ejemplo, β -lactamasa, una enzima que confiere un producto detectable, por ejemplo, β -galactosidasa, cloranfenicol acetiltransferasa (CAT), por ejemplo, un producto tóxico, tal como barnasa, en una posición no esencial (por ejemplo, que todavía produce una barnasa funcional, etc.). La identificación/selección se realiza opcionalmente dejando crecer la población de células en presencia de un agente selectivo (por ejemplo, un antibiótico tal como ampicilina). En una realización, la concentración del agente de selección varía.

25

Por ejemplo, para medir la actividad de ARNt supresores se usa un sistema de selección que se basa en la supresión *in vivo* de un codón selector, por ejemplo, mutaciones sin sentido (p. ej., de terminación) o de desplazamiento del marco de lectura introducidas en un polinucleótido que codifica un marcador de selección negativa, por ejemplo, un gen para β -lactamasa (*bla*). Por ejemplo, se construyen variantes polinucleotídicas, por ejemplo, variantes *bla* con un codón selector en una determinada posición (por ejemplo, A184). Se transforman las células, por ejemplo, bacterias con estos polinucleótidos. En el caso de un ARNt ortogonal que no puede ser cargado de forma eficaz mediante sintetizadas de *E. coli* endógenas, la resistencia a antibióticos, por ejemplo, resistencia a ampicilina, debe ser aproximadamente igual o menor que la de una bacteria transformada sin plásmido. Si el ARNt no es ortogonal, o si se co-expresa una sintetasa heteróloga capaz de cargar el ARNt en el sistema, se observa un mayor nivel de resistencia a antibióticos, por ejemplo, a ampicilina. Se eligen células, por ejemplo, bacterias que son incapaces de crecer en placas de agar LB con concentraciones de antibiótico aproximadamente iguales a las células transformadas sin plásmidos.

30

35

40

En el caso de un producto tóxico (por ejemplo, ribonucleasa o barnasa), cuando un miembro de la pluralidad de ARNt potenciales es aminoacilado por un hospedador endógeno, por ejemplo, sintetizadas de *Escherichia coli* (es decir, no es ortogonal al hospedador, por ejemplo, sintetizadas de *Escherichia coli*), el codón selector se suprime y el producto polinucleotídico tóxico producido conduce a la muerte celular. Las células que alojan ARNt ortogonales o ARNt no funcionales sobreviven.

45

En una realización, el conjunto de ARNt que son ortogonales a un organismo deseado se someten después a una selección positiva en la que se coloca un codón selector en un marcador de selección positiva, por ejemplo, codificado por un gen de resistencia a fármacos, tal como un gen de β -lactamasa. La selección positiva se realiza en una célula que comprende un polinucleótido que codifica o que comprende un miembro del conjunto de ARNt que son ortogonales a la célula, un polinucleótido que codifica un marcador de selección positiva y un polinucleótido que codifica una RS afín. En determinadas realizaciones, la segunda población de células comprende células que no han sido eliminados por la selección negativa. Los polinucleótidos se expresan en la célula y la célula se deja crecer en presencia de un agente de selección, por ejemplo, ampicilina. Después se seleccionan los ARNt por su capacidad para ser aminoacilado mediante la sintetasa afín co-expresada y para insertar un aminoácido en respuesta a este codón selector. Normalmente, estas células muestran una potenciación en la eficacia de supresión en comparación con células que alojan ARNt no funcional(es) o ARNt que no pueden ser reconocidas de forma eficaz por la sintetasa de interés. Las células que alojan los ARNt no funcionales o ARNt que no son reconocidas de forma eficaz por la sintetasa de interés son sensibles al antibiótico. Por lo tanto, los ARNt que: (i) no son sustratos para sintetizadas endógenas del hospedador, por ejemplo, la sintetasa de *Escherichia coli*; (ii) pueden ser aminoacilados por la sintetasa de interés; y (iii) son funcionales en la traducción, sobreviven a ambas selecciones.

50

55

60

En consecuencia, el mismo marcador puede ser un marcador positivo o negativo, dependiendo del contexto en el que se identifica. Es decir, si se consigue la identificación se realiza a favor del marcador es un marcador positivo,

65

pero si se identifica contra él es un marcador negativo.

La rigurosidad de la selección, por ejemplo, la selección positiva, la selección negativa o tanto la selección positiva como la negativa, en los procedimientos que se han descrito anteriormente, incluye opcionalmente variar la rigurosidad de la selección. Por ejemplo, debido a que la barnasa es una proteína extremadamente tóxica, la rigurosidad de la selección negativa se puede controlar introduciendo diferentes números de codones selectores en el gen de la barnasa y/o usando un promotor inducible. En otro ejemplo se varía la concentración del agente de selección o identificación (por ejemplo, concentración de ampicilina). En algunos aspectos de la invención se varía la rigurosidad debido a que la actividad deseada puede ser baja durante las primeras rondas. De esta manera, se aplican criterios de selección menos rigurosos en las primeras rondas y se aplican criterios más rigurosos en rondas de selección posteriores. En determinadas realizaciones, la selección negativa, la selección positiva o tanto la selección negativa como la positiva se pueden repetir múltiples veces. Se pueden usar múltiples marcadores diferentes de selección negativa, marcadores de selección positiva o tanto marcadores de selección negativa como positivas. En determinadas realizaciones, el marcador de selección positiva y negativa puede ser el mismo.

Se pueden usar otros tipos de selección/identificación en la invención para producir componentes de traducción ortogonales, por ejemplo, un O-ARNt, una O-RS, y un par de O-ARNt/O-RS que carga un aminoácido no natural en respuesta a un codón selector. Por ejemplo, el marcador de selección negativa, el marcador de selección positiva o tanto el marcador de selección positiva como negativa pueden incluir un marcador que produce fluorescencia o cataliza una reacción luminiscente en presencia de un reactivo adecuado. En otra realización se detecta un producto del marcador mediante separación de células activadas por fluorescencia (FACS) o por luminiscencia. Opcionalmente, el marcador incluye un marcador de identificación basado en afinidad. Véase también, Francisco, J. A., et al., (1993) Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci USA. 90:10444-8.

Se pueden encontrar procedimientos adicionales para la producción de tRNA ortogonal recombinante en, p. ej., las solicitudes de patente internacional WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS" y el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE" y el documento WO 2005/019415, presentado el 7 de julio de 2004. Véase también Foster et al., (2003) Programming peptidomimetic synthetases by translating genetic codes designed de novo PNAS 100(11):6353-6357; y Feng et al., (2003), Expanding tRNA recognition of a tRNA synthetase by a single amino acid change, PNAS, 100(10): 5676-5681.

Aminoacil-ARNt sintetasa ortogonal (O-RS)

Una O-RS de la invención aminoacila de forma preferente un O-ARNt con un aminoácido no natural *in vitro* o *in vivo*. Se puede proporcionar una O-RS de la invención al sistema de traducción, por ejemplo, una célula, mediante un polipéptido que incluye una O-RS y/o un polinucleótido que codifica una O-RS o una parte de la misma. Por ejemplo, una O-RS ilustrativa comprende una secuencia de aminoácidos conocida en la técnica o una variación conservadora de la misma. En otro ejemplo una O-RS, o una parte de la misma, está codificada por una secuencia polinucleotídica que codifica una secuencia que comprende aminoácidos en la lista de secuencias o los ejemplos de este documento, o una secuencia polinucleotídica complementaria de la misma. Véase, por ejemplo, en la **FIG. 8** las secuencias de moléculas de O-RS útiles.

Se conocen procedimientos para identificar una aminoacil-ARNt sintetasa ortogonal (O-RS), por ejemplo una O-RS para usar con un O-ARNt. Por ejemplo, un procedimiento incluye someter a una población de células de una primera especie a selección, por ejemplo, a selección positiva, en la que las células comprenden individualmente: 1) un miembro de una pluralidad de aminoacil-ARNt sintetazas (RS), (por ejemplo, la pluralidad de RS puede incluir RS mutantes, RS procedentes de una especie diferente de la primera especie o tanto RS mutantes como RS procedentes de una especie diferente de la primera especie); 2) el ARNt ortogonal (O-ARNt) (por ejemplo, de una o más especies); y 3) un polinucleótido que codifica un marcador de selección (por ejemplo, positiva) y comprende al menos un codón selector. Se seleccionan o identifican las células que muestran un aumento en la eficacia de supresión en comparación con células que carecen de o con una cantidad reducida del miembro de la pluralidad de RS. Se puede medir la eficacia de supresión mediante técnicas conocidas en la técnica y como se describe en el presente documento. Las células que tienen un aumento en la eficacia de supresión comprenden una RS activa que aminoacila el O-ARNt. Se compara un nivel de aminoacilación (*in vitro* o *in vivo*) mediante la RS activa de un primer conjunto de ARNt de la primera especie con el nivel de aminoacilación (*in vitro* o *in vivo*) por la RS activa de un segundo conjunto de ARNt de la segunda especie. Se puede determinar el nivel de aminoacilación mediante una sustancia detectable (por ejemplo, un aminoácido no natural marcado). La RS activa que aminoacila de forma más eficaz el segundo conjunto de ARNt en comparación con el primer conjunto de ARNt, proporcionando por lo tanto una aminoacil-ARNt sintetasa ortogonal eficaz (optimizada) para usar con el O-ARNt. También es una característica de la invención una O-RS identificada por el procedimiento.

Para determinar la aminoacilación se puede usar cualquiera de los numerosos ensayos. Estos ensayos se pueden llevar a cabo *in vivo* o *in vitro*. Por ejemplo, ensayos de aminoacilación *in vitro* se describen en, p. ej., Hoben y Soll (1985) Methods Enzymol. 113:55-59. La aminoacilación también puede determinarse usando un indicador junto con

los componentes de traducción ortogonales y detectando el indicador en una célula que expresa un polinucleótido que comprende al menos un codón selector que codifica una proteína. Véase también el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" y el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE."

5 La O-RS identificada se puede manipular adicionalmente para alterar la especificidad de sustrato de la sintetasa, de forma que solamente se carga un aminoácido no natural deseado, pero ninguno de los 20 aminoácidos comunes, en el O-ARNt. Los procedimientos para generar una aminoacil-ARNt sintetasa ortogonal con una especificidad de sustrato para un aminoácido no natural incluyen mutar la sintetasa, por ejemplo, en el sitio activo en la sintetasa, en el sitio del mecanismo de edición en la sintetasa, en diferentes sitios combinando diferentes dominios de sintetasas o similares, y aplicar un proceso de selección. Se usa una estrategia que se basa en la combinación de una selección positiva seguida de una selección negativa. En la selección positiva, la supresión del codón selector introducido en una o más posiciones no esenciales de un marcador positivo permite que las células sobrevivan bajo presión de selección positiva. En presencia de tanto aminoácidos naturales como de aminoácidos no naturales, los supervivientes codifican de esta manera sintetasas activas que cargan el ARNt supresor ortogonal con un aminoácido natural o no natural. En la selección negativa, la supresión de un codón selector introducido en una o más posiciones no esenciales de un marcador negativo retira las sintetasas con especificidades de aminoácidos naturales. Los supervivientes de la selección negativa y positiva codifican sintetasas que aminoacilan (cargan) el ARNt supresor ortogonal solamente con aminoácidos no naturales. Estas sintetasas se pueden someter, a continuación, a mutagénesis adicional, por ejemplo, barajado de ADN u otros procedimientos de mutagénesis recursivos.

Se puede generar una biblioteca de O-RS mutantes usando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, se pueden generar las RS mutantes mediante mutaciones con especificidad de sitio, mutaciones puntuales aleatorias, recombinación homóloga, barajado de ADN u otros procedimientos de mutagénesis recursivos, construcción quimérica o cualquier combinación de los mismos. Por ejemplo, se puede producir una biblioteca de RS mutantes a partir de dos o más "sub-bibliotecas" diferentes, por ejemplo, menores, menos diversas. También se incluyen en la invención bibliotecas quiméricas de RS. Cabe destacar que las bibliotecas de ARNt sintetasas de diversos organismos (por ejemplo, microorganismos tales como eubacterias o arqueobacterias) tales como bibliotecas que comprenden diversidad natural (véase, por ejemplo, la Patente de Estados Unidos Nº 6.238.884 de Short et al; la Patente de Estados Unidos Nº 5.756.316 de Schallenberger et al; la Patente de Estados Unidos Nº 5.783.431 de Petersen et al; la Patente de Estados Unidos Nº 5.824.485 de Thompson et al; la Patente de Estados Unidos Nº 5.958.672 de Short et al), se construyen y exploran opcionalmente por pares ortogonales.

Una vez que las sintetasas se han sometido a la estrategia de selección/identificación positiva y negativa, estas sintetasas se pueden someter después a mutagénesis adicionales. Por ejemplo, se puede aislar un ácido nucleico que codifica la O-RS; se puede generar a partir del ácido nucleico un conjunto de polinucleótidos que codifican O-RS mutadas (por ejemplo, por mutagénesis aleatoria, mutagénesis con especificidad de sitio, recombinación o cualquier combinación de las mismas); y se pueden repetir estas etapas individuales o una combinación de estas etapas hasta que se obtenga una O-RS mutada que aminoacile de forma preferente el O-ARNt con el aminoácido no natural. En algunos aspectos de la invención, las etapas se realizan múltiples veces, por ejemplo, al menos dos veces.

También se pueden usar niveles adicionales de rigurosidad de selección/identificación en los procedimientos de la invención para producir O-ARNt, O-RS o pares de los mismos. La rigurosidad de selección o identificación se puede variar en una o en las dos etapas del procedimiento para producir una O-RS. Esto podría incluir, por ejemplo, variar la cantidad del agente de selección/identificación que se usa, etc. También se pueden realizar rondas adicionales de selecciones positivas y/o negativas. La selección o identificación también puede comprender uno o más de un cambio en la permeabilidad de aminoácidos, un cambio en la eficacia de traducción, un cambio en la fidelidad de traducción, etc. Normalmente, dichos uno o más cambios se basan en una mutación en uno o más genes en un organismo en el que usa un par de ARNt-ARNt sintetasa ortogonal para producir una proteína.

Se pueden encontrar detalles generales adicionales para la producción de O-RS y la alteración de la especificidad de la sintetasa por el sustrato en la población interna número WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS;" y el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE." Véase también, Wang and Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005).

Organismos fuente y hospedadores

60 Los componentes de traducción ortogonales (O-ARNt y O-RS) encuentran uso con la invención se pueden obtener de cualquier organismo (o una combinación de organismos) para usar en un sistema de traducción del huésped de cualquier otra especie, con el inconveniente de que los componentes de O-ARNt-O-RS y el sistema huésped funcionan de un modo ortogonal. No es necesario que el O-ARNt y la O-RS de un par ortogonal procedan del mismo organismo. En algunos aspectos, los componentes ortogonales derivan de genes de *Archaea* (es decir, arqueobacteria) para usar en un sistema huésped eubacteriano.

Por ejemplo, el O-ARNt ortogonal se puede obtener de un organismo Archae, por ejemplo, una arqueobacteria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies NRC-1 de *Halobacterium*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (Mm), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o similares, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares, mientras que la O-RS ortogonal se puede obtener de un organismo o combinación de organismos, por ejemplo una arqueobacteria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies NRC-1 de *Halobacterium*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o similares o una eubacteria, tales como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares. En una realización también se pueden usar fuentes eucariotas, por ejemplo, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos, artrópodos o similares) como fuentes de O-ARNt y O-RS.

Los componentes individuales de un par de O-ARNt/O-RS se pueden obtener a partir del mismo organismo o de diferentes organismos. En una realización, el par O-ARNt/O-RS procede del mismo organismo. Como alternativa, el O-ARNt y la O-RS del par O-ARNt/O-RS son de organismos diferentes.

El O-ARNt, la O-RS o el par de O-ARNt/O-RS se puede seleccionar o identificar *in vivo* o *in vitro* y/o usar en una célula, por ejemplo, una célula eubacteriana, para producir un polipéptido con un aminoácido no natural. La célula eubacteriana usada no se limita a, por ejemplo, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus* o similares. También son una característica de la invención composiciones de células eubacterianas que comprenden componentes de traducción de la invención.

Véase también la publicación de solicitud internacional número WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE," presentada el 16 de abril de 2004, para el cribado de O-tRNA y/o O-RS en una especie para usar en otra especie.

En algunos aspectos, el O-ARNt, la O-RS o el par de O-ARNt/O-RS se puede seleccionar o identificar *in vivo* o *in vitro* y/o usar en una célula, por ejemplo, una célula eucariota, para producir un polipéptido con un aminoácido no natural. La célula eucariota usada no está limitada, por ejemplo, se puede usar cualquier célula de levadura adecuada, tal como *Saccharomyces cerevisiae* (*S. cerevisiae*) o similares. También son una característica de la invención composiciones de células eucariotas que comprenden componentes de traducción de la invención.

Aunque los sistemas de traducción ortogonal (p. ej., que comprenden una O-RS, un O-ARNt y un aminoácido no natural) pueden usar células huésped cultivadas para producir proteínas que tienen aminoácidos no naturales, no se pretende que un sistema de traducción ortogonal de la invención requieran una célula huésped intacta viable. Por ejemplo, un sistema de traducción ortogonal puede usar un sistema sin células en presencia de un extracto celular. De hecho, el uso de sistemas de traducción/transcripción *in vitro* sin células para la producción de proteínas es una técnica bien establecida. La adaptación de estos sistemas *in vitro* para la producción de proteínas que tengan aminoácidos no naturales usando componentes del sistema de traducción ortogonal descritos en el presente documento entra dentro del alcance de la invención.

Codones selectores

Los codones selectores en los sistemas de traducción ortogonales expanden el marco de lectura del codón genético de la maquinaria biosintética de proteínas. Por ejemplo, un codón selector incluye, por ejemplo, un único codón de tres bases, un codón sin sentido tal como un codón de terminación, por ejemplo, un codón ámbar (UAG), o un codón ópalo (UGA), un codón no natural, al menos un codón de cuatro bases, un codón raro o similares. Se pueden introducir varios codones selectores en un gen deseado, por ejemplo, uno o más, dos o más, más de tres, etc. Usando diferentes codones selectores se pueden usar múltiples pares ortogonales de ARNt/sintetasa para permitir la incorporación con especificidad de sitio simultánea de múltiples aminoácidos no naturales diferentes usando estos codones selectores diferentes.

En una realización, los procedimientos implican el uso de un codón selector que es un codón de terminación para la incorporación de un aminoácido no natural *in vivo* en una célula en un polipéptido. Por ejemplo, se produce un O-ARNt que reconoce un codón de terminación y es aminoacilado por una O-RS con un aminoácido no natural. Este O-ARNt no es reconocido por las aminoacil-ARNt sintetasa naturales del huésped. Se puede usar mutagénesis dirigida convencional para introducir el codón de terminación en el sitio de interés en un polinucleótido que codifica un polipéptido de interés. Véase, por ejemplo, Sayers *et al.* (1988), *5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis*. Nucleic Acids Res., 791-802. Cuando la O-RS, el O-ARNt y el ácido nucleico que codifica un polipéptido de interés se combinan, por ejemplo, *in vivo*, el aminoácido no natural se incorpora en respuesta al codón de terminación para dar un polipéptido que contiene el aminoácido no natural en la posición especificada. En una realización de la invención, el codón de terminación usado como codón selector es un codón

ámbar, UAG, y/o un codón ópalo, UGA. En un ejemplo, un código genético en el que UAG y UGA se usan como codón selector puede codificar 22 aminoácidos al tiempo que conserva el codón sin sentido ocre, YAA, que es la señal de terminación más abundante.

- 5 La incorporación de aminoácidos no naturales in vivo se puede realizar sin perturbación significativa de la célula huésped. Por ejemplo, en células no eucariotas, tales como *Escherichia coli*, como la eficacia de supresión de un codón UAG, depende de la competitividad entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y el factor de liberación 1 (RF1) (que se une al codón UAG e inicia la liberación del péptido de crecimiento del ribosoma), se puede modular la eficacia de supresión, por ejemplo, aumentando el nivel de expresión de O-ARNt, por ejemplo, el
- 10 ARNt supresor, o usando una cepa deficiente en RF1. En células eucariotas, debido a que la eficacia de supresión de un codón UAG depende de la competitividad entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y un factor de liberación eucariota (por ejemplo, eRF) (que se une a un codón de terminación e inicia la liberación del péptido de crecimiento del ribosoma), se puede modular la eficacia de supresión, por ejemplo, aumentando el nivel de expresión de O-ARNt, por ejemplo, el ARNt supresor. Además, también pueden estar presentes compuestos
- 15 adicionales, por ejemplo, reduciendo agentes tales como el ditiotretiol (DTT).

También se pueden codificar aminoácidos no naturales con codones raros. Por ejemplo, se ha demostrado que cuando la concentración de arginina se reduce en una reacción de síntesis proteica *in vitro*, el codón raro de arginina, AGG, es eficaz para la inserción de Ala mediante un ARNt sintético acilado con alanina. Véase, por

20 ejemplo, Ma et al., *Biochemistry*, 32:7939 (1993). En este caso, el ARNt sintético compite con el ARNt^{Arg} de origen natural que existe como una especie minoritaria en *Escherichia coli*. Además, algunos organismos no usan todos los codones de triplete. Se ha utilizado un codón no asignado AGA en *Micrococcus luteus* para la inserción de aminoácidos en un extracto de transcripción/traducción *in vitro*. Véase, por ejemplo, Kowal and Oliver, *Nucl. Acid. Res.*, 25:4685 (1997). Se pueden generar componentes de la invención para usar estos codones raros *in vivo*.

25 Los codones selectores también pueden comprender codones extendidos, por ejemplo, codones de cuatro o más bases, tales como codones de cuatro, cinco, seis o más bases. Los ejemplos de codones de cuatro bases incluyen, por ejemplo, AGGA, CUAG, UAGA, CCCU y similares. Los ejemplos de codones de cinco bases incluyen, por ejemplo, AGGAC, CCCCUC, CCCUC, CUAGA, CUACU, UAGGC y similares. Los procedimientos de la invención incluyen usar codones extendidos basados en supresión del desplazamiento del marco de lectura. Los codones de

30 cuatro o más bases pueden insertar, por ejemplo, uno o múltiples aminoácidos no naturales en la misma proteína. En otras realizaciones, los bucles anticodón pueden decodificar, por ejemplo, al menos un codón de cuatro bases, al menos un codón de cinco bases o al menos un codón de seis bases o más. Ya que hay 256 codones de cuatro bases posibles, se pueden codificar múltiples aminoácidos no naturales en la misma célula usando un codón de

35 cuatro o más bases. Véase también Anderson et al., (2002) Exploring the Limits of Codon and Anticodon Size, *Chemistry and Biology*, 9:237-244; y Magliery, (2001) Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in *Escherichia coli*, *J. Mol. Biol.* 307: 755-769.

40 Por ejemplo, se han usado codones de cuatro bases para incorporar aminoácidos no naturales en proteínas usando procedimientos biosintéticos *in vitro*. Véase por ejemplo, Ma et al., (1993) *Biochemistry*, 32:7939; y Hohsaka et al., (1999) *J. Am. Chem. Soc.*, 121:34. Se usaron CGGG y AGGU para incorporar de forma simultánea 2-naftilalanina y un derivado NBD de lisina en estreptavidina *in vitro* con dos ARNt supresores de desplazamiento del marco de lectura acilados químicamente. Véase, por ejemplo, Hohsaka et al., (1999) *J. Am. Chem. Soc.*, 121:12194. En un

45 estudio *in vivo*, Moore et al. examinaron la capacidad de los derivados de ARNt^{Leu} con anticodones NCUA para suprimir codones UAGN (N puede ser U, A, G o C), y observaron que el cuadruplete UAGA puede ser decodificado mediante un ARNt^{Leu} con un anticodón UCUA con una eficacia del 13 al 26 % con poca decodificación en la fase 0 o -1. Véase Moore et al., (2000) *J. Mol. Biol.*, 298:195. En una realización se pueden usar en la invención codones extendidos basados en codones raros o codones sin sentido, que pueden reducir la lectura completa sin sentido y la

50 supresión por desplazamiento del marco de lectura en otros sitios no deseados. En diversos sistemas ortogonales se han usado codones de cuatro bases como codones selectores. Véase por ejemplo, los documentos WO 2005/019415; WO 2005/007870 and WO 2005/07624. Véase también Wang and Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005). Aunque los siguientes ejemplos usan un codón selector ámbar, se pueden usar codones de cuatro o más bases, modificando los ejemplos del presente documento para que

55 incluyan O-ARNt de cuatro bases y sintetasas modificadas para incluir mutaciones similares a las descritas anteriormente para varias O-RS de aminoácidos no naturales.

Para un sistema determinado, un codón selector también puede incluir uno de los codones de tres bases naturales, donde el sistema endógeno no usa (o usa raramente) el codón de bases naturales. Por ejemplo, esto incluye un

60 sistema que carece de un ARNt que reconoce el codón de tres bases naturales y/o un sistema en el que el codón de tres bases es un codón raro.

Los codones selectores incluyen opcionalmente pares de bases no naturales. Estos pares de bases no naturales expanden adicionalmente el alfabeto genético existente. Un par de bases adicional aumenta el número de codones

65 de triplete de 64 a 125. Las propiedades de pares de tercera base incluyen un emparejamiento de bases estable y selectivo, una incorporación enzimática eficaz en ADN con elevada fidelidad mediante una polimerasa y la extensión

por cebador continuada eficaz después de la síntesis del par de bases no natural que se genera. Las descripciones de pares de bases no naturales que se pueden adaptar para procedimientos y composiciones incluyen, por ejemplo, Hirao, et al., (2002) An unnatural base pair for incorporating amino acid analogues into protein, Nature Biotechnology, 20:177-182. Véase también Wu, Y., et al., (2002) J. Am. Chem. Soc. 124:14626-14630. Más adelante se enumeran otras publicaciones relevantes.

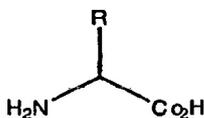
Para el uso in vivo, el nucleósido no natural es permeable a la membrana y se fosforila para formar el trifosfato correspondiente. Además, la información genética aumentada es estable y no es destruida por enzimas celulares. Ciertos esfuerzos previos de Benner y otros aprovecharon patrones de formación de enlaces de hidrógeno que son diferentes de los presentes en pares de Watson-Crick canónicos, de los cuales el ejemplo más notable es el par iso-C:iso-G. Véase, por ejemplo, Switzer et al., (1989) J. Am. Chem. Soc, 111:8322; y Piccirilli et al., (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. Estas bases en general se aparean de forma errónea hasta cierto grado con bases naturales y no se pueden replicar enzimáticamente. Kool y colaboradores demostraron que las interacciones de empaquetamiento hidrófobas entre bases pueden sustituir a las uniones por enlaces de hidrógeno para dirigir la formación de pares de bases. Véase Kool, (2000) Curr. Opin. Chem. Biol., 4:602; y Guckian y Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825. En un esfuerzo por desarrollar un par de bases no natural que satisfaga todos los requisitos anteriores, Schultz, Romesberg y colaboradores han sintetizado y estudiado sistemáticamente una serie de bases hidrófobas no naturales. Se ha observado que un autopar PICS:PICS es más estable que pares de bases naturales y se puede incorporar de forma eficaz en el ADN por el fragmento Klenow de la ADN polimerasa I de Escherichia coli (KF). Véase, por ejemplo, McMinn et al., (1999) J. Am. Chem. Soc, 121:11586; y Ogawa et al., (2000) J. Am. Chem. Soc, 122:3274. Se puede sintetizar un autopar 3MN:3MN por KF con suficiente eficacia y selectividad para la función biológica. Véase, por ejemplo, Ogawa et al., (2000) J. Am. Chem. Soc, 122:8803. Sin embargo, ambas bases actúan como un terminador de cadena para la replicación posterior. Se ha desarrollado recientemente una ADN polimerasa mutante que se puede usar para replicar el autopar PICS. Además, se puede replicar un autopar 7Al. Véase, por ejemplo, Tae et al., (2001) J. Am. Chem. Soc., 123:7439. También se ha desarrollado un nuevo par de bases metálico, Dipic:Py, que forma un par estable después unirse a Cu(II). Véase Meggers et al., (2000) J. Am. Chem. Soc, 122:10714. Debido a que los codones extendidos y los codones no naturales son intrínsecamente ortogonales a codones naturales, los procedimientos de la invención pueden utilizar ventajosamente esta propiedad para generar ARNt ortogonales para los mismos.

También se puede usar un sistema de derivación de la traducción para incorporar un aminoácido no natural en un polipéptido deseado. En un sistema de derivación de traducción se inserta una gran secuencia en un gen pero no se traduce en una proteína. La secuencia contiene una estructura que sirve como clave para inducir el salto del ribosoma sobre la secuencia y reanudar la traducción más adelante de la inserción.

Aminoácidos no naturales

Como se usa en este documento, un aminoácido no natural se refiere a cualquier aminoácido, aminoácido modificado o análogo de aminoácido diferente de selenocisteína y/o pirrolisina y los siguientes veinte alfa-aminoácidos codificados genéticamente: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. La estructura genérica de un alfa-aminoácido se ilustra en la Fórmula I:

I



Un aminoácido no natural es normalmente cualquier estructura que tenga la Fórmula I, en la que el grupo R es cualquier sustituyente diferente de los usados en los veinte aminoácidos naturales. Véase, por ejemplo, Biochemistry de L. Stryer, 3ª ed. 1988, Freeman and Company, New York, para las estructuras de los veinte aminoácidos naturales. Obsérvese que los aminoácidos no naturales de la invención pueden ser compuestos de origen natural diferentes de los veinte alfa-aminoácidos anteriores.

Ya que los aminoácidos no naturales de la invención difieren normalmente de los aminoácidos naturales en la cadena lateral, los aminoácidos no naturales forman enlaces amida con otros aminoácidos, por ejemplo, naturales o no naturales, del mismo modo en el que se forman en proteínas de origen natural. Sin embargo, los aminoácidos no naturales tienen grupos de cadena lateral que los distinguen de los aminoácidos naturales.

La FIG. 1 proporciona las estructuras de aminoácidos no naturales que se usan en los ejemplos de trabajo de la presente invención. Estos aminoácidos no naturales se pueden incorporar en proteínas usando pares de O-RS y O-ARNt adecuados. Por ejemplo, se puede incorporar *p*-benzoil-L-fenilalanina (Bpa) usando un par de traducción ortogonal que comprende el O-ARNt de la SEC ID N°: 1 y la O-RS afín de la SEC ID N°: 4. *para*-acetil-L-fenilalanina (*p*AcPhe) se puede incorporar usando un par de traducción ortogonal que comprende el O-ARNt de la SEC ID N°: 1

y la O-RS afín de la SEC ID N°: 6. *para*-azido-L-fenilalanina (*p*AzPhe) se puede incorporar usando un par de traducción ortogonal que comprende el O-ARNt de la SEC ID N°: 1 y la O-RS afín de la SEC ID N°: 8. *para*-yodo-L-fenilalanina (*p*IzPhe) se puede incorporar usando un par de traducción ortogonal que comprende el O-ARNt de la SEC ID N°: 1 y la O-RS afín de la SEC ID N°: 10.

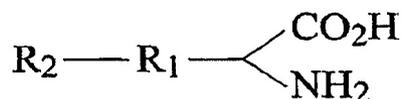
5 No obstante, los aminoácidos no naturales usados en el presente documento solo sirven para ilustrar una aplicabilidad más amplia de la invención y la invención no está limitada al uso de estos aminoácidos mostrados en la FIG. 1.

10 Una pluralidad de diferentes aminoácidos no naturales se pueden incorporar de forma simultánea en un polipéptido de interés, usando, por ejemplo, un segundo par de O-RS/O-ARNt adecuado junto con el primer par ortogonal y en el que los pares ortogonales primero y segundo usan diferentes codones selectores.

15 En otros aminoácidos no naturales, por ejemplo, el grupo R presente en la Fórmula I comprende opcionalmente un alquilo-, arilo-, acilo-, ceto-, azido-, hidroxilo-, hidracina, ciano-, halo-, hidracida, alqueno-, alquino-, éter, tiol, seleno-, sulfonilo-, borato, boronato, fosfo, fosfona, fosfina, heterociclo, enona, imina, aldehído, éster, tioácido, hidroxilamina, amina y similares, o cualquier combinación de los mismos. Otros aminoácidos de interés incluyen, pero no se limitan a, aminoácidos que comprenden un agente de entrecruzamiento fotoactivable, aminoácidos marcados por espín, aminoácidos fluorescentes, aminoácidos de unión a metal, aminoácidos que contienen metal, aminoácidos radiactivos, aminoácidos con nuevos grupos funcionales, aminoácidos que interaccionan de forma covalente o de forma no covalente con otras moléculas, aminoácidos con grupos fotoprotectores y/o fotoisomerizables, aminoácidos que contienen biotina o análogos de biotina, aminoácidos que contienen ceto, aminoácidos glicosilados, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles o fotoescindibles, aminoácidos con una cadena lateral alargada en comparación con aminoácidos naturales (por ejemplo, poliéteres o hidrocarburos de cadena larga, por ejemplo, de más de aproximadamente 5, de más de aproximadamente 10 carbonos, etc.), aminoácidos que contienen azúcar unido a carbono, aminoácidos con actividad redox, aminoácidos que contienen aminotioácido y aminoácidos que contienen uno o más residuos tóxicos.

30 En otro aspecto, la invención proporciona aminoácidos no naturales que tiene la estructura general ilustrada por las Fórmulas IV siguientes:

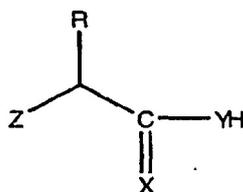
IV



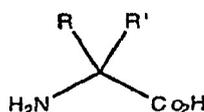
35 Un aminoácido no natural que tiene esta estructura, normalmente es cualquier estructura en la que R₁ es un sustituyente usado en uno de los veinte aminoácidos naturales (p. ej., tirosina o fenilalanina) y R₂ es un sustituyente. Por tanto, este tipo de aminoácido no natural se puede ver como un derivado de un aminoácido natural.

40 Los aminoácidos no naturales también comprenden opcionalmente estructuras de cadena principal modificadas, por ejemplo, como se ilustra mediante las estructuras de las Fórmulas II y III:

II



III



45

en las que Z comprende normalmente OH, NH₂, SH, NH-R' o S-R'; X e Y, que pueden ser iguales o diferentes, comprenden normalmente S u O, y R y R', que son opcionalmente iguales o diferentes, se seleccionan normalmente de la misma lista de constituyentes para el grupo R que se ha descrito anteriormente para los aminoácidos no naturales que tienen la Fórmula I, así como hidrógeno. Por ejemplo, los aminoácidos no naturales de la invención comprenden opcionalmente sustituciones en el grupo amino o carboxilo, como se ilustra por las Fórmulas II y III. Los aminoácidos no naturales de este tipo incluyen, pero no se limitan a, α -hidroxiácidos, α -tioácidos, α -aminotiocarboxilatos, por ejemplo, con cadenas laterales correspondientes a los veinte aminoácidos naturales comunes o cadenas laterales no naturales. Además, las sustituciones en el carbono α incluyen opcionalmente aminoácidos L, D, o α - α disustituidos tales como D-glutamato, D-alanina, D-metil-O-tirosina, ácido aminobutírico y similares. Otras alternativas estructurales incluyen aminoácidos cíclicos, tales como análogos de prolina así como análogos de prolina con anillo de 3, 4, 6, 7, 8 y 9 miembros, β y γ aminoácidos tales como β -alanina sustituida y ácido γ -aminobutírico.

En algunos aspectos, la invención usa aminoácidos no naturales en configuración L. No obstante, no se pretende limitar la invención al uso de la configuración L de los aminoácidos naturales. Se contempla que los enantiómeros D de estos aminoácidos no naturales también encuentran uso con la invención.

Los análogos de tirosina incluyen tirosinas para-sustituidas, tirosinas orto-sustituidas y tirosinas meta-sustituidas, en las que la tirosina sustituida comprende un grupo alquililo, un grupo acetilo, un grupo benzoilo, un grupo amino, una hidracina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo C₆-C₂₀ de cadena lineal o ramificada, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro o similares. Además, también se contemplan anillos de arilo sustituidos de forma múltiple. Los análogos de glutamina de la invención incluyen, pero no se limitan a, derivados α -hidroxi, derivados γ -sustituidos, derivados cíclicos y derivados de glutamina sustituidos con amida. Ejemplos de análogos de fenilalanina incluyen, entre otros, fenilalaninas para-sustituidas, fenilalaninas orto-sustituidas y fenilalaninas meta-sustituidas, en las que el sustituyente comprende un grupo alquililo, un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído o grupo ceto o similares. Ejemplos específicos de aminoácidos no naturales incluyen, pero no se limitan a, *p*-etil-tiocarbonil-L-fenilalanina, *p*-(3-oxobutanoil)-L-fenilalanina, 1,5-dansil-alanina, aminoácido 7-amino-coumarina, aminoácido 7-hidroxi-coumarina, nitrobenzil-serina, O-(2-nitrobenzil)-L-tirosina; *p*-carboximetil-L-fenilalanina, *p*-ciano-L-fenilalanina, *m*-ciano-L-fenilalanina, bifenilalanina, 3-amino-L-tirosina, biperidil alanina, *p*-(2-amino-1-hidroxi-etil)-L-fenilalanina, *p*-isopropiltiocarbonil-L-fenilalanina, 3-nitro-L-tirosina y *p*-nitro-L-fenilalanina. Asimismo, una *p*-propargiloxifenilalanina, una 3,4-dihidroxi-L-fenilalanina (DHP), una 3, 4, 6-trihidroxi-L-fenilalanina, una 3,4,5-trihidroxi-L-fenilalanina, 4-nitro-fenilalanina, una *p*-acetil-L-fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metil-fenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una 3-nitro-tirosina, una 3-tiol-tirosina, una tri-O-acetil-GlcNAc β -serina, una L-Dopa, una fenilalanina fluorada, una isopropil-L-fenilalanina, una *p*-azido-L-fenilalanina, una *p*-acil-L-fenilalanina, una *p*-benzoil-L-fenilalanina, una L-fosfoserina, una fosfonoserina, una fosfonotirosina, una *p*-yodo-fenilalanina. Se conocen las estructuras de diversos aminoácidos no naturales que se pueden incorporar usando sistemas de traducción ortogonales. Véase las referencias citadas en el presente documento.

Síntesis química de aminoácidos no naturales

Muchos de los aminoácidos no naturales proporcionados más arriba están disponibles comercialmente, p. ej., de Sigma (USA) o Aldrich (Milwaukee, WI, USA). Los que no están disponibles comercialmente se sintetizan opcionalmente según se proporciona en varias publicaciones o usando procedimientos estándar conocidos para los expertos en la materia. Para técnicas de síntesis orgánica, véase, p. ej., Organic Chemistry de Fessenden y Fessenden (1982, segunda edición, Willard Grant Press, Boston Mass); Advanced Organic Chemistry de March (tercera edición, partes A y B, 1990, Plenum Press, Nueva York). Publicaciones adicionales que describen la síntesis de aminoácidos no naturales incluyen, p. ej., el documento WO 2002/085923 titulado "In vivo incorporation of Unnatural Amino Acids;" Matsoukas et al., (1995) J. Med. Chem., 38, 4660-4669; King and Kidd (1949) A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319; Friedman and Chatterji (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750-3752, Craig et al. (1988) Absolute Configuration of the Enantiomers of 7-Chloro-4[[4-(diethylamino)-1-methylbutyl] amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170; Azoulay et al., (1991) Glutamine analogues as Potential Antimalarials. Eur. J. Med. Chem. 26, 201-5; Koskinen and Rapoport (1989) Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859-1866; Christie and Rapoport (1985) Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Syntheses of (+)- Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization. J. Org. Chem. 1989:1859-1866; Barton et al. (1987) Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron Lett. 43:4297-4308; y Subasinghe et al., (1992) Quisqualic acid analogues: synthesis, of betaheterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35:4602-7. Véase también la publicación internacional WO 2004/058946, titulada "PROTEIN ARRAYS," presentada el 22 de diciembre de 2003.

Captación celular de aminoácidos no naturales

La captación de aminoácidos no naturales por una célula es una cuestión que se considera normalmente cuando se diseñan y seleccionan aminoácidos no naturales, por ejemplo, para la incorporación en una proteína. Por ejemplo, la gran densidad de carga α -aminoácidos sugiere que estos compuestos probablemente no sean permeables a la célula. Los aminoácidos naturales se introducen en la célula a través de una colección de sistemas de transporte basados en proteínas que a menudo muestran grados variables de especificidad de aminoácidos. Se puede realizar una exploración rápida que evalúa qué aminoácidos no naturales, si hay alguno, se captan por las células. Véase, por ejemplo, ensayos de toxicidad en, por ejemplo, la publicación internacional WO 2004/058946, titulada "PROTEIN ARRAYS," presentada el 22 de diciembre de 2003; y Liu y Schultz (1999) Progress toward the evolution of an organism with an expanded genetic code. PNAS 96:4780-4785. Aunque la captación se analiza de manera sencilla con diversos ensayos, una alternativa para diseñar aminoácidos no naturales que son susceptibles a rutas de captación celulares es proporcionar rutas biosintéticas para crear aminoácidos *in vivo*.

Biosíntesis de aminoácidos no naturales

Ya existen muchas rutas biosintéticas en las células para la producción de aminoácidos y otros compuestos. Mientras que un procedimiento biosintético para un aminoácido no natural concreto puede no existir en la naturaleza, por ejemplo, en una célula, la invención proporciona tales procedimientos. Por ejemplo, se pueden generar rutas biosintéticas para aminoácidos no naturales en células huésped añadiendo nuevas enzimas o modificando rutas existentes de la célula huésped. Las nuevas enzimas adicionales son opcionalmente enzimas de origen natural o enzimas desarrolladas artificialmente. Por ejemplo, la biosíntesis de p-aminofenilalanina (como se presenta en un ejemplo del documento WO 2002/085923, supra) se basa en la adición de una combinación de enzimas conocidas de otros organismos. Se pueden introducir los genes para estas enzimas en una célula transformando la célula con un plásmido que comprende los genes. Los genes, cuando se expresan en la célula, proporcionan una ruta enzimática para sintetizar el compuesto deseado. En los ejemplos que se presentan más adelante se proporcionan ejemplos de los tipos de enzimas que se pueden añadir. Pueden encontrarse secuencias de enzimas adicionales, por ejemplo, en el Genbank. También se pueden añadir enzimas desarrolladas artificialmente a una célula del mismo modo. De esta manera, la maquinaria celular y los recursos de una célula se manipulan para producir aminoácidos no naturales.

De hecho, se puede usar cualquiera de diversos procedimientos para producir nuevas enzimas para usar en rutas biosintéticas, o para la evolución de rutas existentes, para la producción de aminoácidos no naturales, *in vitro* o *in vivo*. Se pueden aplicar muchos procedimientos disponibles para desarrollar enzimas y otros componentes de ruta biosintética a la presente invención para producir aminoácidos no naturales (o, de hecho, para desarrollar sintetetas para que tengan nuevas especificidades de sustrato u otras actividades de interés). Por ejemplo, opcionalmente se usa el barajado de ADN para desarrollar nuevas enzimas y/o rutas de tales enzimas para la producción de aminoácidos no naturales (o producción de nuevas sintetetas), *in vivo* o *in vitro*. Véase, por ejemplo., Stemmer (1994), Rapid evolution of a protein *in vitro* by DNA shuffling, Nature 370(4):389-391; y, Stemmer, (1994), DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution, Proc. Natl. Acad. Sci. USA., 91:10747-10751. Una estrategia relacionada baraja familias de genes relacionados (por ejemplo, homólogos) para desarrollar rápidamente enzimas con características deseadas. Un ejemplo de tales procedimientos de "barajado de genes de familias" se encuentra en Cramer et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature, 391(6664): 288-291. También se pueden generar nuevas enzimas (componentes de rutas biosintéticas o sintetetas) usando un procedimiento de recombinación de ADN conocido como "truncamiento incremental para la creación de enzimas híbridas" ("ITCHY"), por ejemplo, como se describe en Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" Nature Biotech 17:1205. Esta estrategia también se puede usar para generar una biblioteca de enzimas u otras variantes de ruta que pueden servir como sustratos para uno o más procedimientos de recombinación *in vitro* o *in vivo*. Véase también, Ostermeier et al.(1999) "Combinatorial Protein Engineering by Incremental Truncation," Proc. Natl. Acad. Sci. USA, 96: 3562-67 y Ostermeier et al. (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts," Biological and Medicinal Chemistry, 7: 2139-44. Otra estrategia usa mutagénesis de conjunto exponencial para producir bibliotecas de enzimas u otras variantes de ruta que se seleccionan, por ejemplo, por la capacidad de catalizar una reacción biosintética pertinente para la producción de un aminoácido no natural (o una nueva sintetasa). En esta estrategia se aleatorizan pequeños grupos de residuos en una secuencia de interés en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales. Pueden encontrarse ejemplos de tales procedimientos, que se pueden adaptar a la presente invención para producir nuevas enzimas para la producción de aminoácidos no naturales (nuevas sintetetas), en Delegrave & Youvan (1993) Biotechnology Research 11:1548-1552. En otra estrategia más, se puede usar mutagénesis aleatoria o semi-aleatoria usando oligonucleótidos dopados o degenerados para la producción por ingeniería genética de enzimas y/o componentes de ruta, por ejemplo, usando los procedimientos de mutagénesis generales de, por ejemplo, Arkin y Youvan (1992) "Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis" Biotechnology 10:297300; o Reidhaar-Olson et al. (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" Methods Enzymol. 208:564-86. Otra estrategia más, denominada a menudo mutagénesis "no estocástica", que usa el reensamblaje de polinucleótidos y mutagénesis de saturación de sitio, se puede usar para producir enzimas y/o componentes de ruta que después se

pueden explorar con respecto a una capacidad para realizar una o más funciones de sintetasa o de ruta biosintética (por ejemplo, para la producción de aminoácidos no naturales *in vivo*). Véase, por ejemplo el documento WO 2000/046344, titulado "NON-STOCHASTIC GENERATION OF GENETIC VACCINES AND ENZYMES" de Short.

5 Una alternativa a tales procedimientos mutacionales implica recombinar genomas enteros de organismos y seleccionar la progenie resultante para funciones de ruta particulares (lo cual se denomina a menudo "barajado de genoma completo"). Esta estrategia se puede aplicar a la presente invención, por ejemplo, mediante recombinación genómica y selección de un organismo (por ejemplo, una célula *E. coli* u otra célula) con respecto a la capacidad de producir un aminoácido no natural (o intermedio del mismo). Por ejemplo, se pueden aplicar procedimientos
10 descritos en las siguientes publicaciones para el diseño de rutas para la evolución de rutas existentes y/o nuevas en células para producir aminoácidos no naturales *in vivo*: Patnaik et al. (2002) "Genome shuffling of lactobacillus for improved acid tolerance" *Nature Biotechnology*, 20(7): 707-712; y Zhang et al. (2002) "Genome shuffling leads to rapid phenotypic improvement in bacteria" *Nature*, 7 de febrero, 415(6872): 644-646.

15 También están disponibles otras técnicas para la producción por ingeniería genética de organismos y rutas metabólicas, por ejemplo, para la producción de compuestos deseados, y también se pueden aplicar a la producción de aminoácidos no naturales. Los ejemplos de publicaciones que describen estrategias de producción por ingeniería genética de rutas incluyen: Nakamura y White (2003) "Metabolic engineering for the microbiol production of 1,3 propanediol" *Curr. Opin. Biotechnol.* 14(5):454-9; Berry et al. (2002) "Application of Metabolic Engineering to improbé
20 both the production and use of Biotech Indigo" *J. Industrial Microbiology and Biotechnology* 28:127-133; Banta et al. (2002) "Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of *Corynebacterium* 2,5-diketo D-gluconic acid reductase for use in vitamin C biosynthesis" *Biochemistry*, 41(20), 6226-36; Selivonova et al. (2001) "Rapid Evolution of Novel Traits in Microorganisms" *Applied and Environmental Microbiology*, 67:3645, y muchas otras.

25 Sin tener en cuenta el procedimiento usado, normalmente el aminoácido no natural producido con una ruta biosintética creada por ingeniería genética de la invención se produce en una concentración suficiente para la biosíntesis eficaz de proteínas, por ejemplo, una cantidad celular natural, pero no hasta tal grado como para afectar de forma significativa a la concentración de otros aminoácidos celulares o hasta agotar los recursos celulares. Las
30 concentraciones típicas producidas de este modo *in vivo* son de aproximadamente 10 mM a aproximadamente 0,05 mM. Una vez que se produce una célula por ingeniería genética para producir enzimas deseadas para una ruta específica y se genera un aminoácido no natural, se usan opcionalmente selecciones *in vivo* para optimizar adicionalmente la producción del aminoácido no natural tanto para la síntesis de proteínas ribosómicas como para el crecimiento celular.

35 Componentes ortogonales de uso en la invención

La incorporación del aminoácido no natural en la proteína se consigue mediante pares ortogonales que incorporan el aminoácido natural en respuesta a la señal genética del codón selector, en *E. coli*, en el que los componentes
40 ortogonales no sufren reacciones cruzadas con los componentes endógenos de *E. coli* de la maquinaria traduccional de la célula huésped, pero reconocen el aminoácido no natural deseado y lo incorporan en proteínas en respuesta al codón selector (p. ej., un codón sin sentido ámbar, TAG). Los componentes ortogonales que encuentran uso con la invención incluyen aminoacil-ARNt sintetetas ortogonales (O-RS) derivadas de procedente del tiosil-ARNt de *Methanococcus jannaschii* y el tiosil-ARNt_{CUA} mutante supresor ámbar, que funcionan como par ortogonal en una
45 célula huésped eubacteriana, tal como *E. coli*. En este sistema, las aminoacil-ARNt sintetetas mutantes aminoacilan el ARNt supresor con su respectivo aminoácido no natural y no con cualquiera de los veinte aminoácidos comunes.

Procedimientos para producir componentes ortogonales encuentran uso con la invención, en la que estos procedimientos tienen como resultado la incorporación de aminoácidos no naturales, por ejemplo, entre otros, los
50 aminoácidos no naturales proporcionados en la FIG. 1, en una cadena polipeptídica en crecimiento en respuesta a un codón selector, por ejemplo, codón de terminación cambiar, un codón sin sentido, un codón de cuatro o más bases etc., por ejemplo *in vivo*. Por ejemplo, ARNt ortogonales (O-ARNt), aminoacil-ARNt sintetetas ortogonales (O-RS) y pares de los mismos encuentran uso en la invención.

55 En algunas realizaciones, estos pares se pueden usar para incorporar un aminoácido no natural en cadenas polipeptídicas en crecimiento y, después, el polipéptido se modifica postraduccionalmente. Para información adicional sobre aminoácidos no naturales que se pueden modificar postraduccionalmente véase, por ejemplo, los sistemas ortogonales de aminoácidos no naturales descritos en Chin et al., *Science* (2003) 301:964-967; Zhang et al., *Proc. Natl. Acad. Sci. U.S.A.* 2004, 101:8882-8887; Anderson et al., *Proc. Natl. Acad. Sci. U.S.A.* 2004,
60 101:7566-7571; Wang et al., (2001) *Science* 292:498-500; Chin et al (2002) *Journal of the American Chemical Society* 124:9026-9027; Chin and Schultz, (2002) *ChemBioChem* 11:1135-1137; Chin, et al., (2002) *PNAS United States of America* 99:11020-11024; Wang and Schultz, (2002) *Chem. Comm.*, 1-10; Wang and Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005); Xie and Schultz, "An Expanding Genetic Code," *Methods* 36:227-238 (2005); and Deiters et al, *Bioorganic & Medicinal Chemistry Letters* 15:1521-1524
65 (2005).

Véase también los sistemas ortogonales de aminoácidos no naturales en las publicaciones internacionales WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS;" el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" el documento WO 2004/094593, titulado ""EXPANDING THE EUKARYOTIC GENETIC CODE;" el documento WO 2005/019415, presentado el 7 de julio de 2004; el documento WO2005/007870, presentado el 7 de julio de 2004; el documento WO 2005/007624, presentado el 7 de julio de 2004; el documento WO 2006/034332, presentado el 20 de septiembre de 2005; y el documento WO 2006/110182 titulado "ORTHOAGONAL TRANSLATION COMPONENTS .

10 En determinadas realizaciones, la O-RS que encuentra uso en la invención aminoacila de forma preferente el O-ARNt antes que cualquier ARNt endógeno con un aminoácido no natural, en la que la O-RS tiene un sesgo por el O-ARNt y en la que la proporción entre el O-ARNt cargado con un aminoácido no natural y el ARNt endógeno cargado con el mismo aminoácido no natural es mayor que 1:1 y, más preferentemente, en la que O-RS carga el O-ARNt exclusivamente o casi exclusivamente.

15 La invención también usa ARN ortogonales (O-ARNt), en los que el O-ARNt reconoce un codón selector. Normalmente, un O-ARNt de la invención incluye al menos aproximadamente, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 % o un 90 % o más de eficacia de supresión en presencia de una sintetasa afín en respuesta a un codón selector en comparación con el O-ARNt que comprende o está codificado por una secuencia polinucleotídica como se indica en la lista de secuencias (p. ej., la SEC ID N^o: 1). En una realización, la eficacia de supresión de la O-RS y el O-ARNt conjuntamente es, por ejemplo, 5 veces, 10 veces, 15 veces, 20 veces, 25 veces o más veces mayor que la eficacia de supresión del O-ARNt que carece de la O-RS. En algunos aspectos, la eficacia de supresión de la O-RS y del O-ARNt conjuntamente es al menos un 45 % de la eficacia de supresión de un par ortogonal de tirosil-ARNt sintetasa procedente de *Methanococcus jannaschii*.

25 La invención usa células (p. ej., *E. coli*) que comprenden un sistema de traducción y secuencias de nucleótidos que programan la producción de proteínas, en la que el sistema de traducción incluye un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS) y un aminoácido no natural. Normalmente, la O-RS aminoacila de forma preferente el O-ARNt antes que cualquier ARNt endógeno con un aminoácido no natural, en la que la O-RS tiene un sesgo por el O-ARNt y en la que la proporción entre el O-ARNt cargado con el aminoácido no natural y el ARNt endógeno cargado con el aminoácido no natural es mayor que 1:1 y, más preferentemente, en la que O-RS carga el O-ARNt exclusivamente o casi exclusivamente. El O-ARNt reconoce el primer codón selector y la O-RS aminoacila de forma preferente el O-ARNt con el aminoácido no natural.

35 Varios polinucleótidos también encuentran uso con la invención. Estos polinucleótidos incluyen un polinucleótido artificial (por ejemplo, hecho por el hombre y no de origen natural, p. ej., recombinante) que comprende una secuencia de nucleótidos que codifica una O-RS. Un polinucleótido que encuentra uso con la invención también puede incluir un ácido nucleico que hibrida con un polinucleótido descrito anteriormente, en condiciones de alta rigurosidad, a lo largo de sustancialmente toda la longitud del ácido nucleico. Vectores que comprenden polinucleótidos también encuentran uso con la invención. Por ejemplo, un vector puede incluir un plásmido, un cósmido, un fago, un virus, un vector de expresión y/o similares. Procedimientos para producir componentes de un par de O-ARNt/O-RS se conocen y encuentran uso en la invención. Véase la presente divulgación y las referencias citadas en el presente documento.

45 Secuencias de ácido nucleico y polipeptídica y variantes

Como se describe en el presente documento, las secuencias polinucleotídicas que codifican, por ejemplo, O-ARNt y O-RS, encuentran uso en la invención, como también lo hacen las respectivas secuencias de aminoácidos codificadas por los polinucleótidos. La divulgación proporciona ejemplos y referencias de secuencias de polipéptidos y polinucleótidos que encuentran uso en la invención. Sin embargo, se apreciará que el uso de la invención no está limitado a las secuencias divulgadas en el presente documento. Un experto en la técnica que la invención también proporciona muchas secuencias relacionadas con las funciones descritas en el presente documento, por ejemplo polinucleótidos y polipéptidos que codifican variantes conservadoras de una O-RS divulgada en el presente documento.

55 Un polinucleótido que encuentra uso en la invención también incluye un polinucleótido artificial que tiene, por ejemplo, una identidad de al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 % o mayor que un ARNt de origen natural, (pero es diferente de un ARNt de origen natural). Un polinucleótido que encuentra uso en la invención también incluye un polinucleótido artificial que tiene, por ejemplo, una identidad de al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 % (pero no una identidad del 100 %) que un ARNt de origen natural.

65 En determinadas realizaciones, un vector que encuentra uso en la invención (por ejemplo un plásmido, un cósmido, un fago, un virus, etc.) comprende un polinucleótido que encuentra uso la invención. En algunas realizaciones, el vector es un vector de expresión. En otras realizaciones, el vector de expresión incluye un promotor unido de manera operativa a uno o más de los polinucleótidos de la invención. En otras realizaciones, una célula comprende

un vector que incluye un polinucleótido que encuentra uso en la invención.

Un experto también entenderá que en la invención también encuentran uso muchas variantes de las secuencias divulgadas. Por ejemplo, en la invención encuentran uso variaciones conservadoras de las secuencias divulgadas que proporcionan una secuencia funcionalmente idéntica. Las variantes de las secuencias polinucleotídicas de ácido nucleico, en las que las variantes hibridan con al menos una secuencia divulgada encuentran uso en la invención.

Variantes conservadoras

Debido a la degeneración del código genético, las "sustituciones silenciosas" (es decir, sustituciones en una secuencia de ácido nucleico que no dan como resultado una alteración en un polipéptido codificado) son una característica implícita de todas las secuencias de ácido nucleico que codifican un aminoácido. De forma similar, en las "sustituciones de aminoácidos conservadoras", uno o unos pocos aminoácidos de una secuencia de aminoácidos se han sustituido por aminoácidos diferentes con propiedades muy similares, también se identifican de forma sencilla como muy similares a una construcción descrita. Tales variaciones conservadoras de cada secuencia descrita son una característica de la presente invención.

Las "variaciones conservadoras" de una secuencia de ácido nucleico particular se refieren a los ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas o, cuando el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas. Un experto reconocerá que las sustituciones, deleciones o adiciones individuales que alteran, añaden o eliminan un solo aminoácido o un pequeño porcentaje de aminoácidos (normalmente menos del 5 %, más normalmente menos del 4 %, 2 % o 1 %) en una secuencia codificada son variaciones modificadas de forma conservadora", resultan las alteraciones en la delección de un aminoácido, la adición de un aminoácido o la sustitución de un aminoácido por un aminoácido químicamente similar. Por lo tanto, las "variaciones conservadoras" de una secuencia polipeptídica enumerada de la presente invención incluyen sustituciones de un pequeño porcentaje, normalmente menos del 5 %, más normalmente menos del 2 % o del 1 %, de los aminoácidos de la secuencia polipeptídica, por un aminoácido del mismo grupo de sustitución conservadora. Finalmente, la adición de secuencias que no alteran la actividad codificada de una molécula de ácido nucleico, tal como la adición de una secuencia no funcional, es una variación conservadora del ácido nucleico básico.

Se conocen bien en la técnica tablas de sustitución conservadora que proporcionan aminoácidos funcionalmente similares, donde un residuo aminoacídico se sustituye por otro residuo aminoacídico que tiene propiedades químicas similares (por ejemplo, cadenas laterales aromáticas o cadenas laterales cargadas positivamente) y, por lo tanto, no cambia sustancialmente las propiedades funcionales de la molécula polipeptídica. A continuación se indican grupos ilustrativos que contienen aminoácidos naturales con propiedades químicas similares, donde la sustitución dentro de un grupo es una "sustitución conservadora".

TABLA 1

Sustituciones conservadoras de aminoácidos				
Cadenas Laterales No Polares y/o Alifáticas	Cadenas Laterales Polares SIN CARGA	Cadenas laterales aromáticas	Cadenas laterales cargadas positivamente	Cadenas laterales cargadas negativamente
Glicina Alanina	Serina Treonina	Fenilalanina	Lisina	Aspartato
Valina	Cisteína	Tirosina	Arginina	Glutamato
Leucina	Metionina	Triptófano	Histidina	
Isoleucina Prolina	Asparagina Glutamina			

Hibridación de ácidos nucleicos

Se puede usar hibridación comparativa para identificar ácidos nucleicos que encuentran uso en la invención, incluyendo variaciones conservadoras de ácidos nucleicos proporcionados en el presente documento y este procedimiento de hibridación comparativo es un procedimiento preferido para distinguir ácidos nucleicos que encuentran uso en la invención. Los ácidos nucleicos diana que hibridan con un ácido nucleico representado por los

de la lista de secuencias en condiciones de rigurosidad alta, ultra-alta y ultraultra-alta encuentran uso en la invención. Los ejemplos de tales ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones de ácido nucleico conservadoras o silenciosas en comparación con una secuencia de ácido nucleico dada.

- 5 Se dice que un ácido nucleico de ensayo hibrida específicamente con un ácido nucleico sonda cuando hibrida al menos la mitad de bien con la sonda como con la diana complementaria perfectamente coincidente, es decir, con una relación señal-ruido de al menos la mitad de la hibridación de la sonda con la diana en condiciones en las que la sonda perfectamente coincidente se une a la diana complementaria perfectamente coincidente con una relación entre señal e interferencias que es al menos de aproximadamente 5 veces a 10 veces la observada para la hibridación con cualquiera de los ácidos nucleicos diana no coincidentes.

15 Los ácidos nucleicos "hibridan" cuando se asocian, normalmente en solución. Los ácidos nucleicos hibridan debido a diversas fuerzas fisicoquímicas bien caracterizadas, tales como formación de enlaces de hidrógeno, exclusión de disolvente, adhesión de bases y similares. Se encuentra una guía extensa para la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* parte I capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), así como en Ausubel, supra. Hames y Higgins (1995) *Gene Probes 1* IRL Press en Oxford University Press, Oxford, Inglaterra, (Hames y Higgins 1) y Hames y Higgins (1995) *Gene Probes 2* IRL Press en Oxford University Press, Oxford, Inglaterra (Hames y Higgins 2) proporcionan detalles de la síntesis, marcaje, detección y cuantificación de ADN y ARN, incluyendo oligonucleótidos.

25 Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia Southern o Northern es formalina al 50 % con un 1 mg de heparina a 42 °C, realizándose la hibridación durante la noche. Un ejemplo de condiciones de lavado rigurosas es un lavado con SSC 0,2x a 65 °C durante 15 minutos (véase Sambrook, supra para una descripción de tampón SSC). A menudo, el lavado de alta rigurosidad está precedido por un lavado de baja rigurosidad para retirar la señal de fondo de la sonda. Un ejemplo de lavado de baja rigurosidad es SSC 2x a 40 °C durante 15 minutos. En general, una relación señal-ruido 5 veces mayor (o superior) que la observada para una sonda no relacionada en el ensayo de hibridación particular indica detección de una hibridación específica.

30 Las "condiciones rigurosas de lavado de hibridación" en el contexto de los experimentos de hibridación de ácidos nucleicos tales como hibridaciones Southern y Northern son dependientes de la secuencia y son diferentes con diferentes parámetros ambientales. Se encuentra una guía extensa sobre la hibridación de ácidos nucleicos en Tijssen (1993), supra y en Hames y Higgins, 1 y 2. Las condiciones de hibridación y lavado rigurosas se pueden determinar de forma sencilla empíricamente para cualquier ácido nucleico de ensayo. Por ejemplo, para determinar condiciones de hibridación y lavado rigurosas, las condiciones de hibridación y lavado se aumentan gradualmente (por ejemplo, aumentando la temperatura, disminuyendo la concentración de sal, aumentando la concentración de detergente y/o aumentando la concentración de disolventes orgánicos tales como formalina en la hibridación o el lavado), hasta que se cumpla un conjunto seleccionado de criterios. Por ejemplo, en condiciones de hibridación y lavado de alta rigurosidad, las condiciones de hibridación y lavado se aumentan gradualmente hasta que una sonda se une a una diana complementaria perfectamente coincidente con una relación entre señal e interferencias que es al menos 5 veces la observada para la hibridación de la sonda con una diana no coincidente.

45 Se seleccionan condiciones "muy rigurosas" para que sean iguales al punto de fusión térmico (T_m) de una sonda particular. La T_m es la temperatura (con fuerza iónica y pH definidos) a la que el 50 % de la secuencia de ensayo hibrida con una sonda perfectamente coincidente. Para los propósitos de la presente invención, generalmente se seleccionan las condiciones de hibridación y lavado "altamente rigurosas" para que sean aproximadamente 5 °C menores que la T_m para la secuencia específica a una fuerza iónica y pH definidos.

50 Las condiciones de hibridación y lavado "de rigurosidad ultra-alta" son aquellas en las que la rigurosidad de las condiciones de hibridación y lavado se aumentan hasta que la relación entre señal e interferencias para unir la sonda al ácido nucleico diana complementario perfectamente coincidente sea al menos 10 veces mayor que la observada para la hibridación con cualquiera de los ácidos nucleicos diana no coincidentes. Se dice que un ácido nucleico diana que hibrida con una sonda en tales condiciones, con una relación entre señal e interferencias de al menos la mitad de la del ácido nucleico diana complementario perfectamente coincidente, se une a la sonda en condiciones de rigurosidad ultra-alta.

60 De forma similar, se pueden determinar niveles incluso mayores de rigurosidad aumentando gradualmente las condiciones de hibridación y/o lavado del ensayo de hibridación pertinente. Por ejemplo, aquellos en los que la rigurosidad de las condiciones de hibridación y lavado se aumentan hasta que la relación señal-ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente coincidente sea al menos 10 veces, 20 veces, 50 veces, 100 veces, 500 veces o más veces la observada para la hibridación con cualquiera de los ácidos nucleicos diana no coincidentes. Se dice que un ácido nucleico diana que hibrida con una sonda en tales condiciones, con una relación entre señal e interferencias de al menos la mitad de la del ácido nucleico diana complementario perfectamente coincidente, se une a la sonda en condiciones de rigurosidad ultra-ultra-alta.

65

Los ácidos nucleicos que no hibridan entre sí en condiciones rigurosas todavía son sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto sucede, por ejemplo, cuando se crea una copia de un ácido nucleico usando la máxima degeneración de codón permitida por el código genético.

5 Subsecuencias únicas

En algunos aspectos, la invención usa un ácido nucleico que comprende una subsecuencia única en un ácido nucleico seleccionado de las secuencias de O-ARNt y O-RS divulgadas en el presente documento. La subsecuencia única es única en comparación con un ácido nucleico correspondiente a cualquier secuencia de ácido nucleico de ARNt o RS conocida previamente. Se puede realizar el alineamiento usando, por ejemplo, un ajuste BLAST para parámetros por defecto. Cualquier subsecuencia única es útil, por ejemplo, como una sonda para identificar los ácidos nucleicos de la invención.

De forma similar, la invención incluye un polipéptido que comprende una subsecuencia única en un polipéptido seleccionado de las secuencias de las O-RS divulgadas en el presente documento. En este caso, la subsecuencia única es única en comparación con un polipéptido correspondiente a cualquier secuencia polipeptídica conocida.

La invención también proporciona ácidos nucleicos diana que hibridan en condiciones rigurosas con un único oligonucleótido codificante que codifica una subsecuencia única en un polipéptido seleccionado de las secuencias de O-RS donde la subsecuencia única es única en comparación con un polipéptido correspondiente a cualquiera de los polipéptidos de control (por ejemplo, secuencias parentales de las que se obtuvieron las sintetasas de la invención, por ejemplo, mediante mutación). Las secuencias únicas se determinan como se ha señalado anteriormente.

25 Comparación, identidad y homología de secuencia

Los términos "idéntico" o porcentaje de "identidad", en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas, se refieren a dos o más secuencias o subsecuencias que son iguales o que tienen un porcentaje especificado de residuos aminoácidos o nucleótidos que son iguales cuando se comparan y alinean para la máxima correspondencia, como se mide usando uno de los algoritmos de comparación de secuencia que se describen a continuación (u otros algoritmos disponibles para los expertos en la materia) o mediante inspección visual.

La expresión "sustancialmente idéntico", en el contexto de dos ácidos nucleicos o polipéptidos (por ejemplo, ADN que codifica un O-ARNt o una O-RS o la secuencia de aminoácidos de una O-RS), se refiere a dos o más secuencias o subsecuencias que tienen al menos aproximadamente un 60 %, aproximadamente un 80 %, aproximadamente un 90-95 %, aproximadamente un 98 %, aproximadamente un 99 % o más de identidad de nucleótidos o residuos aminoácidos, cuando se comparan y se alinean para la máxima correspondencia, como se mide usando un algoritmo de comparación de secuencias o mediante inspección visual. Tales secuencias "sustancialmente idénticas" normalmente se consideran "homólogas", sin referencia a la ascendencia real. Preferentemente, la "identidad sustancial" existe a lo largo de una región de las secuencias que tiene una longitud de aproximadamente al menos 50 residuos, más preferentemente a lo largo de una región de aproximadamente al menos 100 residuos y mucho más preferentemente las secuencias son sustancialmente idénticas a lo largo de aproximadamente al menos 150 residuos o a lo largo de la longitud completa de las dos secuencias que se tienen que comparar.

Las proteínas y/o secuencias de proteínas son "homólogas" cuando se obtienen, de forma natural o de forma artificial, a partir de una proteína o secuencia de proteína ancestral común. De forma similar, los ácidos nucleicos y/o las secuencias de ácido nucleico son homólogas cuando se obtienen, de forma natural o de forma artificial, a partir de un ácido nucleico o una secuencia de ácido nucleico ancestral común. Por ejemplo, cualquier ácido nucleico de origen natural se puede modificar mediante cualquier procedimiento de mutagénesis disponible para incluir uno o más codones selectores. Cuando se expresa, este ácido nucleico mutado codifica un polipéptido que comprende una o más aminoácidos no naturales. El proceso de mutación puede, por supuesto, alterar adicionalmente uno o más codones convencionales, cambiando también de esta manera uno o más aminoácidos convencionales en la proteína mutante resultante. La homología generalmente se deduce a partir de la similitud de secuencia entre dos o más ácidos nucleicos o proteínas (o secuencias de los mismos). El porcentaje preciso de similitud entre secuencias que es útil para establecer la homología varía con el ácido nucleico y la proteína en cuestión, pero de forma rutinaria se usa únicamente un 25 % de similitud de secuencia para establecer la homología. También se pueden usar niveles mayores de similitud de secuencia, por ejemplo, un 30 %, un 40 %, un 50 %, un 60 %, un 70 %, un 80 %, un 90 %, un 95 % o un 99 % o más, para establecer la homología. En este documento se describen procedimientos para determinar porcentajes de similitud de secuencia (por ejemplo, BLASTP y BLASTN usando parámetros por defecto) y están generalmente disponibles.

Para la comparación de secuencias y determinación de homología, normalmente una secuencia actúa como una secuencia de referencia con la que se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencias, se introducen las secuencias de ensayo y de referencia en un ordenador, se diseñan

las coordenadas de subsecuencia, si es necesario, y se diseñan parámetros de programa de algoritmo de secuencia. Después, el algoritmo de comparación de secuencia calcula el porcentaje de identidad de secuencia para la o las secuencias de ensayo con respecto a las secuencias de referencia, basándose en los parámetros de programa diseñados.

5 Se puede realizar un alineamiento óptimo de secuencias para la comparación, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), mediante el algoritmo de alineamiento de homología de Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), mediante el procedimiento de búsqueda de similitud de Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), mediante aplicaciones informáticas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Paquete de Software de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, WI) o mediante inspección visual (véase en general *Current Protocols in Molecular Biology*, Ausubel et al., eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., suplemento hasta 2004).

15 Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y de similitud de secuencia es el algoritmo BLAST, que se describe en Altschul et al., *J. Mol. Biol.* 215:403410 (1990). El software para realizar análisis BLAST está disponible públicamente por el National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). Este algoritmo implica identificar en primer lugar pares de secuencias con alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia problema, que se ajustan o satisfacen alguna puntuación umbral T valorada positivamente cuando se alinean con una palabra de la misma longitud en una secuencia de base de datos. T se denomina el umbral de puntuación de la palabra vecina (Altschul et al., supra). Estos aciertos de palabra vecina iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largos que los contengan. Después, los aciertos de palabra se extienden en ambas direcciones a lo largo de cada secuencia hasta que se pueda aumentar la puntuación de alineamiento acumulativa. Las puntuaciones acumulativas se calculan usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y N (puntuación de penalización para residuos no coincidentes; siempre < 0). Para secuencias de aminoácidos se usa una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabra en cada dirección se detiene cuando: la puntuación de alineamiento acumulativa cae en la cantidad X desde su máximo valor conseguido; la puntuación acumulativa tiende a cero o a un valor inferior debido a la acumulación de uno o más alineamientos de residuos de puntuación negativa; o se alcanza el extremo de cualquier secuencia. Los parámetros de algoritmo BLAST W, T y X determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, una esperanza (E) de 10, un límite de 100, M=5, N=-4 y una comparación de las dos cadenas. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una esperanza (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5877 (1993)). Una medición de la similitud proporcionada por el algoritmo BLAST es la menor probabilidad total (P(N)), que proporciona una indicación de la probabilidad con la que ocurriría por casualidad una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, se considera que un ácido nucleico es similar a una secuencia de referencia si la menor probabilidad total, en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia, es menor de aproximadamente 0,1, más preferentemente menor de aproximadamente 0,01 y mucho más preferentemente menor de aproximadamente 0,001..

Mutagénesis y otras técnicas de biología molecular

50 Los polinucleótidos y polipéptidos de la invención y usados en la invención se pueden manipular usando técnicas de biología molecular. Los textos generales que describen técnicas de biología molecular incluyen Berger y Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* volumen 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., *Molecular Cloning - A Laboratory Manual* (3a Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 2001 ("Sambrook") y *Current Protocols in Molecular Biology*, F. M. Ausubel et al., eds., *Current Protocols*, una sociedad conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (complementado por 2003) ("Ausubel"). Estos textos describen mutagénesis, el uso de vectores, promotores y muchos otros aspectos pertinentes relacionados con, por ejemplo, la generación de genes que incluyen codones selectores para la producción de proteínas que incluyen aminoácidos no naturales, ARNt ortogonales, sintetasas ortogonales y pares de los mismos.

60 En la invención se usan diversos tipos de mutagénesis, por ejemplo, para mutar moléculas de ARNt, para producir bibliotecas de ARNt, para producir bibliotecas de sintetasas, o para insertar codones selectores que codifican aminoácidos no naturales en una proteína o un polipéptido de interés. Estos incluyen, pero no se limitan a, mutagénesis puntual aleatoria dirigida, recombinación homóloga, barajado de ADN u otros procedimientos de mutagénesis recursiva, construcción quimérica, mutagénesis usando plantillas que contienen uracilo, mutagénesis dirigida por oligonucleótido, mutagénesis de ADN modificado por fosforotioato, mutagénesis usando ADN dúplex discontinuo o similares o cualquier combinación de los mismos. Otros procedimientos adecuados incluyen

reparación de emparejamiento erróneo, puntual, mutagénesis usando cepas de hospedador deficientes en reparación, selección por restricción y purificación por restricción, mutagénesis por delección, mutagénesis por síntesis génica total, reparación de rupturas de cadena doble o similares. También se incluye en la presente invención, por ejemplo, la mutagénesis que implica construcciones químicas. En una realización, la mutagénesis se puede guiar por una información conocida de la molécula de origen natural o la molécula natural alterada o mutada, por ejemplo, por información de la secuencia, de comparaciones de secuencias, de propiedades físicas, de la estructura cristalina o similares.

Las células huésped se crean mediante ingeniería genética (por ejemplo, se transforman, transducen o transfectan) con los polinucleótidos de la invención o con construcciones que incluyen un polinucleótido de la invención, por ejemplo, un vector de la invención, que puede ser, por ejemplo, un vector de clonación o un vector de expresión. Por ejemplo, las regiones codificantes para los ARNt ortogonales, la ARNt sintetasa ortogonal y la proteína que se tiene que derivatizar están unidas de manera operativa a elementos de control de la expresión génica que son funcionales en la célula huésped deseada. Los vectores típicos contienen terminadores de la transcripción y traducción, secuencias de inicio de la transcripción y traducción y promotores útiles en la regulación de la expresión del ácido nucleico diana particular. Los vectores comprenden opcionalmente casetes de expresión genéricos que contienen al menos una secuencia terminadora independiente, secuencias que permiten la replicación del casete en eucariotas o procariotas, o ambas cosas (por ejemplo, vectores lanzadera) y marcadores de selección tanto para sistemas procariotas como para sistemas eucariotas. Los vectores son adecuados para replicación y/o integración en procariotas, eucariotas o preferentemente en ambos tipos celulares. Véase Gillman & Smith, *Gene* 8:81 (1979); Roberts, et al., *Nature*, 328: 731 (1987); Schneider, B., et al., *Protein Expr. Purif.* 6:435:10 (1995); Ausubel, Sambrook, Berger (todos supra). El vector puede estar, por ejemplo, en forma de un plásmido, una bacteria, un virus, un polinucleótido desnudo o un polinucleótido conjugado. Los vectores son introducidos en células y/o microorganismos por procedimientos convencionales incluyendo electroporación (From et al., *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infección por vectores virales, penetración balística a gran velocidad con partículas pequeñas con el ácido nucleico en el interior de la matriz de pequeñas perlas o partículas o sobre la superficie (Klein et al., *Nature* 327, 70-73 (1987)) y/o similares.

Se ha desarrollado un sistema de plásmido único muy eficiente y versátil para la incorporación específica de sitio de aminoácidos no naturales en proteínas en respuesta al codón de terminación ámbar (UAG) en *E. coli*. En el nuevo sistema, el par de tRNA^{tyr}(CUA) supresor y tirosil-ARNt sintetasa de *M. jannaschii* está codificado en un plásmido único, que es compatible con la mayoría de los vectores de expresión en *E. coli*. Se construyó un operón de ARNt monocistrónico bajo el control del promotor proK y el terminador para una estructura secundaria óptima y el procesamiento del ARNt. La introducción de una forma mutada del promotor glnS para la sintetasa tuvo como resultado un incremento significativo de la eficacia de la supresión y de la fidelidad. También se obtuvieron incrementos de la supresión mediante múltiples copias del gen del ARNt así como mediante una mutación específica (D286R) en la sintetasa (Kobayashi et al., "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion," *Nat. Struct. Biol.*, 10(6):425-432 [2003]). La generalidad del sistema optimizado también se demostró mediante la incorporación precisa y altamente eficaz de varios aminoácidos no naturales diferentes cuyas utilidades únicas en el estudio de la función y estructura de las proteínas ya se habían demostrado anteriormente.

Se proporciona un catálogo de Bacterias y Bacteriófagos útiles para la clonación, por ejemplo, en la ATCC, por ejemplo, *The ATCC Catalogue of Bacteria and Bacteriophage* (1996) Gherna et al. (eds) publicado por la ATCC. También se encuentran procedimientos básicos adicionales para la secuenciación, la clonación y otros aspectos de biología molecular y consideraciones teóricas subyacentes en Sambrook (supra), Ausubel (supra, y en Watson et al. (1992) *Recombinant DNA Second Edition Scientific American Books*, NY. Además, esencialmente cualquier ácido nucleico (y prácticamente cualquier ácido nucleico marcado, bien convencional o no convencional) se puede encargar a petición del cliente o de forma convencional en cualquiera de diversas fuentes comerciales, tales como la Midland Certified Reagent Company (Midland, TX mcr.com), The Great American Gene Company (Ramona, CA disponible en la World Wide Web en genco.com), ExpressGen Inc. (Chicago, IL disponible en la World Wide Web en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchas otras.

Las células huésped creadas por ingeniería genética se pueden cultivar en medios nutrientes convencionales modificados de forma apropiada para dichas actividades tales como, por ejemplo, etapas de identificación, activación de promotores o selección de transformantes. Estas células se pueden cultivar opcionalmente en organismos transgénicos. Otras referencias útiles, por ejemplo, para el aislamiento y cultivo de células (por ejemplo, para el aislamiento posterior de ácidos nucleicos) incluyen Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, New York y las referencias citadas en ese documento; Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, NY; Gamburg y Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Editorial Springer (Berlin Heidelberg New York) y Atlas y Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL.

Proteínas y polipéptidos de interés

También constituyen una característica de la invención procedimientos de producir una proteína que comprenden un aminoácido no natural en una posición especificada. Por ejemplo, un procedimiento puede incluir dejar crecer, en un medio apropiado, una célula (p. ej., en una célula de *E.coli*), en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína; y proporcionar el aminoácido natural, en el que la célula comprende además: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector; y una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila de forma preferente el OARNt con un aminoácido no natural. La proteína producida de este modo en *E. coli* comprende un aminoácido no natural en la posición correspondiente al codón selector. Después, la proteína se puede hacer reaccionar opcionalmente en condiciones en las que el aminoácido no natural sufre modificación covalente de modo que produce una proteína postraduccionalmente.

En determinadas realizaciones, la O-RS comprende un sesgo para la aminoacilación del O-ARNt afín sobre cualquier ARNt endógeno en un sistema de expresión. La proporción relativa entre el O-ARNt y el ARNt endógeno cargado por la O-RS cuando el O-ARNt y la O-RS están presentes en concentraciones molares iguales, es superior a 1:1, preferentemente de al menos aproximadamente 2:1, más preferentemente de 5:1, todavía más preferentemente de 10:1, todavía más preferentemente 20:1, todavía más preferentemente 50:1, todavía más preferentemente 75:1, todavía más preferentemente 95:1, 98:1, 99:1, 100:1, 500:1, 1,000:1, 5,000:1 o superior.

También constituye una característica de la invención la proteína que comprende un aminoácido no natural en una posición especificada que está modificada postraduccionalmente. La proteína se produce en una célula, por ejemplo una célula de *E. coli*. Los pares de O-ARNt/O-RS también residen en la célula y usan la maquinaria de traducción de la célula, lo que tiene como resultado *in vivo* la incorporación de un aminoácido no natural en una proteína de fusión en respuesta a un codón selector. Se conoce la capacidad de un sistema de O-ARNt/O-RS para funcionar en una célula huésped para incorporar una amplia variedad de aminoácidos no naturales que se pueden modificar postraduccionalmente. Véase, por ejemplo, Chin et al (2002) Journal of the American Chemical Society 124:9026-9027; Chin and Schultz, (2002) ChemBioChem 11:1135-1137; Chin, et al., (2002) PNAS United States of America 99:11020-11024; Wang and Schultz, (2002) Chem. Comm., 1-10; Wang and Schultz "Expanding the Genetic Code," Angewandte Chemie Int. Ed., 44(1):34-66 (2005); Xie and Schultz, "An Expanding Genetic Code," Methods 36:227-238 (2005); and Deiters et al, Bioorganic & Medicinal Chemistry Letters 15:1521-1524 (2005).

Véase también los sistemas ortogonales de aminoácidos no naturales en las publicaciones internacionales WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS;" el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" el documento WO 2004/094593, titulado ""EXPANDING THE EUKARYOTIC GENETIC CODE;" el documento WO 2005/019415, presentado el 7 de julio de 2004; el documento WO2005/007870, presentado el 7 de julio de 2004; el documento WO 2005/007624, presentado el 7 de julio de 2004; el documento WO 2006/034332, presentado el 20 de septiembre de 2005; y el documento WO 2006/110182 titulado "ORTHOAGONAL TRANSLATION COMPONENTS .

La incorporación de un aminoácido no natural se puede realizar para, por ejemplo, adaptar cambios en la estructura y/o función de la proteína, por ejemplo, para cambiar el tamaño, acidez, nucleofilia, formación de enlaces de hidrógeno, hidrofobia, accesibilidad a los sitios diana de la proteasa, llegar a un residuo (por ejemplo, para una serie de proteínas), incorporación de marcadores o de grupos reactivos etc. Las proteínas que incluyen un aminoácido no natural pueden tener propiedades catalíticas o físicas mejoradas o incluso completamente nuevas. Por ejemplo, mediante la inclusión de un aminoácido no natural en una proteína se modifican opcionalmente las siguientes propiedades: toxicidad, biodistribución, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, capacidad catalítica, semivida (por ejemplo, semivida en suero), capacidad para reaccionar con otras moléculas, por ejemplo, de forma covalente o de forma no covalente, y similares. Las composiciones que incluyen proteínas que incluyen al menos un aminoácido no natural son útiles para, por ejemplo, nuevas enzimas terapéuticas, de diagnóstico, catalíticas, enzimas industriales, proteínas de unión (por ejemplo, anticuerpos) y, por ejemplo, para el estudio de la estructura y función de proteínas. Véase, por ejemplo, Dougherty, (2000) Unnatural Amino Acids as Probes of Protein Structure and Function, Current Opinion in Chemical Biology, 4:645-652. Las proteínas pueden comprender un aminoácido no natural que se puede modificar postraduccionalmente (p. ej., por una cicloadición [3+2] o una modificación de Staudinger) se pueden someter a ingeniería genética para contener cualquier funcionalidad deseada que se pueda acoplar a la pareja de reacción. La naturaleza de la pareja de reacción no está limitada de ningún modo, a excepción únicamente de que comprende un resto reactivo adecuado que tiene como resultado una unión covalente al residuo de aminoácido ni natural en el polipéptido.

En algunos aspectos, una composición incluye al menos una proteína con al menos una, por ejemplo al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve o al menos diez o más aminoácidos no naturales. Los aminoácidos no naturales pueden ser iguales o diferentes, por ejemplo puede haber 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10, o más sitios diferentes en la proteína que comprenden 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 o más aminoácidos no naturales diferentes. En otro aspecto, una composición incluye una proteína

con al menos uno, pero menos que todos, de un aminoácido concreto presente en la proteína es un aminoácido no natural. Para una proteína dada con más de un aminoácido no natural, los aminoácidos no naturales pueden ser idénticos o diferentes (p. ej., la proteína puede incluir dos o más tipos diferentes de aminoácidos no naturales o puede incluir dos de los mismos aminoácidos no naturales). Para una proteína dada con más de dos aminoácidos no naturales los aminoácidos no naturales pueden ser iguales, diferentes o una combinación de un múltiple aminoácido no natural del mismo tipo con al menos aminoácido no natural diferente.

Usando las composiciones y los procedimientos del presente documento se puede producir esencialmente cualquier proteína (o parte de la misma) que incluya un aminoácido no natural (y cualquier ácido nucleico codificante correspondiente, por ejemplo, que incluya uno o más codones selectores). No se ha intentado identificar los cientos de miles de proteínas conocidas que se pueden modificar para incluir uno o más aminoácidos no naturales, por ejemplo, adaptando cualquier procedimiento de mutación disponible para incluir uno o más codones selectores apropiados en un sistema de traducción pertinente. Los depósitos de secuencia comunes para proteínas conocidas incluyen GenBank EMBL, DDBJ y NCBI. Se pueden identificar otros depósitos de forma sencilla buscando en internet.

Normalmente, las proteínas son, p. ej., al menos un 60 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, o al menos un 99 % o más idénticas a cualquiera de las proteínas disponibles (p.ej., una proteína terapéutica, una proteína de diagnóstico, una enzima industrial o una porción de la misma, y similares), y comprenden uno o más aminoácidos no naturales. Ejemplos de proteínas terapéuticas, de diagnóstico y otras proteínas que se pueden modificar para que comprendan uno o más aminoácidos no naturales incluyen, entre otras, por ejemplo, Alfa-1 antitripsina, Angiostatina, factor antihemolítico, anticuerpos (más adelante se encuentran detalles adicionales sobre anticuerpos), Apoproteína, factor natriurético atrial, polipéptido natriurético atrial, péptidos atriales, quimiocinas C-X-C (p. ej., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP 4, SDF-1, PF4, MIG), Calcitonina, quimiocinas CC β p. ej., proteína-1 quimioatrayente de monocitos, proteína-2 quimioatrayente de monocitos, proteína-3 quimioatrayente de monocitos, proteína-1 alfa inflamatoria de monocitos, proteína-1 beta inflamatoria de monocitos, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, ligando CD40, ligando C-kit, colágeno, factor estimulador de colonias (CSF), factor 5a de complemento, inhibidor de complemento, receptor 1 de complemento, citoquinas, (p. ej., péptido 78 activador de neutrófilos epiteliales, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP1), factor de crecimiento epidérmico (EGF), eritropoietina ("EPO"), toxinas exfoliantes A y B, Factor IX, Factor VII, Factor VIII, Factor X, factor de crecimiento de fibroblastos (FGF), fibrinógeno, fibronectina, G-CSF, GM-CSF, glucocerebrosidasa, gonadotropina, factores de crecimiento, proteínas Hedgehog (p. ej., Sonic, Indian, Desert), hemoglobina, factor de crecimiento de hepatocitos (HGF), hirudina, albúmina de suero humano, insulina, factor de crecimiento similar a insulina (IGF), interferones (p. ej., IFN- α , IFN- β , IFN- γ), interleucinas (p. ej., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, etc.), factor de crecimiento de queratinocitos (KGF), lactoferrina, factor inhibidor de leucemia, luciferasa, neurturina, factor inhibidor de neutrófilos (NIF), oncostatina M, proteína osteogénica, hormona paratiroide, PD-ECSF, PDGF, hormonas peptídicas (p. ej., hormona de crecimiento humano), pleiotropina, proteína A, proteína G, exotoxinas pirogénicas A, B y C, relaxina, renina, SCF, receptor I de complemento soluble, I-CAM 1 soluble, receptores solubles de interleucina (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), receptor soluble de TNF, somatomedina, somatostatina, somatotropina, estreptoquinasa, superantígenos, es decir, enterotoxinas de *Staphylococcus* (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), Superóxido dismutasa (SOD), toxina del síndrome de shock tóxico (TSST-1), timosina alfa 1, activador de plasminógeno de tejido, factor beta de necrosis tumoral (TNF beta), receptor del factor de necrosis tumoral (TNFR), factor-alfa de necrosis tumoral (TNF alpha), factor de crecimiento del endotelio vascular (VEGEG), uroquinasa y muchas otras.

Una clase de proteínas que se puede obtener usando las composiciones y los procedimientos para la incorporación in vivo de aminoácidos no naturales descritos en este documento incluye moduladores de la transcripción o una parte de los mismos. Ejemplos de moduladores de la transcripción incluyen genes y proteínas modificadoras de la transcripción que modulan el crecimiento, la diferenciación, o la regulación celular o similares. Los moduladores de la transcripción se encuentran en procariotas, virus y eucariotas, incluyendo hongos, vegetales, levaduras, insectos y animales, incluyendo mamíferos, proporcionando una amplia serie de dianas terapéuticas. Se entenderá que los activadores de la expresión y de transcripción regulan la transcripción mediante muchos mecanismos, por ejemplo, uniéndose a receptores, estimulando una cascada de transducción de señales, regulando la expresión de factores de transcripción, uniéndose a promotores y potenciadores, uniéndose a proteínas que se unen a promotores y potenciadores, desenrollando el ADN, por corte y empalme del pre-ARNm, poliadenilando el ADN y degradando el ARN.

Una clase de proteínas de la invención (por ejemplo, proteínas con uno o más aminoácidos no naturales) incluye activadores de expresión tales como citoquinas, moléculas inflamatorias, factores de crecimiento, sus receptores y productos oncogénicos, por ejemplo, interleucinas (por ejemplo, IL-1, IL-2, IL-8, etc.), interferones, FGF, IGF-I, IGF II, FGF, PDGF, TNF, TGF- α TGF- β , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1 e hialurina/CD44; moléculas de transducción de señales y productos oncogénicos correspondientes, por ejemplo, Mos, Ras, Raf y Met; y activadores y supresores de la transcripción, por ejemplo, p53, Tat, Fos, Myc, Jun, Myb, Rel y receptores de hormonas esteroideas tales como los receptores de estrógenos, progesterona, testosterona, aldosterona, el ligando del receptor de LDL y corticosterona.

La invención también proporciona enzimas (por ejemplo, enzimas industriales) o partes de las mismas con al menos un aminoácido no natural. Los ejemplos de enzimas incluyen, pero no se limitan a, por ejemplo, amidasas, aminoácido racemasas, acilasas, deshalogenasas, dioxigenasas, diarilpropano peroxidadas, epimerasas, epóxido hidrolasas, esteradas, isomerasas, quinasas, glucosa isomerasas, glicosidasas, glicosil transferasas, haloperoxidasas, monooxigenasas (por ejemplo, p450), lipasas, lignina peroxidadas, nitrilo hidratadas, nitrilasas, proteasas, fosfatasas, subtilisinas, transaminasas y nucleasas.

Muchas de estas proteínas están disponibles en el mercado (véase, por ejemplo, el catálogo y la lista de precios de Sigma BioSciences 2003) y las secuencias de proteínas y genes correspondientes y, normalmente, muchas variantes de las mismas son bien conocidas (véase, por ejemplo, Genbank). Cualquiera de ellas se puede modificar mediante la inserción de uno o más aminoácidos no naturales de acuerdo con la invención, por ejemplo, para alterar la proteína con respecto a una o más propiedades terapéuticas, de diagnóstico o enzimáticas de interés. Los ejemplos de propiedades terapéuticamente pertinentes incluyen la semivida sérica, semivida de almacenamiento, estabilidad, inmunogenicidad, actividad terapéutica, detectabilidad (por ejemplo, mediante la inclusión de grupos indicadores (por ejemplo, marcadores o sitios de unión a marcadores) en los aminoácidos no naturales, reducción de DL50 u otros efectos secundarios, capacidad de entrar en el cuerpo a través del tracto gástrico (por ejemplo, disponibilidad oral) o similares. Los ejemplos de las propiedades de diagnóstico incluyen semivida en almacenamiento, estabilidad, actividad de diagnóstico, detectabilidad o similares. Los ejemplos de propiedades enzimáticas pertinentes incluyen semivida en almacenamiento, estabilidad, actividad enzimática, capacidad de producción o similares.

También se puede modificar diversas proteínas diferentes para incluir uno o más aminoácidos no naturales u otro aminoácido no natural de la invención. Por ejemplo, la invención puede incluir sustituir uno o más aminoácidos naturales en una o más proteínas de vacuna con un aminoácido no natural, por ejemplo, en proteínas de hongos infecciosos, por ejemplo, *Aspergillus*, especie de *Candida*; bacterias, particularmente *E. coli*, que sirve como modelo para bacterias patógenas, así como bacterias médicamente importantes tales como *Staphylococci* (por ejemplo, *aureus*) o *Streptococci* (por ejemplo, *pneumoniae*); protozoos tales como esporozoos (por ejemplo, *Plasmodia*), rizópodos (por ejemplo, *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); virus tales como virus de ARN (+) (los ejemplos incluyen Poxvirus, por ejemplo, *vaccinia*; Picornavirus, por ejemplo, polio; Togavirus, por ejemplo, rubéola; Flavivirus, por ejemplo, VCH; y Coronavirus), virus de ARN (-) (por ejemplo, Rhabdovirus, por ejemplo, VSV; Paramixovirus, por ejemplo, RSV; Orthomyxovirus, por ejemplo, el virus de la gripe; Bunyavirus; y Arenavirus), virus ADNbc (por ejemplo, Reovirus), virus de ARN a ADN, es decir, Retrovirus, por ejemplo, VIH y HTLV, y determinados virus de ADN a ARN tales como el virus de la Hepatitis B.

Las proteínas relacionadas con la agricultura tales como las proteínas de resistencia a insectos (por ejemplo, las proteínas Cry), enzimas de producción de almidón y lípidos, toxinas de plantas e insectos, proteínas de resistencia a toxinas, proteínas de destoxicación de micotoxina, enzimas de crecimiento de plantas (por ejemplo, ribulosa 1,5-Bisfosfato Carboxilasa/Oxigenasa, "RUBISCO"), lipoxigenasa (LOX) y Fosfoenolpiruvato (PEP) carboxilasa son también dianas apropiadas para la modificación con un aminoácido no natural.

En determinadas realizaciones, la proteína (o una parte de la misma) de interés está codificada por un ácido nucleico. Normalmente, el ácido nucleico comprende al menos un codón selector, al menos dos codones selectores, al menos tres codones selectores, al menos cuatro codones selectores, al menos cinco codones selectores, al menos seis codones selectores, al menos siete codones selectores, al menos ocho codones selectores, al menos nueve codones selectores, diez o más codones selectores.

Los genes que codifican proteínas o polipéptidos de interés se pueden mutar usando procedimientos bien conocidos por los expertos en la materia y descritos en este documento en "Mutagénesis y otras Técnicas de Biología Molecular" para incluir, por ejemplo, uno o más codones selectores para la incorporación de un aminoácido no natural. Por ejemplo, se muta un ácido nucleico para una proteína de interés con el fin de que incluya uno o más codones selectores, asegurando la inserción de uno o más aminoácidos no naturales. La invención incluye cualquiera de dichas variantes, por ejemplo, mutantes, versiones de cualquier proteína, por ejemplo, que incluye al menos un aminoácido no natural. De forma similar, la invención también incluye ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifique uno o más aminoácidos no naturales.

Para fabricar una proteína que incluye un aminoácido no natural modificado postraduccionalmente se pueden usar células y organismos hospedadores que estén adaptados para la incorporación *in vivo* del aminoácido no natural mediante pares ortogonales de ARNt/RS. Las células huésped se crean mediante ingeniería genética (por ejemplo, se transforman, transducen o transfectan) con uno o más vectores que expresan el ARNt ortogonal, la ARNt sintetasa ortogonal y un vector que codifica la proteína que se tiene que derivatizar. Cada uno de estos componentes pueden estar en el mismo vector o cada uno puede estar en un vector separado, o los dos componentes pueden estar en un vector y el tercer componente en un segundo vector. El vector puede estar, por ejemplo, en forma de un plásmido, una bacteria, un virus, un polinucleótido desnudo o un polinucleótido conjugado.

Definición de polipéptidos mediante inmunorreactividad

Como los polipéptidos de la invención proporcionan diversas nuevas secuencias polipeptídicas (comprendiendo, por ejemplo, aminoácido no naturales en el caso de proteínas sintetizadas en los sistemas de traducción de este documento o, por ejemplo, en el caso de las nuevas sintetasas, nuevas secuencias de aminoácidos convencionales), los polipéptidos también proporcionan nuevas características estructurales que se pueden reconocer, por ejemplo, en ensayos inmunológicos. La generación de antisueros que se unen específicamente a los polipéptidos de la invención, así como los polipéptidos que se unen por dichos antisueros, constituyen una característica de la invención. El término "anticuerpo", como se usa en este documento, incluye, pero no se limita a, un polipéptido codificado sustancialmente por un gen de inmunoglobulina o genes de inmunoglobulina o fragmentos de los mismos que se unen específicamente y reconocen un analito (antígeno). Los ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos, de cadena única y similares. Los fragmentos de inmunoglobulinas, incluyendo fragmentos Fab y fragmentos producidos mediante una biblioteca de expresión, incluyendo presentación en fagos, también están incluidos en el término "anticuerpo" como se usa en este documento. Véase, por ejemplo, en Paul, *Fundamental Immunology*, 4^a Ed. 1999, Raven Press, New York, la estructura y terminología de anticuerpos.

Con el fin de producir los antisueros para uso en un inmunoensayo, se producen uno o más polipéptidos inmunogénicos y se purifican según se describe en el presente documento. Por ejemplo, la proteína recombinante se puede producir en una célula recombinante. Una cepa innata de ratón usada en este ensayo ya que los resultados son más reproducibles debido a la identidad genética virtual del ratón) se inmuniza con la(s) proteína(s) inmunogénica(s) en combinación con un adyuvante estándar, tal como adyuvante de Freund, y un protocolo de inmunización de ratón estándar (véase, p. ej., Harlow y Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, para una descripción estándar de generación de anticuerpos, formatos de inmunoensayo y condiciones que se pueden usar para determinar la inmunorreactividad específica. Detalles adicionales sobre proteínas, anticuerpos, antisueros, etc. se pueden encontrar en las publicaciones internacionales números WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE;" WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" WO 2004/035605, titulada "GLYCOPROTEIN SYNTHESIS;" y WO 2004/058946, titulada "PROTEIN ARRAYS."

Ejemplos

Los ejemplos siguientes se ofrecen para ilustrar, pero no para limitar, la invención reivindicada.

Ejemplo 1

Construcción de un sistema de un solo plásmido para la expresión de polipéptidos que comprenden aminoácidos no naturales

El presente ejemplo describe la construcción de un plásmido que codifica ambos miembros de un par de aminoacil-ARNt ortogonal y aminoacil-ARNt sintetasa para la incorporación de *p*-benzoil-L-fenilalanina.

Se construyó un plásmido (denominado pYR-BpaRS1) que contenía secuencias de nucleótidos que codifican ambos componentes de un par de traducción ortogonal que funciona en una célula huésped de *E. coli*. A saber, estos dos componentes son los ARNt ortogonales *Mj*tRNA-Tyr(CUA) y la *Mj*TyrRS sintetasa mutante (BpaRS) que aminoacila específicamente el ARNt ortogonal con el aminoácido de fotoreticulación, *p*-benzoil-L-fenilalanina (Bpa, véase la **FIG. 1**). (Véase, Chin et al., *Proc. Natl. Acad. Sci. USA* 99:11020-11024 (2002)).

Para determinar la eficacia de supresión de este sistema PLASMÍDICO pYR-BpaRS1 se midió la actividad de la β -galactosidasa para células TOP410 *E. coli* (Invitrogen™) cotransformadas con pYR-BpaRS1 y un plásmido indicador de *lacZ* que codifica la β -galactosidasa con una mutación ámbar en un sitio permisivo en la secuencia líder del gen *lacZ*. Por desgracia, cuando las células se cultivaron en presencia de Bpa 1mM se observaron niveles muy bajos de actividad β -galactosidasa (**FIG. 2**). Los intentos de aumentar la eficacia de supresión modificando las secuencias flanqueantes del gen de ARNt no fueron infructuosos.

Con el fin de mejorar esta eficacia de la supresión se construyó un nuevo operón del ARNt supresor ámbar con un promotor y terminador del ARNt de *E. coli* natural. Una visión de los genes de ARNt de *E. coli* reveló que los ARNt de prolina de *E. coli* tienen el mismo par C1-G72 que los ARNt de Archaea; este par de bases es un determinante principal de la identidad para el reconocimiento selectivo de *Mj*tRNA-Tyr(CUA) por *Mj*TyrRS en *E. coli* (Wang y Schultz, *Chem. Chem. Biol.* 8:883-890 (2001)). A la luz de esta observación se construyó un gen del ARNt supresor ámbar sintético de un modo tal que el gen de *Mj*tRNA-Tyr(CUA) sustituya al gen de *proK* de *E. coli* de la misma longitud (77 nucleótidos) en el operón *proK* monocistrónico. Dado que el gen *proK* codifica el ARNt que reconoce el codón de la prolina más usado (CCG) en *E. coli* (Nakamura et al., *Nucleic Acids Res.* 28:292 (2000)), los inventores esperaban que el gen *Mj*tRNA-Tyr(CUA) se transcribiera con eficacia bajo el control del promotor de *proK*. También se incluyó en el constructo génico sintético un sitio de unión FOS localizado de forma natural cadena arriba del promotor de *proK* para potenciar la transcripción (Muskhelishvili et al., *EMBO J.* 16:3655-3665 (1997)). El vector de

expresión final (pYR-BpaRS5) se generó sustituyendo el operón del ARNt supresor original en pYR-BpaRS 1 con el gen *MjtRNA-Tyr(CUA)* bajo el control del promotor *proK* y el terminador. Cuando se transforma en *E. coli*, este plásmido produjo un incremento de dos veces (con respecto al gen del ARNt original bajo el control del promotor 1pp y el terminador *rrnC*) en la expresión de *MjtRNA-Tyr(CUA)* como se observa mediante análisis de tipo northern (véase la **FIG. 3**). Este incremento en la expresión corresponde a una eficacia de la supresión del 8 % (respecto a la expresión de la β -galactosidasa de tipo silvestre) determinado mediante el ensayo de la actividad de la β -galactosidasa (**FIG. 2**).

Ejemplo 2

Generación de genes y promotores de la sintetasa mejorados

El presente ejemplo describe la construcción de sistemas mejorados para la expresión de componentes de sistemas de traducción ortogonal, en la que se determinan los efectos de las mutaciones en el gen de la sintetasa y el promotor de la sintetasa.

Kobayashi *et al.* Notificaron anteriormente que una única sustitución de un aminoácido de Asp286 en Arg (D286R) en *MjTyrRS* aumentó la tasa de aminoacilación (kcat/Km 67 veces mayor) de *MjtRNA-Tyr(CUA)* *in vitro*, principalmente debido al aumento del reconocimiento (Km 57 veces menor) del anti-codon (CUA) en el ARNt supresor ámbar por la sintetasa afin (Kobayashi *et al.*, Nat. Struct. Biol. 10:425-432 (2003)). Usando esta información se introdujo esta misma mutación D286R en el gen *BpaRS* en pYR-BpaRS5 mediante mutagénesis dirigida a sitio. De hecho, el mutante D286R de *BpaRS* condujo a un incremento de 4,5 veces en la actividad de la β -galactosidasa (véase la **FIG. 2**).

Anteriormente se ha demostrado que un promotor de *glnS* mutante (SEC ID N°: 12) que tiene una secuencia TATC en lugar de GATC en la región -10 que aumentaba la expresión génica (Plumbridge and Söll, Biochimie 69:539-541 (1987)). Sabiendo esto, el promotor de *glnS* de tipo silvestre en pYR-BpaRS5 se sustituyó por la forma mutada del promotor de *glnS* descrito en Plumbridge y Söll en un intento de mejorar la eficiencia del sistema. No obstante, tras la inserción de esta secuencia promotora en pYR-BpaRS5, la secuenciación reveló que además de la mutación pretendida también se identificaron mutaciones adicionales de tipo delección no intencionadas. De los mutantes con delecciones en los que se analizó la actividad de la β -galactosidasa, un mutante concreto, pYR-BpaRS-TRN, que tiene una única delección de nucleótidos (residuo A en la posición -15) además de la sustitución deseada de un nucleótido (G a T en la posición -11) en el promotor de *glnS*, exhibió un incremento de 5 veces de la actividad de la β -galactosidasa en comparación con pYR-BpaRS5 (véase la **FIG. 2**). La secuencia de nucleótidos completa de este nuevo dominio del promotor de *glnS* mutante, denominado *glnS*-TRN, identificada tras la secuenciación de los inventores se proporciona en la SEC ID N°: 13.

Descripción	Secuencia	SEC ID N°:
Promotor de <i>glnS</i> mutante descrito en Plumbridge y Söll	CGATTATCAATTTTAAAAAATAACAGTTGTCAGCCTGTCC CGCTTATAATATCATACGCC	12
Promotor TRN de <i>glnS</i>	CGATTATCAATTTTAAAAAATAACAGTTGTCAGCC TGTCGCCGCTTTAATATCATACGCC	13

La expresión de *BpaRS* bajo el control de la forma mutada del promotor de *glnS* se multiplicó por 2 según se determinó mediante análisis de transferencia de tipo Western (véase la **FIG. 4**). Se observó un incremento adicional de 1,5 veces en la actividad de la β -galactosidasa mediante la combinación de la sustitución D286R de *BpaRS* y el nuevo promotor de *glnS* mutante, correspondiente a una eficacia de supresión global del 57 % para pYR-BpaRS-TRN(D286R).

Ejemplo 3

Generación de sistemas de expresión de ARNt mejorados

El presente ejemplo describe la construcción de sistemas mejorados para la expresión de componentes de sistemas de traducción ortogonal, en la que se determinan los efectos de la introducción de múltiples copias de *MjtRNA-Tyr(CUA)* en un operón policistrónico.

Se observó el efecto de múltiples copias del ARNt supresor ámbar sobre la eficacia de la supresión. Se construyó un operón *MjtRNA-Tyr(CUA)* policistrónico que contiene tres copias del gen del ARNt supresor ámbar bajo el control de un promotor sencillo de *proK* y un terminador.

Las tres secuencias del O-ARNt en tándem en el operón policistrónico se separaron unas de otras mediante secuencias adaptadoras del ARNt derivadas de secuencias adaptadoras de ARNt de *E. coli*. Estas secuencias de adaptadores concretas se escogieron porque contienen los nucleótidos T(-1) and A(77). Se ha demostrado que

estas dos posiciones de nucleótidos en los adaptadores de ARNt son óptimas para el procesamiento 5' y 3' eficiente de los precursores de ARNt cuando están en su contexto nativo (es decir, endógeno).

El primero y el segundo gen *MjtRNA-Tyr(CUA)* en el operón policistrónico recombinante están separados por un adaptador de ARNt derivado del adaptador de origen natural entre los genes de ARNt de *valU* y *valX* de *E. coli* (SEC ID N°: 14). El segundo y el tercer gen *MjtRNA-Tyr(CUA)* en el operón policistrónico recombinante están separados por un adaptador de ARNt derivado del adaptador de origen natural entre los genes de ARNt de *ileT* y *alaT* de *E. coli* (SEC ID N°: 15). El uso de los adaptadores tiene una ventaja práctica adicional en cuanto a que estos polinucleótidos contienen sitios de restricción convenientes.

Dos copias idénticas del operón del ARNt policistrónico sintético que contienen tres copias del gen del ARNt supresor se ligaron para generar clúster génicos con seis copias del gen *MjtRNA-Tyr(CUA)*. Después, los clúster génicos ensamblados con tres y seis copias de los ARNt se clonaron en pYR-BpaRS-TRN(D286R) para generar pYR-BpaRS-3TRN(D286R) y pYR-BpaRS-6TRN(D286R), respectivamente. Cuando se expresan en *E. coli*, estos plásmidos proporcionaron un incremento del 30 % y del 50 % de la expresión de *MjtRNA-Tyr(CUA)*, respectivamente, determinado mediante análisis de tipo Northern (véase la FIG. 3). Este incremento en la expresión del mensaje produjo un incremento del 30-40 % de la actividad de la β -galactosidasa (véase la FIG. 2).

Dado que los ARNt de codones raros de *E. coli* codificados en estos plásmidos pueden ser innecesarios para la expresión de la mayoría de las proteínas, estos genes de ARNt de *E. coli* se eliminaron del plásmido pYR-BpaRS-6TRN(D286R) para dar pSup-BpaRS-6TRN(D268R) (mostrado en la FIG. 5). Como cabía esperar, la eficacia de supresión de este plásmido determinada mediante el ensayo de la actividad de la β -galactosidasa *in vivo* permaneció igual que la de su plásmido parental (véase la FIG. 6).

Ejemplo 4

Expresión de un modelo de proteína que comprende un aminoácido no natural usando sistemas de expresión mejorados

El presente ejemplo describe la expresión de un modelo de proteína mioglobina de esperma de ballena que comprende un aminoácido no natural usando los sistemas de expresión mejorados de la invención.

Para estudiar adicionalmente las mejoras en el rendimiento y la fidelidad de la incorporación de aminoácidos no naturales en proteínas usando los sistemas de la invención, un mutante de Ser-4 en Bpa de la mioglobina de ballena de esperma (descrito en Chin et al., Proc. Natl. Acad. Sci. USA 99:11020-11024 (2002)) se expresó en *E. coli*. Las células TOP10 *E. coli* (Invitrogen™) cotransformadas con pBAD/Myc-His/MB(S4TAG) y pSup-BpaRS-6TRN(D286R) se cultivaron en medio de Luria-Bertani a 37 °C en presencia de Bpa 1mM. Consistente con los datos del ensayo *in vivo* de la β -galactosidasa anterior se produjo la mioglobina mutante de longitud completa en un rendimiento global purificado de 40 mg/ml, mientras que el sistema anterior proporcionaba 2 mg/ml de la proteína mutante (Chin et al., Proc. Natl. Acad. Sci. USA 99:11020-11024 (2002)). No se observó la proteína mutante mediante gel de SDS-PAGE en ausencia del aminoácido (FIG. 6). La espectrometría de masas MALDI-TOF de la mioglobina mutante que contiene Bpa en lugar de Ser-4 dio una masa promedio de 18521, que concuerda con la masa predicha calculada de 18520. En el espectro de masas no se detectaron signos de incorporación de ningún aminoácido natural en la posición 4.

Ejemplo 5

Aplicabilidad amplia de los sistemas de expresión mejorados de la invención

El presente ejemplo describe la expresión de la β -galactosidasa que comprende cuatro aminoácidos no naturales diferentes cuando los sistemas de expresión usan diferentes sintetetas mutantes que tienen especificidades de carga para diferentes aminoácidos no naturales.

Para comprobar la generalidad de estos sistemas de expresión de la invención se analizaron en el sistema tres aminoacil-ARNt sintetetas ortogonales. Estas O-RS aminoacilan (es decir, cargan) de forma específica el O-ARNt *MjtRNA-Tyr(CUA)* con, como alternativa, *p*-acetil-L-fenilalanina (*pAcPhe*), *p*-azido-L-fenilalanina (*pAzPhe*) y *p*-yodo-L-fenilalanina (*pIPhe*) (FIG. 1). Estos aminoácidos no naturales son útiles para experimentos de marcaje químico (Wang et al., Proc. Natl. Acad. Sci. USA 100:56-61 (2003)), fotoreticulación (Chin et al., J. Am. Chem. Soc., 124:9026-9027 (2002)), y fase cristalográfica (Xie et al., Nat. Biotechnol., 22: 1297-1301 (2004)). Os vectores de expresión de la invención que codifican estos genes *MjTyrRS* se construyeron mediante subclonación de la O-RS respectiva en los sitios *NdeI/PstI* de pSug-BpaRS-6TRN(D286R), para dar los vectores pSup-pAcPheRS-6TRN, pSup-pAzPheRS-6TRN y pSup-pIPheRS-6TRN. Como en el caso de Bpa, cuando las células de *E. coli* que alojan estos plásmidos se cultivan en presencia de estos aminoácidos (1mM), el nivel de actividad de la β -galactosidasa es similar al de la β -galactosidasa de tipo silvestre (FIG. 7), lo que indica producción de β -galactosidasa que comprende el correspondiente aminoácido no natural.

Ejemplo 6

Construcciones plasmídicas

5 pYR-BpaRS1

pYR-BpaRS1, un replicón p15A que contiene un marcador de Resistencia al cloranfenicol, *MjtRNA-Tyr*(CUA) bajo el control de BpaRS y *MjtRNA-Tyr*(CUA) en los sitios *SacI* y *SpeI* del plásmido pRARE2 (Novagen).

10 Gen *MjtRNA-Tyr*(CUA) con el promotor y el terminador *proK*

Se construyó un operón *MjtRNA-Tyr*(CUA) monocistrónico que contiene el promotor y el terminador *proK* mediante PCR usando cuatro oligonucleótidos sintéticos en una estrategia de PCR solapante para construir todo el operón del ARN en una única reacción de PCR:

15

Cebador	Secuencia	SEC ID Nº:
proK P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAA AGGC	37
proK T1	CTTTCTCGCCGATCAAAAAGGCATTTTGCTATTAAGGGATTGACGA GGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCCGGCGGTAGT TCAGCAGGGC	38
proK T2	CTTTCTCGCCGATCAAAAAGGCATTTTGCTATTAAGGGATTGACGA GGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCCGGCGGTAGT TCAGCAGGGC	39
proK P2	GCATAAGCTTATGCAAAAAAGCCTGCTCGTTGAGCAGGCTTTTCG	40

El amplicón de PCR se amplificó después mediante PCR usando dos cebadores.

Cebador	Secuencia	SEC ID Nº:
proK-F	AGTCTGATCAGTCACGGCTAACTAAGCGG	41
proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG	42

20 El amplicón resultante se insertó entre los sitios *BclI* y *XhoI* (subrayados) de pYR-BpaRS1 para generar el plásmido pYR-BpaRS5.

Promotor de *glnS* mutante

25 Se construyó un promotor de *glnS* mutante mediante PCR usando cuatro oligonucleótidos sintéticos.

Cebador	Secuencia	SEC ID Nº:
<i>glnS</i> P1	CCGAGCTCCCGGGTCATC	43
<i>glnS</i> T1	CCGAGCTCCCGGGTCATCAATCATCCCCATAATCCTTGTTAGATTATCAAT TTTAAAAAACTAACAGTTGTCAGCCTGTC	44
<i>glnS</i> T2	GTCCATATGGGATTCTCAAAGCGTAAACAACGTATAACGGCGTATGATAT TATAAGCGGGACAGGCTGACAACTGTTAG	45
<i>glnS</i> P2	GTCCATATGGGATTCTC	46

30 El producto se insertó entre los sitios *XmaI* y *NdeI* de pYR-BpaRS5. Una serie de clones se sometieron a detección selectiva con un ensayo de actividad *in vivo* de LacZ y se identificó un mutante concreto de delección de una sola base con actividad mejorada y se secuenció (denominado pYR-BpaRS-TRN).

Casetes génicos de ARNt en tándem

35 Dos secuencias adaptadoras de ARNt que se producen de forma natural entre los genes *valU* y *valX* (SEC ID Nº: 14) y entre los genes *ileT* y *alaT* (SEC ID Nº: 15) en el genoma de *E. coli* se usaron como espaciadores entre los genes *MjtRNA-Tyr*(CUA). Estas secuencias adaptadoras contienen los sitios de restricción *BsmAI* y *EatI* a los que se

unieron los genes *M*tRNA-Tyr(CUA). Estas secuencias flanqueantes también contienen residuos T(-1) y A(77) que son óptimos para la eficacia del procesamiento en 5' y 3' de los precursores del ARNt. El casete génico del ARNt se amplificó mediante PCR usando tres conjuntos de cebadores.

	Conjunto 1	
Cebador	Secuencia	SEC ID Nº:
Adaptador P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAA GGC	47
Adaptador P2	TACACGGCGGAGACTACATAAAGTAGTTGGTCCGGCGGGCCGGATT TG	48

5

	Conjunto 2	
Cebador	Secuencia	SEC ID Nº:
Adaptador P3	GTAGTCTCCGCCGTGTAGCAAGAAATTGAGAAGTCCGGCGGTAGTT CAGCAG	49
Adaptador P4	AAQCCTCTTCAAATTTGCCGTGCAAATTTGGTCCGGCGGGCCGGATTG	50
	Conjunto 3	
Cebador	Secuencia	SEC ID Nº:
Adaptador P5	GCAAATTTGAAGAGGTTTTAACTACATGTTATCCGGCGGTAGTTCA GCAG	51
proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG	52

10

El producto de cada conjunto se digirió con *Bsm*AI (Set 1), *Eal*I (Set 2) o ambos (Conjunto 3). La unión de estos tres fragmentos de restricción produjo un operón de ARNt policistrónico que contiene tres copias del gen de ARNt conectado por dos secuencias adaptadoras diferentes de origen natural. El clúster génico resultante se amplificó mediante PCR usando tres conjuntos de cebadores:

	Conjunto 4	
Cebador	Secuencia	SEC ID Nº:
Tándem P1	ATCAGTGCACGGCTAACTAAGCGG	53
Tándem P2	GCTGGCATGCATGCAAAAAAGCCTGCTCGTTGAGC	54

	Conjunto 5	
Cebador	Secuencia	SEC ID Nº:
Tándem P3	ATCAGCATGCGGCTAACTAAGCGGCCTGCTG	55
Tándem P4	GCTGCTCGAGATGCAAAAAAGCCTGC	56

15 Los productos de la PCR de los conjuntos 4 y 5 se digirieron con *Sph*I y se ligaron entre sí para generar un ensamblaje génico del ARNt en tándem unidireccional, que consiste en dos operones de ARNt policistrónicos idénticos, codificando cada uno tres genes de ARNt bajo el control de un único promotor y terminador de *proK*. Cada clúster génico de ARNt que contiene una o dos copias idénticas del operón del ARNt policistrónico se clonó en los sitios *Apa*LI y *Xho*I de pYR-BpaRS-TRN, para dar pYR-BpaRS-3TRN y pYR-BpaRS-6TRN, respectivamente.

20

Plásmidos pSup

25 Cada uno de los doce genes de ARNt de *E. coli*, inicialmente codificados en el plásmido pRARE2 se eliminaron de pYR-BpaRS-6TRN(D286R) mediante digestión con *Spe*I y *Drd*I, seguido del tratamiento con la nucleasa Mung bean. Se volvieron a ligar los vectores linealizados, lo que generó pSup-BpaRS-6TRN(D286R). Los genes *M*tRNA-TyrRS mutantes para *p*-acetil-L-fenilalanina, *p*-azido-L-fenilalanina y *p*-yodo-L-fenilalanina se subclonaron de sus correspondientes plásmidos pBK en los sitios *Nde*I y *Pst*I de pSup-BpaRS-6TRN(D286R) para generar pSup-

pAcPheRS-6TRN, pSup-pAzPheRS-6TRN y pSup-plodoPheRS-6TRN, respectivamente.

Plásmido indicador de LacZ y ensayo de actividad de β -galactosidasa *in vivo*

5 El codón de fenilalanina (TTC) en el residuo 13 (subrayado) de la secuencia líder (MDPLVTAASVLEFGLFET; SEC ID N°: 57) localizado cadena arriba del gen *lacZ* de pBAD/Myc-His/LacZ (Invitrogen™) se mutó en un codón ámbar (TAG) mediante mutagénesis dirigida a sitio para producir un plásmido indicador de LacZ pBAD/Myc-His/LacZ(TAG). Este plásmido se co-transformó con cada plásmido supresor en células TOP10 de *E. coli* (Invitrogen™). Las células se incubaron a 37 °C durante la noche en medio de Luria-Bertani (LB) que contiene 0,02 % de arabinosa y aminoácido no natural 1 mM. La actividad de LacZ (β -galactosidasa) se midió de acuerdo con el procedimiento descrito por Miller (Miller, J.H. Experiments in Molecular Genetics. (Cold Spring Harbor Laboratory, New York, 1972)).

Ejemplo 7

15

Metodologías generales

Se usaron células XL1-Blue de *E. coli* (Stratagene) para clonar y mantener los plásmidos. La ADN polimerasa de alta fidelidad PfuUltra™ (Stratagene®) se usó para la reacción en cadena de la polimerasa (PCR). El kit para mutagénesis dirigida a sitio QuikChange® II (Stratagene®) se usó para la mutagénesis dirigida a sitio. Las secuencias de todos los plásmidos construidos se verificaron mediante secuenciación.

20

Expresión de proteínas

25 Un gen de la mioglobina mutante de esperma de ballena marcado con hexahistidina en el extremo C (Ser-4) del pBAD-JYAMB-4TAG entre los sitios NcoI y KpnI de pBAD/Myc-His (Invitrogen™) para generar pBAD/Myc-His/MB(S4TAG). El plásmido se co-transformó con pSup-BpaRS-6TRN(D286R) en células TOP10 de *E. coli* (Invitrogen™). Las células se incubaron a 37 °C en medio LB que contiene 1400 mg/ml de carbenicilina, 50 mg/ml de cloranfenicol y Bpa 1 mM. A DO₆₀₀= 0,6, las células fueron inducidas mediante la adición de 0,2 % de arabinosa y se incubaron durante 12 horas. Las células se recogieron mediante centrifugación y se lisaron con reactivo BugBuster® (Novagen®). La proteína de los cuerpos de inclusión se purificó con una resina de afinidad metálica TALON® (Clontech®) en condiciones desnaturizantes de acuerdo con el protocolo del fabricante. La proteína purificada se concentró mediante ultrafiltración y se analizó mediante espectrometría de masas MALDI-TOF. La conjugación de proteínas se midió mediante el procedimiento de Bradford.

30

35

Análisis de tipo Northern

Las células TOP10 de *E. coli* (Invitrogen™) transformadas con cada plásmido de supresión se incubaron en LB a 37 °C. A una DO₆₀₀= 0,8 se recogieron las células. El ARN total se aisló mediante extracción con fenol y fraccionamiento con isopropanol como se ha descrito anteriormente (Deutscher and Hilderman, "Isolation and partial characterization of Escherichia coli mutants with low levels of transfer ribonucleic acid nucleotidyltransferase," J. Bacteriol., 118:621-627 (1974)). Las muestras de ARN se separaron en un gel de poliacrilamida desnaturizante al 15 % y se transfirieron a una membrana de GeneScreen Plus® (PerkinElmer®). La membrana se hibridó durante la noche a 55 °C con:

45

5'-biotina-CCCTGCTGAACTACCGCC-3' (SEC ID N°: 58).

La sonda biotinilada hibridada se detectó usando el kit de hibridación y detección por quimioluminiscencia North2South (Pierce) de acuerdo con el protocolo del fabricante.

50

Análisis de tipo Western de la expresión de bpa

Un gen de BpaRS marcado con hexahistidina en C terminal se construyó mediante PCR y se insertó entre los sitios NdeI y PstI de pYR-BpaRS5 and pYR-BpaRS-TRN para generar pYR-BpaRS5(C-His) y pYR-BpaRS-TRN(C-His), respectivamente. Las células Top 10 de *E. coli* transformadas con cada plásmido se incubaron en LB a 37 °C. Las células se recogieron a DO₆₀₀= 1 y se lisaron con reactivo de Bugbuster. La proteína total se separó en gel del 10 % de poliacrilamida y se transfirió a la membrana de PVDF (Invitrogen). La membrana se hibridó con el conjugado anticuerpo-HRP anti-His (extremo C) (Invitrogen) y se detectó mediante quimioluminiscencia.

60

REIVINDICACIONES

1. Una composición que comprende un constructo de ácido nucleico, comprendiendo dicho constructo:

5 secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una primera secuencia de nucleótidos expresable que codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en donde dicho O-ARNt de *Archaea* comprende un par C1-G72 y en donde dichas secuencias del promotor y del terminador están unidas operativamente a dicha primera secuencia de nucleótidos expresable y en donde dicha primera secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, y en donde dicha secuencia de nucleótidos del promotor tiene la secuencia de nucleótidos SEC ID N°: 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36.

2. La composición de la reivindicación 1, en la que:

15 (i) el constructo de ácido nucleico comprende además un promotor de *glnS* de *E. coli* modificado que tiene una secuencia de nucleótidos de SEC ID N°: 13 y una segunda secuencia de nucleótidos expresable, en donde dicha secuencia de nucleótidos del promotor *glnS* de *E. coli* está unida operativamente a dicha segunda secuencia de nucleótidos expresable, de modo que la segunda secuencia de nucleótidos expresable codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en donde dicha O-RS aminoacila, preferentemente, dicha O-ARNt con un aminoácido no natural.

3. La composición de la reivindicación 1, en la que dicho O-ARNt está codificado por una secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*); o en la que dicha secuencia de nucleótidos expresable es un operón policistrónico que comprende una pluralidad de O-ARNt cada uno de los cuales tiene la secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*).

4. La composición de la reivindicación 3, en la que dicha primera secuencia de nucleótidos expresable comprende una pluralidad de dicho operón policistrónico.

5. La composición de la reivindicación 2, en la que:

35 (i) dicha O-RS es una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o (ii) dicha O-RS es una tirosil-ARNt sintetasa de *Methanococcus jannaschii*; o (iii) dicha O-RS tiene una sustitución de ácido aspártico en arginina en la posición de aminoácidos 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre proporcionada en la SEC ID N°: 2 (*Mj-tRNA-Tyr RS* de tipo silvestre).

6. Una célula huésped que comprende la composición de una cualquiera de las reivindicaciones precedentes.

7. La célula huésped de la reivindicación 6, en la que dicha célula huésped es una célula huésped eubacteriana o dicha célula huésped es una célula de *E. coli*.

8. Un sistema de traducción para la expresión de un polipéptido de interés que comprende al menos un aminoácido no natural en una posición especificada; comprendiendo el sistema:

- (a) un aminoácido no natural;
- (b) un constructo de ácido nucleico, comprendiendo dicho constructo:

50 (i) secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una secuencia de nucleótidos expresable, en la que dicha secuencia de nucleótidos expresable codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en donde dicho O-ARNt de *Archaea* comprende un par C1-G72 y en donde dichas secuencias del promotor y del terminador están unidas operativamente a dicha secuencia de nucleótidos expresable y en donde dicha secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, en donde dicha secuencia de nucleótidos tiene una secuencia de nucleótidos de SEC ID N° 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos seleccionada de SEC ID N° 33, 35 y 36; y

55 (ii) una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural; y

60 (c) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt, en el que la posición del codón selector en el polinucleótido controla la posición especificada del aminoácido no natural en el polipéptido de interés tras la expresión del polinucleótido para producir el polipéptido.

65

9. El sistema de traducción de la reivindicación 8, en el que dicho constructo de ácido nucleico comprende además al menos uno de:

(i) una secuencia de nucleótidos que tiene un promotor de *glnS* de *E. coli* que tiene una secuencia de nucleótidos de SEC ID N° 13, en la que dicha secuencia de nucleótidos de *glnS* de *E. coli* está unida operativamente a dicha secuencia de nucleótidos que codifica dicha O-RS; y

(ii) un operón policistrónico que comprende una pluralidad de dichas secuencias de nucleótidos del gen de O-ARNt, en el que al menos un gen de O-ARNt está separado de al menos un gen de O-ARNt adyacente por un adaptador polinucleotídico heterólogo de un adaptador polinucleotídico de un operón de ARNt.

10. El sistema de traducción de la reivindicación 9, en el que dichas secuencias del promotor y del terminador del ARNt de prolina de *E. coli* se proporcionan en las SEC ID N° 32 (promotor) y 33 (terminador), respectivamente; o en el que dicho operón policistrónico comprende una pluralidad de adaptadores polinucleotídicos heterólogos idénticos.

o en el que dicho operón policistrónico comprende una pluralidad de adaptadores polinucleotídicos heterólogos, en el que al menos dos de los adaptadores polinucleotídicos heterólogos son diferentes;

o en el que dicho adaptador polinucleotídico heterólogo comprende un nucleótido de timidina en el extremo 5' o un nucleótido de adenosina en el extremo 3' o ambos, un nucleótido de timidina en el extremo 5' y un nucleótido de adenosina en el extremo 3'.

o en el que dicho adaptador polinucleotídico heterólogo es el adaptador polinucleotídico localizado entre los genes de ARNt endógenos de *Escherichia coli* seleccionados de: *valU* y *valX*, *ileT* y *alaT*; *serV* y *argV*; *valV* y *valW*; *glyT* y *thrT*; *metT* y *leuW*; *glnW* y *metU*; *hisR* y *leuT*; *glnU* y *glnW*; *leuP* y *leuV*; *glnV* y *glnX*; *alaW* y *alaX*; *ileU* y *alaU*; *ileV* y *alaV*; *metU* y *glnV*; *glyW* y *cysT*; *argX* y *hisR*; y *argY* y *argZ*;

o en el que dicho adaptador polinucleotídico heterólogo comprende la secuencia de nucleótidos de SEC ID N° 14 (adaptador *valU/valX*) o 15 (adaptador *ileT/alaT*).

11. El sistema de traducción de la reivindicación 9, en el que dicha secuencia de nucleótidos que codifica un O-ARNt es un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos que codifican un O-ARNt.

o en el que dicha secuencia de nucleótidos que codifica un O-ARNt comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*);

o en el que dicha secuencia de nucleótidos que codifica un O-ARNt es un operón policistrónico que comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*);

o en el que dicha O-RS es una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*;

o en el que dicha O-RS es una tirosil-ARNt sintetasa de *Methanococcus jannaschii*;

o en el que dicha O-RS tiene una sustitución de ácido aspártico en arginina en la posición del aminoácido 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre que se proporciona en la SEC ID N°: 2 (*Mj-tRNA^{Tyr} RS* de tipo silvestre).

12. El sistema de traducción de la reivindicación 8, que comprende una célula huésped que comprende (a), (b) y (c).

13. El sistema de traducción de la reivindicación 12, en el que dicha célula huésped es una célula huésped eubacteriana.

14. Un procedimiento para producir en una célula huésped un polipéptido de interés, que comprende un aminoácido no natural en una posición especificada; comprendiendo el procedimiento:

(a) proporcionar:

(i) un aminoácido no natural;

(ii) un constructo de ácido nucleico que comprende:

secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una secuencia de nucleótidos expresable, en donde dicha secuencia de nucleótidos expresable codifica un ARNt ortogonal de *Archaea* (O-ARNt).

en donde dicho O-ARNt de *Archaea* comprende un par C1-G72 y en el que dichas secuencias del promotor y del terminador está unidos operativamente a dicha secuencia de nucleótidos expresable y en donde dicha secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, en donde dicha secuencia de nucleótidos del promotor tiene una secuencia de nucleótidos de SEC ID N° 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36; y

una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural; y

(iii) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt y en el que la posición del codón selector se correlaciona con la posición especificada del aminoácido no natural en el polipéptido de interés.

(iv) una célula huésped que comprende (i), (ii) y (iii); y

- (b) cultivar dicha célula huésped; y
 (c) incorporar dicho aminoácido no natural en dicha posición especificada en dicho polipéptido durante la traducción de dicho polipéptido en dicha célula huésped, de modo que se produce dicho polipéptido de interés que comprende dicho aminoácido no natural en la posición especificada.

15. El procedimiento de la reivindicación 14, en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos que codifican una o más especies de O-ARNt; o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifican un O-ARNt que comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de la secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una tirosil-ARNt sintetasa de *Methanococcus jannaschii*; o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifica una O-RS que tiene una sustitución de ácido aspártico en arginina en la posición de aminoácidos 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre proporcionada en la SEC ID N° 2 (Mj-tRNATyrRS de tipo silvestre); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar al menos uno de:

- (I) secuencias del promotor y del terminador que comprenden las secuencias del promotor y del terminador de proK de *E. coli* proporcionadas en las SEC ID N° 32 (promotor) y 33 (terminador), respectivamente;
 (II) una secuencia de nucleótidos del promotor que tiene un promotor de glnS modificado de *E. coli* que tiene una secuencia de nucleótidos de SEC ID N° 13, en donde dicha secuencia de nucleótidos de glnS de *E. coli* está unida operativamente a dicha secuencia de nucleótidos que codifica dicha O-RS; y
 (III) un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos del gen de O-ARNt, en el que al menos un gen de O-ARNt está separado de al menos un gen de O-ARNt adyacente por un adaptador polinucleotídico heterólogo de un adaptador polinucleotídico de un operón de ARNt.

16. El procedimiento de la reivindicación 15, en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar una pluralidad de adaptadores polinucleotídicos heterólogos idénticos; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar una pluralidad de adaptadores polinucleotídicos heterólogos, en el que al menos dos de los adaptadores polinucleotídicos heterólogos son diferentes; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar un adaptador polinucleotídico heterólogo que comprende un nucleótido timidina en el extremo 5', un nucleótido adenosina en el extremo 3' o ambos un nucleótido timidina en el extremo 5' y un nucleótido adenosina en el extremo 3'; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar un adaptador polinucleotídico heterólogo del adaptador polinucleotídico localizado entre los genes de ARNt endógenos de *Escherichia coli* seleccionados de: *valU* y *valX*; *ileT* y *alaT*; *serV* y *argV*; *valV* y *valW*; *glyT* y *thrT*; *metT* y *leuW*; *glnW* y *metU*; *hisR* y *leuT*; *glnU* y *glnW*; *leuP* y *leuV*; *glnV* y *glnX*; *alaW* y *alaX*; *ileU* y *alaU*; *ile V* y *alaV*; *metU* y *glnV*; *glyW* y *cysT*; *argX* y *hisR*; y *argY* y *argZ*; o en el que dicha provisión de un operón policistrónico de (III) comprende proporcionar un adaptador polinucleotídico heterólogo que tiene la secuencia de nucleótidos de SEC ID N°: 14 (adaptador *valU/valX*) o 15 (adaptador *ileT/alaT*).

17. El procedimiento de la reivindicación 14, en el que dicha provisión de una célula huésped comprende proporcionar una célula huésped eubacteriana o una célula huésped de *Escherichia coli*.

18. Un procedimiento para producir en una célula huésped un polipéptido de interés, que comprende un aminoácido no natural en una posición especificada; comprendiendo el procedimiento:

(a) proporcionar:

- (i) un aminoácido no natural;
 (ii) un constructo de ácido nucleico que comprende una secuencia de nucleótidos que codifica un ARNt ortogonal (O-ARNt) de *Archaea*, en el que el O-ARNt de *Archaea* comprende un par C1-G72.
 (iii) un constructo de ácido nucleico que comprende una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural;
 (iv) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt y en el que la posición del codón selector se correlaciona con la posición especificada del aminoácido no natural en el polipéptido de interés; y
 (v) una célula huésped que comprende (i), (ii), (iii) y (iv);
 en el que dicho constructo de ácido nucleico de (ii) comprende secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli*, en donde dichas secuencias de nucleótidos del promotor y del terminador están unidas operativamente a dicha secuencia de nucleótidos que codifica dicho O-ARNt, y en donde dicha secuencia de nucleótidos que codifica dicho O-ARNt es

heteróloga de dichas secuencias del promotor y del terminador, en donde dicha secuencia de nucleótidos del promotor tiene una secuencia de nucleótidos de SEC ID N° 32 o 34 y dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos de SEC ID N° 33, 35 y 36;

(b) cultivar dicha célula huésped; y

5 (c) incorporar dicho aminoácido no natural en dicha posición especificada en dicho polipéptido durante la traducción de dicho polipéptido en dicha célula huésped, de modo que se produce dicho polipéptido de interés que comprende dicho aminoácido no natural en la posición especificada.

10 19. El procedimiento de la reivindicación 18, en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de secuencias de nucleótidos que codifican una o más especies de O-ARNt: o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifican un O-ARNt que comprende una secuencia de nucleótidos de SEC ID N° 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar un operón policistrónico que comprende una pluralidad de la secuencia de nucleótidos de SEC ID N°: 1 (*MjtRNA-Tyr(CUA)*); o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una tirosil-ARNt sintetasa de *Methanococcus jannaschii*;

15 o en el que dicha provisión de un constructo de ácido nucleico comprende proporcionar una secuencia de nucleótidos que codifica una O-RS que tiene una sustitución de ácido aspártico en arginina en la posición del aminoácido 286 o en una posición análoga a la posición 286 respecto a la secuencia de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre que se proporciona en la SEC ID N° 2 (*Mj-tRNATyrRS* de tipo silvestre); o en el que dicha provisión de un constructo de ácido nucleico de (I) comprende secuencias del promotor y del terminador que comprenden las secuencias del promotor y del terminador de proK de *E. coli* proporcionadas en las SEC ID N° 32 (promotor) y 33 (terminador), respectivamente; o en el que dicha provisión de una célula huésped comprende proporcionar una célula huésped eubacteriana; o en el que dicha provisión de una célula huésped comprende proporcionar una célula huésped de *Escherichia coli*.

20

25

30 20. Un procedimiento para producir un sistema de traducción para la expresión de un polipéptido de interés que comprende al menos un aminoácido no natural en una posición especificada; comprendiendo el procedimiento proporcionar:

(a) un aminoácido no natural;

(b) un constructo de ácido nucleico, comprendiendo dicho constructo:

35 (i) secuencias de nucleótidos del promotor y del terminador de un gen de ARNt de prolina de *Escherichia coli* y una secuencia de nucleótidos expresable, codificando dicha secuencia de nucleótidos expresable un ARNt ortogonal (O-ARNt) de *Archaea*, en donde dicho O-ARNt de *Archaea* comprende un par C1-G72 y en donde dichas secuencias del promotor y del terminador están unidas operativamente a dicha secuencia de nucleótidos expresable y en donde dicha secuencia de nucleótidos expresable es heteróloga de dichas secuencias de nucleótidos del promotor y del terminador, y en donde dicha secuencia de nucleótidos del promotor tiene una secuencia de nucleótidos SEC ID N° 32 o 34; y en donde dicha secuencia de nucleótidos del terminador tiene una secuencia de nucleótidos seleccionada de SEC ID N° 33, 35 y 36; y

40 (ii) una secuencia de nucleótidos que codifica una aminoacil-ARNt sintetasa ortogonal (O-RS), en donde dicha O-RS aminoacila preferentemente dicho O-ARNt con dicho aminoácido no natural; y

45

(c) un polinucleótido que codifica dicho polipéptido de interés, comprendiendo dicho polinucleótido al menos un codón selector que es reconocido por dicho O-ARNt, en el que la posición del codón selector en el polinucleótido controla la posición especificada del aminoácido no natural en el polipéptido de interés tras la expresión del polinucleótido para producir el polipéptido.

50

21. El procedimiento de la reivindicación 20, en el que (a), (b) y (c) se proporcionan en una célula huésped.

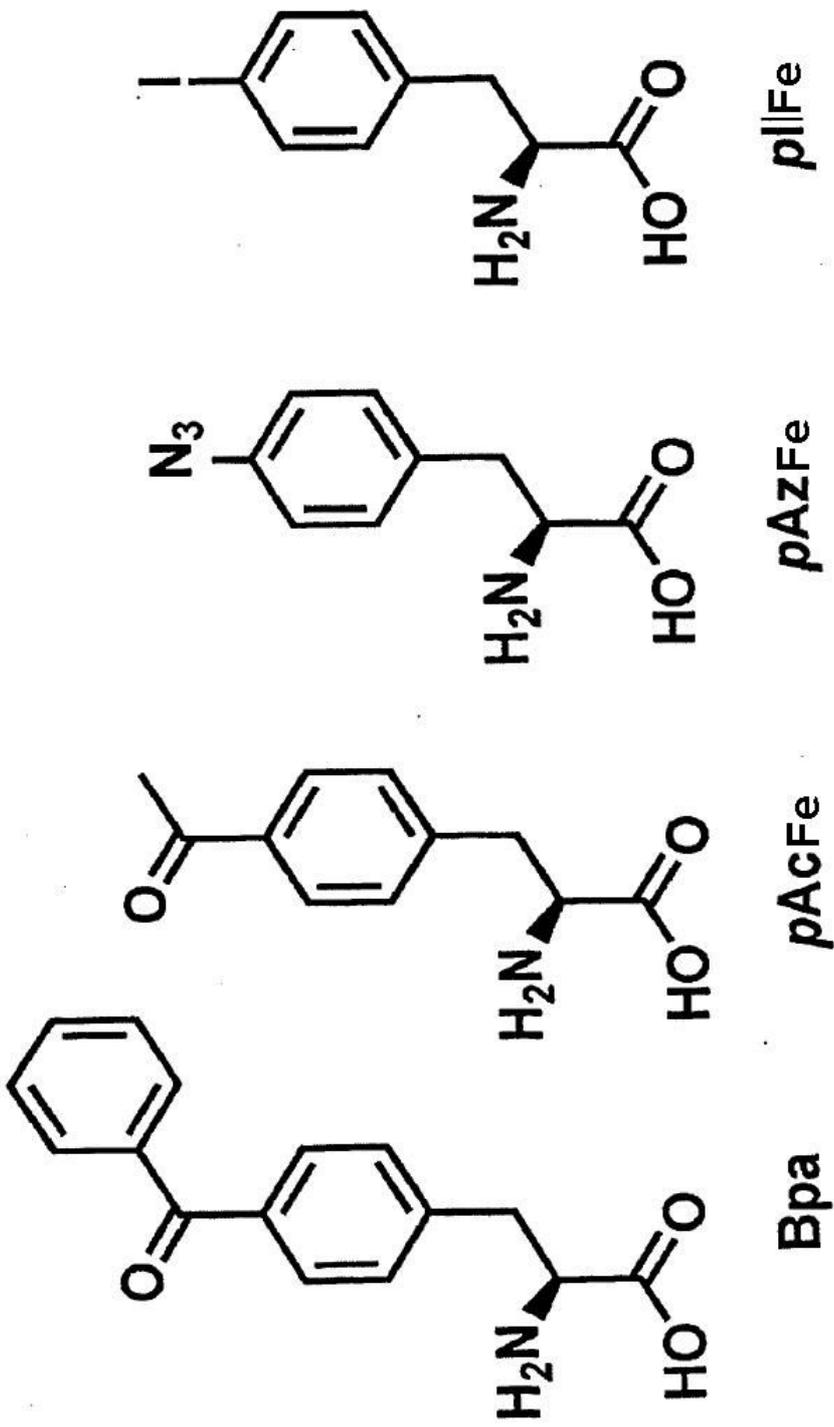


Fig. 1

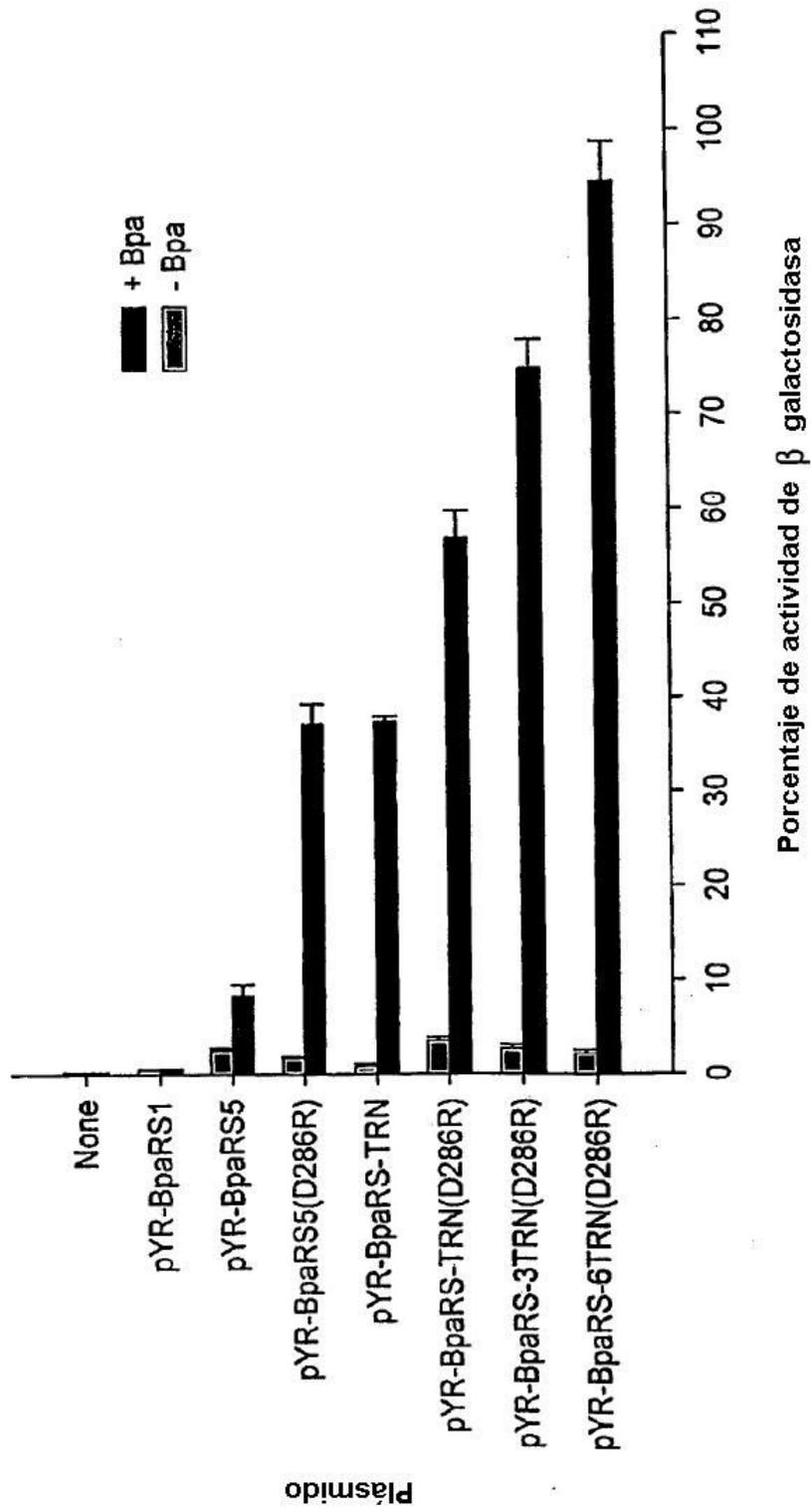


Fig. 2

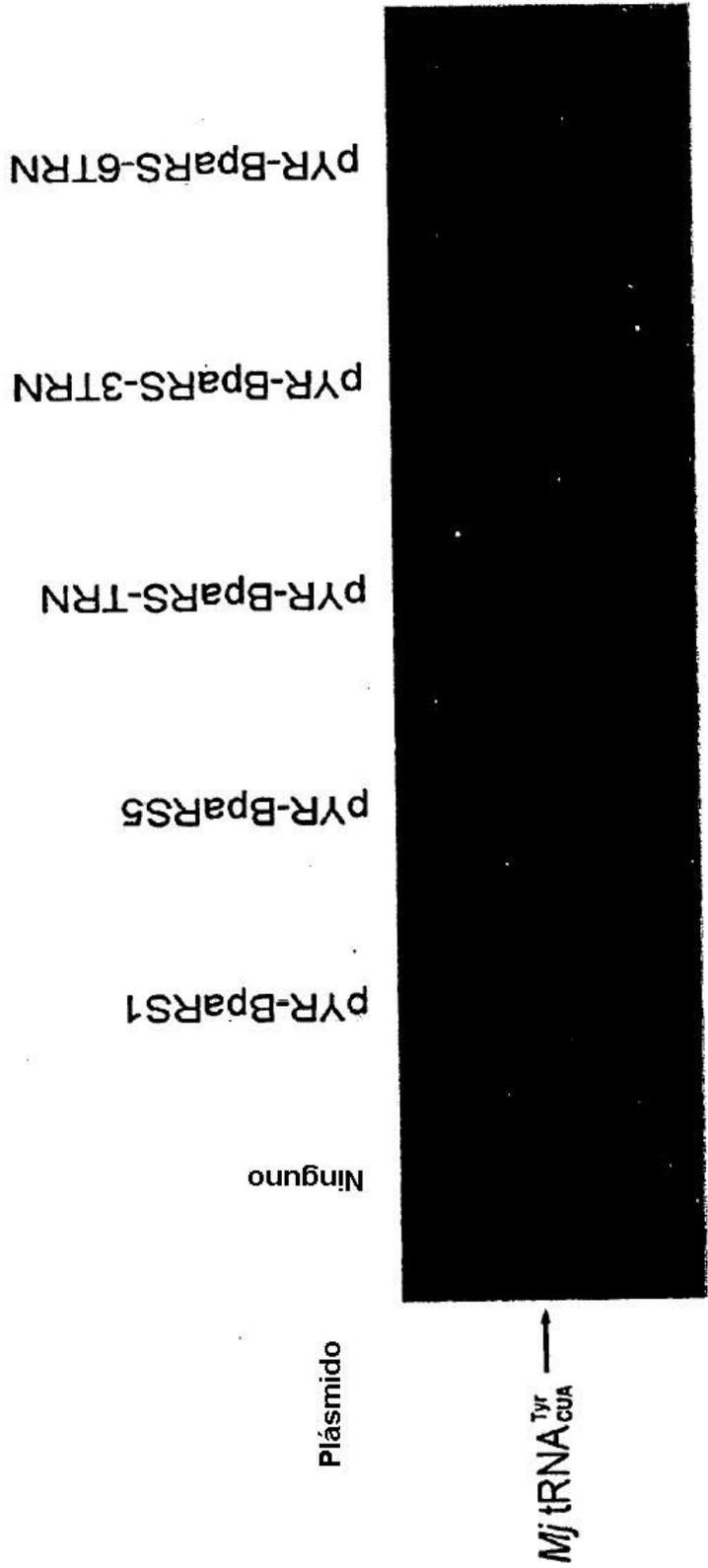


Fig. 3

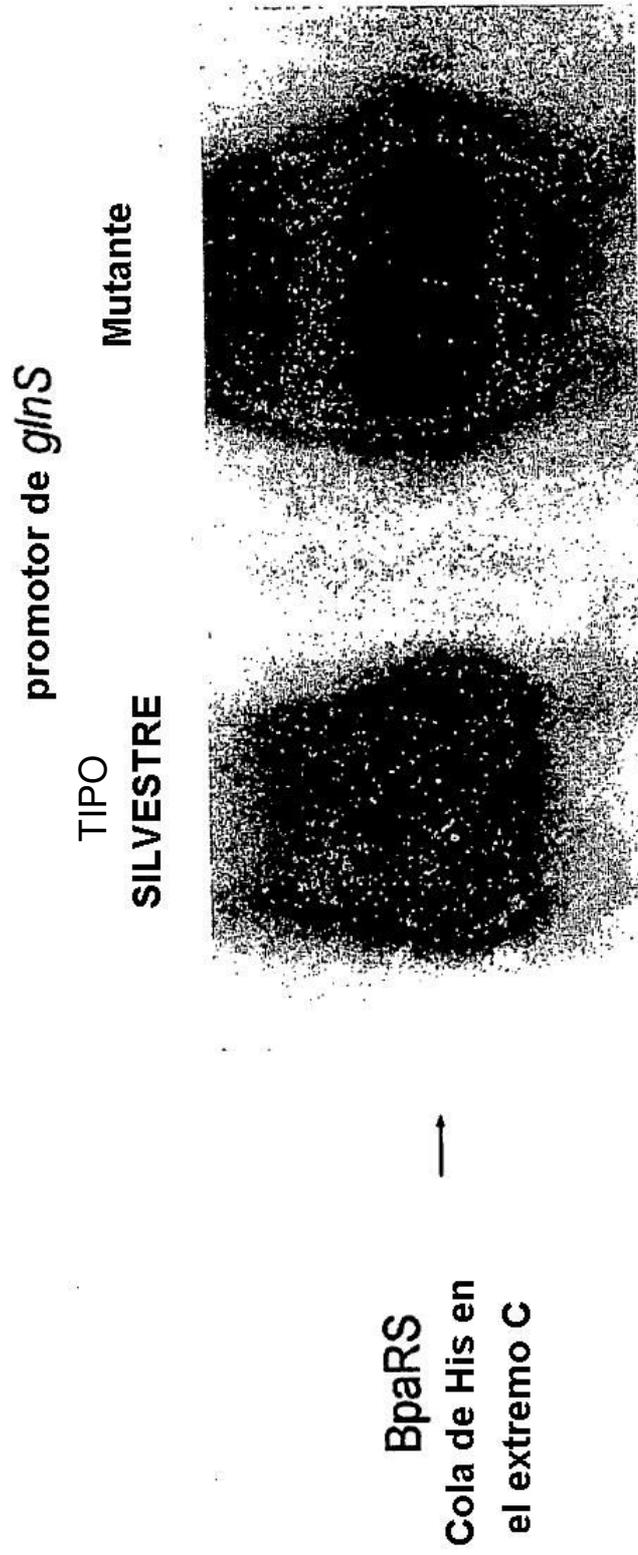


Fig. 4

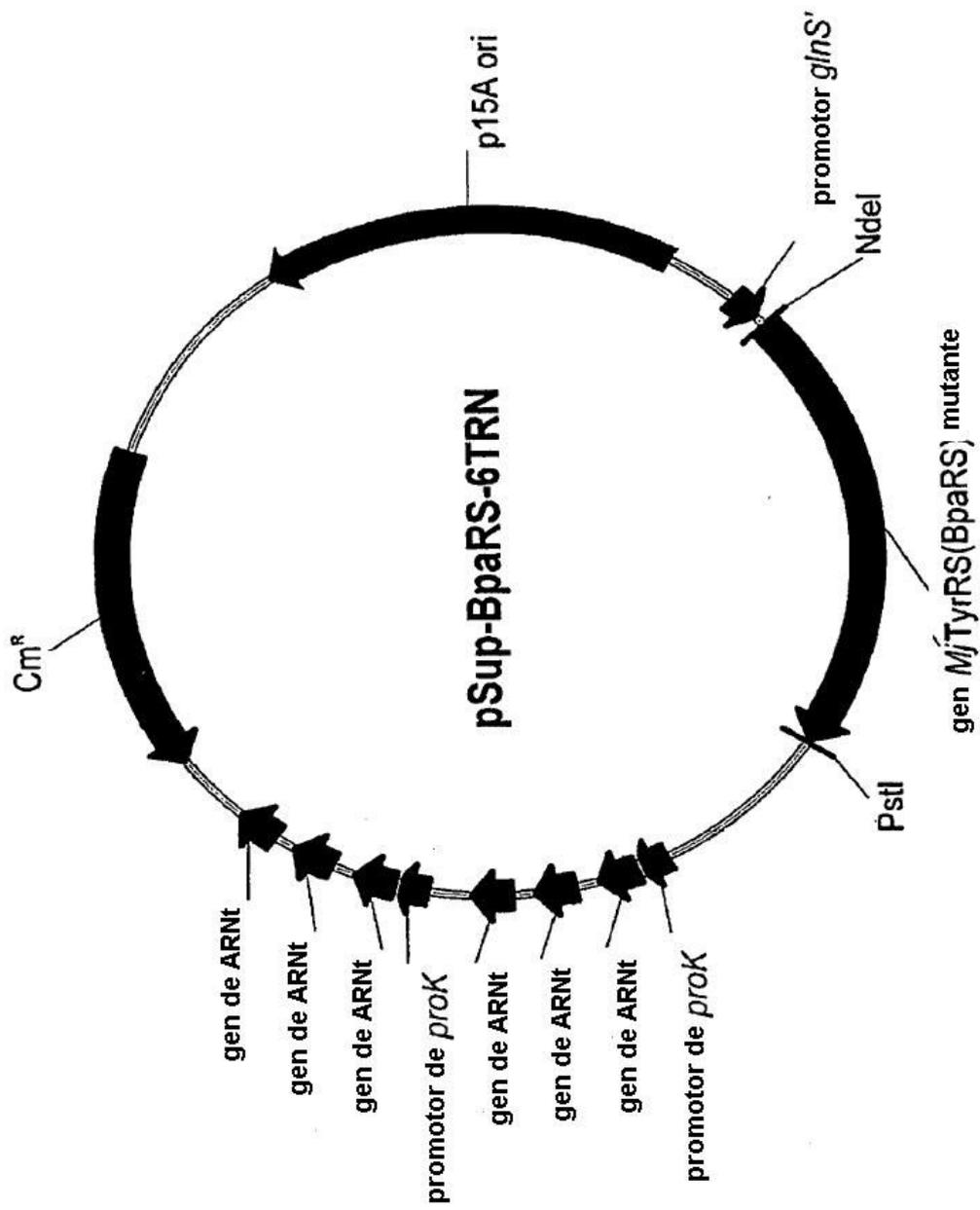


Fig. 5

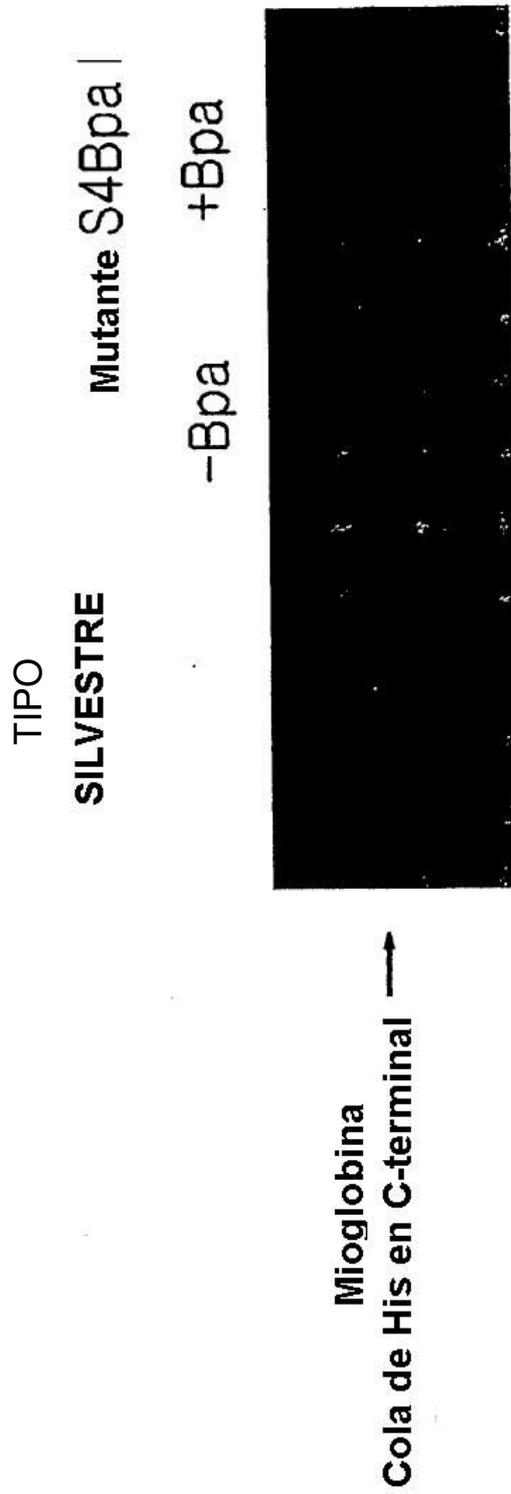


Fig. 6

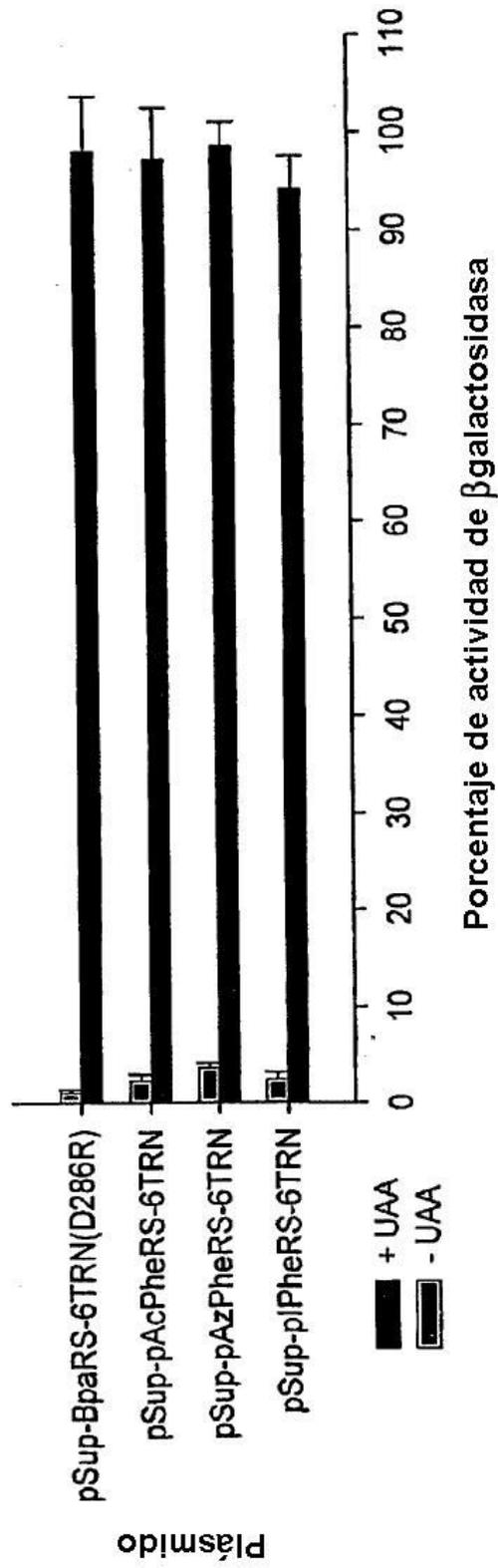


Fig. 7

Secuencias de nucleótidos y de aminoácidos

SEC ID N°	Descripción	SECUENCIA
1	tirosil-ARNt ^{CUA} supresor de <i>Methanococcus jannaschii</i> también conocido como <i>MjtRNA-Tyr(CUA)</i> o <i>mutRNA^{Tyr}_{CUA}</i>	CCGGCGGUAGUUCAGCAGGGCAGAACGGCGGACUCUAAAUCCG CAUGGCGCUGGUUCAAUCCGGCCCCGGACCA
2	Secuencia de aminoácidos de la tirosil-ARNt sintetasa (MjTyrRS) de tipo silvestre de <i>Methanococcus jannaschii</i>	MDEFEMIKRNTSEI ISEEELREVLKKDEKSAYIGFEPGSKIHL GHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLLKR ARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESFLFNKELHPMDLKNVAEELIKILEP IRKRL
3	Secuencia de nucleótidos de la tirosil-ARNt sintetasa (MjTyrRS) de tipo silvestre de <i>Methanococcus jannaschii</i>	ATGGACGAATTTGAAATGATAAAGAGAAAACACATCTGAAATTCAGC GAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATCTGCTTAC ATAGGTTTTGAACCAAGTGGTAAAAATACATTTAGGGCATTATCTCAA ATAAAAAAGATGATGATTTACAAAATGCTGGATTTGATATAATATA TTGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAGTTTTTGAAGCAATG GGTTAAAGGCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAG GATTATACACTGAATGTCATAGATTGGCTTTAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGATATTCAT TATTTAGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGAGCTTTTACCAAAAAAGGTTGTTGTATTCAC AACCTGTCTTAACGGSTTGGATGGAGAAGGAAAGATGAGTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAA AGCCAGAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTCATCCAATGGATTTAAAA AATGCTGTAGCTGAAGAACCTATAAAGATTTTAGAGCCAATTAGAAA AGATTA
4	Secuencia de aminoácidos de p-benzoil-L-fenilalanina aminoacil-ARNt sintetasa (derivada de la tirosil-ARNt (derivada de tirosil-ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre	MDEFEMIKRNTSEI ISEEELREVLKKDEKSAGIGFEPGSKIHL GHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYLYGSPFQLDKDYTLNVYRLALKTTLLKR ARRSMELIAREDENPKVAEVIYPIMQVNTSHYLGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESFLFNKELHPMRLKNVAEELIKILEP IRKRL

Fig. 8

SEC ID N°:	Descripción	SECUENCIA
5	Secuencia de nucleótidos de p-benzoil-L-fenilalanina aminoacil-ARNt sintetasa	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTGGTATAGGTTTTGAACCAAGTGGTAAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTGTTGGCTGATTTACACGCCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCCAA AATATCTTTATGGAAGTCCTTTCCAGCTTGATAAGGATTATAC ACTGAATGCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATAC GAGTCATTATTTAGGCGTTGATGTTGCGAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTCAACAACCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCTTGAATATCCTTAAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACCTATAAAGATTTTLAGACCA ATTAGAAAGAGATTA</p>
6	Secuencia de aminoácidos de P-acetil-L-fenilalanina aminoacil-ARNt sintetasa (pAcPheRS) (derivada de tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre)	<p>MDEFEMIKRNTSEI ISEEELREVLKDKESALIGFEPGSKIHL GHYLQIKKMIDLQNAAGFDII ILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTLKR ARRSMELIAREDENPKVAEVIYPI MQVNGCHYRGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGDDLTVNSYEELESFLKNEKELHPMDLKNVAEELIKILEP IRKRL</p>
7	Secuencia de nucleótidos P-acetil-L-fenilalanina aminoacil-ARNt sintetasa	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTGTTGGCTGATTTACACGCCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCCAA AATATGTTTATGGAAGTGAATTCAGCTTGATAAGGATTATAC ACTGAATGCTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGG TTGTCATTATAGGGCGTTGATGTTGCTGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTGTATTCAACAACCTGTCTTAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCTTGAATATCCTTAAACCATAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACCTATAAAGATTTTLAGACCA ATTAGAAAGAGATTA</p>

Fig. 8 cont.

SEC ID Nº:	Descripción	SECUENCIA
8	<p>secuencia de aminoácidos de p-azido-L-fenilalanina ino a aminoacil-ARNt sintetasa (pAzPheRS): (derivada de tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre)</p>	<p>MDEFEMIKRNTSEIISEEELREVLKKEKSALIGFEPGKIHL GHYLQIKKMIDLQNAQFDIIILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNQIHSSGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSKSNFNFIIV DDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESFLFKNKELHPMDLKNVAEELIKILEP IRKRL</p>
9	<p>secuencia de nucleótidos de p-azido-L-fenilalanina aminoacil-ARNt sintetasa</p>	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTAAAAAAGATGAAAA ATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAATTATATTGTTGGCTGATTTACACGCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCCA AATATGTTTATGGAAGTCCGTTCCAGCTTGATAAGGATTATAC ACTGAATGCTTATAGATTGGCTTTAAAAACTACCTTAAAAAGA GCAAGAAGGAGTATGGAACCTTATAGCAAGAGAGGATGAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATCA GATTCATTC TAGTGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGAGCTTTTACCAAAAA AGGTTGTTTTGTATTACAACCCCTGCTTAAACGGGTTTGGATGG AGAAGGAAAGATGAGTTCTTCAAAGGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGCTAAGATAAAGAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACCTCTTGAATATCTTTAACCATAAAAAAGGCCA GAAAAATTTGGTGGAGATTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTA</p>
10	<p>secuencia de aminoácidos de p-yodo-L-fenilalanina aminoacil-ARNt sintetasa (pIPheRS) (derivada de tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre)</p>	<p>MDEFEMIKRNTSEIISEEELREVLKKEKSALIGFEPGKIHL GHYLQIKKMIDLQNAQFDIIILLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNPLHYEGVDVAVGGME QRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSKSNFNFIIV DDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGGDLTVNSYEELESFLFKNKELHPMDLKNVAEELIKILEP IRKRL</p>

Fig. 8 cont.

SEC ID N°	Descripción	SECUENCIA
11	secuencia de nucleótidos de p-yodo-L-fenilalanina aminoacil-ARNt sintetasa	ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTA TCAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAA ATCTGCTCTGATAGGTTTTGAACCAAGTGGTAAAATACATTTA GGGCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATG CTGGATTTGATATAAATTATATTGTTGGCTGATTTACACGCCTA TTTAAACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGA GATTATAACAAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAA AATATGTTTTATGGAAGTTCGTTCCAGCTTGATAAGGATTATAC ACTGAATGCTATAGATTGGCTTTAAAAACACCTTAAAAAAGA GCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAAATC CAAAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATCC TCTTCATTATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAG CAGAGAAAAATACACATGTTAGCAAGGGGAGCTTTTACAAAAA AGTTGTTTTGTATTCAACAACCCTGTCTTAACGGGTTTGGATGG AGAAGGAAGATGAGTTCTTCAAAGGGGAATTTTATAGCTGTT GATGACTCTCCAGAAGAGATTAGGGCTAAGATAAAGAAAAGCAT ACTGCCAGCTGGAGTTGTTGAAGGAAATCCAATAATGGAGAT AGCTAAATACTTCCTTGAATATCCTTTAACCATAAAAAAGGCCA GAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAGT TAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTT AAAAAATGCTGTAGCTGAAGAACCTATAAAGATTTTAGAGCCA ATTAGAAAGAGATTA
12	promotor de glnS mutante de E. coli que tiene una secuencia TATC en lugar de GATC en la región -10 (identificada por Plumbridge and Söll, <i>Biochimie</i> 69:539-541 [1987])	CGATTATCAATTTTAAAAAATAACAGTTGTCAGCCTGTCCCG CTTATAATATCATACGCC
13	promotor de glnS mutante de E. coli modificado TRN	CGATTATCAATTTTAAAAAATAACAGTTGTCAGCCTGTCCCG CTTATAATATCATACGCC
14	secuencia de unión de ARNt entre los genes valU y valX de E.coli	ACTACTTTATGTAGTCTCCGCCGTAGCAAGAAATTGAGAAG T
15	secuencia de unión de ARNt entre los genes ileT y alaT de E.coli	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT
16	secuencia de unión de ARNt entre los genes serV y argV E.coli	TTT
17	secuencia de unión de ARNt entre los genes valV y valW de E.coli	TCCT
18	secuencia de unión de ARNt entre los genes glyt t thrT de E.coli	AGATGT
19	secuencia de unión de ARNt entre los genes metI y leuW de E.coli	TCTTTTTTTT
20	secuencia de unión de ARNt entre los genes gnW y metU de E.coli	TCGAAGAAACAATCT
21	secuencia de unión de ARNt entre los genes hisR y leuT de E.coli	TTATTAGAAGTTGTGACAAT
22	secuencia de unión de ARNt entre los genes glnU y glnW de E.coli	TCTTCTTCGAGTAAAGCGGTTCCACGCCCGGTTAT
23	secuencia de unión de ARNt entre los genes leuP y leuV de E.coli	AACGAGGCGATATCAAAAAAAGTAAGATGACTGT

Fig. 8 cont.

SEC ID N°	Descripción	SECUENCIA
24	secuencia de unión del ARNt entre los genes glnV y glnX de E.coli	ATTTATTCAAGACGCTTACCTTGTAAGTGCACCCAGT
25	secuencia de unión del ARNt entre los genes alaW y alaX de E.coli	AATTTTGCACCCAGCAAACCTTGGTACGTAAACGCATCGT
26	secuencia de unión del ARNt entre los genes ileU y alaU de E.coli	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT
27	secuencia de unión del ARNt entre los genes ileV y alaV de E.coli	AATTTGCACGGCAAATTTGAAGAGGTTTTAACTACATGTTAT
28	secuencia de unión del ARNt entre los genes metU y glnV de E.coli	AATCTGAATGTATCGAATATGTTCCGGCAAATTCAAAACCAATTTGT
29	secuencia de unión del ARNt entre los genes glyW y cyst de E.coli	GTTTAAAAGACATCGCGCTCAAGCGGATGTCTGGCTGAAAGGCCTGAAGAATTT
30	secuencia de unión del ARNt entre los genes argX y hisR de E.coli	TTTAGTCCC GGCGCTTGAGCTGCGGTGGTAGTAATACCGCGTACAAGATTTGTAGT
31	secuencia de unión del ARNt entre los genes argY y argZ de E.coli	TCTCTACTTGATATGGCTTTAGTAGCGGTATCAATATCAGCAGTAAATAAATTTCCCGAT
32	promotor de proK de E.coli	AGGCATTTTGCTATTAAGGGATTGACGAGGGCGTATCTGCGCATTAAGATGCGCCCCGCATT
33	terminador de proK de E.coli	AATTCGAAAAGCCTGCTCAACGAGCAGGCTTTTTT
34	promotor de proL de E.coli	ATCAGTTAGCGAAAATATCTTACTTGCAATCGGTGTGGAAAACGGTAGTATTAGCAGCCACGAGT
35	terminador de proL de E.coli	AAAATCCCAAGAAAAACCAACCTTACGGTTGGTTTTTTTT
36	terminador de proM de E.coli	ATTTTGAACCCCGCTTCGGCGGGTTTTTT
37	Cebador de proK P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAGGC
38	Cebador de proK T1	CTTTTCTCGCCGATCAAAGGCATTTTGCATTAAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCGGCGGTAGTTCAGCAGGGC
39	Cebador de proK T2	CTTTTCTCGCCGATCAAAGGCATTTTGCATTAAAGGGATTGACGAGGGCGTATCTGCGCAGTAAGATGCGCCCCGCATTCGGCGGTAGTTCAGCAGGGC
40	Cebador de proK P2	GCATAAGCTTATGCAAAAAAGCCTGCTCGTTGAGCAGGCTTTTCG
41	Cebador de proK-F	AGTCTGATCAGTGCACGGCTAACTAAGCGG
42	Cebador de proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG
43	Cebador de glnS P1	CCGAGCTCCCGGGTCATC
44	Cebador de glnS T1	CCGAGCTCCCGGGTCATCAATCATCCCATAATCCTTGTTAGATTATCAATTTTAAAAAATAACAGTTGTCAGCCTGTC
45	Cebador de glnS T2	GTCCATATGGGATTCTCAAAGCGTAAACAACGTATAACGGCGTATGATATTATAAGCGGGACAGGCTGACAACCTGTTAG
46	Cebador de glnS P2	GTCCATATGGGATTCTC
47	Adaptador del cebador P1	GTGCACGGCTAACTAAGCGGCCTGCTGACTTTCTCGCCGATCAAAGGC
48	Adaptador del cebador P2	TACACGGCGGAGACTACATAAAGTAGTTGGTCCGGCGGGCCGGATTG
49	Adaptador del cebador P3	GTAGTCTCCGCGGTAGCAAGAAATTGAGAAGTCCGGCGGTA GTTCAGCAG
50	Adaptador del cebador P4	AAACCTCTTCAAATTTGCCGTGCAAATTTGGTCCGGCGGGCCG GATTTG

Fig. 8 cont.

SEC ID N°	Descripción	SECUENCIA
51	Adaptador del cebador P5	GCAAATTTGAAGAGGTTTTAACTACATGTTATCCGGCGGTAGT TCAGCAG
52	Cebador de proK-R	GCATCTCGAGATGCAAAAAAGCCTGCTCGTTG
53	Cebador en tándem P1	ATCAGTGCACGGCTAACTAAGCGG
54	Cebador en tándem P2	GCTGGCATGCATGCAAAAAAGCCTGCTCGTTGAGC
55	Cebador en tándem P3	ATCAGCATGCGGCTAACTAAGCGGCCTGCTG
56	Cebador en tándem P4	GCTGCTCGAGATGCAAAAAAGCCTGC
57	Secuencia líder	MDPLVTAASVLEFGLFET
58	Sonda de análisis Northern	CCCTGCTGAACTACCGCC-

Fig. 8 cont.