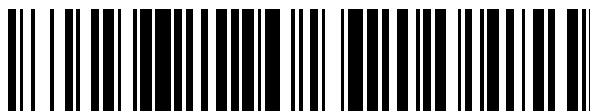


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 396 712**

51 Int. Cl.:

**C12N 9/22** (2006.01)

**C12N 15/12** (2006.01)

**C12N 9/10** (2006.01)

**C12N 15/54** (2006.01)

**C12P 21/02** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **13.10.2004 E 04795276 (7)**

97 Fecha y número de publicación de la concesión europea: **21.11.2012 EP 1687401**

54 Título: **Incorporación específica para sitio de aminoácidos activos por rédox en proteínas**

30 Prioridad:

**14.10.2003 US 511532 P**

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

**25.02.2013**

73 Titular/es:

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

72 Inventor/es:

**ALFONTA, LITAL;  
SCHULTZ, PETER, G. y  
ZHANG, ZHIWEN**

74 Agente/Representante:

**PONTI SALES, Adelaida**

**ES 2 396 712 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

Incorporación específica para sitio de aminoácidos activos por rédox en proteínas.

## 5 CAMPO DE LA INVENCION

[0001] La invención pertenece al campo de la bioquímica de traducción. La invención se refiere a composiciones y procedimientos para preparar y usar ARNt ortogonales, aminoacil-ARNt sintetasa ortogonales y pares de los mismos que incorporan aminoácidos activos por rédox en proteínas. La invención también se refiere a procedimientos de producción de proteínas en células usando tales pares y composiciones relacionadas.

## ANTECEDENTES DE LA INVENCION

[0002] Entre los veinte aminoácidos comunes genéticamente codificados, sólo la cisteína experimenta una química rédox superficial y, como resultado, puede participar en una amplia variedad de reacciones de oxidación y reducción catalizadas por enzimas (Surdhar y Armstrong (1987) J. Phys. Chem., 91: 6532 – 6537; Licht y col. (1996) Science 271: 477 – 481). Por consiguiente, la mayoría de los procesos rédox biológicos requieren cofactores tales como flavinas, nicotinamidas e iones metálicos. En casos raros, las quinonas derivadas de la modificación postraduccional de cadenas laterales de tirosina y triptófano se usan como el cofactor rédox (Stubbe y Van der Donk (1998) Chem. Rev., 98: 705 – 762). Por ejemplo, la cobre amina oxidasa de plasma bovino usa 3,4,6-trihidroxil-L-fenilalanina (TOPA) en la conversión de aminas primarias y oxígeno molecular en aldehídos y peróxido de hidrógeno, respectivamente (Janes y col. (1990) Science 248: 981 – 987). Estos catalizadores rédox derivados de aminoácidos se generan por tanto reacciones mediadas por radicales como enzimáticas (Rodgers y Dean (2000) Int. J. Biochem. Cell Biol., 32: 945 – 955). Claramente, la capacidad para codificar genéticamente aminoácidos activos por rédox adicionales, en vez de generarlos por mecanismos postraducionales complejos, potenciaría significativamente la capacidad para tanto estudiar como manipular los procesos de transferencia de electrones en proteínas. La presente invención satisface estas y otras necesidades, como serán evidentes tras la revisión de la siguiente divulgación.

[0003] Wang, L. y col., 2003, Proc. Natl. Acad. Sci. USA, vol. 100, pág. 56 – 61, describen la adición del grupo funcional ceto al código genético de *Escherichia coli*.

[0004] El documento WO 02/086075 A2 describe procedimientos y composiciones para la producción de pares de ARNt-aminoacil ARNt sintetasa ortogonales.

## RESUMEN DE LA INVENCION

[0005] La invención proporciona, en un primer aspecto, una composición que comprende una aminoacil-ARNt sintetasa ortogonal (O-RS) en la que la O-RS tiene al menos el 90 % de identidad de secuencias de aminoácidos con la longitud completa de la secuencia de SEQ ID NO: 1 y en la que dicha O-RS tiene una eficiencia que es al menos el 50 % de la eficiencia de traducción observada para la aminoacil-ARNt sintetasa que consiste en la secuencia de aminoácidos de SEQ ID NO: 1 en el sistema de traducción que comprende 3,4-dihidroxi-L-fenilalanina, el ARNt de SEQ ID NO: 2 y la aminoacil-ARNt sintetasa de SEQ ID NO: 1, y en la que dicha O-RS aminoacila el O-ARNt con 3,4-dihidroxi-L-fenilalanina preferencialmente.

[0006] En algunas realizaciones, la O-RS comprende una secuencia de aminoácidos de SEQ ID NO: 1.

[0007] En un segundo aspecto, la presente invención proporciona una célula que comprende la composición del primer aspecto. En algunas realizaciones, la célula es una célula de *E. coli*.

[0008] En algunas realizaciones del primer aspecto de la invención la composición comprende un sistema de traducción.

[0009] En algunas realizaciones del segundo aspecto de la invención, la célula comprende:

un ARNt ortogonal (O-ARNt) que está codificado por una secuencia de polinucleótidos que tiene al menos el 90 % de identidad de secuencias de nucleótidos con la secuencia de longitud completa de SEQ ID NO: 2;

dicha O-RS; y,

un primer aminoácido no natural que es 3,4-dihidroxi-L-fenilalanina (DHP).

[0010] En algunas realizaciones del segundo aspecto de la invención, el O-ARNt está codificado por la secuencia de polinucleótidos que se expone en SEQ ID NO: 2.

- 5 **[0011]** En algunas realizaciones del segundo aspecto de la invención, la célula comprende además un segundo par de O-ARNt/O-RS y un segundo aminoácido no natural, en la que el segundo O-ARNt reconoce un segundo codón selector y la segunda O-RS aminoacila el segundo O-ARNt con el segundo aminoácido no natural preferencialmente.
- [0012]** En algunas realizaciones del segundo aspecto de la invención, la célula es una célula no eucariota. En algunas realizaciones, la célula no eucariota es una célula de *E. coli*.
- 10 **[0013]** En algunas realizaciones del segundo aspecto de la invención, la célula comprende además un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en la que el polinucleótido comprende un codón selector que es reconocido por dicho O-ARNt.
- 15 **[0014]** En un tercer aspecto, la presente invención proporciona un polipéptido que comprende SEQ ID NO. 1.
- [0015]** En un cuarto aspecto, la presente invención proporciona un polinucleótido que codifica un polipéptido del tercer aspecto de la invención.
- 20 **[0016]** En un quinto aspecto, la presente invención proporciona un vector que comprende o que codifica un polinucleótido del cuarto aspecto de la invención. En algunas realizaciones, el vector comprende un plásmido, un cósmido, un fago o un virus.
- [0017]** En algunas realizaciones del quinto aspecto de la invención, el vector es un vector de expresión.
- 25 **[0018]** En un sexto aspecto, la presente invención proporciona una célula que comprende el vector del quinto aspecto de la invención.
- [0019]** En un séptimo aspecto, la presente invención proporciona un procedimiento para identificar una aminoacil-ARNt sintetasa ortogonal que aminoacila específicamente un O-ARNt con 3,4-dihidroxi-L-fenilalanina (DHP), comprendiendo el procedimiento:
- 30 a) someter a selección una población de células de una primera especie, en la que las células comprenden cada una:
- 35 i) un miembro de una pluralidad de aminoacil-ARNt sintetasas (RS);
- ii) el ARNt ortogonal (O-ARNt);
- 40 iii) un polinucleótido que codifica un marcador de selección y comprende al menos un codón selector que es reconocido por dicho O-ARNt; y,
- iv) dicha DHP;
- 45 en la que células en dicha población que están potenciadas en la eficiencia de supresión comprenden un miembro de la pluralidad de RS que es una RS activa que aminoacila específicamente el O-ARNt con dicha DHP; y,
- b) seleccionar la RS activa que aminoacila el O-ARNt con la DHP,
- 50 en el que dicha RS activa tiene una eficiencia que es al menos el 50 % de la eficiencia observada para la aminoacil-ARNt sintetasa que consiste en la secuencia de aminoácidos de SEQ ID NO: 1 en un sistema de traducción que comprende 3,4-dihidroxi-L-fenilalanina, un ARNt de SEQ ID NO: 2 y la aminoacil-ARNt sintetasa de SEQ ID NO: 1; identificándose así la aminoacil-ARNt sintetasa ortogonal que aminoacila específicamente el O-ARNt.
- 55 **[0020]** En algunas realizaciones del séptimo aspecto de la invención, la selección comprende una selección positiva y el marcador de selección comprende un marcador de selección positiva.
- [0021]** En algunas realizaciones del séptimo aspecto de la invención, la pluralidad de RS comprende RS mutantes, o RS derivadas de una o más especies distintas de la primera especie o tanto RS mutantes como RS derivadas de una especie distinta de la primera especie.
- 60 **[0022]** En un octavo aspecto, la presente invención proporciona un procedimiento de producción de una proteína en una célula con 3,4-dihidroxi-L-fenilalanina (DHP) en una posición especificada, comprendiendo el procedimiento:

a) cultivar, en un medio apropiado, la célula del segundo aspecto de la invención,

b) proporcionar la DHP; y

5 c) incorporar la DHP en la posición especificada en la proteína durante la traducción del ácido nucleico con el al menos un codón selector, produciéndose así la proteína con dicha DHP en la posición especificada.

[0023] En algunas realizaciones del octavo aspecto de la invención, la O-RS comprende una secuencia de aminoácidos que comprende SEQ ID NO: 1.

10 [0024] En algunas realizaciones del octavo aspecto de la invención, la célula es una célula no eucariota. En algunas realizaciones, la célula no eucariota es una célula de *E. coli*.

## DEFINICIONES

15 [0025] Ante de describir la invención en detalle debe entenderse que la presente invención no se limita a sistemas biológicos particulares que, por supuesto, pueden variar. También debe entenderse que la terminología usada en el presente documento es sólo con el fin de describir realizaciones particulares, y no pretende ser limitante. Como se usa en esta memoria descriptiva y las reivindicaciones adjuntas, las formas en singular “un”, “una”, “el” y “la” incluyen referentes plurales, a menos que el contenido dicte claramente de otro modo. Por tanto, por ejemplo, referencia a “una célula” incluye una combinación de dos o más células; referencia a “bacteria” incluye mezclas de bacterias, y similares.

20 [0026] A menos que se defina en el presente documento y más adelante en el resto de la memoria descriptiva, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que comúnmente es entendido por un experto en la materia a la que se refiere la invención.

25 [0027] Ortogonal: como se usa en el presente documento, el término “ortogonal” se refiere a una molécula (por ejemplo, un ARNt ortogonal (O-ARNt) y/o una aminoacil-ARNt sintetasa ortogonal (O-RS)) que funciona con componentes endógenos de una célula con eficiencia reducida con respecto a una molécula correspondiente que es endógena a la célula o sistema de traducción, o que fracasa en funcionar con componentes endógenos de la célula. En el contexto de ARNt y aminoacil-ARNt sintetetasas, ortogonal se refiere a una incapacidad o eficiencia reducida, por ejemplo, inferior al 20 % de eficiencia, inferior al 10 % de eficiencia, inferior al 5 % de eficiencia o inferior al 1 % de eficiencia, de un ARNt ortogonal que funciona con una ARNt sintetasa endógena en comparación con un ARNt endógeno que funciona con la ARNt sintetasa endógena, o de una aminoacil-ARNt sintetasa ortogonal que funciona con un ARNt endógeno en comparación con una ARNt sintetasa endógena que 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 está aminoacilado por cualquier RS endógena de la célula con eficiencia reducida o incluso cero cuando se compara con la aminoacilación de un ARNt endógeno por la RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno de una célula de interés con eficiencia reducida o incluso cero con respecto a la aminoacilación del ARNt endógeno por una RS endógena. Una segunda molécula ortogonal puede introducirse en la célula que funciona con la primera molécula ortogonal. Por ejemplo, un par de ARNt ortogonal/RS incluye componentes complementarios introducidos que funcionan juntos en la célula con una eficiencia (por ejemplo, 45 % de eficiencia, 50 % de eficiencia, 60 % de eficiencia, 70 % de eficiencia, 75 % de eficiencia, 80 % de eficiencia, 90 % de eficiencia, 95 % de eficiencia o 99 % o más eficiencia) con respecto a la de un control, por ejemplo, un par de ARNt/RS endógeno correspondiente, o un par ortogonal activo (por ejemplo, un par de tirosil ARNt/RS ortogonal).

30 [0028] 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 ARNt es: (1) idéntico o sustancialmente similar a un tirosil-ARNt que se produce naturalmente, (2) derivado de un tirosil-ARNt que se produce naturalmente por mutagénesis natural o artificial, (3) derivado por un procedimiento que tiene en cuenta una secuencia de una secuencia de tirosil-ARNt natural o mutante de (1) o (2), (4) homólogo a un tirosil-ARNt natural o mutante; (5) homólogo a cualquier ARNt de ejemplo que está diseñado como un sustrato para una tirosil-ARNt sintetasa en la Tabla 1, o (6) una variante conservativa de cualquier ARNt de ejemplo que está diseñado como un sustrato para una tirosil-ARNt sintetasa en la Tabla 1. El tirosil-ARNt pueden existir cargado con un aminoácido, o en un estado sin cargar. También debe entenderse que un “tirosil-O-ARNt” es opcionalmente cargado (aminoacilado) por una sintetasa semejante con un aminoácido distinto de lisina, por ejemplo, con el aminoácido homoglutamina. De hecho, se apreciará que un tirosil-O-ARNt de la invención se usa ventajosamente para insertar esencialmente cualquier aminoácido, tanto si es natural como artificial, en un polipéptido en cultivo, durante la traducción, en respuesta a un codón selector.

55 [0029] Tirosil aminoácido sintetasa ortogonal: como se usa en el presente documento, una tirosil aminoácido sintetasa ortogonal (tirosil-O-RS) es una enzima que aminoacila preferencialmente el tirosil-O-ARNt con un aminoácido

en un sistema de traducción de interés. El aminoácido que la tirosil-O-RS carga sobre el tirosil-O-ARNt puede ser cualquier aminoácido, tanto si es natural como artificial, y no está limitado en el presente documento. La sintetasa es opcionalmente la misma que u homóloga a una tirosil aminoácido sintetasa que se produce naturalmente, o la misma que u homóloga a una sintetasa diseñada como una tirosil-O-RS en la Tabla 1. Por ejemplo, la tirosil-O-RS pueden ser una variante conservativa de una tirosil-O-RS de la Tabla 1 y/o puede ser al menos el 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o más idéntica en secuencia a una tirosil-O-RS de la Tabla 1.

**[0030]** Semejante: el término “semejante” se refiere a componentes que funcionan juntos, por ejemplo, un ARNt ortogonal y una aminoacil-ARNt sintetasa ortogonal. También puede denominarse que los componentes son complementarios.

**[0031]** Aminoacila preferencialmente: el término “aminoacila preferencialmente” se refiere a una eficiencia, por ejemplo, del 70 % de eficiencia, 75 % de eficiencia, 85 % de eficiencia, 90 % de eficiencia, 95 % de eficiencia o 99 % o más de eficiencia, a la que una O-RS aminoacila un O-ARNt con un aminoácido activo por rédox con respecto a la O-RS que aminoacila un ARNt que se produce naturalmente o un material de partida usado para generar el O-ARNt.

**[0032]** Codón selector: el término “codón selector” se refiere a codones reconocidos por el O-ARNt en el proceso de traducción y no reconocidos por un ARNt endógeno. El bucle anticodón de O-ARNt reconoce el codón selector sobre el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido no natural, tal como un aminoácido activo por rédox, en este sitio en el polipéptido. Los codones selectores pueden incluir, por ejemplo, codones antisentido tales como codones de terminación, por ejemplo, codones ámbar, ocre y ópalo; codones de cuatro o más bases; codones raros; codones derivados de pares de bases naturales o no naturales y/o similares.

**[0033]** ARNt supresor: un ARNt supresor es un ARNt que altera la lectura de un ARN mensajero (ARNm) en un sistema de traducción dado, por ejemplo, proporcionándose un mecanismo para incorporar un aminoácido en una cadena de polipéptidos en respuesta a un codón selector. Por ejemplo, un ARNt supresor puede ultraleer, por ejemplo, un codón de terminación, un codón de cuatro bases, un codón raro, etc.

**[0034]** Actividad de supresión: como se usa en el presente documento, el término “actividad de supresión” se refiere, en general, a la capacidad de un ARNt (por ejemplo, un ARNt supresor) para permitir la ultralectura traduccional de un codón (por ejemplo, un codón selector que es un codón ámbar o un codón de 4 o más bases) que de otro modo produciría la terminación de la traducción o la traducción errónea (por ejemplo, desplazamiento del marco). La actividad de supresión de un ARNt supresor puede expresarse como un porcentaje de actividad de ultralectura traduccional observada en comparación con un segundo ARNt supresor, o con respecto a un sistema de control, por ejemplo, un sistema de control que carece de una O-RS.

**[0035]** La presente invención proporciona diversos medios por los que puede cuantificarse la actividad de supresión. El porcentaje de supresión de un O-ARNt y O-RS particular contra un codón selector (por ejemplo, un codón ámbar) de interés se refiere al porcentaje de actividad de un marcador de prueba expresado dado (por ejemplo, LacZ), que incluye un codón selector, en un ácido nucleico que codifica el marcador de prueba expresado, en un sistema de traducción de interés, en el que el sistema de traducción de interés incluye una O-RS y un O-ARNt, con respecto a una construcción de control positivo, en el que el control positivo carece del O-ARNt, la O-RS y el codón selector. Por tanto, por ejemplo, si una construcción de marcador de 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 de marcador en cuestión, entonces el porcentaje de supresión de una construcción de prueba que comprende el codón selector es el porcentaje de X que la construcción del marcador de prueba muestra bajo esencialmente las mismas condiciones medioambientales bajo las que se expresó el marcador de control positivo, excepto que la construcción de 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 O-ARNt. Opcionalmente, la medición del porcentaje de supresión puede refinarse por comparación del marcador de prueba con una construcción de marcador de control de “referencia” o “negativo” que incluye el mismo codón selector que el marcador de prueba, pero en un sistema que no incluye el O-ARNt, O-RS y/o aminoácido relevante reconocido por el O-ARNt y/o O-RS. Este control negativo es útil en la normalización de las mediciones del porcentaje de supresión para explicar efectos de señales de referencia del marcador en el sistema de traducción de interés.

**[0036]** La eficiencia de supresión puede determinarse por distintos ensayos conocidos en la técnica. Por ejemplo, puede usarse un ensayo indicador de  $\beta$ -galactosidasa, por ejemplo, un plásmido lacZ derivado (en el que la construcción tiene un codón selector en la secuencia de ácidos nucleicos de lacZ) se introduce en células de un organismo apropiado (por ejemplo, un organismo en el que pueden usarse componentes ortogonales) junto con plásmido que comprende un O-ARNt de la invención. También puede introducirse una sintetasa semejante (tanto como un polipéptido o un polinucleótido que codifica la sintetasa semejante cuando se expresa). Las células se cultivan en medios a una densidad deseada, por ejemplo, a una  $DO_{600}$  de aproximadamente 0,5, y se realizan ensayos de  $\beta$ -galactosidasa, por ejemplo, usando el kit de ensayo de  $\beta$ -galactosidasa BetaFluor™ (Novagen). El

porcentaje de supresión puede calcularse como el porcentaje de actividad para una muestra con respecto a un control comparable, por ejemplo, el valor observado de la construcción de lacZ derivatizado, en la que la construcción tiene un codón sentido correspondiente en la posición deseada en vez de un codón selector.

5 **[0037]** Sistema de traducción: el término “sistema de traducción” se refiere a los componentes que incorporan un aminoácido en una cadena de polipéptidos en cultivo (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetasas, ARNm y similares. El O-ARNt y/o las O-RS de la invención pueden añadirse a o ser parte de un sistema de traducción *in vitro* o *in vivo*, por ejemplo, en una célula no eucariota, por ejemplo, una bacteria (tal como *E. coli*), o en una célula eucariota, por ejemplo, una célula de levadura, una célula de mamífero, una célula de planta, una célula de alga, una célula de hongo, una célula de insecto y/o similares.

15 **[0038]** Aminoácido no natural: como se usa en el presente documento, el término “aminoácido no natural” se refiere a cualquier aminoácido, aminoácido modificado y/o análogo de aminoácido, tal como un aminoácido activo por rédox, que no es uno de los 20 aminoácidos comunes que se producen naturalmente o selenocisteína o pirrolisina.

20 **[0039]** Derivado de: como se usa en el presente documento, el término “derivado de” se refiere a un componente que se aísla de o se prepara usando una molécula u organismo específico, o información de la molécula u organismo específico.

25 **[0040]** Selección positiva o marcador de cribado: como se usa en el presente documento, el término “selección positiva o marcador de cribado” se refiere a un marcador que, cuando está presente, por ejemplo, expresado, activado o similares, produce la identificación de una célula que comprende el rasgo, por ejemplo, células con el marcador de selección positiva, de aquellos sin el rasgo.

30 **[0041]** Selección negativa o marcador de cribado: como se usa en el presente documento, el término “selección negativa o marcador de cribado” se refiere a un marcador que, cuando está presente, por ejemplo, expresado, activado o similares, permite la identificación de una célula que no comprende una propiedad o rasgo seleccionado (por ejemplo, con respecto a una célula que no posee la propiedad o rasgo).

35 **[0042]** Indicador: como se usa en el presente documento, el término “indicador” se refiere a un componente que puede usarse para identificar y/o seleccionar componentes diana de un sistema de interés. Por ejemplo, un indicador puede incluir una proteína, por ejemplo, una enzima que confiere resistencia o sensibilidad a antibióticos (por ejemplo,  $\beta$ -lactamasa, cloranfenicol acetiltransferasa (CAT) y similares), un marcador de cribado fluorescente (por ejemplo, proteína fluorescente verde (por ejemplo, GFP), YFP, EGFP, RFP, etc.), un marcador luminiscente (por ejemplo, una proteína de luciferasa de luciérnaga), un marcador de cribado basado en afinidad o gen positivo o negativo de los marcadores de selección tales como lacZ,  $\beta$ -gal/lacZ ( $\beta$ -galactosidasa), Adh (alcohol deshidrogenasa), his3, ura3, leu2, lys2 o similares.

40 **[0043]** Eucariota: como se usa en el presente documento, el término “eucariota” se refiere a organismos que pertenecen al dominio filogenético Eucarya tal como animales (por ejemplo, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (por ejemplo, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidios, protistas, etc.

45 **[0044]** No eucariotas: como se usa en el presente documento, el término “no eucariota” se refiere a organismos no eucariotas. Por ejemplo, un organismo no eucariota puede pertenecer al dominio filogenético eubacteria (por ejemplo, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, etc.), o el dominio filogenético arquea (por ejemplo, *Methanococcus jannaschii* (Mj), *Methanosarcina mazei* (Mm), *Methanobacterium thermoautotrophicum* (Mt), *Methanococcus marisaludis*, *Methanopyrus kandleri*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus* (Af), *Pyrococcus furiosus* (Pf), *Pyrococcus horikoshii* (Ph), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Aeropyrum pernix* (Ap), *Thermoplasma acidophilum*, *Thermoplasma volcanium*, etc.).

55 **[0045]** Variante conservativa: como se usa en el presente documento, el término “variante conservativa” en el contexto de un componente de traducción se refiere a un componente de traducción, por ejemplo, un O-ARNt de variante conservativa o un O-RS de variante conservativa que responde funcionalmente similar a un componente base al que la variante conservativa es similar, por ejemplo, un O-ARNt o O-RS, que tiene variaciones en la secuencia con respecto a un O-ARNt o O-RS de referencia. Por ejemplo, una O-RS aminoacilará un O-ARNt complementario o un O-ARNt de variante conservativa con un aminoácido no natural, por ejemplo, un aminoácido activo por rédox tal como 3,4-dihidroxi-L-fenilalanina (DHP), aunque el O-ARNt y el O-ARNt de variante conservativa no tienen la misma secuencia. La variante conservativa puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones o cinco o más variaciones en la secuencia, en tanto que la variante conservativa sea complementaria a los O-ARNt o O-RS correspondientes.

[0046] Agente de selección o de cribado: como se usa en el presente documento, el término “agente de selección o de cribado” se refiere a un agente que, cuando está presente, permite la selección/cribado de ciertos componentes de una población. Por ejemplo, un agente de selección o de cribado puede ser, pero no se limita a, por ejemplo, un nutriente, un antibiótico, una longitud de onda de luz, un anticuerpo, un polinucleótido expresado o similares. El agente de selección puede variarse, por ejemplo, por concentración, intensidad, etc.

[0047] En respuesta a: como se usa en el presente documento, el término “en respuesta a” se refiere al procedimiento en el que un ARNt de la invención reconoce un codón selector y media en la incorporación del aminoácido activo por rédox, que está unido a un ARNt, en la cadena de polipéptidos en cultivo.

[0048] Codificar: como se usa en el presente documento, el término “codificar” se refiere a cualquier proceso por el cual la información en una macromolécula polimérica o cadena de secuencias se usa para dirigir la producción de una segunda molécula o cadena de secuencias que es diferente de la primera molécula o cadena de secuencias. Como se usa en el presente documento, el término se usa ampliamente y puede tener una variedad de aplicaciones. En un aspecto, el término “codificar” describe el procedimiento de replicación de ADN semi-conservativo, en el que una cadena de una molécula de ADN bicatenario se usa como molde para codificar una cadena hermana complementaria recientemente sintetizada por una ADN polimerasa dependiente de ADN.

[0049] En otro aspecto, el término “codificar” se refiere a cualquier proceso por el cual la información en una molécula se usa para dirigir la producción de una segunda molécula que tiene una naturaleza química diferente de la primera molécula. Por ejemplo, una molécula de ADN puede codificar una molécula de ARN (por ejemplo, mediante el proceso de transcripción que incorpora una enzima ARN polimerasa dependiente de ADN). Por tanto, una molécula de ARN puede codificar un polipéptido, como en el proceso de traducción. Cuando se usa para describir el proceso de translación, el término “codificar” también se extiende al codón de triplete que codifica un aminoácido. En algunos aspectos, una molécula de ARN puede codificar una molécula de ADN, 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 el que se entiende que “codificar” como se usa en ese caso incorpora tanto los procedimientos de transcripción como de traducción.

## BREVE DESCRIPCIÓN DE LAS FIGURAS

[0050] La **FIG. 1** proporciona una ilustración esquemática de los productos oxidativos de 3,4-dihidroxi-L-fenilalanina (DHP; estructura 1) en el radical 2 de DHP-semiquinona, que se oxida fácilmente a DHP-quinona 3.

[0051] Las **FIGS. 2A y 2B** proporcionan ilustraciones de expresión dependiente de DHP de mioglobina de esperma de ballena como una respuesta a un codón ámbar en la posición 4 en el gen Mb. La **FIG. 2A** proporciona un gel teñido con plata y transferencia Western. La **FIG. 2B** proporciona un análisis del espectro de masas de ESI-QqTOF de DHPMb.

[0052] Las **FIGS. 3A y 3B** proporcionan voltamogramas cíclicos. La **FIG. 3A** proporciona voltamogramas cíclicos del mismo grupo hemo en wtMb y DHPMb. La **FIG. 3B** proporciona voltamogramas cíclicos de DHP para diferentes disoluciones que contienen: DHP 100  $\mu$ M, wtMb o DHPMb. Todos los voltamogramas se registraron en tampón fosfato 0,1 M, pH 7,4, bajo argón; tasa de barrido: 1 V s<sup>-1</sup> frente a SCE.

[0053] La **FIG. 4** proporciona una ilustración de la oxidación de DHP electroquímicamente dentro de una proteína.

[0054] La **FIG. 5** proporciona las secuencias de nucleótidos y de aminoácidos de tirosil-ARNt sintetasa de *Methanococcus jannaschii* (MjTyrRS).

[0055] La **FIG. 6** proporciona las secuencias de nucleótidos y de aminoácidos de 3,4-dihidroxi-L-fenilalanina (DHP)-ARNt sintetasa (DHPRS), basadas en tirosil-ARNt sintetasa de *Methanococcus jannaschii* (MjTyrRS) que tienen los cambios de aminoácidos: Tyr32 → Leu, Ala67 → Ser, His70 → Asn y Ala167 → Gln. Están recuadrados los aminoácidos cambiados y los codones de triplete correspondientes (con respecto a la secuencia natural).

## DESCRIPCIÓN DETALLADA

[0056] Con el fin de añadir aminoácidos sintéticos adicionales tales como un aminoácido activo por rédox al código genético, *in vivo*, se necesitan nuevos pares ortogonales de una aminoacil-ARNt sintetasa y un ARNt que puedan funcionar eficientemente en la maquinaria de traducción, pero que sean “ortogonales”, al sistema de traducción en cuestión, que significa que funciona independientemente de las sintetetas y ARNt endógenos al sistema de traducción. Características deseadas del par ortólogo incluyen ARNt que descodifican o reconocen sólo un nuevo codón específico, por ejemplo, un codón selector que no está codificado por ningún ARNt endógeno, y

aminoacil-ARNt sintetasas que aminoacilan (o cargan) preferencialmente su ARNt semejante con sólo un aminoácido activo por rédox específico. El O-ARNt tampoco se aminoacila normalmente por sintetasas endógenas. Por ejemplo, en *E. coli*, un par ortogonal incluirá una aminoacil-ARNt sintetasa que no reacciona de forma cruzada con ninguno del ARNt endógeno, por ejemplo, de los que hay 40 en *E. coli*, y un ARNt ortogonal que no está aminoacilado por ninguna de las sintetasas endógenas, por ejemplo, de las que hay 21 en *E. coli*.

**[0057]** La presente invención proporciona composiciones de y procedimientos como se definen en las reivindicaciones.

**[0058]** Por ejemplo, el aminoácido activo por rédox 3,4-dihidroxi-L-fenilalanina (DHP), que puede someterse a oxidación de dos electrones en una quinona, se ha incorporado selectivamente y eficientemente en proteínas en un organismo, por ejemplo, *Escherichia coli* (*E. coli*) en respuesta a un codón selector, por ejemplo, el codón TAG. Véase la **FIG. 1**. DHP puede oxidarse electroquímicamente dentro de la proteína. La capacidad para incorporar un aminoácido activo por rédox específicamente para sitio en proteínas puede facilitar el estudio de la transferencia de electrones en proteínas, además de permitir la manipulación de una proteína rédox con novedosas propiedades. Véase la **FIG. 4**. Por ejemplo, la expresión de proteínas activas por rédox puede facilitar el estudio y la capacidad para alterar las rutas de transferencia de electrones en proteínas, alterar la función catalítica de enzimas, reticular la proteína con moléculas pequeñas y biomoléculas, etc.

#### ARNt ORTOGONAL / AMINOACIL-ARNt SINTETASAS ORTOGONALES Y PARES DE LOS MISMOS

**[0059]** Los sistemas de traducción que son adecuados para preparar proteínas que incluyen uno o más aminoácidos no naturales, por ejemplo, aminoácidos activos por rédox, se describen en las publicaciones internacionales número WO 2002/086075 titulada "PROCEDIMIENTOS Y COMPOSICIÓN PARA LA PRODUCCIÓN DE PARES DE ARNt-AMINOACIL-ARNt SINTETASA ORTOGONALES" y WO 2002/085923 titulada "INCORPORACIÓN IN VIVO DE AMINOÁCIDOS NO NATURALES". Además, véase la solicitud internacional número PCT/US2004/011786 presentada el 16 de abril de 2004. Tales sistemas de traducción generalmente comprenden células (por ejemplo, células no eucariotas o células eucariotas) que incluyen un ARNt ortogonal (O-ARNt), una aminoacil ARNt sintetasa ortogonal (O-RS) y un aminoácido activo por rédox en las que la O-RS aminoacila el O-ARNt con el aminoácido activo por rédox. Un par ortogonal de la invención incluye un O-ARNt, por ejemplo, un ARNt supresor, un ARNt de desplazamiento de marco o similares, y una O-RS. Componentes individuales también se proporcionan en la invención.

**[0060]** El O-ARNt reconoce un codón selector e incluye al menos aproximadamente, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 % o un 90 % o más de eficiencia de supresión en presencia de una sintetasa semejante en respuesta a un codón selector con respecto al O-ARNt que comprende o está codificado por una secuencia de polinucleótidos que se expone en SEQ ID NO: 2. La O-RS aminoacila el O-ARNt con el aminoácido activo por rédox. La célula usa los componentes para incorporar el aminoácido activo por rédox en una cadena de polipéptidos en cultivo, por ejemplo, mediante 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. En ciertas realizaciones de la invención, una célula tal como una célula de *E. coli* que incluye un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), un aminoácido activo por rédox; y un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en la que el polinucleótido comprende el codón selector que es reconocido por el O-ARNt. El sistema de traducción también puede ser un sistema *in vitro*.

**[0061]** En una realización, la eficiencia de supresión de la O-RS y el O-ARNt juntos es aproximadamente, por ejemplo, 5 veces, 10 veces, 15 veces, 20 veces o 25 veces o más superior que la eficiencia de supresión del O-ARNt que carece de la O-RS. En un aspecto, la eficiencia de supresión de la O-RS y el O-ARNt juntos es al menos aproximadamente, por ejemplo, el 35 %, 40 %, 45 %, 50 %, 60 %, 75 %, 80 % o el 90 % o más de la eficiencia de supresión de un par de tirosil-ARNt sintetasa ortogonal derivado de *Methanococcus jannaschii*.

**[0062]** La invención incluye opcionalmente múltiples pares de O-ARNt/O-RS en una célula, que permiten la incorporación de más de un aminoácido no natural, por ejemplo, un aminoácido activo por rédox y otro aminoácido no natural. Por ejemplo, la célula pueden incluir adicionalmente un par de O-ARNt/O-RS diferente adicional y un segundo aminoácido no natural, en la que este O-ARNt adicional reconoce un segundo codón selector y esta O-RS adicional aminoacila preferencialmente 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, por ejemplo, leucilo, lisilo, glutamilo, etc., (en el que el segundo O-ARNt reconoce un codón selector diferente, por ejemplo, un ópalo, cuatro bases o similares).

**[0063]** El O-ARNt y/o la O-RS pueden producirse naturalmente o pueden derivarse por mutación de un ARNt y/o RS que se producen naturalmente, por ejemplo, que genera bibliotecas de ARNt y/o bibliotecas de RS, de una variedad de organismos. Por ejemplo, una estrategia para producir un par de ARNt / aminoacil-ARNt sintetasa ortogonal implica importar un par de ARNt/sintetasa heterólogo (a la célula huésped) de, por ejemplo, una fuente



distinta de la célula huésped, o múltiples fuentes; en la célula huésped. Las propiedades del candidato de sintetasa heterólogo incluyen, por ejemplo, que no carga ningún ARNt de célula huésped, y las propiedades del candidato de ARNt heterólogo incluyen, por ejemplo, que no es aminoacilado por ninguna sintetasa de célula huésped. Además, el ARNt heterólogo es ortogonal a todas las sintetasas de célula huésped.

5 **[0064]** Una segunda estrategia para generar un par ortogonal implica generar bibliotecas mutantes de las que cribar y/o seleccionar un O-ARNt o O-RS. Estas estrategias también pueden combinarse.

ARNt ortogonal (O-ARNt)

10 **[0065]** Un ARNt ortogonal (O-ARNt) media en la incorporación de un aminoácido activo por rédox en una proteína que está codificada por un polinucleótido que comprende un codón selector que es reconocido por el O-ARNt, por ejemplo, *in vivo* o *in vitro*. En ciertas realizaciones, 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 eficiencia de supresión en presencia de una sintetasa semejante en respuesta a un codón selector con respecto al O-ARNt que comprende o está codificado por una secuencia de polinucleótidos que se expone en SEQ ID NO: 2.

15 **[0066]** La eficiencia de supresión puede determinarse por distintos ensayos conocidos en la técnica. Por ejemplo, puede usarse un ensayo indicador de  $\beta$ -galactosidasa, por ejemplo, un plásmido lacZ derivatizado (en el que la construcción tiene un codón selector en la secuencia de ácidos nucleicos de lacZ) se introduce en células de un organismo apropiado (por ejemplo, un organismo en el que pueden usarse los componentes ortogonales) junto con plásmido que comprende un O-ARNt de la invención. Una sintetasa semejante también puede introducirse (tanto como un polipéptido como un polinucleótido que codifica la sintetasa semejante cuando se expresa). Las células se cultivan en medios a una densidad deseada, por ejemplo, a una DO<sub>600</sub> de aproximadamente 0,5, y se realizan ensayos de  $\beta$ -galactosidasa, por ejemplo, usando el kit de ensayo de  $\beta$ -galactosidasa BetaFluor™ (Novagen). El porcentaje de supresión puede calcularse como el porcentaje de actividad para una muestra con respecto a un control comparable, por ejemplo, el valor observado de la construcción de lacZ derivada, en la que la construcción tiene un codón sentido correspondiente en la posición deseada en vez de un codón selector.

20 **[0067]** Un ejemplo de O-ARNt de la invención es SEQ ID NO: 2. Véase la Tabla 1 y el Ejemplo 2, en el presente documento, para secuencias de moléculas de O-ARNt y O-RS a modo de ejemplo. Véase también la sección titulada “Secuencia de ácidos nucleicos y de polipéptidos y variantes” en el presente documento. En la molécula de ARNt, la timina (T) se sustituye con uracilo (U). También pueden estar presentes modificaciones adicionales a las bases. La invención también incluye variaciones conservativas de O-ARNt. Por ejemplo, variaciones conservativas de O-ARNt incluyen aquellas moléculas que funcionan como O-ARNt de SEQ ID NO: 2 y mantienen la estructura con forma de L de ARNt, pero no tienen la misma secuencia (y son distintas de moléculas de ARNt naturales). Véase también la sección en el presente documento titulada “Secuencia de ácidos nucleicos y de polipéptidos y variantes”.

25 **[0068]** La composición que comprende un O-ARNt pueden incluir adicionalmente una aminoacil-ARNt sintetasa ortogonal (O-RS) en la que la O-RS aminoacila preferencialmente el O-ARNt con un aminoácido activo por rédox. En ciertas realizaciones, una composición que incluye un O-ARNt pueden incluir adicionalmente un sistema de traducción (por ejemplo, *in vitro* o *in vivo*). 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 éstos, también puede estar presente en la célula. Véase también la sección en el presente documento titulada “Aminoacil-ARNt sintetasas ortogonales”.

30 **[0069]** Los O-ARNt pueden producirse generando una biblioteca de mutantes. La biblioteca de ARNt mutantes pueden generarse usando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, los ARNt mutantes pueden generarse por mutaciones específicas para sitio, mutaciones puntuales aleatorias, recombinación de homólogos, barajado de ADN u otros procesos de mutagénesis recurrentes, construcción quimérica o cualquier combinación de los mismos.

35 **[0070]** Pueden introducirse mutaciones adicionales en una posición (posiciones) específica (s), por ejemplo, en una posición (posiciones) no conservativa (s), o en una posición conservativa, en una posición (posiciones) aleatoria (s), o una combinación de tanto en un bucle deseado como región de un ARNt, por ejemplo, un bucle anticodón, el tallo aceptor, brazo o bucle D, bucle variable, brazo o bucle T $\psi$ C, 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 pueden incluir adicionalmente una secuencia adicional (CCA) a un extremo del O-ARNt. Normalmente, un O-ARNt posee una mejora de ortogonalidad para un organismo deseado en comparación con el material de partida, por ejemplo, la pluralidad de secuencias de ARNt, mientras que se preserva su afinidad hacia una RS deseada.

**[0071]** Los procedimientos incluyen analizar opcionalmente la homología de secuencias de ARNt y/o aminoacil-ARNt sintetasas para determinar posibles candidatos para un O-ARNt, O-RS y/o pares de los mismos, que parecen ser ortogonales para un organismo específico. Programas informáticos conocidos en la técnica y descritos en el presente documento pueden usarse para el análisis, por ejemplo, BLAST y pueden usarse programas de compilación. En un ejemplo, para elegir posibles componentes traduccionales ortogonales para su uso en *E. coli*, se elige un organismo procarionta, una sintetasa y/o un ARNt que no muestra homología inusual con organismos procariontas.

**[0072]** Normalmente, un O-ARNt se obtiene sometiendo a, por ejemplo, selección negativa una población de células de una primera especie, en la que las células comprenden un miembro de la pluralidad de posibles O-ARNt. La selección negativa elimina células que comprenden un miembro de la biblioteca de posibles O-ARNt que es aminoacilado por 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.

**[0073]** En ciertas realizaciones, en la selección negativa, un codón (codones) selector (es) se introduce (n) en el polinucleótido que codifica un marcador de selección negativa, por ejemplo, una enzima que confiere resistencia a antibióticos, por ejemplo,  $\beta$ -lactamasa, una enzima que confiere un producto detectable, por ejemplo,  $\beta$ -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. El cribado/selección se hace opcionalmente cultivando la población de células en presencia de un agente selectivo (por ejemplo, un antibiótico, tal como ampicilina). En una realización se varía la concentración del agente de selección.

**[0074]** Por ejemplo, para medir la actividad de supresores de ARNt se usa un sistema de selección que se basa en la supresión *in vivo* de codón selector, por ejemplo, mutaciones antisentido o de desplazamiento de marco introducidas en un polinucleótido que codifica un marcador de selección negativa, por ejemplo, un gen para  $\beta$ -lactamasa (*bla*). Por ejemplo, se construyen variantes de polinucleótido, por ejemplo, variantes de *bla*, con un codón selector en una cierta posición. Las células, por ejemplo, bacterias, se transforman con estos polinucleótidos. En el caso de un ARNt ortogonal, que no puede cargarse eficazmente por sintetasas de *E. coli* endógenas, la resistencia a antibióticos, por ejemplo, resistencia a ampicilina, debe ser aproximadamente o inferior a aquella para la que una bacteria se transformó sin plásmido. Si el ARNt es no ortogonal, o si una sintetasa heteróloga que puede cargar el ARNt se co-expresa en el sistema, se observa un mayor nivel de antibiótico, por ejemplo, resistencia a ampicilina. Se eligen células, por ejemplo, bacterias, que no pueden cultivarse en placas de LB agar con concentraciones de antibiótico aproximadamente iguales a células transformadas sin plásmidos.

**[0075]** En el caso de un producto tóxico (por ejemplo, ribonucleasa o barnasa), cuando un miembro de la pluralidad de ARNt posibles está aminoacilado por huésped endógeno, por ejemplo, sintetasas de *Escherichia coli* (es decir, no es ortogonal al huésped, por ejemplo, sintetasas de *Escherichia coli*), el codón selector es suprimido y el producto de polinucleótido tóxico producido conduce a muerte celular. Sobreviven las células que albergan ARNt ortogonales o ARNt no funcionales.

**[0076]** En una realización, el conjunto de ARNt que son ortogonales a un organismo deseado se someten entonces a una selección positiva en la que un codón selector está situado en un marcador de selección positiva, por ejemplo, codificado por un gen de resistencia a fármaco, como un gen  $\beta$ -lactamasa. La selección positiva se realiza sobre 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 semejante. En ciertas realizaciones, la segunda población de células comprende células que no se eliminaron por la selección negativa. Los polinucleótidos se expresan en la célula y la célula se cultiva en presencia de un agente de selección, por ejemplo, ampicilina. Entonces, los ARNt están seleccionados para su capacidad para ser aminoacilados por la sintetasa semejante coexpresada y para insertar un aminoácido en respuesta a este codón selector. Normalmente, estas células muestran un potenciamiento en la eficiencia de supresión en comparación con células que alojan ARNt no funcional (es), o ARNt que no pueden ser eficientemente reconocidos por la sintetasa de interés. La célula que aloja los ARNt no funcional (es) o ARNt que no son eficientemente reconocidos por la sintetasa de interés son sensibles al antibiótico. Por tanto, los ARNt que: (i) no son sustratos para sintetasas del huésped endógenas, por ejemplo, *Escherichia coli*; (ii) pueden aminoacilarse por la sintetasa de interés; y (iii) son funcionales en la traducción, sobreviven a ambas selecciones.

**[0077]** La rigurosidad de la selección, por ejemplo, la selección positiva, la selección negativa o tanto la selección positiva como la negativa, en los procedimientos anteriormente descritos, 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 puede controlarse introduciendo diferentes números de codones selectores en el gen barnasa y/o usando un promotor inducible. En otro ejemplo, la concentración del agente de selección o de cribado varía (por ejemplo, concentración de ampicilina). En un aspecto de la invención, la rigurosidad varía debido a que la actividad deseada puede ser baja durante rondas tempranas. Por tanto, criterios de selección menos rigurosos se aplican en rondas tempranas y criterios más rigurosos se aplican en posteriores rondas de selección.

En ciertas realizaciones, la selección negativa, la selección positiva o tanto la selección negativa como la positiva pueden repetirse múltiples veces. Pueden usarse múltiples marcadores de selección negativa, marcadores de selección positiva o tanto marcadores de selección negativa como positiva diferentes. En ciertas realizaciones, el marcador de selección positiva y negativa puede ser el mismo.

5 **[0078]** Pueden usarse otros tipos de selecciones/cribado 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 utiliza un aminoácido activo por rédox. Por ejemplo, el marcador de selección negativa, el marcador de selección positiva o tanto los marcadores de selección positiva como negativa pueden incluir un marcador que fluoresce o cataliza una reacción luminiscente en presencia de un reactivo adecuado. En otra realización, un producto del marcador se detecta por citometría de flujo activada por fluorescencia (FACS) o por luminiscencia. Opcionalmente, el marcador incluye un marcador de cribado basado en afinidad. Véase Francisco, J. A. y col., (1993) Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A. 90: 10444-8.

15 **[0079]** Procedimientos adicionales para producir un ARNt ortogonal recombinante pueden encontrarse, por ejemplo, en las solicitudes de patente internacional WO 2002/086075 titulada "Procedimientos y composiciones para la producción de pares de ARNt-aminoacil-ARNt sintetasa ortogonales" y USSN 60/479.931, y 60/496.548 titulada "EXPANSIÓN DEL CÓDIGO GENÉTICO EUCARIOTA". Véase también Forster y col., (2003) Programming peptidomimetics synthetases by translating genetic codes designed de novo PNAS 100 (11): 6353 – 6357; y Feng y col., (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)

25 **[0080]** Una O-RS de la invención es como se define en las reivindicaciones. Una O-RS de la invención puede proporcionarse al sistema de traducción, por ejemplo, una célula, por un polipéptido que incluye una O-RS y/o por un polinucleótido que codifica una O-RS o una parte de la misma. Por ejemplo, una O-RS comprende una secuencia de aminoácidos que se expone en SEQ ID NO: 1 o una variación conservativa de la misma. En otro ejemplo, una O-RS, o una parte de la misma, está codificada por una secuencia de polinucleótidos que codifica un aminoácido que comprende SEQ ID NO: 1, o una secuencia de polinucleótidos complementaria de la misma. Véase, por ejemplo, la Tabla 1 y el Ejemplo 2 en el presente documento para secuencias de moléculas de O-RS a modo de ejemplo. Véase también la sección titulada "Secuencia de ácidos nucleicos y de polipéptidos y variantes" en el presente documento.

35 **[0081]** Los procedimientos para identificar una aminoacil-ARNt sintetasa ortogonal (O-RS), por ejemplo, una O-RS, para su uso con un O-ARNt, también son un rasgo de la invención. Por ejemplo, un procedimiento incluye someter a selección, por ejemplo, selección positiva, una población de células de una primera especie, en la que las células comprenden individualmente: 1) un miembro de una pluralidad de aminoacil-ARNt sintetastas (RS), (por ejemplo, la pluralidad de RS puede incluir RS mutantes, RS derivadas de una especie distinta de la primera especie o tanto RS mutantes como RS derivadas de una especie distinta 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. Las células están seleccionadas o cribadas para aquellas que muestran un potenciamiento en la eficiencia de supresión en comparación con células que carecen o con una cantidad reducida del miembro de la pluralidad de RS. La eficiencia de supresión puede medirse por técnicas conocidas en la materia y como se describen en el presente documento. Las células que tienen un potenciamiento en la eficiencia de supresión comprenden una RS activa que aminoacila el O-ARNt. Un nivel de aminoacilación (*in vitro* o *in vivo*) por la RS activa de un primer conjunto de ARNt de la primera especie se compara con el nivel de aminoacilación (*in vitro* o *in vivo*) por la RS activa de un segundo conjunto de ARNt de la segunda especie. El nivel de aminoacilación puede determinarse por una sustancia detectable (por ejemplo, un aminoácido marcado o aminoácido no natural, por ejemplo, un aminoácido activo por rédox tal como DHP). La RS activa que aminoacila más eficientemente el segundo conjunto de ARNt en comparación con el primer conjunto de ARNt está normalmente seleccionada, proporcionándose así una aminoacil-ARNt sintetasa ortogonal eficiente (optimizada) para su uso con el O-ARNt. Una O-RS, identificada mediante el procedimiento, también es un rasgo de la invención.

55 **[0082]** Pueden usarse distintos ensayos para determinar la aminoacilación. Estos ensayos pueden realizarse *in vitro* o *in vivo*. Por ejemplo, los ensayos de aminoacilación *in vitro* se describen en, por ejemplo, Hoben y Soll (1985) Methods Enzymol. 113: 55 – 59. La aminoacilación también puede determinarse usando un indicador junto con 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. Véanse también el documento WO 2002/085923 titulado "INCORPORACIÓN IN VIVO DE AMINOÁCIDOS NO NATURALES"; y la solicitud internacional número PCT/US2004/011786 presentada el 16 de abril de 2004.

60 **[0083]** La O-RS puede manipularse para alterar la especificidad por sustrato de la sintetasa de manera que sólo un aminoácido no natural deseado, por ejemplo, un aminoácido activo por rédox tal como DHP, pero no ninguno

de los 20 aminoácidos comunes, sean cargados al O-ARNt. Los procedimientos para generar una aminoacil-ARNt sintetasa ortogonal con una especificidad por sustrato por 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 posición (posiciones) no esencial (es) de un marcador positivo permite que las células sobrevivan bajo presión de selección positiva. En presencia de aminoácidos tanto naturales como no naturales, los supervivientes codifican así sintetasas activas que cargan el ARNt supresor ortogonal con tanto un aminoácido natural como no natural. En la selección negativa, la supresión de un codón selector introducido en una posición (posiciones) no esencial (es) de un marcador negativo elimina sintetasas con especificidades por aminoácidos naturales. Los supervivientes de la selección negativa y positiva codifican sintetasas que aminoacilan (cargan) el ARNt supresor ortogonal con aminoácidos no naturales solos. Estas sintetasas pueden luego someterse a mutagénesis adicional, por ejemplo, barajado de ADN u otros procedimientos de mutagénesis recurrentes.

**[0084]** Puede generarse una biblioteca de O-RS mutantes usando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, las RS mutantes pueden generarse por mutaciones específicas para sitio, mutaciones puntuales aleatorias, recombinación de homólogos, barajado de ADN u otros procedimientos de mutagénesis recurrentes, construcción quimérica o cualquier combinación de los mismos. Por ejemplo, puede producirse una biblioteca de RS mutantes a partir de dos o varias de otras "sub-bibliotecas", por ejemplo, más pequeñas, menos diversas. Las bibliotecas quiméricas de RS también están incluidas en la invención. Debe observarse que 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 EE.UU. n° 6.238.884 a Short y col.; la patente de EE.UU. n° 5.756.316 a Schallenberger y col.; la patente de EE.UU. n° 5.783.431 a Petersen y col.; la patente de EE.UU. n° 5.824.485 a Thompson y col.; la patente de EE.UU. n° 5.958.672 a Short y col.) son opcionalmente construidas y se criban para pares ortogonales.

**[0085]** Una vez las sintetasas se han sometido a la estrategia de selección/cribado positivo y negativo, estas sintetasas pueden luego someterse a mutagénesis adicional. Por ejemplo, puede aislarse un ácido nucleico que codifica la O-RS; puede generarse un conjunto de polinucleótidos que codifica O-RS mutadas (por ejemplo, por mutagénesis aleatoria, mutagénesis específica para sitio, recombinación o cualquier combinación de las mismas) a partir del ácido nucleico; y estas etapas individuales o una combinación de estas etapas puede repetirse hasta que se obtenga una O-RS mutada que preferencialmente aminoacila el O-ARNt con el aminoácido no natural, por ejemplo, el aminoácido activo por rédox. En un aspecto de la invención, las etapas se realizan múltiples veces, por ejemplo, al menos dos veces.

**[0086]** También pueden usarse niveles adicionales de rigurosidad de selección/cribado en los procedimientos de la invención para producir O-ARNt, O-RS, o pares de los mismos. La rigurosidad de la selección o cribado puede variarse en una o ambas etapas del procedimiento para producir una O-RS. Esto podría incluir, por ejemplo, variar la cantidad de agente de selección/cribado que se usa, etc. También pueden realizarse rondas adicionales de selecciones positivas y/o negativas. La selección o cribado también puede comprender uno o más de un cambio en la permeabilidad de aminoácidos, un cambio en la eficiencia de traducción, un cambio en la fidelidad de traducción, etc. Normalmente, el uno o más cambios se basan en una mutación en uno o más genes en un organismo en el que un par de ARNt-ARNt sintetasa ortogonal se usa para producir proteína.

**[0087]** Detalles generales adicionales para producir O-RS, y alterar la especificidad por sustrato de la sintetasa, pueden encontrarse en el documento WO 2002/086075 titulado "Procedimientos y composiciones para la producción de pares de ARNt-aminoacil-ARNt sintetasa ortogonales"; y solicitud internacional número PCT/US2004/011786 presentada el 16 de abril de 2004.

#### FUENTE Y ORGANISMOS HUÉSPED

**[0088]** Los componentes traduccionales de la invención pueden derivarse de organismos no eucariotas. Por ejemplo, el O-ARNt ortogonal puede derivarse de un organismo no eucariota (o una combinación de organismos), por ejemplo, una arqueobacteria tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeropyrum 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 puede derivarse de un organismo no eucariota (o una combinación de organismos), por ejemplo, una arqueobacteria tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Pyrobaculum aerophilum*,

*Pyrococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium* o similares, o una eubacteria tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus* o similares. En una realización también pueden usarse fuentes eucariotas, por ejemplo, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos, artrópodos, etc.) o similares como fuentes de O-ARNt y O-RS.

**[0089]** Los componentes individuales de un par de O-ARNt/O-RS pueden derivarse del mismo organismo u organismos diferentes. En una realización, el par de O-ARNt/O-RS es del mismo organismo. Alternativamente, el O-ARNt y la O-RS del par de O-ARNt/O-RS son de organismos diferentes.

**[0090]** El O-ARNt, O-RS o par de O-ARNt/O-RS puede seleccionarse o cribarse *in vivo* o *in vitro* y/o usarse en una célula, por ejemplo, una célula no eucariota, o células eucariotas, para producir un polipéptido con un aminoácido activo por rédox. Una célula no eucariota puede ser de una variedad de fuentes, por ejemplo, una eubacteria tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus* o similares, o una arqueobacteria tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloflexax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeropyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (Mm), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium* o similares. Una célula eucariota puede ser de una variedad de fuentes, por ejemplo, una planta (por ejemplo, planta compleja tal como monocotiledóneas, o dicotiledóneas), un alga, un protista, un hongo, una levadura (por ejemplo, *Saccharomyces cerevisiae*), un animal (por ejemplo, un mamífero, un insecto, un artrópodo, etc.), o similares. Composiciones de células con componentes traducionales de la invención también son un rasgo de la invención.

**[0091]** Véase, por tanto, la solicitud internacional número PCT/US2004/011786 presentada el 16 de abril de 2004 para cribar O-ARNt y/o O-RS en una especie para su uso en otra especie.

#### CODONES SELECTORES

**[0092]** Los codones selectores de la invención expanden la región estructural del cogón genético de la maquinaria biosintética de proteínas. Por ejemplo, un codón selector incluye, por ejemplo, un único codón de tres bases, un codón antisentido, 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. Varios codones selectores pueden introducirse en un gen deseado, por ejemplo, uno o más, dos o más, más de tres, etc. Usando diferentes codones selectores pueden usarse múltiples pares de ARNt/sintetasa ortogonales que permiten la incorporación específica para sitio simultánea de múltiples aminoácidos activos por rédox, por ejemplo, aminoácidos no naturales, usando estos codones selectores diferentes.

**[0093]** En una realización, los procedimientos implican el uso de un codón selector que es un codón de terminación para la incorporación de un aminoácido activo por rédox *in vivo* en una célula. Por ejemplo, se produce un O-ARNt que reconoce el codón de terminación y es aminoacilado por una O-RS con un aminoácido activo por rédox. Este O-ARNt no es reconocido por las aminoacil-ARNt sintetasa del huésped que se producen naturalmente. La mutagénesis dirigida a sitio convencional pueden usarse para introducir el codón de terminación en el sitio de interés en un polinucleótido que codifica un polipéptido de interés. Véase, por ejemplo, Sayers, J.R. y col. (1988), 5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis. Nucleic Acids Res, 791 – 802. Cuando la O-RS, O-ARNt y el ácido nucleico que codifica un polipéptido de interés se combinan, por ejemplo, *in vivo*, el aminoácido activo por rédox se incorpora en respuesta al codón de terminación dando un polipéptido que contiene el aminoácido activo por rédox en la posición especificada. En una realización de la invención, un 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 ambos como codón selector puede codificar 22 aminoácidos a la vez que preserva el codón antisentido ocre, UAA, que es la señal de terminación más abundante.

**[0094]** La incorporación de aminoácidos activos por rédox *in vivo* puede hacerse sin perturbación significativa de la célula huésped. Por ejemplo, en células no eucariotas, tales como *Escherichia coli*, debido a que la eficiencia de supresión para el codón UAG depende de la competencia 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 en cultivo del ribosoma), la eficiencia de supresión puede modularse, por ejemplo, tanto aumentando el nivel de expresión de O-ARNt, por ejemplo, el ARNt supresor, como usando una cepa deficiente en RF1. En células eucariotas, debido a que la eficiencia de supresión para el codón UAG depende de la competencia 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 en cultivo del ribosoma), la eficiencia de supresión puede modularse por, por ejemplo, aumentando el nivel de expresión de O-ARNt, por ejemplo, el ARNt supresor. Además, también pueden estar presentes compuestos adicionales, por ejemplo, agentes reductores tales como ditiotreitól (DTT).

**[0095]** Aminoácidos activos por rédox también puede estar codificados con codones raros. Por ejemplo, si se reduce la concentración de arginina en una reacción de síntesis de proteína *in vitro*, el codón de arginina raro, AGG, ha demostrado ser eficiente para la inserción de Ala por un ARNt sintético acilado con alanina. Véase, por ejemplo, Ma y col., *Biochemistry*, 32: 7939 (1993). En este caso, el ARNt sintético compite con el ARNtArg que se produce naturalmente, que existe como una especie minoritaria en *Escherichia coli*. Además, algunos organismos no usan todos los codones de triplete. Se ha utilizado un codón AGA sin asignar en *Micrococcus luteus* para la inserción de aminoácidos en un extracto de transcripción/traducción *in vitro*. Véase, por ejemplo, Kowal y Oliver, *Nucl. Acid. Res.*, 25: 4685 (1997). Los componentes de la invención pueden generarse para usar estos codones raros *in vivo*.

**[0096]** Los codones selectores también pueden comprender codones extendidos, por ejemplo, cuatro o más codones base tal como, cuatro, cinco, seis o más codones base. Ejemplos de cuatro codones base incluyen, por ejemplo, AGGA, CUAG, UAGA, CCCU y similares. Ejemplos de cinco codones base incluyen, por ejemplo, AGGAC, CCCC, CCCUC, CUAGA, CUACU, UAGGC y similares. Los procedimientos de la invención incluyen usar codones extendidos basándose en la supresión del desplazamiento de marco. Cuatro o más codones base pueden insertar, por ejemplo, uno o múltiples aminoácidos no naturales tales como un aminoácido activo por rédox, en la misma proteína. En otras realizaciones, los bucles anticodón pueden descodificar, 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. Como hay 256 codones de cuatro bases posibles, múltiples aminoácidos no naturales pueden codificarse en la misma célula usando un codón de cuatro o más bases. Véase, por tanto, Anderson y col., (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.

**[0097]** Por ejemplo, se han usado codones de cuatro bases para incorporar aminoácidos no naturales en proteínas usando procedimientos biosintéticos *in vitro*. Véanse, por ejemplo, Ma y col., (1993) *Biochemistry*, 32: 7939; y Hohsaka y col., (1999) *J. Am. Chem. Soc.*, 121: 34. Se usaron CGGG y AGGU para incorporar simultáneamente 2-naftilalanina y un derivado de NBD de lisina en estreptavidina *in vitro* con dos ARNt supresores de desplazamiento de marco químicamente acilados. Véase, por ejemplo, Hohsaka y col., (1999) *J. Am. Chem. Soc.*, 121: 12194. En un estudio *in vivo*, Moore y col. examinaron la capacidad de derivados de ARNt<sup>Leu</sup> con anticodones NCUA para suprimir codones UAGN (N puede ser U, A, G o C) y encontraron que el cuadruplete UAGA podría ser codificado por un ARNt<sup>Leu</sup> con un anticodón UCUA con una eficiencia del 13 al 26 % con poca descodificación en el marco 0 ó -1. Véase Moore y col., (2000) *J. Mol. Biol.*, 298: 195. En una realización, en la invención pueden usarse codones extendidos basados en codones raros o codones antisentido que pueden reducir la ultralectura con cambio de sentido y supresión del desplazamiento de marco en otros sitios no deseados.

**[0098]** Para un sistema dado, un codón selector también puede incluir uno de los codones de tres bases naturales, en los que el sistema endógeno no usa (o raramente usa) el codón base natural. Por ejemplo, esto incluye un 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.

**[0099]** 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 adicionales aumenta el número de codones de triplete de 64 a 125. Las propiedades de los terceros pares de bases incluyen apareamiento de bases estable y selectivo, incorporación enzimática eficiente en ADN con alta fidelidad por una polimerasa y la extensión de cebadores continuada eficiente después de la síntesis del par de bases no naturales nacientes. Las descripciones de pares de bases no naturales que pueden adaptarse a procedimientos y composiciones incluyen, por ejemplo, Hirao y col., (2002) An unnatural base pair for incorporating amino acid analogues into protein, *Nature Biotechnology*, 20: 177 – 182. Véase también Wu, Y. y col., (2002) *J. Am. Chem. Soc.* 124: 14626 – 14630. Otras publicaciones relevantes se enumeran a continuación.

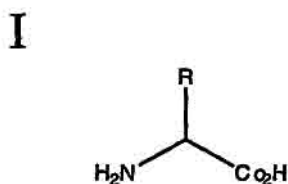
**[0100]** Para uso *in vivo*, el nucleósido no natural es permeable a la membrana y está fosforilado para formar el trifosfato correspondiente. Además, la elevada información genética es estable y no es destruida por enzimas celulares. Esfuerzos previos por Benner y otros aprovecharon la ventaja de los patrones de enlaces de hidrógeno que son diferentes de aquellos en pares de Watson-Crick canónicos, siendo el ejemplo más notorio el par iso-C: iso-G. Véanse, por ejemplo, Switzer y col., (1989) *J. Am. Chem. Soc.*, 111: 8322; y Piccirilli y col., (1990) *Nature*, 343: 33; Kool, (2000) *Curr. Opin. Chem. Biol.*, 4: 602. Estas bases se desaparean en general a cierto grado con bases naturales y no pueden replicarse enzimáticamente. Kool y colaboradores demostraron que las interacciones de encapsidación hidrófobas entre bases pueden sustituir el enlace de hidrógeno para accionar la formación del par 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 naturales que satisfagan todos los requisitos anteriores, Schultz, Romesberg y colaboradores han sintetizado y estudiado sistemáticamente una serie de bases hidrófobas no naturales. Se encuentra que un auto-par de PICS: PICS es más estable que los pares de bases naturales, y puede incorporarse eficientemente en ADN por fragmento de Klenow de ADN polimerasa I de *Escherichia coli* (KF). Véanse, por ejemplo, McMinn y col., (1999) *J. Am. Chem. Soc.*, 121: 11586; y Ogawa y col., (2000) *J. Am. Chem.*

Soc., 122: 3274. Un auto-par de 3MN: 3MN puede sintetizarse por KF con eficiencia y selectividad suficiente para la función biológica. Véase, por ejemplo, Ogawa y col., (2000) J. Am. Chem. Soc., 122: 8803. Sin embargo, ambas bases actúan de terminador de cadena para la posterior replicación. Recientemente se ha desarrollado una ADN polimerasa mutante que puede usarse para replicar el auto-par de PICS. Además, un auto-par de 7AI puede replicarse. Véase, por ejemplo, Tae y col., (2001) J. Am. Chem. Soc., 123: 7439. También se ha desarrollado un par de metalobases novedoso, Dipic: Py, que forma un par estable tras la unión a Cu (II). Véase Meggers y col., (2000) J. Am. Chem. Soc., 122: 10714: Debido a que los codones extendidos y codones no naturales son intrínsecamente ortogonales a codones naturales, los procedimientos de la invención pueden aprovechar esta propiedad para generar ARNt ortogonales para ellos.

**[0101]** También puede usarse un sistema de derivación traduccional para incorporar un aminoácido activo por rédox en un polipéptido deseado. En un sistema de derivación traduccional, una secuencia grande se inserta en un gen, pero no es traducida en proteína. La secuencia contiene una estructura que sirve de entrada para inducir el salto del ribosoma sobre la secuencia y reanuda la traducción en la dirección 3' de la inserción.

### AMINOÁCIDOS NO NATURALES

**[0102]** Como se usa en el presente documento, un aminoácido no natural se refiere a cualquier aminoácido, aminoácido modificado o análogo de aminoácido distinto de selenocisteína y/o pirrolisina y los veinte siguientes alfa-aminoácidos genéticamente codificados: 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 por la fórmula I:



**[0103]** Un aminoácido no natural es normalmente cualquier estructura que tiene la fórmula I en la que el grupo R es cualquier sustituyente distinto de uno usado en los veinte aminoácidos naturales. Véase, por ejemplo, Biochemistry por L. Stryer, 3ª ed. 1988, Freeman and Company, Nueva York, para estructuras de los veinte aminoácidos naturales. Obsérvese que los aminoácidos no naturales de la invención pueden ser compuestos que se producen naturalmente distintos de los veinte alfa-aminoácidos anteriores.

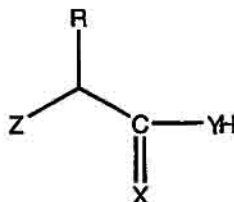
**[0104]** Debido a que los aminoácidos no naturales de la invención normalmente se diferencian de los aminoácidos naturales en la cadena lateral, los aminoácidos no naturales forman enlaces amida con otros aminoácidos, por ejemplo, naturales o no naturales, del mismo modo en el que se forman en proteínas que se producen naturalmente. Sin embargo, los aminoácidos no naturales tienen grupos de cadena lateral que los distinguen de los aminoácidos naturales.

**[0105]** Es de particular interés incorporar aminoácidos no naturales en proteínas que tienen la capacidad de incorporar un aminoácido activo por rédox, por ejemplo, un aminoácido no natural que comprende un resto que permite transferir electrones y/o protones dentro y fuera de la molécula, en proteínas. Por ejemplo, en un aminoácido activo por rédox, R en la fórmula I incluye, pero no se limita a, por ejemplo, ceto, azido, hidroxilo, halo (por ejemplo, yodo), nitro, tiol, seleno, sulfonilo, heterocíclico, aldehído, tioácido y similares, o cualquier combinación de los mismos. El aminoácido activo por rédox de la invención es 3,4-dihidroxi-L-fenilalanina. Véase también la **FIG. 1**.

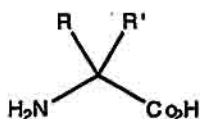
**[0106]** En otros aminoácidos no naturales, por ejemplo, R en la fórmula I comprende opcionalmente un alquilo, arilo, acilo, hidracina, ciano, halo, hidrazida, alquenilo, alquinilo, éter, borato, boronato, fosfo, fosfono, fosfina, enona, imina, éster, hidroxilamina, amina y similares, o cualquier combinación de los mismos. Otros aminoácidos no naturales de interés incluyen, pero no se limitan a, aminoácidos que comprenden un reticulante fotoactivable, aminoácidos marcados con espín, aminoácidos fluorescentes, aminoácidos de unión a metal, aminoácidos que contienen metal, ácidos radioaminoactivos, aminoácidos con grupos funcionales novedosos, aminoácidos que interactúan covalentemente o no covalentemente con otras moléculas, aminoácidos fotoenjaulados y/o fotoisomerizables, aminoácidos que contienen biotina o análogos de biotina, aminoácidos que contienen ceto, aminoácidos glucosilados, un resto de sacárido unido a la cadena lateral de aminoácido, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles o fotoescindibles, aminoácidos con una cadena lateral alargada con respecto a aminoácidos naturales (por ejemplo, poliéteres o hidrocarburos de cadena larga, por ejemplo, más de aproximadamente 5, más de aproximadamente 10 carbonos, etc.), aminoácidos que contienen azúcares unidos a carbono, aminoácidos que contienen aminotioácidos y aminoácidos que contienen uno o más restos tóxicos.

**[0107]** Además de aminoácidos no naturales que contienen cadenas laterales novedosas, los aminoácidos no naturales también comprenden opcionalmente estructuras de esqueleto modificadas, por ejemplo, como se ilustra por las estructuras de fórmula II y III:

II



III



5 en las que Z comprende normalmente OH, NH<sub>2</sub>, SH, NH-R' o S-R'; X e Y, que pueden ser iguales o diferentes, comprenden normalmente S o O, y R y R', que son opcionalmente los mismos o diferentes, normalmente se seleccionan de la misma lista de constituyentes para el grupo R descrito anteriormente para los aminoácidos no naturales que tienen la fórmula I, además de 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,  $\alpha$ -hidroxiácidos,  $\alpha$ -tioácidos,  $\alpha$ -aminotiocarboxilatos, por ejemplo, con cadenas laterales correspondientes a los veinte aminoácidos naturales comunes o cadenas laterales no naturales. Además, las sustituciones en el  $\alpha$ -carbono incluyen opcionalmente aminoácidos disustituídos en L, D o  $\alpha,\alpha$  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, además de análogos de prolina de anillo de 3, 4, 6, 7, 8 y 9 miembros, aminoácidos  $\beta$  y  $\gamma$  tales como  $\beta$ -alanina sustituida y ácido  $\gamma$ -aminobutírico.

**[0108]** Por ejemplo, muchos aminoácidos no naturales se basan en aminoácidos naturales tales como tirosina, glutamina, fenilalanina y similares. Los análogos de tirosina incluyen tirosinas para-sustituídas, tirosinas orto-sustituídas y tirosinas meta-sustituídas, en las que la tirosina sustituida comprende un grupo acetilo, un grupo benzoílo, un grupo amino, una hidracina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo C<sub>6</sub> - C<sub>20</sub> de cadena lineal o ramificado, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro o similares. Además, también se contemplan anillos arilo múltiplemente sustituidos. Análogos de glutamina de la invención incluyen, pero no se limitan a, derivados de  $\alpha$ -hidroxi, derivados  $\gamma$ -sustituídos, derivados cíclicos y derivados de glutamina sustituidos con amida. Ejemplos de análogos de fenilalanina incluyen, pero no se limitan a, fenilalaninas para-sustituídas, fenilalaninas orto-sustituídas y fenilalaninas meta-sustituídas, en las que el sustituyente comprende un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído, un nitro, un grupo tiol o grupo ceto o similares. Ejemplos específicos de aminoácidos no naturales incluyen, pero no se limitan a, 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, una *p*-propargiloxifenilalanina, 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 $\beta$ -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 fosfontirosina, una *p*-yodo-fenilalanina, una *p*-bromofenilalanina, una *p*-amino-L-fenilalanina y una isopropil-L-fenilalanina, y similares. Las estructuras de una variedad de aminoácidos no naturales se proporcionan en, por ejemplo, la **FIG. 1** en el presente documento y las FIG. 16, 17, 18, 19, 26 y 29 del documento WO 2002/085923 titulado "Incorporación *in vivo* de aminoácidos no naturales".

#### Síntesis química de aminoácidos no naturales

**[0109]** Muchos de los aminoácidos no naturales proporcionados anteriormente están comercialmente disponibles, por ejemplo, de Sigma (EE.UU.) o Aldrich (Milwaukee, WI, EE.UU.). Aquellos que no están



comercialmente disponibles se sintetizan opcionalmente como se proporciona en diversas publicaciones o usando procedimientos convencionales conocidos para aquellos expertos en la materia. Para técnicas de síntesis orgánica véase, por ejemplo, Organic Chemistry por Fessenden y Fessenden (1982, segunda edición, Willard Grant Press, Boston Mass); Advanced Organic Chemistry por March (tercera edición, 1985, Wiley and Sons, Nueva York); y Advanced Organic Chemistry por Carey y Sundberg (tercera edición, Partes A y B, 1990, Plenum Press, Nueva York). Otras publicaciones que describen la síntesis de aminoácidos no naturales incluyen, por ejemplo, el documento WO 2002/085923 titulado "Incorporación in vivo de aminoácidos no naturales"; Matsoukas y col., (1995) J. Med. Chem., 38, 4660 – 4669; King, F.E. & Kidd, D.A.A. (1949) A New Synthesis of Glutamine and of  $\gamma$ -Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315 – 3319; Friedman, O.M. & Chatterji, R. (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750 – 3752; Craig, J.C. y col. (1988) Absolute Configuration of the Enantiomers of 7-Chloro-4[[4- (diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167 – 1170; Azoulay, M., Vilmont, M. & Frappier, F. (1991) Glutamine analogues as Potential Antimalarials, Eur. J. Med. Chem. 26, 201 – 5; Kapielen, A.M.P. & Rapoport, H. (1989) Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859 – 1866; Christie, B.D. & Rapoport, H. (1985) Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of ( + )-Apovincamine through Amino Acid Decarboxylation and Iminium Ion Cyclization. J. Org. Chem. 1989: 1859 – 1866; Barton y col., (1987) Synthesis of Novel  $\alpha$ -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- $\alpha$ -Amino-Adipic Acids, L- $\alpha$ -aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron Lett. 43: 4297 – 4308; y Subasinghe y col., (1992) Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35: 4602 – 7. Véase también la solicitud internacional número PCT/US03/41346 titulada "Matrices de proteínas" presentada el 22 de diciembre de 2003.

#### Captación celular de aminoácidos no naturales

**[0110]** La captación de aminoácidos no naturales por una célula es un asunto que es normalmente considerado cuando se diseñan y seleccionan aminoácidos no naturales, por ejemplo, para la incorporación en una proteína. Por ejemplo, la alta densidad de carga de  $\alpha$ -aminoácidos sugiere que es poco probable que estos compuestos sean permeables a células. Los aminoácidos naturales se recogen en la célula mediante un conjunto de sistemas de transporte basados en proteínas que frecuentemente muestran grados variables de especificidad de aminoácidos. Puede hacerse un cribado rápido que evalúa qué aminoácidos no naturales, si los hay, son recogidos por células. Véanse, por ejemplo, los ensayos de toxicidad en, por ejemplo, la solicitud internacional número PCT/US03/41346 titulada "Matrices de proteínas" 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 es fácilmente analizada con diversos ensayos, una alternativa a diseñar aminoácidos no naturales que son aceptados por rutas de captación celular es proporcionar rutas biosintéticas para crear aminoácidos *in vivo*.

#### Biosíntesis de aminoácidos no naturales

**[0111]** Muchas rutas biosintéticas ya existen en células para la producción de aminoácidos y otros compuestos. Aunque un procedimiento biosintético para un aminoácido no natural particular puede no existir en la naturaleza, por ejemplo, en una célula, la invención proporciona tales procedimientos. Por ejemplo, las rutas biosintéticas para aminoácidos no naturales se generan opcionalmente en célula huésped añadiendo nuevas enzimas o modificando rutas de células huésped existentes. Nuevas enzimas adicionales son enzimas que se producen opcionalmente naturalmente o enzimas artificialmente desarrolladas. Por ejemplo, la biosíntesis de *p*-aminofenilalanina (como se presenta en un ejemplo en el documento WO 2002/085923, arriba) se basa en la adición de una combinación de enzimas conocidas de otros organismos. Los genes para estas enzimas pueden introducirse 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. Ejemplos de los tipos de enzimas que se añaden opcionalmente se proporcionan en los ejemplos más adelante. Secuencias de enzimas adicionales se encuentran, por ejemplo, en Genbank. Enzimas artificialmente desarrolladas también se añaden opcionalmente a una célula del mismo modo. De este modo, la maquinaria celular y los recursos de una célula se manipulan para producir aminoácidos no naturales.

**[0112]** De hecho, puede usarse cualquiera de una variedad de procedimientos para producir enzimas novedosas para su uso en rutas biosintéticas, o para la evolución de rutas existentes, para la producción de aminoácidos no naturales, *in vitro* o *in vivo*. Muchos procedimientos disponibles de enzimas en desarrollo y otros componentes de la ruta biosintética pueden aplicarse a la presente invención para producir aminoácidos no naturales (o, de hecho, para desarrollar sintetetasas que tienen nuevas especificidades por sustrato u otras actividades de interés). Por ejemplo, el barajado de ADN se usa opcionalmente para desarrollar enzimas novedosas y/o rutas de tales enzimas para la producción de aminoácidos no naturales (o producción de nuevas sintetetasas), *in vitro* o *in vivo*. 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. Un enfoque relacionado baraja familias de

genes relacionados (por ejemplo, homólogos) para desarrollar rápidamente enzimas con características deseadas. Un ejemplo de tales procedimientos de "barajado de familias de genes" se encuentra en Cramer y col. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature*, 391 (6664): 288 – 291. También pueden generarse nuevas enzimas (tanto componentes de rutas biosintéticas como sintetetas) usando un procedimiento de recombinación de ADN conocido como "truncación incremental para la creación de enzimas híbridas" ("ITCHY"), por ejemplo, como se describe en Ostermeier y col. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" *Nature Biotech* 17: 1205. Este enfoque también puede usarse para generar una biblioteca de enzima u otras variantes de ruta que puedan servir de sustratos para uno o más procedimientos de recombinación *in vitro* o *in vivo*. Véase, por tanto, Ostermeier y col. (1999) "Combinatorial Protein Engineering by Incremental Truncation", *Proc. Natl. Acad. Sci. USA*, 96: 3562 – 67, y Ostermeier y col. (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts", *Biological and Medicinal Chemistry*, 7: 2139 – 44. Otro enfoque usa mutagénesis de conjunto exponencial para producir bibliotecas de enzima u otras variantes de ruta que se seleccionan, por ejemplo, para una capacidad para catalizar una reacción biosintética relevante para producir un aminoácido no natural (o una nueva sinteteta). En este enfoque, grupos pequeños de residuos en una secuencia de interés se aleatorizan en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales. Ejemplos de tales procedimientos, que pueden adaptarse a la presente invención para producir nuevas enzimas para la producción de aminoácidos no naturales (o nuevas sintetetas), se encuentran en Delegrave & Youvan (1993) *Biotechnology Research* 11: 1548 – 1552. En otro enfoque más puede usarse la mutagénesis aleatoria o semialeatoria usando oligonucleótidos dopados o degenerados para la manipulación de componentes de enzima y/o 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: 297 – 300; o Reidhaar-Olson y col. (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" *Methods Enzymol.* 208: 564 – 86. Puede usarse otro enfoque más, frecuentemente llamado una mutagénesis "no estocástica", que usa reensamblaje de polinucleótidos y mutagénesis de saturación de sitio, para producir enzimas y/o componentes de ruta, que luego pueden cribarse para una capacidad para realizar una o más sintetetas o función de ruta biosintética (por ejemplo, para la producción de aminoácidos no naturales *in vivo*). Véase, por ejemplo, Short "Non-Stochastic Generation of Genetic Vaccines and Enzymes", documento WO 00/46344.

**[0113]** Una alternativa a tales procedimientos mutaciones implica recombinar genomas enteros de organismos y seleccionar progenie resultante para funciones de ruta particulares (frecuentemente denominado "barajado del genoma completo"). Este enfoque puede aplicarse a la presente invención, por ejemplo, por recombinación genómica y selección de un organismo (por ejemplo, una célula de *E. coli* u otra) para una capacidad para producir un aminoácido no natural (o producto intermedio del mismo). Por ejemplo, procedimientos enseñados en las siguientes publicaciones pueden aplicarse al diseño de rutas para la evolución de rutas existentes y/o nuevas en células para producir aminoácidos no naturales *in vivo*: Patnaik y col. (2002) "Genome shuffling of lactobacillus for improved acid tolerance" *Nature Biotechnology*, 20 (7): 707 – 712; y Zhang y col. (2002) "Genome shuffling leads to rapid phenotypic improvement in bacteria" *Nature*, 7 de febrero, 415 (6872): 644 – 646.

**[0114]** Otras técnicas para la manipulación de rutas de organismos y metabólicas, por ejemplo, para la producción de los compuestos deseados, también están disponibles y también pueden aplicarse a la producción de aminoácidos no naturales. Ejemplos de publicaciones que enseñan enfoques de manipulación de rutas útiles incluyen: Nakamura y White (2003) "Metabolic engineering for the microbial production of 1,3 propanediol" *Curr. Opin. Biotechnol.* 14 (5): 454 – 9; Berry y col. (2002) "Application of Metabolic Engineering to improve both the production and use of Biotech Indigo" *J. Industrial Microbiology and Biotechnology* 28: 127 – 133; Banta y col. (2002) "Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of Corynebacterium 2,5-diketo-D-gluconic acid reductasa for use in vitamin C biosynthesis" *Biochemistry*, 41 (20), 6226 – 36; Selivonova y col. (2001) "Rapid Evolution of Novel Traits in Microorganisms" *Applied and Environmental Microbiology*, 67: 3645, y muchos otros.

**[0115]** Independientemente del procedimiento usado, normalmente, el aminoácido no natural producido con una ruta biosintética manipulada de la invención se produce en una concentración suficiente para la biosíntesis eficiente de proteínas, por ejemplo, una cantidad celular natural, pero no a tal grado que afecte significativamente la concentración de otros aminoácidos celulares o para agotar recursos celulares. Concentraciones típicas producidas *in vivo* de este modo son aproximadamente 10 mM a aproximadamente 0,05 mM. Una vez se ha manipulado una célula para producir enzimas deseadas para una ruta específica y se genera un aminoácido no natural, las selecciones *in vivo* se usan opcionalmente para optimizar adicionalmente la producción del aminoácido no natural para tanto la síntesis de proteínas ribosómicas como el crecimiento celular.

#### 60 Componentes ortogonales para incorporar 3,4-dihidroxi-L-fenilalanina (DHP)

**[0116]** La invención proporciona composiciones y procedimientos de producción de componentes ortogonales para incorporar un aminoácido activo por rédox, 3,4-dihidroxi-L-fenilalanina (DHP), en una cadena de polipéptidos en cultivo en respuesta a un codón selector, por ejemplo, codón de terminación, un codón antisentido, un codón de

cuatro o más bases, etc., por ejemplo, *in vivo*. Por ejemplo, la invención proporciona ARNt ortogonales (O-ARNt), aminoacil-ARNt sintetasa ortogonales (O-RS) y pares de los mismos. Estos pares pueden usarse para incorporar DHP en cadenas de polipéptidos en cultivo y son como se definen en las reivindicaciones.

5 **[0117]** Una composición de la invención es como se define en la reivindicación 1.

10 **[0118]** Una composición que incluye una O-RS puede incluir opcionalmente adicionalmente un ARNt ortogonal (O-ARNt) en el 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 eficiencia de supresión en presencia de una sintetasa semejante en respuesta a un codón selector con respecto al O-ARNt que comprende o es codificado por una secuencia de polinucleótidos que se expone en los listados de secuencias y ejemplos en el presente documento. En una realización, la eficiencia de supresión de la O-RS y el O-ARNt juntos es, por ejemplo, 5 veces, 10 veces, 15 veces, 20 veces, 25 veces o más superior que la eficiencia de supresión del O-ARNt que carece de O-RS. En un aspecto, la eficiencia de supresión de la O-RS y el O-ARNt juntos es al menos el 45 % de la eficiencia de supresión de un par de tirosil-ARNt sintetasa ortogonal derivado de *Methanococcus jannaschii*.

15 **[0119]** Una composición que incluye un O-ARNt puede incluir opcionalmente una célula (por ejemplo, una célula no eucariota tal como una célula de *E. coli* y similares, o una célula eucariota), y/o un sistema de traducción.

20 **[0120]** Una célula (por ejemplo, una célula no eucariota, o una célula eucariota) que comprende un sistema de traducción también se proporciona por la invención, en la que el sistema de traducción incluye como se define en las reivindicaciones un ARNt ortogonal (O-ARNt); una aminoacil-ARNt sintetasa ortogonal (O-RS); y un aminoácido activo por rédox, 3,4-dihidroxi-L-fenilalanina (DHP). La O-RS aminoacila preferencialmente el O-ARNt con una eficiencia de al menos el 50 % de la eficiencia de un polipéptido que comprende una secuencia de aminoácidos de SEQ ID NO: 1. El O-ARNt reconoce el primer codón selector, y la O-RS aminoacila preferencialmente el O-ARNt con la 3,4-dihidroxi-L-fenilalanina (DHP). En una realización, el O-ARNt comprende o está codificado por una secuencia de polinucleótidos que se expone en SEQ ID NO: 2, o una secuencia de polinucleótidos complementaria de la misma. En una realización, la O-RS comprende una secuencia de aminoácidos que se expone en una cualquiera de SEQ ID NO: 1.

25 **[0121]** Una célula de la invención puede comprender opcionalmente además un par de O-ARNt/O-RS diferente adicional y un segundo aminoácido no natural, por ejemplo, en el que O-ARNt reconoce un segundo codón selector y esta O-RS aminoacila preferencialmente el O-ARNt con el segundo aminoácido no natural. Opcionalmente, una célula de la invención incluye 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.

30 **[0122]** En ciertas realizaciones, una célula de la invención incluye una célula de *E. coli* que incluye un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), un aminoácido activo por rédox y un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en la que el polinucleótido comprende el codón selector que es reconocido por el O-ARNt. En ciertas realizaciones de la invención, la O-RS aminoacila preferencialmente el O-ARNt con una eficiencia de al menos el 50 % de la eficiencia de un polipéptido que comprende una secuencia de aminoácidos de cualquier secuencia de O-RS enumerada en el presente documento.

35 **[0123]** En ciertas realizaciones de la invención, un O-ARNt de la invención comprende o está codificado por una secuencia de polinucleótidos que se expone en los listados de secuencias o ejemplos en el presente documento, o una secuencia de polinucleótidos complementaria del mismo. En ciertas realizaciones de la invención, una O-RS comprende una secuencia de aminoácidos que se expone en los listados de secuencias, o una variación conservativa de la misma. En una realización, la O-RS o una parte de la misma está codificada por una secuencia de polinucleótidos que codifica un aminoácido que se expone en los listados de secuencias o ejemplos en el presente documento, o una secuencia de polinucleótidos complementaria de la misma.

40 **[0124]** El O-ARNt y/o la O-RS de la invención pueden derivarse de cualquiera de una variedad de organismos (por ejemplo, organismos eucariotas y/o no eucariotas).

45 **[0125]** Los polinucleótidos también son un rasgo de la invención. Un polinucleótido de la invención incluye un polinucleótido artificial (por ejemplo, preparado por el hombre y que no se produce naturalmente) que comprende una secuencia de nucleótidos que codifica un polipéptido que se expone en los listados de secuencias en el presente documento y/o es complementaria a o esa secuencia de polinucleótidos. Un polinucleótido de la invención también puede incluir un ácido nucleico que se hibrida con un polinucleótido descrito anteriormente, bajo condiciones altamente rigurosas, sobre sustancialmente la longitud entera del ácido nucleico. Un polinucleótido de la invención también incluye un polinucleótido que es, por ejemplo, al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o más idéntico al de un ARNt que se produce naturalmente o ácido nucleico

codificante correspondiente (pero un polinucleótido de la invención es distinto de un ARNt que se produce naturalmente o ácido nucleico codificante correspondiente), en el que el ARNt reconoce un codón selector, por ejemplo, un codón de cuatro bases. Polinucleótidos artificiales que son, por ejemplo, al menos el 80 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o más idénticos a cualquiera de los anteriores y/o un polinucleótido que comprende una variación conservativa de cualquiera de los anteriores también están incluidos en los polinucleótidos de la invención.

**[0126]** Vectores que comprenden un polinucleótido de la invención también son un rasgo de la invención. Por ejemplo, un vector de la invención puede incluir un plásmido, un cósmido, un fago, un virus, un vector de expresión y/o similares. Una célula que comprende un vector de la invención también es un rasgo de la invención.

**[0127]** Procedimientos de producción de componentes de un par de O-ARNt/O-RS también son rasgos de la invención. Componentes producidos por estos procedimientos también son un rasgo de la invención. Por ejemplo, procedimientos de producción de al menos un ARNt que son ortogonales a una célula (O-ARNt) incluyen generar una biblioteca de ARNt mutantes; mutar un bucle anticodón de cada miembro de la biblioteca de ARNt mutantes para permitir el reconocimiento de un codón selector, proporcionándose así una biblioteca de posibles O-ARNt, y someter a selección negativa una primera población de células de una primera especie, en la que las células comprenden un miembro de la biblioteca de posibles O-ARNt. La selección negativa elimina células que comprenden un miembro de la biblioteca de posibles O-ARNt que está aminoacilado por 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, proporcionándose así al menos un O-ARNt. También se proporciona un O-ARNt producido mediante los procedimientos de la invención.

**[0128]** En ciertas realizaciones, los procedimientos comprenden además someter a selección positiva una segunda población de células de la primera especie, en la que las células comprenden un miembro del conjunto de ARNt que son ortogonales a la célula de la primera especie, una aminoacil-ARNt sintetasa semejante y un marcador de selección positiva. Usando la selección positiva, las células se seleccionan o se criban para aquellas células que comprenden un miembro del conjunto de ARNt que está aminoacilado por la aminoacil-ARNt sintetasa semejante y que muestra una respuesta deseada en presencia del marcador de selección positiva, proporcionándose así un O-ARNt. En ciertas realizaciones, la segunda población de células comprende células que no se eliminaron por la selección negativa.

**[0129]** También se proporcionan procedimientos para identificar una aminoacil-ARNt sintetasa ortogonal que carga un O-ARNt con un aminoácido activo por rédox. Por ejemplo, los procedimientos incluyen someter una población de células de una primera especie a una selección, en la que las células comprenden cada una: 1) un miembro de una pluralidad de aminoacil-ARNt sintetasa (RS) (por ejemplo, la pluralidad de RS puede incluir RS mutantes, RS derivadas de una especie distinta de una primera especie o tanto RS mutantes como RS derivadas de una especie distinta de una 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 positiva y comprende al menos un codón selector.

**[0130]** Las células (por ejemplo, una célula huésped) son seleccionadas o cribadas para aquellas que muestran un potenciamiento en la eficiencia de supresión en comparación con células que carecen o que tienen una cantidad reducida del miembro de la pluralidad de RS. Estas células seleccionadas/cribadas comprenden una RS activa que aminoacila el O-ARNt. Una aminoacil-ARNt sintetasa ortogonal identificada mediante el procedimiento también es un rasgo de la invención.

**[0131]** Los procedimientos de producción de una proteína en una célula (por ejemplo, una célula no eucariota tal como una célula de *E. coli* o similares, o una célula eucariota) con una 3,4-dihidroxi-L-fenilalanina (DHP) en una posición especificada también son un rasgo de la invención. Por ejemplo, un procedimiento incluye cultivar, en un medio apropiado, una célula, en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codificar una proteína, proporcionar la DHP e incorporar la DHP en la posición especificada en la proteína durante la traducción del ácido nucleico con el al menos un codón selector, produciéndose así la proteína. 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 preferencialmente el O-ARNt con la DHP.

## SECUENCIA DE ÁCIDOS NUCLEICOS Y DE POLIPÉPTIDOS Y VARIANTES

**[0132]** Como se ha descrito anteriormente y más adelante, la invención proporciona secuencias de polinucleótidos de ácidos nucleicos que codifican, por ejemplo, O-ARNt y O-RS, y secuencias de aminoácidos de polipéptidos, por ejemplo, O-RS, y, por ejemplo, composiciones, sistemas y procedimientos que comprenden dichas secuencias. Ejemplos de dichas secuencias, por ejemplo, secuencias de aminoácidos y de nucleótidos de O-ARNt y O-RS se desvelan en el presente documento (véase la Tabla 1, por ejemplo, SEQ ID NOS: 1 a 3). Sin embargo, un experto en la materia apreciará que la invención no se limita a aquellas secuencias desveladas en el presente documento, por ejemplo, como en los ejemplos y listados de secuencias. Un experto apreciará que la invención

también proporciona, por ejemplo, muchas secuencias y secuencias sin relacionar con las funciones descritas en el presente documento, por ejemplo, que codifican un O-ARNt o una O-RS.

**[0133]** La invención proporciona polipéptidos (O-RS) y polinucleótidos, por ejemplo, O-ARNt, polinucleótidos que codifican O-RS o porciones de los mismos, oligonucleótidos usados para aislar clones de aminoacil-ARNt sintetasa, etc. Los polinucleótidos de la invención incluyen aquellos que codifican proteínas o polipéptidos de interés de la invención con uno o más codones selectores. Además, los polinucleótidos de la invención incluyen, por ejemplo, un polinucleótido que comprende una secuencia de nucleótidos que se expone en SEQ ID NO: 2; un polinucleótido que es complementario a o que codifica una secuencia de polinucleótidos de la misma. Un polinucleótido de la invención también incluye un polinucleótido que codifica una secuencia de aminoácidos que comprende SEQ ID NO: 1. Un polinucleótido de la invención también incluye un polinucleótido que codifica un polipéptido de la invención. Similarmente, un ácido nucleico artificial que se hibrida con un polinucleótido indicado anteriormente bajo condiciones altamente rigurosas sobre sustancialmente la longitud entera del ácido nucleico (y es distinto de un polinucleótido natural) es un polinucleótido de la invención. En una realización, una composición incluye un polipéptido de la invención y un excipiente (por ejemplo, tampón, agua, excipiente farmacéuticamente aceptable, etc.). La invención también proporciona un anticuerpo o antisuero específicamente inmunorreactivo con un polipéptido de la invención. Un polinucleótido artificial es un polinucleótido que está hecho por el hombre y que no se produce naturalmente.

**[0134]** Un polinucleótido de la invención también incluye un polinucleótido artificial que es, por ejemplo, al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o más idéntico al de un ARNt que se produce naturalmente (pero es distinto de un ARNt que se produce naturalmente). Un polinucleótido también incluye un polinucleótido artificial que es, por ejemplo, al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o más idéntico al de un ARNt que se produce naturalmente.

**[0135]** En ciertas realizaciones, un vector (por ejemplo, un plásmido, un cósmido, un fago, un virus, etc.) comprende un polinucleótido de la invención. En una realización, el vector es un vector de expresión. En otra realización, el vector de expresión incluye un promotor operativamente ligado a uno o más de los polinucleótidos de la invención. En otra realización, una célula comprende un vector que incluye un polinucleótido de la invención.

**[0136]** Un experto también apreciará que muchas variantes de las secuencias desveladas están incluidas en la invención. Por ejemplo, variaciones conservativas de las secuencias desveladas que dan una secuencia funcionalmente idéntica están incluidas en la invención. Se considera que variantes de las secuencias de polinucleótidos de ácidos nucleicos, en las que las variantes se hibridan con al menos una secuencia desvelada, se incluyen en la invención. También están incluidas en la invención subsecuencias únicas de las secuencias desveladas en el presente documento, como se ha determinado por, por ejemplo, técnicas de comparación de secuencias convencionales.

#### Variaciones conservativas

**[0137]** Debido a la degeneración del código genético, "sustituciones silenciosas" (es decir, sustituciones en una secuencia de ácidos nucleicos que no producen una alteración en un polipéptido codificado) son un rasgo implícito en *cada* secuencia de ácidos nucleicos que codifica un aminoácido. Similarmente, "sustituciones de aminoácidos conservativas", en uno o algunos aminoácidos en una secuencia de aminoácidos están sustituidos con diferentes aminoácidos con propiedades altamente similares, también se identifican fácilmente como que son altamente similares a una construcción desvelada. Tales variaciones conservativas de cada secuencia desvelada son un rasgo de la presente invención.

**[0138]** "Variaciones conservativas" de una secuencia de ácidos nucleicos particular se refiere a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o, si el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas. Un experto reconocerá que sustituciones, deleciones o adiciones individuales que alteran, añaden o delecionan un único 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 conservativamente modificadas" en las que las alteraciones producen la deleción de un aminoácido, adición de un aminoácido o sustitución de un aminoácido con un aminoácido químicamente similar. Por tanto, "variaciones conservativas" de una secuencia de polipéptidos 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 de polipéptidos, con un aminoácido conservativamente activo por rédox del mismo grupo de sustitución conservativo. 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 conservativa del ácido nucleico básico.

**[0139]** Las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares son muy conocidas en la técnica, en las que un residuo de aminoácido está sustituido con otro residuo de aminoácido

que tiene propiedades químicas similares (por ejemplo, cadenas laterales aromáticas o cadenas laterales positivamente cargadas) y, por tanto, no cambia sustancialmente las propiedades funcionales de la molécula de polipéptido. Lo siguiente expone grupos de ejemplo que contienen aminoácidos naturales de propiedades químicas similares en las que las sustituciones dentro de un grupo es una “sustitución conservativa”.

5

Cadenas laterales no polares y/o alifáticas	Cadenas laterales no cargadas polares	Cadenas laterales aromáticas	Cadenas laterales positivamente cargadas	Cadenas laterales negativamente cargadas
Glicina Alanina Valina Leucina Isoleucina Prolina	Serina Treonina Cisteína Metionina Asparagina Glutamina	Fenilalanina Tirosina Triptófano	Lisina Arginina Histidina	Aspartato Glutamato

#### Hibridación de ácidos nucleicos

10 **[0140]** Puede usarse hibridación comparativa para identificar ácidos nucleicos de la invención, tales como SEQ ID NO: 2, que incluyen variaciones conservativas de ácidos nucleicos de la invención, y este procedimiento de hibridación comparativa es un procedimiento preferido de distinguir ácidos nucleicos de la invención. Además, los ácidos nucleicos diana que se hibridan con un ácido nucleico representado por SEQ ID NO: 2 bajo condiciones de rigurosidad alta, ultra-alta y ultra-ultra-alta son un rasgo de la invención. Ejemplos de tales ácidos nucleicos incluyen aquellos con una o algunas sustituciones de ácidos nucleicos silenciosas o conservativas con respecto a una secuencia de ácidos nucleicos dada.

15

20 **[0141]** Un ácido nucleico de prueba se dice que se hibrida específicamente con un ácido nucleico de sonda cuando se hibrida tanto al menos  $\frac{1}{2}$  con la sonda como con diana complementaria perfectamente apareada, es decir, con una relación de señal con respecto a ruido de al menos  $\frac{1}{2}$  tan alta como la hibridación de la sonda con la diana en condiciones en las que la sonda perfectamente apareada se une a la diana complementaria perfectamente apareada con una relación de señal con respecto a ruido que es al menos aproximadamente 5x-10x tan alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana sin aparear.

20

25 **[0142]** Los ácidos nucleicos “se hibridan” cuando se asocian, normalmente en disolución. Los ácidos nucleicos se hibridan debido a una variedad de fuerzas fisicoquímicas bien caracterizadas tales como enlace de hidrógeno, exclusión de disolventes, apilamiento de bases y similares. Una amplia guía a la hibridación de ácidos nucleicos se encuentra en Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology - Hybridization with Nucleic Acid, parte I capítulo 2, “Overview of principles of hybridization and the strategy of nucleic acid probe assays” (Elsevier, Nueva York), además de en Ausubel, arriba. Hames y Higgins (1995) Gene Probes 1 IRL Press at Oxford University Press, Oxford, Inglaterra (Hames y Higgins 1) y Hames y Higgins (1995) Gene Probes 2 IRL Press at Oxford University Press, Oxford, Inglaterra (Hames y Higgins 2) proporcionan detalles sobre la síntesis, marcado, detección y cuantificación de ADN y ARN, que incluyen oligonucleótidos.

25

30

35 **[0143]** Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios sobre un filtro en una transferencia Southern o Northern es 50 % de formalina con 1 mg de heparina a 42 °C, llevándose a cabo la hibridación durante la noche. Un ejemplo de condiciones de lavado rigurosas es un lavado con 0,2x SSC a 65 °C durante 15 minutos (véase, Sambrook, arriba, para una descripción del tampón SSC). Frecuentemente, el lavado de alta rigurosidad va precedido de un lavado de baja rigurosidad para eliminar la señal de la sonda de referencia. Un ejemplo de lavado de baja rigurosidad es 2x SSC a 40 °C durante 15 minutos. En general, una relación de señal con respecto a ruido de 5x (o superior) a la observada para una sonda no relacionada en el ensayo de hibridación particular indica detección de una hibridación específica.

40

45 **[0144]** “Condiciones de hibridación y lavado rigurosas” en el contexto de los experimentos de hibridación de ácidos nucleicos tales como hibridaciones Southern y Northern dependen de la secuencia, y son diferentes bajo diferentes parámetros medioambientales. Una amplia orientación sobre la hibridación de ácidos nucleicos se encuentra en Tijssen (1993), arriba, y en Hames y Higgins, 1 y 2. Condiciones de hibridación y lavado rigurosas pueden determinarse fácilmente empíricamente para cualquier ácido nucleico de prueba. Por ejemplo, en la determinación de las condiciones de hibridación y lavado rigurosas, las condiciones de hibridación y lavado se aumentan gradualmente (por ejemplo, aumentando la temperatura, disminuyendo la concentración de sales, aumentando la concentración de detergente y/o aumentando la concentración de disolventes orgánicos tales como formalina en la hibridación o lavado) hasta que se cumple un conjunto seleccionado de criterios. Por ejemplo, en condiciones de hibridación y lavado altamente rigurosas, las condiciones de hibridación y lavado se aumentan gradualmente hasta que una sonda se une a una diana complementaria perfectamente apareada con una relación de señal con respecto a ruido que es al menos 5x tal alta como la observada para la hibridación de la sonda con una diana sin aparear.

50

55

**[0145]** Condiciones “muy rigurosas” se seleccionan para ser iguales al punto de fusión térmico ( $T_m$ ) para una sonda particular. La  $T_m$  es la temperatura (bajo fuerza iónica y pH definidos) a la que el 50 % de la secuencia de prueba se hibrida con una sonda perfectamente apareada. Para los fines de la presente invención, generalmente, condiciones de hibridación y lavado “altamente rigurosas” se seleccionan para ser aproximadamente 5 °C inferior a la  $T_m$  para la secuencia específica a una fuerza iónica y pH definidos.

**[0146]** Condiciones de hibridación y lavado de “rigurosidad ultra-alta” son aquellas en las que la rigurosidad de las condiciones de hibridación y lavado es aumentada hasta que la relación de señal con respecto a ruido para la unión de la sonda con el ácido nucleico diana complementario perfectamente apareado es al menos 10x tal alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana sin aparear. Un ácido nucleico diana que se hibrida con una sonda bajo tales condiciones, con una relación de señal con respecto a ruido de al menos  $\frac{1}{2}$  de la del ácido nucleico diana complementario perfectamente apareado, se dice que se une a la sonda bajo condiciones de rigurosidad ultra-alta.

**[0147]** Similarmente, pueden determinarse niveles de rigurosidad incluso mayores aumentando gradualmente las condiciones de hibridación y/o lavado del ensayo de hibridación relevante. Por ejemplo, aquellos en los que la rigurosidad de las condiciones de hibridación y lavado es aumentada hasta que la relación de señal con respecto a ruido para la unión de la sonda con el ácido nucleico diana complementario perfectamente apareado es al menos 10x, 20X, 50X, 100X o 500X o más tan alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana sin aparear. Un ácido nucleico diana que se hibrida con una sonda bajo tales condiciones, con una relación de señal con respecto a ruido de al menos  $\frac{1}{2}$  de la del ácido nucleico diana complementario perfectamente apareado, se dice que se une a la sonda bajo condiciones de rigurosidad ultra-ultra-alta.

**[0148]** Los ácidos nucleicos que no se hibridan entre sí bajo condiciones rigurosas todavía son sustancialmente idénticos si los polipéptidos que codifican son sustancialmente idénticos. Esto se produce, por ejemplo, cuando una copia de un ácido nucleico es creada usando la máxima degeneración de codones permitida por el código genético.

#### Subsecuencias únicas

**[0149]** En un aspecto, la invención proporciona un ácido nucleico que comprende una subsecuencia única en un ácido nucleico seleccionado de las secuencias de O-ARNt y O-RS desveladas en el presente documento. La subsecuencia única es única con respecto a un ácido nucleico correspondiente a cualquier secuencia de ácidos nucleicos de O-ARNt o O-RS conocida. El alineamiento puede realizarse usando, por ejemplo, el conjunto BLAST para los parámetros por defecto. Cualquier subsecuencia única es útil, por ejemplo, como sonda para identificar los ácidos nucleicos de la invención.

**[0150]** Similarmente, la invención incluye un polipéptido que comprende una única subsecuencia en un polipéptido seleccionada de las secuencias de O-RS desveladas en el presente documento. Aquí, la subsecuencia única es única con respecto a un polipéptido correspondiente a cualquiera de la secuencia de polipéptidos conocida.

**[0151]** La invención también proporciona ácidos nucleicos diana que se hibridan bajo condiciones rigurosas con un único oligonucleótido codificante que codifica una subsecuencia única en un polipéptido seleccionado de las secuencias de O-RS en las que la subsecuencia única es única con respecto a un polipéptido correspondiente a cualquiera de los polipéptidos de control (por ejemplo, secuencias parentales a partir de las cuales se derivaron las sintetisas de la invención, por ejemplo, por mutación). Las secuencias únicas se determinan como se observa anteriormente.

#### Comparación de secuencias, identidad y homología

**[0152]** Los términos “idéntico” o porcentaje de “identidad,” en el contexto de dos o más secuencias de ácidos nucleicos o de polipéptidos, se refieren a dos o más secuencias o subsecuencias que son las mismas o tienen un porcentaje especificado de residuos de aminoácidos o nucleótidos que son los mismos, cuando se comparan y alinean para máxima correspondencia, como se mide usando uno de los algoritmos de comparación de secuencias descritos más adelante (u otros algoritmos disponibles para expertos habituales) o por inspección visual.

**[0153]** El término “sustancialmente idéntico” en el contexto de dos ácidos nucleicos o polipéptidos (por ejemplo, ADN que codifican un O-ARNt o 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 el 60 %, aproximadamente el 80 %, aproximadamente el 90 – 95 %, aproximadamente el 98 %, aproximadamente el 99 % o más de identidad de residuos de nucleótidos o de aminoácidos cuando se comparan y se alinean para correspondencia máxima, como se mide usando un algoritmo de comparación de secuencias o por inspección visual. Tales secuencias “sustancialmente idénticas” se consideran normalmente que son “homólogas”, sin referencia al linaje real.

Preferentemente, la "identidad sustancial" existe sobre una región de las secuencias que tiene al menos aproximadamente 50 residuos de longitud, más preferentemente sobre una región de al menos aproximadamente 100 residuos y, lo más preferentemente, las secuencias son sustancialmente idénticas sobre al menos aproximadamente 150 residuos, o sobre la longitud completa de las dos secuencias a comparar.

5 **[0154]** Proteínas y/o secuencias de proteínas son "homólogas" cuando se derivan, naturalmente o artificialmente, de una proteína o secuencia de proteínas ancestral común. Similarmente, ácidos nucleicos y/o secuencias de ácidos nucleicos son homólogos cuando se derivan, naturalmente o artificialmente, de un ácido nucleico o secuencia de ácidos nucleicos ancestral. Por ejemplo, cualquier ácido nucleico que se produce naturalmente puede modificarse por cualquier procedimiento de mutagénesis disponible que incluya uno o más codones selectores. Cuando se expresan, este ácido nucleico mutagenizado codifica un polipéptido que comprende uno o más aminoácidos activos por rédox, por ejemplo, aminoácido no natural. Por supuesto, el procedimiento de mutación puede alterar adicionalmente uno o más codones convencionales, cambiando también de esta forma uno o más aminoácidos convencionales en la proteína mutante resultante. La homología se deduce generalmente de la similitud de secuencias entre dos o más ácidos nucleicos o proteínas (o secuencias de los mismos). El porcentaje de similitud preciso entre secuencias que es útil en el establecimiento de la homología varía con el ácido nucleico y la proteína en cuestión, pero tan sólo el 25 % de la similitud de secuencias se usa rutinariamente para establecer homología. También pueden usarse mayores niveles de similitud de secuencias, por ejemplo, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o 99 % o más, para establecer homología. Los procedimientos para determinar los porcentajes de similitud de secuencias (por ejemplo, BLASTP y BLASTN usando parámetros por defecto) se describen en el presente documento y generalmente están disponibles.

25 **[0155]** Para comparación de secuencias y determinación de homología, normalmente una secuencia actúa de secuencia de referencia con la que se comparan las secuencias de prueba. Si se usa un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se entran en un ordenador, se designan coordenadas de subsecuencia, si fuera necesario, y se designan los parámetros del programa de algoritmo de secuencias. Entonces, el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencias para el (las) secuencia (s) de prueba con respecto a la secuencia de referencia, basándose en los parámetros del programa designados.

30 **[0156]** Puede realizarse el alineamiento óptimo de secuencias para comparación, por ejemplo, por el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2: 482 (1981), por el algoritmo de alineamiento por homología de Needleman & Wunsch, J. Mol. Biol. 48: 443 (1970), por la búsqueda del procedimiento de similitud de Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85: 2444 (1988), por implementaciones computerizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI) o por inspección visual (véase generalmente Ausubel y col., más adelante).

35 **[0157]** Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencias y similitud de secuencias es el algoritmo BLAST, que se describe en Altschul y col., J. Mol. Biol. 215: 403 – 410 (1990). El software para realizar los análisis de BLAST está públicamente disponible mediante el Centro nacional para información biotecnológica ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)). Este algoritmo implica identificar primero pares de secuencias de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia de búsqueda, que tanto coinciden como satisfacen alguna puntuación de valor umbral positivo T cuando se alinea con una palabra de la misma longitud en una secuencia de base de datos. T se denomina el umbral de puntuación de palabras en la vecindad (Altschul y col., arriba). Estas palabras comunes en la vecindad iniciales actúan de semillas para iniciar búsquedas para encontrar HSP más largos que las contengan. Entonces, las palabras comunes se extienden en ambas direcciones a lo largo de cada secuencia para que, en la medida de lo posible, pueda puntuarse el alineamiento acumulativo. Las puntuaciones acumuladas 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 acumulada. La extensión de las palabras comunes en cada dirección se detiene cuando: la puntuación de alineamiento acumulada se sale por la cantidad X de su máximo valor alcanzado; la puntuación acumulada tiende a cero o por debajo debido a la acumulación de uno o más alineamientos de residuos de puntuación negativos; o se alcanza el extremo de cualquier secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad de un alineamiento. El programa BLASTN (para secuencias de nucleótidos) usa como valores por defecto una longitud de palabra (W) de 11, una esperanza (E) de 10, un corte de 100, M = 5, N = -4 y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP usa como valores 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).

60 **[0158]** Además de calcular el porcentaje de identidad de secuencias, el algoritmo BLAST también 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 medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de sumas más pequeña (P (N)) que proporciona una indicación de la probabilidad a la que una coincidencia entre dos secuencias de nucleótidos o de aminoácidos se produciría por casualidad. Por ejemplo, un ácido nucleico se



considera similar a una secuencia de referencia si la probabilidad de sumas más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es inferior a aproximadamente 0,1, más preferentemente inferior a aproximadamente 0,01, y lo más preferentemente inferior a aproximadamente 0,001.

## 5 Mutagénesis y otras técnicas de biología molecular

**[0159]** El polinucleótido y polipéptidos de la invención y usados en la invención pueden manipularse usando técnicas biológicas moleculares. Textos generales que describen técnicas biológicas moleculares incluyen Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook y col., Molecular Cloning – A Laboratory Manual (3ª ed.), vol. 1 – 3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 2001 (“Sambrook”) y Current Protocols in Molecular Biology, F.M. Ausubel y col., eds., Current Protocols, una operación conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (complementado a través de 2003) (“Ausubel”). Estos textos describen mutagénesis, el uso de vectores, promotores y muchos otros tópicos relevantes relacionados con, por ejemplo, la generación de genes que incluyen codones selectores para la producción de proteínas que incluyen aminoácidos activos por rédox (por ejemplo, DHP), ARNt ortogonales, sintetasas ortogonales y pares de los mismos.

**[0160]** Se usan diversos tipos de mutagénesis en la invención, por ejemplo, para mutar moléculas de ARNt, para producir bibliotecas de ARNt, para producir bibliotecas de sintetasas, para insertar codones selectores que codifican un aminoácido activo por rédox en una proteína o polipéptido de interés. Incluyen, pero no se limitan a, mutagénesis puntual aleatoria dirigida a sitio, recombinación de homólogos, barajado de ADN u otros procedimientos de mutagénesis recurrentes, construcción química, mutagénesis usando moldes que contienen uracilo, mutagénesis dirigida por oligonucleótidos, mutagénesis de ADN modificada por fosforitoato, mutagénesis usando ADN dúplex con huecos o similares, o cualquier combinación de los mismos. Procedimientos adecuados adicionales incluyen reparación de desapareamientos puntuales, mutagénesis usando cepas huésped deficientes en la reparación, restricción-selección y restricción-purificación, mutagénesis por delección, mutagénesis por síntesis de genes total, reparación de roturas bicatenarias y similares. En la presente invención también se incluye mutagénesis, por ejemplo, que implica construcciones químicas. En una realización, la mutagénesis pueden orientarse por información conocida de la molécula que se produce naturalmente o molécula que se produce naturalmente alterada o mutada, por ejemplo, secuencia, comparaciones de secuencias, propiedades físicas, estructura cristalina o similares.

**[0161]** Las células huésped son genéticamente manipuladas (por ejemplo, transformadas, transducidas o transfectadas) con los polinucleótidos de la invención o 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 el ARNt ortogonal, el ARNt ortogonal sintetasa y la proteína que va a derivatizarse están operativamente ligadas a elementos de control de la expresión génica que son funcionales en la célula huésped deseada. Vectores típicos contienen terminadores de la transcripción y la traducción, secuencias de transcripción e iniciación de la traducción y promotores útiles para 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 (por ejemplo, vectores lanzadera) y marcadores de selección para sistemas tanto procariotas como eucariotas. Los vectores son adecuados para la replicación y/o integración en procariotas, eucariotas, o preferentemente ambos. Véase Giliman & Smith, Gene 8: 81 (1979); Roberts y col., Nature, 328: 731 (1987); Schneider, B. y col., Protein Expr. Purif. 6435: 10 (1995); Ausubel, Sambrook, Berger (todos arriba). El vector puede estar, por ejemplo, en forma de un plásmido, una bacteria, un virus, un polinucleótido desnudo o un polinucleótido conjugado. Los vectores se introducen en células y/o microorganismos mediante procedimientos convencionales que incluyen electroporación (From y col., Proc. Natl. Acad. Sci. USA 82, 5824 (1985)), infección por vectores víricos, penetración balística a alta velocidad por partículas pequeñas con el ácido nucleico tanto dentro de la matriz de perlas o partículas pequeñas como sobre la superficie (Klein y col., Nature 327, 70 – 73 (1987)) y/o similares.

**[0162]** Se proporciona un catálogo de bacterias y bacteriófagos útiles para clonación, por ejemplo, por la ATCC, por ejemplo, The ATCC Catalogue of Bacteria y Bacteriophage (1996) Gherna y col. (eds.) publicado por la ATCC. Procedimientos básicos adicionales para secuenciación, clonación y otros aspectos de biología molecular y consideraciones teóricas subyacentes también se encuentran en Sambrook (arriba), Ausubel (arriba) y en Watson y col. (1992) Recombinant DNA Second Edition Scientific American Books, NY. Además, esencialmente cualquier ácido nucleico (y prácticamente cualquier ácido nucleico marcado, tanto si es convencional como si no es convencional) puede ser pedido por el cliente o convencionalmente a partir de cualquiera de una variedad de fuentes comerciales tales como The Midland Certified Reagent Company (Midland, TX mcr.com), The Great American Gene Company (Ramona, CA disponible en la malla mundial en genco.com), ExpressGen Inc. (Chicago, IL disponible en la malla mundial en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchos otros.

**[0163]** Las células huésped manipuladas pueden cultivarse en medios nutritivos convencionales modificados según convenga para tales actividades como, por ejemplo, etapas de cribado, promotores activantes o transformantes de selección. Estas células pueden cultivarse opcionalmente en organismos transgénicos no humanos. Otras referencias útiles, por ejemplo, para el aislamiento y cultivo de células (por ejemplo, para el posterior aislamiento de ácidos nucleicos) incluyen Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, Nueva York y las referencias citadas en su interior; Payne y col. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. Nueva York, NY; Gamborg y Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods* Springer Lab Manual, Springer-Verlag (Berlin Heidelberg Nueva York) y Atlas y Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL.

#### PROTEÍNAS Y POLIPÉPTIDOS DE INTERÉS

**[0164]** Una ventaja de los aminoácidos activos por rédox es que pueden usarse para manipular los procedimientos de transferencia de electrones en proteína. Otras ventajas incluyen, pero no se limitan a, que la expresión de proteínas activas por rédox puede facilitar el estudio y la capacidad para alterar las rutas de transferencia de electrones en proteínas, alterar la función catalítica de enzimas, reticular proteína con moléculas pequeñas y biomoléculas, etc.

**[0165]** Los procedimientos de producción de una proteína en una célula con un aminoácido activo por rédox en una posición especificada también son un rasgo de la invención. Por ejemplo, un procedimiento incluye cultivar, en un medio apropiado, la célula, en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína; y proporcionar el aminoácido activo por rédox; 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 preferencialmente el O-ARNt con el aminoácido activo por rédox. En ciertas realizaciones, el O-ARNt comprende al menos aproximadamente, por ejemplo, un 45 %, un 50 %, un 60 %; un 75 %, un 80 % o un 90 % o más de eficiencia de supresión en presencia de una sintetasa relacionada en respuesta al codón selector con respecto al O-ARNt que comprende o está codificado por una secuencia de polinucleótidos que se expone en SEQ ID NO: 2. Una proteína producida por este procedimiento también es un rasgo de la invención.

**[0166]** Las composiciones de la invención y composiciones hechas mediante los procedimientos de la invención están opcionalmente en una célula. Los pares de O-ARNt/O-RS o componentes individuales de la invención pueden luego usarse en una maquina de traducción del sistema huésped que produce un aminoácido activo por rédox que se incorpora en una proteína. La solicitud internacional número PCT/US2004/011786 presentada el 16 de abril de 2004 titulada "Expansión del código genético eucariota"; y el documento WO 2002/085923 titulado "INCORPORACIÓN IN VIVO DE AMINOÁCIDOS NO NATURALES" describe este procedimiento. Por ejemplo, cuando un par de O-ARNt/O-RS se introduce en un huésped, por ejemplo, *Escherichia coli*, el par conduce a la incorporación *in vivo* de aminoácido activo por rédox, tal como DHP, por ejemplo, un aminoácido sintético, tal como derivado de un aminoácido de tirosina o fenilalanina, que puede añadirse exógenamente al medio de crecimiento, en una proteína, en respuesta a un codón selector. Opcionalmente, las composiciones de la presente invención puede estar en un sistema de traducción *in vitro*, o en un sistema (s) *in vivo* no humano (s).

**[0167]** Una célula de la invención proporciona la capacidad para sintetizar proteínas que comprenden aminoácidos no naturales en cantidades útiles grandes. En un aspecto, la composición incluye opcionalmente, por ejemplo, al menos 10 microgramos, al menos 50 microgramos, al menos 75 microgramos, al menos 100 microgramos, al menos 200 microgramos, al menos 250 microgramos, al menos 500 microgramos, al menos 1 miligramo, al menos 10 miligramos o más de la proteína que comprende un aminoácido activo por rédox, o una cantidad que puede lograrse con procedimientos de producción de proteína *in vivo* (en el presente documento se proporcionan detalles sobre producción y purificación de proteína recombinante). En otro aspecto, la proteína está opcionalmente presente en la composición a una concentración de, por ejemplo, al menos 10 microgramos de proteína por litro, al menos 50 microgramos de proteína por litro, al menos 75 microgramos de proteína por litro, al menos 100 microgramos de proteína por litro, al menos 200 microgramos de proteína por litro, al menos 250 microgramos de proteína por litro, al menos 500 microgramos de proteína por litro, al menos 1 miligramo de proteína por litro, o al menos 10 miligramos de proteína por litro o más, en, por ejemplo, un lisado celular, un tampón, un tampón farmacéutico u otra suspensión líquida (por ejemplo, en un volumen de, por ejemplo, en cualquier parte de aproximadamente 1 nl a aproximadamente 100 l). La producción de grandes cantidades (por ejemplo, superiores a las normalmente posibles con otros procedimientos, por ejemplo, traducción *in vitro*) de una proteína en una célula que incluye al menos un aminoácido activo por rédox es un rasgo de la invención.

**[0168]** La incorporación de un aminoácido activo por rédox puede hacerse, por ejemplo, confeccionando cambios en la estructura y/o función de proteínas, por ejemplo, para cambiar el tamaño, acidez, nucleofilia, enlaces de hidrógeno, hidrofobia, accesibilidad de sitios diana de proteasas, diana para un resto (por ejemplo, para una matriz de proteínas), etc. Las proteínas que incluyen un aminoácido activo por rédox pueden tener propiedades

catalíticas o físicas potenciadas o incluso nuevas en su totalidad. Por ejemplo, las siguientes propiedades se modifican opcionalmente por inclusión de un aminoácido activo por rédox en una proteína: 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, covalentemente o no covalentemente, y similares. Las composiciones que incluyen proteínas que incluyen al menos un aminoácido activo por rédox son útiles para, por ejemplo, agentes terapéuticos novedosos, diagnósticos, enzimas catalíticas, enzimas industriales, proteínas de unión (por ejemplo, anticuerpos) y, por ejemplo, el estudio de 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.

**[0169]** Una composición incluye al menos una proteína con al menos uno, 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, por ejemplo, aminoácidos activos por rédox y/u otros 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 ó 10 o más sitios diferentes en la proteína que comprenden 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más aminoácidos no naturales diferentes. Una composición que incluye una proteína con al menos uno, pero menos de todos, de un aminoácido particular presente en la proteína está sustituido con el aminoácido activo por rédox. Para una proteína dada con más de un aminoácido no natural, los aminoácidos no naturales pueden ser idénticos o diferentes (por ejemplo, la proteína puede incluir dos o más tipos diferentes de aminoácidos no naturales, o puede incluir dos del mismo aminoácido no natural). Para una proteína dada con más de dos aminoácidos no naturales, los aminoácidos no naturales pueden ser los mismos, diferentes o una combinación de un aminoácido no natural múltiple del mismo tipo con al menos un aminoácido no natural diferente.

**[0170]** Esencialmente cualquier proteína (o porción de la misma) que incluye un aminoácido activo por rédox (y cualquier ácido nucleico codificante correspondiente, por ejemplo, que incluye uno o más codones selectores) puede producirse usando las composiciones y procedimientos en el presente documento. No se hace intento por identificar los cientos de miles de proteínas conocidas, cualquiera de las cuales puede modificarse para incluir uno o más aminoácidos no naturales, por ejemplo, confeccionando cualquier procedimiento de mutación disponible para incluir uno o más codones selectores apropiados en un sistema de traducción relevante. Repositorios de secuencias comunes para proteínas conocidas incluyen GenBank, EMBL, DDBJ y NCBI. Otros repositorios pueden identificarse fácilmente buscando en internet.

**[0171]** Normalmente, las proteínas son, por ejemplo, al menos el 60 %, al menos el 70 %, al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 % o al menos el 99 % o más idénticas a cualquier proteína disponible (por ejemplo, una proteína terapéutica, una proteína de diagnóstico, una enzima industrial, o porción de las mismas, y similares), y comprenden uno o más aminoácidos no naturales. Ejemplos de proteínas terapéuticas, de diagnóstico y otras proteínas que pueden modificarse para comprender uno o más aminoácidos activos por rédox pueden encontrarse, pero no se limitan a aquellas, en los documentos WO 2004/094593 y WO 2002/085923 titulado "INCORPORACIÓN IN VIVO DE AMINOÁCIDOS NO NATURALES". Ejemplos de proteínas terapéuticas, de diagnóstico y otras proteínas que pueden modificarse para comprender uno o más aminoácidos activos por rédox incluyen, pero no se limitan a, por ejemplo, antitripsina alfa-1, angiostatina, factor antihemolítico, anticuerpos (más detalles sobre anticuerpos se encuentran más adelante), apolipoproteína, apoproteína, factor natriurético auricular, polipéptido natriurético auricular, péptidos auriculares, quimiocinas C-X-C (por ejemplo, T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), calcitonina, quimiocinas CC (por ejemplo, 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, T64262), ligando CD40, ligando C-kit, colágeno, factor estimulante de colonias (CSF), factor 5a del complemento, inhibidor del complemento, receptor 1 del complemento, citocinas, (por ejemplo, péptido-78 activante de neutrófilos epiteliales, GRO $\alpha$ /MGSA, GRO $\beta$ , GRO $\gamma$ , MIP-1 $\alpha$ , MIP-1 $\delta$ , MCP-1), factor de crecimiento epidérmico (EGF), eritropoyetina ("EPO"), toxinas A y B exfoliantes, 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 (por ejemplo, Sonic, Indian, Desert), hemoglobina, factor de crecimiento de hepatocitos (HGF), hirudina, albúmina de suero humano, insulina, factor de crecimiento similar a la insulina (IGF), interferones (por ejemplo, IFN- $\alpha$ , IFN- $\beta$ , IFN- $\gamma$ ), interleucinas (por ejemplo, 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, neuriturina, factor inhibidor de neutrófilos (NIF), oncostatina M, proteína osteogénica, hormona paratiroidea, PD-ECSF, PDGF, hormonas peptídicas (por ejemplo, hormona de crecimiento humana), pleiotropina, proteína A, proteína G, exotoxinas pirogénicas A, B y C, relaxina, renina, SCF, receptor I del complemento soluble, I-CAM 1 soluble, receptores de interleucina solubles (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), receptor de TNF soluble, somatomedina, somatostatina, somatotropina, estreptocinas, superantígenos, es decir, enterotoxinas estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutasa (SOD), toxina de síndrome de choque tóxico (TSST-1), timosina alfa 1, activador tisular del plasminógeno, factor de necrosis tumoral beta (TNF beta), receptor del factor de necrosis tumoral (TNFR), factor de necrosis tumoral-alfa (TN-alfa), factor de crecimiento endotelial vascular (VEGEF), urocinasa y muchas otras.

**[0172]** Una clase de proteínas que pueden prepararse usando las composiciones y procedimientos para incorporación *in vivo* de aminoácidos activos por rédox descritos en el presente documento incluye moduladores transcripcionales o una parte de los mismos. Ejemplos de moduladores transcripcionales incluyen genes y proteínas modificadoras transcripcionales que modulan el crecimiento, diferenciación, regulación celular o similares. Moduladores transcripcionales se encuentran en procariotas, virus y eucariotas, que incluyen hongos, plantas, levaduras, insectos y animales, que incluyen mamíferos, proporcionando un amplio intervalo de dianas terapéuticas. Se apreciará que los activadores de la expresión y transcripcionales regulan la transcripción por muchos mecanismos, por ejemplo, por unión a receptores, estimulando una cascada de transducción de señales, regulando la expresión de factores de transcripción, unión a promotores y potenciadores, unión a proteínas que se unen a promotores y potenciadores, desenrollando ADN, cortando y empalmado pre-ARNm, poliadenilando ARN y degradando ARN.

**[0173]** Una clase de proteínas (por ejemplo, proteínas con uno o más aminoácidos activos por rédox) incluyen activadores de la expresión tales como citocinas, 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- $\alpha$ , TGF- $\beta$ , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1 y 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 la hormona esteroidea tales como aquellos para estrógeno, progesterona, testosterona, aldosterona, el ligando para el receptor de LDL y corticosterona.

**[0174]** Enzimas (por ejemplo, enzimas industriales) o porciones de las mismas con al menos un aminoácido activo por rédox también se describen en el presente documento. Ejemplos de enzimas incluyen, pero no se limitan a, por ejemplo, amidasas, aminoácido racemasas, acilasas, deshalogenasas, dioxigenasas, diarilpropano peroxidadas, epimerasas, epóxido hidrolasas, estererasas, isomerasas, cinasas, glucosa isomerasas, glucosidasas, glicosil transferasas, haloperoxidasas, monooxigenasas (por ejemplo, p450s), lipasas, lignina peroxidadas, nitrilo hidratadas, nitrilasas, proteasas, fosfatasas, subtilisinas, transaminasa y nucleasas.

**[0175]** Muchas de estas proteínas están comercialmente disponibles (véase, por ejemplo, el catálogo y la lista de precios de Sigma BioSciences 2002), y las secuencias y genes de proteínas correspondientes y, normalmente, muchas variantes de las mismas, son muy conocidas (véase, por ejemplo, Genbank). Cualquiera de ellas puede modificarse por la inserción de uno o más aminoácidos activos por rédox según la invención, por ejemplo, para alterar la proteína con respecto a una o más propiedades terapéuticas, de diagnóstico o enzimáticas de interés. Ejemplos de propiedades terapéuticamente relevantes incluyen semivida en suero, semivida en almacén, estabilidad, inmunogenicidad, actividad terapéutica, detectabilidad (por ejemplo, por la inclusión de grupos indicadores (por ejemplo, marcas o sitios de unión a marcas) en los aminoácidos no naturales, por ejemplo, aminoácidos activos por rédox), reducción de LD<sub>50</sub> u otros efectos secundarios, capacidad para entrar en el cuerpo por el tubo gástrico (por ejemplo, disponibilidad oral) o similares. Ejemplos de propiedades de diagnóstico incluyen semivida en almacén, estabilidad, actividad de diagnóstico, detectabilidad o similares. Ejemplos de propiedades enzimáticas relevantes incluyen semivida en almacén, estabilidad, actividad enzimática, capacidad de producción o similares.

**[0176]** También puede modificarse una variedad de otras proteínas para incluir uno o más aminoácidos activos por rédox 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 activo por rédox, por ejemplo, en proteínas de hongos infecciosos, por ejemplo, especies de *Aspergillus*, *Candida*; bacterias, particularmente *E. coli*, que sirve de modelo para bacterias patógenas, además de 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 ( + ) (ejemplos incluyen poxvirus, por ejemplo, variolovacuna; picornavirus, por ejemplo, poliomielititis; togavirus, por ejemplo, *rubella*; flavivirus, por ejemplo, VHC; y virus corona), virus de ARN ( - ) (por ejemplo, rabdovirus, por ejemplo, VSV; paramixovirus, por ejemplo, RSV; ortomixovirus, por ejemplo, gripe; bunyavirus; y arenavirus), virus de ADNbc (por ejemplo, reovirus), virus de ARN a ADN, es decir, retrovirus, por ejemplo, VIH y HTLV, y ciertos virus de ADN a ARN tales como hepatitis B.

**[0177]** Proteínas agrícolamente relacionadas tales como proteínas de resistencia a insectos (por ejemplo, las proteínas Cry), enzimas de producción de almidón y lípido, toxinas de plantas e insectos, proteínas de resistencia a toxinas, proteínas de desintoxicación de microtoxinas, enzimas de crecimiento de plantas (por ejemplo, ribulosa 1,5-disfosfato carboxilasa/oxigenasa, "RUBISCO"), lipoxigenasa (LOX) y fosfoenolpiruvato (PEP) carboxilasa también son dianas adecuadas para modificaciones de aminoácidos activos por rédox.

**[0178]** En ciertas realizaciones, la proteína o polipéptido de interés (o porción del mismo) en los procedimientos y/o composiciones de la invención está codificado 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.

5  
**[0179]** Los genes que codifican proteínas o polipéptidos de interés puede mutagenizarse usando procedimientos muy conocidos para un experto en la materia y descritos en el presente documento bajo “Mutagénesis y otras técnicas de biología molecular” para incluir, por ejemplo, uno o más codones selectores para la incorporación de un aminoácido activo por rédox. Por ejemplo, un ácido nucleico para una proteína de interés se mutageniza para incluir uno o más codones selectores, proporcionando la inserción del uno o más aminoácidos activos por rédox. La invención incluye cualquiera de tales variantes, por ejemplo, versiones mutantes de cualquier proteína, por ejemplo, que incluyen al menos un aminoácido activo por rédox. Similarmente, la invención también incluye ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifican uno o más aminoácidos activos por rédox.

15  
**[0180]** Para preparar una proteína que incluye un aminoácido activo por rédox pueden usarse células huésped y organismos no humanos que están adaptados para la incorporación *in vivo* del aminoácido activo por rédox mediante pares de ARNt/RS ortogonales. Las células huésped son genéticamente manipuladas (por ejemplo, transformadas, transducidas o transfectadas) con uno o más vectores que expresan el ARNt ortogonal, la ARNt-sintetasa ortogonal y un vector que codifica la proteína a derivatizar. Cada uno de estos componentes puede estar sobre el mismo vector, o cada uno puede estar sobre un vector separado, o dos componentes pueden estar sobre un vector y el tercer componente sobre 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.

#### 25 Definición de polipéptidos por inmunorreactividad

**[0181]** Debido a que los polipéptidos proporcionan una variedad de nuevas secuencias de polipéptidos (por ejemplo, que comprenden aminoácidos activos por rédox en el caso de proteínas sintetizadas en los sistemas de traducción en el presente documento, o, por ejemplo, en el caso de las sintetisas novedosas, secuencias novedosas de aminoácidos convencionales), los polipéptidos también proporcionan nuevos rasgos estructurales que pueden reconocerse, por ejemplo, en ensayos inmunológicos. La generación de antisueros, que se unen específicamente a los polipéptidos de la invención, además de los polipéptidos que están unidos por tales antisueros, son un rasgo de la invención. El término “anticuerpo”, como se usa en el presente documento, incluye, pero no se limita a, un polipéptido sustancialmente codificado por un gen inmunoglobulina o genes inmunoglobulina, o fragmentos de los mismos que se unen específicamente y reconocen un analito (antígeno). Ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos y monocatenarios, y similares. Fragmentos de inmunoglobulinas, que incluyen fragmentos Fab y fragmentos producidos por una biblioteca de expresión, que incluyen expresión en fago, también están incluidos en el término “anticuerpo” como se usa en el presente documento. Véase, por ejemplo, Paul, *Fundamental Immunology*, 4<sup>a</sup> ed., 1999, Raven Press, Nueva York, para estructura y terminología de anticuerpos.

**[0182]** Con el fin de producir antisueros para su uso en un inmunoensayo, uno o más de los polipéptidos inmunogénicos se produce y purifica como se describe en el presente documento. Por ejemplo, puede producirse proteína recombinante en una célula recombinante. Una cepa endogámica de ratones (usada en este ensayo debido a que los resultados son más reproducibles debido a la identidad genética virtual de los ratones) se inmuniza con la (s) proteína (s) inmunogénica (s) en combinación con un adyuvante convencional tal como adyuvante de Freund, y un protocolo de inmunización de ratones convencional (véase, por ejemplo, Harlow y Lane (1988) *Antibodies*, A Laboratory Manual, Cold Spring Harbor Publications, Nueva York, para una descripción convencional de generación de anticuerpos, formatos de inmunoensayo y condiciones que pueden usarse para determinar inmunorreactividad específica. Detalles adicionales sobre proteínas, anticuerpos, antisueros, etc. pueden encontrarse en los documentos USSN 60/479.931, 60/463.869 y 60/496.548 titulado “Expansión del código genético eucariota”; documento WO 2002/085923 titulado “INCORPORACIÓN IN VIVO DE AMINOÁCIDOS NO NATURALES”; solicitud de patente titulada “Síntesis de glucoproteínas” presentada el 16 de enero de 2003, documento USSN 60/441.450; y solicitud de patente titulada “Matrices de proteínas”, expediente del agente número P1001US00 presentado el 22 de diciembre de 2002.

#### USO DE O-ARNt Y O-RS Y PARES DE O-ARNt/O-RS

**[0183]** Las composiciones de la invención y composiciones preparadas mediante los procedimientos de la invención son opcionalmente en una célula. Los pares de O-ARNt/O-RS o componentes individuales de la invención pueden luego usarse en una maquinaria de traducción del sistema huésped que produce un aminoácido activo por rédox que se incorpora en una proteína. La solicitud de patente correspondiente “Incorporación in vivo de aminoácidos no naturales”, WO 2002/085923, por Schultz y col. describe este procedimiento. Por ejemplo, cuando un par de O-ARNt/O-RS se introduce en un huésped, por ejemplo, *Escherichia coli*, el par conduce a la

incorporación *in vivo* de un aminoácido activo por rédox, que puede añadirse exógenamente al medio de crecimiento, en una proteína, por ejemplo, mioglobina o una proteína terapéutica, en respuesta a un codón selector, por ejemplo, un codón antisentido ámbar. Opcionalmente, las composiciones de la invención puede estar en un sistema de traducción *in vitro*, o en un sistema (s) *in vivo*. Proteínas con el aminoácido activo por rédox pueden usarse como proteínas terapéuticas y pueden usarse para alterar la función catalítica de enzimas y/o las rutas de transferencia de electrones en proteínas, para reticular proteína con moléculas pequeñas y/o biomoléculas y para facilitar estudios sobre la estructura de proteínas, interacciones con otra proteína, procedimientos de transferencia de electrones en proteínas y similares.

## 10 EJEMPLOS

[0184] Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar la invención reivindicada. Un experto reconocerá una variedad de parámetros no críticos que pueden alterarse sin apartarse del alcance de la invención reivindicada.

### 15 EJEMPLO 1: INCORPORACIÓN ESPECÍFICA PARA SITIO DE UN AMINOÁCIDO ACTIVO POR RÉDOX EN PROTEÍNAS

[0185] Recientemente se ha mostrado informado que varios aminoácidos no naturales pueden incorporarse selectivamente en proteínas en *E. coli* y levadura (Wang y col. (2001) *Science* 292: 498 – 500; Zhang y col. (2003) *Biochemistry* 42: 6735 – 6746; Chin y col. (2003) *Science* 301: 964 – 967) usando pares de ARNt-aminoacil ARNt sintetasa ortogonales. Estos pares ortogonales no reaccionan de forma cruzada con componentes endógenos de la máquina de traducción de la célula huésped, pero reconocen el aminoácido no natural deseado y lo incorporan en proteínas en respuesta al codón antisentido ámbar, TAG (Wang y col. (2000) *J. Am. Chem. Soc.*, 122: 5010 – 5011; Wang y Schultz (2001) *Chem. Biol.*, 8: 83 – 89). Para codificar genéticamente 3,4-dihidroxi-L-fenilalanina (DHP; véase el compuesto 1 en la FIG. 1) en *E. coli*, la especificidad de una ARNt-sintetasa de *Methanococcus jannaschii* ortogonal (MjTyrRS; proporcionada en la FIG. 5 y la Tabla 1 y, por tanto, la secuencia de aminoácidos proporcionada en SEQ ID NO: 4 y la secuencia de nucleótidos proporcionada en SEQ ID NO: 5) se alteró de manera que la

sintetasa aminoacilara el supresor ámbar de tirosina ARNt mutante ( $ARNmut_{CUA}^{Tyr}$ ) con DHP y no con cualquiera de los veinte aminoácidos comunes. Estas sintetasas mutantes se seleccionaron de dos bibliotecas de MjTyrRS mutantes (Wang y col. (2001) *Science* 292: 498 – 500; Zhang y col. (2002) *Angew. Chem. Int. Ed.*, 41: 2840 – 2842). En la primera biblioteca, que se basa en un análisis de la estructura cristalina de TyrRS homóloga de *Bacillus stearothermophilus* (Brick y col. (1989) *J. Mol. Biol.*, 208: 83 – 98), cinco residuos (Tyr 32, Glu 107, Asp 158, Ile 159 y Leu 162) en el sitio activo de MjTyrRS que están dentro de 6,5 Å de la posición *para* del anillo de arilo de tirosina se mutaron aleatoriamente (codificados sobre el plásmido pBK-lib). En la segunda biblioteca, seis residuos (Tyr32, Ala 67, His 70, Gln 155, Asp 158, Ala 167) dentro de 6,9 Å de la posición *meta* del anillo de arilo de tirosina se mutaron aleatoriamente (codificados sobre el plásmido pBK-lib-m).

[0186] Para alterar la especificidad de TyrRS, de forma que incorporara específicamente DHP y ninguno de los otros aminoácidos naturales, se aplicó una selección genética que consistió en varias rondas de selección positiva y negativa. En la selección positiva, ambas bibliotecas de TyrRS mutante se sometieron a un esquema de selección basado en la supresión de un codón ámbar introducido en una posición no esencial (Asp112) en el gen cloranfenicol acetil transferasa (CAT) (pRep (2)/YC). Células transformadas con las bibliotecas de TyrRS mutante, el

$ARNmut_{CUA}^{Tyr}$  y el gen CAT mutante ámbar se cultivaron en medio mínimo que contenía DHP 1 mM y 70 µg/ml de cloranfenicol en condiciones anaerobias para evitar la oxidación de DHP. Las células supervivientes contienen TyrRS

mutantes que aminoacilan el  $ARNmut_{CUA}^{Tyr}$  con tanto DHP como aminoácidos endógenos. A continuación se aplicó una selección negativa para eliminar las TyrRS mutantes que cargan aminoácidos naturales basándose en la supresión de tres codones ámbar introducidos en posiciones no esenciales (Gln2, Asp44, Gly55) en el gen barnasa tóxico

(pLWJ17B3). Las células que alojan las TyrRS mutantes de la selección positiva previa, el  $ARNmut_{CUA}^{Tyr}$  y el gen barnasa mutante ámbar se cultivaron en medio de Luria-Bertani (LB) en ausencia de DHP. Bajo estas condiciones, las células que codifican TyrRS mutantes con especificidad por aminoácidos endógenos producirán barnasa de longitud completa y morirán. Sólo aquellas células que contienen TyrRS mutantes con especificidad por DHP pueden sobrevivir. Después de tres rondas de selección positiva alternando con dos rondas de selección negativa se desarrolló un clon cuya supervivencia a altas concentraciones de cloranfenicol (90 mg/l) dependía de la presencia de DHP, el gen TyrRS

mutante seleccionado (DHPRS),  $ARNmut_{CUA}^{Tyr}$  y el gen Asp112TAG CAT. Sin embargo, en ausencia de DHP, las mismas células sólo sobrevivieron en 20 mg/l de cloranfenicol. Este resultado sugiere que la enzima DHPRS seleccionada tiene mayor especificidad por DHP que por aminoácidos naturales. La secuenciación reveló los siguientes mutantes en DHPRS seleccionada: Tyr32 → Leu, Ala67 → Ser, His70 → Asn, Ala167 → Gln. La DHPRS sintetasa se muestra en la FIG. 6 y la Tabla 1. Por tanto, la secuencia de aminoácidos se proporciona en SEQ ID NO: 1 y la secuencia de nucleótidos se proporciona en SEQ ID NO: 3.

[0187] Para medir la fidelidad y eficiencia de la incorporación de DHP, usando el clon seleccionado pDHPRS, los presentes inventores incorporaron DHP en respuesta a un codón ámbar en el cuarto residuo expuesto en la superficie en la mioglobina de esperma de ballena mutante marcada con hexahistidina en el extremo C (Mb; véase Chin y col. (2002) Proc. Natl. Acad. Sci. U.S.A., 99: 11020 – 11024). La DHP de longitud completa que contiene mioglobina (DHPMb) se expresó usando GMML (medio mínimo de glicerol con leucina) como medio de crecimiento y bajo condiciones reductoras (ditiotreitól 100  $\mu$ M (DTT)) con el fin de prevenir la oxidación de DHP antes de la incorporación en la proteína. El rendimiento de proteína mutante fue aproximadamente 1 mg/litro (el rendimiento de Mb natural (wtMb) bajo las mismas condiciones es indetectable). No se expresó Mb de longitud completa en ausencia de DHP; en ausencia de DTT, la mayoría de las células murieron debido a la toxicidad de la quinona oxidada (véase el compuesto 3 en la FIG. 1). Se purificó DHPMb de longitud completa usando resina IMAC basada en cobalto (cromatografía de afinidad por metal inmovilizado). Las muestras purificadas de las proteínas mutantes expresadas en presencia y en ausencia de DHP se cargaron sobre un gel de SDS-PAGE, para tinción con plata, y transferencia Western del gel. Usando anticuerpo anti-marca His6 no se expresó Mb de longitud completa en ausencia de tanto DHPRS como *ARNmut<sup>DHP</sup><sub>CUA</sub>* (mostrado en la FIG. 2A). Se usó ionización por electropulverización (ESI) con un espectrómetro de masas de tiempo de vuelo de cuadrupolo-cuadrupolo (QqTOF) para medir el peso molecular de la proteína. La FIG. 2B muestra el espectro de masas de ESI-QqTOF de DHPMb con una masa de 18.448,5 Dalton. Ésta está dentro de 70 p.p.m. de la masa calculada de 18447,2 Dalton para Mb que contiene DHP (un pico vecino muestra una masa de 18.432,3 Dalton debido a una pérdida de oxígeno, u oxígeno y protón producido, según experimentos de control, por la técnica de medición).

[0188] Se usó voltametría cíclica para determinar si pudo observarse la onda rédox de la hidroquinona oxidada cuando un electrodo de oro desnudo se sumergió en una disolución que contenía DHPMb. La FIG. 3A muestra una respuesta voltamétrica irreversible de una disolución que contiene wtMb y la de DHPMb en condiciones anaerobias (Bard y Faulkner, en *Electrochemical Methods*; John W. Wiley & Sons, Inc.: Nueva York, 1980; pp 213 – 248, 429 – 487 y 675 – 698). El potencial de pico reductor que se origina a partir de wtMbFe (III) se observa a E = -320 mV, mientras que el potencial de pico reductor de la proteína mutante está desplazado a un potencial más negativo de E = -400 mV. Este desplazamiento es atribuido a la presencia de DHP, que puede facilitar la reducción de Fe (III) a un potencial mucho menor que en ausencia de DHP. Los voltamogramas observados irreversibles son debidos a una lenta tasa de transferencia de electrones, que es probable que se derive de la accesibilidad limitada del electrón al electrodo. La FIG. 3B muestra la respuesta voltamétrica de una disolución que contiene 100  $\mu$ M de DHP, wtMb y DHPMb. La corriente que se origina a partir de la oxidación de DHP sólo aparece en presencia de Mb mutado o en una disolución de DHP libre con E = 580 mV y E = 385 mV, respectivamente. Estos resultados muestran claramente que hay una influencia significativa de la presencia de DHP en Mb sobre el potencial rédox del grupo hemo de Fe (III) y viceversa.

[0189] La descripción proporcionada en el presente documento demuestra que aminoácidos activos por rédox, por ejemplo, DHP, pueden incorporarse eficientemente y selectivamente en proteínas en un organismo, por ejemplo, *E. coli*. Estos aminoácidos pueden oxidarse electroquímicamente dentro de la proteína. La capacidad para incorporar específicamente para sitio aminoácidos activos por rédox en proteínas puede facilitar el estudio de la transferencia de electrones en proteínas, además de permitir la manipulación de proteínas rédox con novedosas propiedades. La incorporación específica para sitio de aminoácidos activos por rédox, por ejemplo, DHP, en diversos sitios en proteínas modelo, por ejemplo, Mb y otras proteínas, puede usarse para estudiar las rutas de transferencia de electrones en esta proteína y otras (Mayo y col. (1986) *Science* 233: 948 – 952; Gray y Malmstrom (1989) *Biochemistry* 28: 7499 – 7505).

#### **EJEMPLO 2: O-RS Y O-ARNt A MODO DE EJEMPLO PARA LA INCORPORACIÓN DE AMINOÁCIDOS ACTIVOS POR RÉDOX.**

[0190] Un O-ARNt a modo de ejemplo comprende SEQ ID NO: 2 (véase la Tabla 1). O-RS de ejemplo incluyen la secuencia de aminoácidos proporcionada en SEQ ID NO: 1 (véase la Tabla 1) y la FIG. 6. Ejemplos de polinucleótidos que codifican O-RS o porciones de las mismas incluyen polinucleótidos que codifican una secuencia de aminoácidos que comprende SEQ ID NO: 1. Por ejemplo, el polinucleótido proporcionado en la FIG. 6 y SEQ ID NO: 3 codifica O-RS a modo de ejemplo.



TABLA 1: SECUENCIAS

SEQ ID NO:	Descripción	SECUENCIA
1	Secuencia de aminoácidos de DHPRS (sintetasa) que tiene los cambios de aminoácidos: Tyr32 → Leu, Ala67 → Ser, His70 → Asn, Ala167 → Gln basándose en tirosina ARNt-sintetasa de <i>Methanococcus jannaschii</i> (MjTyrRS)	MDEFEMIKRNTSEI ISEEELREVLKKDEKSA LIGFEPGKIHLGHYLQIKKMIDLQNA GFDI IILLSDLNAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSEFQLDKDYTLNVYRLALK TTLKRARRSMELIAREDENPKVAEVIYPI MQ VNDIHYLGVDVQVGGMEQRKIHMLARELLPK KVVCIHNPVLTGLDGEKMSSSKGNFIAVDD SPEEIRAKIKKAYCPAGVVEGNPIMEIAKYF LEYPLTIKRPEKFGGDLTVNSYEELESLEFKN KELHPMDLKNVAEELIKILEPIRKRL
2	<i>ARN<sup>Tyr</sup>mut<sub>CUA</sub></i>	CCGGCGGUAGUUCAGCAGGGCAGAACGGCGG ACUCUAAAUCGCAUGGCGCUGGUUCAAAUC CGGCCCGCGGACCA
3	Secuencia de nucleótidos de DHPRS (sintetasa) que codifica los cambios de aminoácido: Tyr32 → Leu, Ala67 → Ser, His70 → Asn, Ala167 → Gln basándose en tirosina ARNt-sintetasa de <i>Methanococcus jannaschii</i> (MjTyrRS)	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTCTCATAGGT TTTGAACCAAGTGGTAAAAATACATTTAGGGCATT ATCTCCAAATAAAAAAGATGATTGATTTACAAAA TGCTGGATTTGATAATAATTATATTGTTGAGCGAT TTAAACGCCATTAAACAGAAAGGAGAGTTGG ATGAGATTAGAAAAATAGGAGATTATAACAAAA AGTTTTTGAAGCAATGGGGTTAAAGGCAAAATAT GTTTATGGAAGTGAATTCAGCTTGATAAGGATT ATACACTGAATGCTATAGATTGGCTTTAAAAAC TACCTTAAAAAGAGCAAGAAGGAGTATGGAACCTT ATAGCAAGAGAGGATGAAAAATCCAAGTTGCTG AAGTTATCTATCCAATAATGCAGGTTAATGATAT TCATTATTTAGGCGTTGATGTTTCAAGTTGGAGGG ATGGAGCAGAGAAAAATACACATGTTAGCAAGGG AGCTTTTACCAAAAAAGGTTGTTTGTATTCACAA CCCTGTCTTACGGGTTTGGATGGAGAAGGAAAG ATGAGTTC TCAAAAGGGAATTTTATAGCTGTTG ATGACTCTCCAGAAGAGATTAGGGCTAAGATAAA GAAAGCATACTGCCAGCTGGAGTTGTTGAAGGA AATCCAATAATGGAGATAGCTAAATACTTCCTTG AATATCCTTTAACCATAAAAAAGGCCAGAAAAATT TGGTGGAGATTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTCATC CAATGGATTTAAAAATGCTGTAGCTGAAGAACT TATAAAGATTTTAGAGCCAATTAGAAAGAGATTA
4	Secuencia de aminoácidos de tirosina ARNt-sintetasa de <i>Methanococcus jannaschii</i> (MjTyrRS)	MDEFEMIKRNTSEI ISEEELREVLKKDEKSA YIGFEPGKIHLGHYLQIKKMIDLQNA GFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSEFQLDKDYTLNVYRLALK TTLKRARRSMELIAREDENPKVAEVIYPI MQ VNDIHYLGVDVAVGGMEQRKIHMLARELLPK KVVCIHNPVLTGLDGEKMSSSKGNFIAVDD SPEEIRAKIKKAYCPAGVVEGNPIMEIAKYF LEYPLTIKRPEKFGGDLTVNSYEELESLEFKN KELHPMDLKNVAEELIKILEPIRKRL
5	Secuencia de nucleótidos de tirosina ARNt-sintetasa de <i>Methanococcus jannaschii</i> (MjTyrRS)	



SEQ ID NO:	Descripción	SECUENCIA
		<p>ATGGACGAATTTGAAATGATAAAGAGAAACACAT            CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT            TTTAAAAAAGATGAAAAATCTGCTTACATAGGT            TTTGAACCAAGTGGTAAAATACATTTAGGGCATT            ATCTCCAATAAAAAAGATGATTGATTTACAAAA            TGCTGGATTTGATAAATTATATTGTTGGCTGAT            TTACACGCCATTTAAACCAGAAAGGAGAGTTGG            ATGAGATTAGAAAAATAGGAGATTATAACAAAA            AGTTTTTGAAGCAATGGGGTTAAAGGCAAAATAT            GTTTATGGAAGTGAATTCAGCTTGATAAGGATT            ATACACTGAATGTCATAGATTGGCTTTAAAAAC            TACCTTAAAAAGAGCAAGAAGGAGTATGGAACCTT            ATAGCAAGAGAGGATGAAAAATCCAAGGTTGCTG            AAGTTATCTATCCAATAATGCAGGTTAATGATAT            TCATTATTTAGGCGTTGATGTTGCAGTTGAGGG            ATGGAGCAGAGAAAAATACACATGTTAGCAAGGG            AGCTTTTACCAAAAAGGTTGTTGTAATCACAA            CCCTGTCTTACGGGTTTGGATGGAGAAGGAAAG            ATGAGTTC'TCAAAGGGAATTTATAGCTGTTG            ATGACTCTCCAGAAGAGATTAGGGCTAAGATAAA            GAAAGCATACTGCCAGCTGGAGTTGTTGAAGGA            AATCCAATAATGGAGATAGCTAAATACTTCCTTG            AATATCCTTTAACCATAAAAAGGCCAGAAAAATT            TGGTGGAGATTTGACAGTTAATAGCTATGAGGAG            TTAGAGAGTTTATTTAAAAATAAGGAATTGCATC            CAATGGATTTAAAAATGCTGTAGCTGAAGAACT            TATAAAGATTTTAGAGCCAATTAGAAAGAGATTA</p>

**REIVINDICACIONES**

1. Una composición que comprende una aminoacil-ARNt sintetasa ortogonal (O-RS) en la que la O-RS tiene al menos el 90 % de identidad de secuencias de aminoácidos con la longitud completa de la secuencia de SEQ ID NO: 1 y en la que dicha O-RS tiene una eficiencia que es al menos el 50 % de la eficiencia de traducción observada para la aminoacil-ARNt sintetasa que consiste en la secuencia de aminoácidos de SEQ ID NO: 1 en el sistema de traducción que comprende 3,4-dihidroxi-L-fenilalanina, el ARNt de SEQ ID NO: 2 y la aminoacil-ARNt sintetasa de SEQ ID NO: 1,
- 5 y en la que dicha O-RS aminoacila el O-ARNt con 3,4-dihidroxi-L-fenilalanina preferencialmente.
2. La composición de la reivindicación 1, en la que la O-RS comprende una secuencia de aminoácidos de SEQ ID NO: 1.
- 15 3. Una célula que comprende la composición de la reivindicación 1.
4. La célula de la reivindicación 3, en la que la célula es una célula de *E. coli*.
5. La composición de la reivindicación 1 que comprende un sistema de traducción.
- 20 6. La célula de la reivindicación 3, en la que dicha célula comprende:
- un ARNt ortogonal (O-ARNt) que está codificado por una secuencia de polinucleótidos que tiene al menos el 90 % de identidad de secuencias de nucleótidos con la secuencia de longitud completa de SEQ ID NO: 2;
- 25 dicha O-RS; y,
- un primer aminoácido no natural que es 3,4-dihidroxi-L-fenilalanina (DHP).
- 30 7. La célula de la reivindicación 6, en la que el O-ARNt está codificado por la secuencia de polinucleótidos que se expone en SEQ ID NO: 2.
8. La célula de la reivindicación 6, en la que la célula comprende además un segundo par de O-ARNt/O-RS y un segundo aminoácido no natural, en la que el segundo O-ARNt reconoce un segundo codón selector y la segunda O-RS aminoacila el segundo O-ARNt con el segundo aminoácido no natural preferencialmente.
- 35 9. La célula de la reivindicación 6, en la que la célula es una célula no eucariota.
10. La célula de la reivindicación 9, en la que la célula no eucariota es una célula de *E. coli*.
- 40 11. La célula de la reivindicación 6, en la que la célula comprende además un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en la que el polinucleótido comprende un codón selector que es reconocido por dicho O-ARNt.
- 45 12. Un polipéptido que comprende SEQ ID NO. 1.
13. Un polinucleótido que codifica un polipéptido de la reivindicación 12.
14. Un vector que comprende o que codifica un polinucleótido de la reivindicación 13.
- 50 15. El vector de la reivindicación 14, en el que el vector comprende un plásmido, un cósmido, un fago o un virus.
16. El vector de la reivindicación 14, en el que el vector es un vector de expresión.
- 55 17. Una célula que comprende el vector de la reivindicación 14.
18. Un procedimiento para identificar una aminoacil-ARNt sintetasa ortogonal que aminoacila específicamente un O-ARNt con 3,4-dihidroxi-L-fenilalanina (DHP), comprendiendo el procedimiento:
- 60 a) someter a selección una población de células de una primera especie, en la que las células comprenden cada una:
- i) un miembro de una pluralidad de aminoacil-ARNt sintetastas (RS);

ii) el ARNt ortogonal (O-ARNt);

iii) un polinucleótido que codifica un marcador de selección y comprende al menos un codón selector que es reconocido por dicho O-ARNt; y,

5

iv) dicha DHP;

en el que células en dicha población que están potenciadas en la eficiencia de supresión comprenden un miembro de la pluralidad de RS que es una RS activa que aminoacila específicamente el O-ARNt con dicha DHP; y,

10

b) seleccionar la RS activa que aminoacila el O-ARNt con la DHP, en el que dicha RS activa tiene una eficiencia que es al menos el 50 % de la eficiencia observada para la aminoacil-ARNt sintetasa que consiste en la secuencia de aminoácidos de SEQ ID NO: 1 en un sistema de traducción que comprende 3,4-dihidroxi-L-fenilalanina, un ARNt de SEQ ID NO: 2 y la aminoacil-ARNt sintetasa de SEQ ID NO: 1;

15

identificándose así la aminoacil-ARNt sintetasa ortogonal que aminoacila específicamente el O-ARNt.

19. El procedimiento de la reivindicación 18, en el que la selección comprende una selección positiva y el marcador de selección comprende un marcador de selección positiva.

20

20. El procedimiento de la reivindicación 18, en el que la pluralidad de RS comprende RS mutantes, o RS derivadas de una o más especies distintas de la primera especie o tanto RS mutantes como RS derivadas de una especie distinta de la primera especie.

25

21. Un procedimiento de producción de una proteína en una célula con 3,4-dihidroxi-L-fenilalanina (DHP) en una posición especificada, comprendiendo el procedimiento:

a) cultivar, en un medio apropiado, la célula de la reivindicación 11,

30

b) proporcionar la DHP; e

c) incorporar la DHP en la posición especificada en la proteína durante la traducción del ácido nucleico con el al menos un codón selector, produciéndose así la proteína con dicha DHP en la posición especificada.

35

22. El procedimiento de la reivindicación 21, en el que la O-RS comprende una secuencia de aminoácidos que comprende SEQ ID NO: 1.

23. El procedimiento de la reivindicación 21, en el que la célula es una célula no eucariota.

40

24. El procedimiento de la reivindicación 23, en el que la célula no eucariota es una célula de *E. coli*.

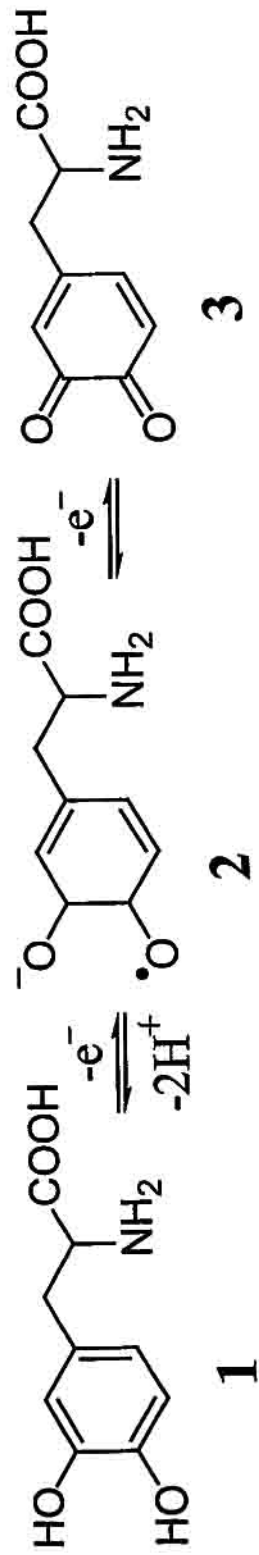


Fig. 1

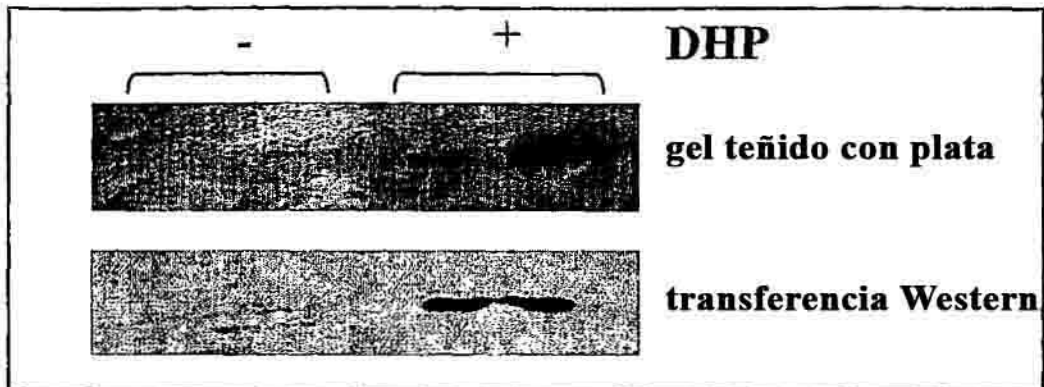


Fig. 2A

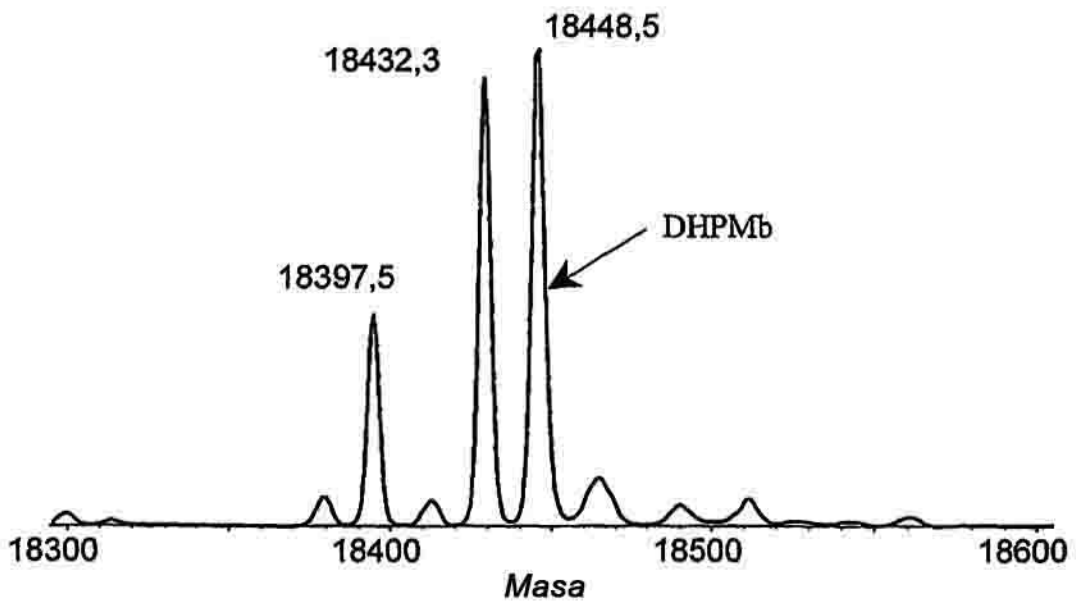


Fig. 2B

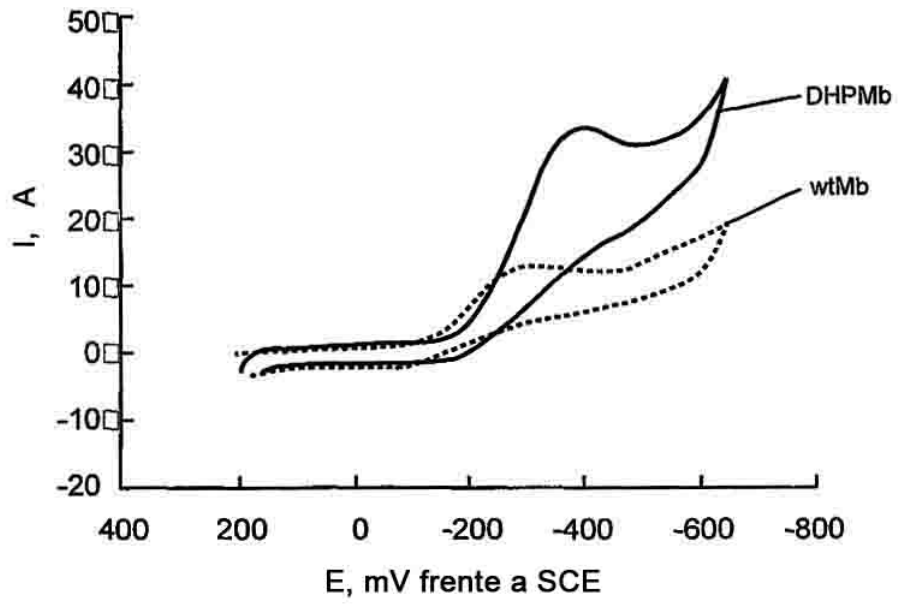


Fig. 3A

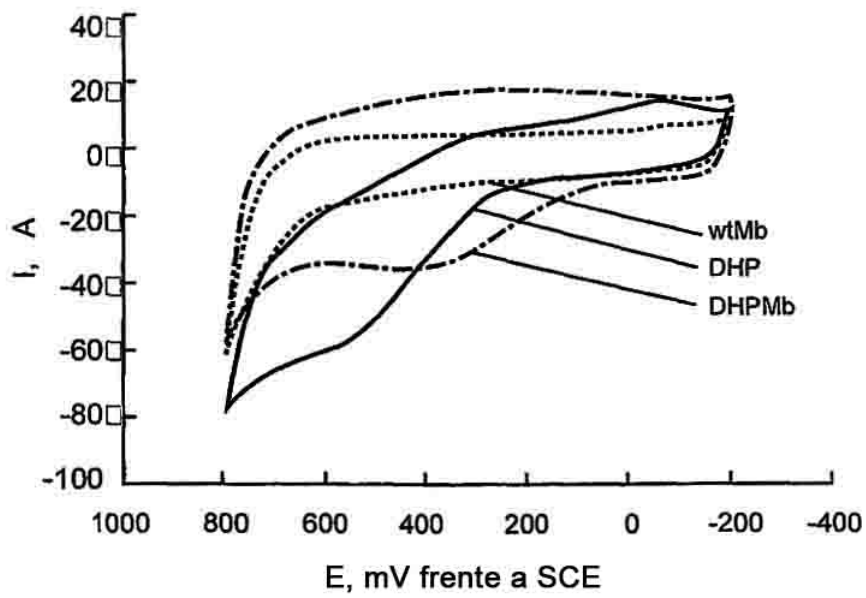


Fig. 3B

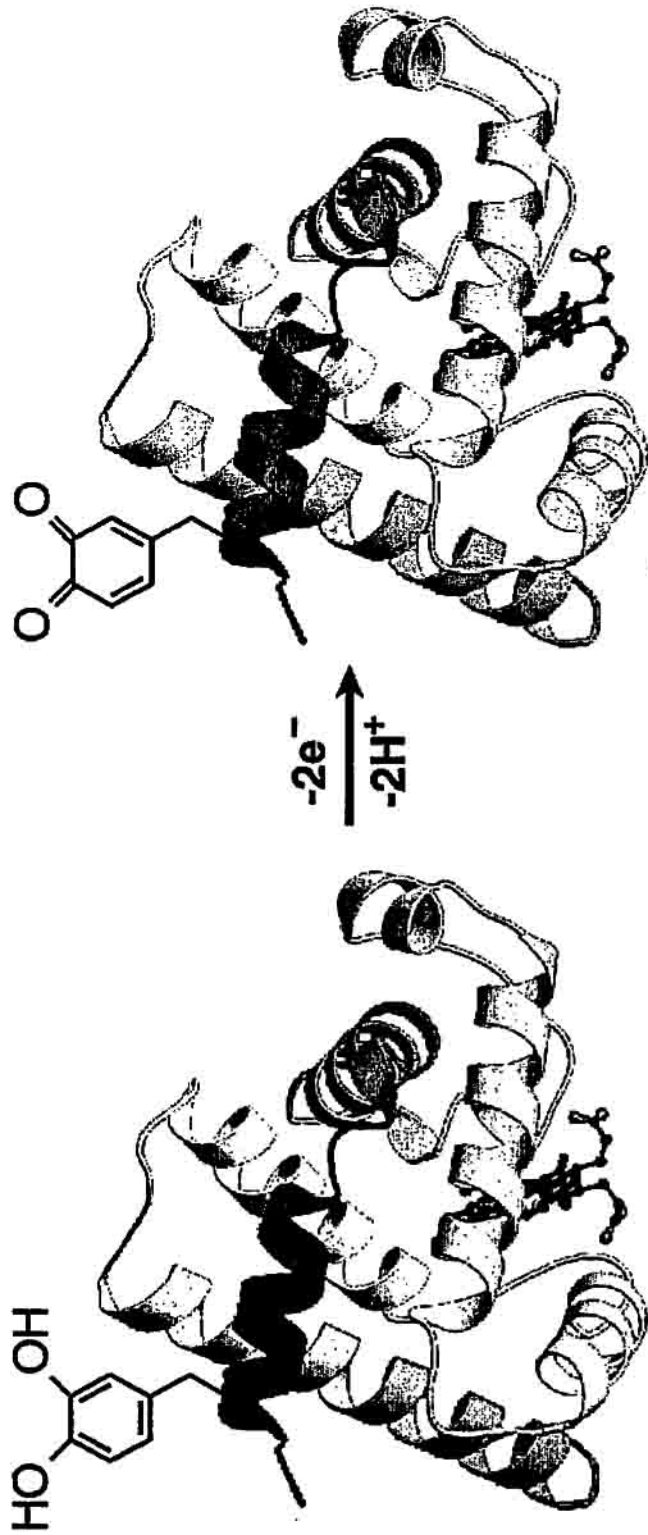


Fig. 4

1/1 31/11 91/31  
 ATG GAC GAA TTT GAA ATG ATA AAG AGA AAC ACA TCT GAA ATC ATC ATC GCT TAC ATG GGT TTT GAA CCA AGT GGT AAA  
 Met asp glu phe glu met ile lys arg asn thr ser glu ile ile ser glu glu glu leu arg glu val leu lys lys asp glu lys ser ala tyr ile gly phe glu pro ser gly lys  
 121/41 151/51 181/61 211/71  
 ATG CAT TTA GGG CAG TAT CTC CAA ATA AAA AAG ATG ATG GAT TTA CAA AAT GCT GCA TTT GAT ATG ATA TTT GAT TTA CAC GCC TAT TTA AAC CAG AAA GGA GAG TTG GAT  
 ile his leu gly his tyr leu gln ile lys lys met ile asp leu gln asn ala gly phe asp ile ile ile leu leu ala asp leu his ala tyr leu asn gln lys gly glu leu asp  
 241/81 271/91 301/101 331/111  
 GAG ATT AGA AAA ATA GGA GAT TAT AAC AAA AAA GTT TTT GAA GCA ATG GGG TTA AAG GCA AAA TAT GTT TAT GGA AGT GAA TTC CAG CTT GAT AAG GAT TAT ACA CTG AAT GTC TAT AGA  
 glu ile arg lys ile gly asp tyr asn lys lys val phe glu ala met gly leu lys ala lys tyr val tyr gly ser glu phe gln leu asp lys asp tyr thr leu asn val tyr arg  
 361/121 391/131 421/141 451/151  
 TTG GCT TTA AAA ACT ACC TTA AAA AGA GCA ATG AGG AGT ATG GAA CTT ATA GCA AGA GAG GAT GAA AAT CCA AAG GTT GCT GAA GTT ATC TAT CCA ATA ATG CAG GTT AAT GAT ATT CAT  
 leu ala leu lys thr thr leu lys arg ala arg arg ser met glu leu ile ala arg glu asp glu asn pro lys val ala glu val ile tyr pro ile met gln val asn asp ile his  
 481/161 511/171 541/181 571/191  
 TAT TTA GGC GTT GAT CTT CCA GTT GGA GGG ATG GAG CAG AGA AAA ATA CAC ATG TTA GCA AGG GAG CTT TTA CCA AAA AAG GTT GTT TGT ATT CAC AAC CCT GTC TTA ACG GGT TTC GAT  
 tyr leu gly val asp val ala val gly gly met glu met glu arg lys ile his met leu ala arg glu leu leu pro lys lys val val cys ile his asn pro val leu thr gly leu asp  
 601/201 631/211 661/221 691/231  
 GGA GAA GGA AAG ATG AGT TCT TCA AAA GGG AAT TTT ATA GCT GTT GAT GAC TCT CCA GAA GAG AAT AGG GCT AAG ATA AAG AAA GCA TAC TGC CCA GCT GGA GTT GTT GAA GGA AAT CCA  
 gly glu gly lys met ser ser lys gly asn phe ile ala val asp asp ser pro glu glu ile arg ala lys ile lys lys ala tyr cys pro ala gly val val glu gly asn pro  
 721/241 751/251 781/261 811/271  
 ATA ATG GAG ATA GCT AAA TAC TTC CTT GAA TAT CCT TTA ACC ATA AAA AGG CCA GAA AAA TTT GGT GGA GAT TTT GCA GAT TTT AAT AGC TAT GAG GAG TTA GAG AGT TTA TTT AAA AAT AAG  
 ile met glu ile ala lys tyr phe leu glu tyr pro leu thr ile lys arg pro glu lys phe gly gly asp leu thr val asn ser tyr glu glu leu ser leu phe lys asn lys  
 841/281 871/291 901/301  
 GAA TTG CAT CCA ATG GAT TTA AAA AAT GCT GTA GCT GAA GAA CTT ATA AAG ATT TTA GAG CCA ATT AGA AAG AGA TTA (SEQ ID NO:5)  
 glu leu his pro met asp leu lys asn ala val ala glu glu leu ile lys lys ile leu glu pro ile arg lys arg leu (SEQ ID NO:4)

Fig. 5



1/1 241/81 31/11 61/21 91/31  
 ATG GAC GAA TTT GAA ATG ATA AAG AGA AAC ACA TCT GAA ATT ATC AGC GAG GAA GAG TTA AGA GAG GTT TTA AAA AAA GAT GAA AAA TCT GCT CTC ATA GGT TTT GAA CCA AGT GGT AAA  
 Met asp glu phe glu met ile lys arg asn thr ser glu ile ile ser glu glu glu leu  
 121/41 151/51 181/61 211/71  
 ATA CNT TTA GGG CAF TAT CTC CAA ATA AAA ANG ATG ATT GAT TTA CAA AAT GCT GGA TTT GAT ATA ATT ATA TTG TTG AGC GAT TTA AAC GCC TAT TTA AAC CAG AAA GGA GAG TTG GAT  
 ile his leu gly his tyr leu leu gin ile lys lys met ile asp leu gin asn ala gly phe asp ile ile ile leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu  
 241/81 271/91 301/101 331/111  
 GAG ATT AGA AAA ATA GGA GAT TAT AAC AAA GAT TTT GAA GCA ATG GCG TTA AAG CCA AAA TAT GTT TAT GGA AGT GAA TTC CAG CTT GAT AAG GAT TAT ACA CTG AAT GTC TAT AGA  
 glu ile arg lys ile gly asp tyr asn lys lys val phe glu ala met gly leu lys ala lys lys tyr val tyr gly ser glu phe gin leu leu asp lys asp tyr thr leu asn val tyr arg  
 361/121 391/131 421/141 451/151  
 TTG GCT TTA AAA ACT ACC TTA AAA AGA SCA AGA AGG AGT ATG GRA CTT ATA GCA AGA GAG GAT GAA AAT CCA AAG GGT GCT GAA GTT ATC TAT CCA ATA ATG CAG GTT AAT GAT ATT CNT  
 leu ala leu lys thr thr leu lys arg ala arg arg ser met glu leu ile ala arg glu asp glu asn pro lys val ala glu val ile tyr pro ile met gin val asn asp ile his  
 481/161 511/171 541/181 571/191  
 TAT TTA GGC GTT GAT GTT CAG GTT GGA GGG ATG GAG CAG AGA AAA ATA CAC ATG TTA GCA AGG GAG CTT TTA CCA AAA AAG GGT GTT TGT ATT CAC AAC CCT GTC TTA ACG GGT TTG GAT  
 tyr leu gly val asp val gin val gly gly met glu gin arg lys ile his met leu ala arg glu leu leu pro lys lys val val cys ile his asn pro val leu thr gly leu asp  
 601/201 631/211 661/221  
 GGA GAA GGA ANG ATG TCT TCA AAA GGG AAT TTT ATA GCT GTT GAT GAC TCT CCA GAA GAG ATT AGG GCT AAG ATA AAG AAA GCA TAC TGC CCA GCT GGA GTT GTT GAA GGA AAT CCA  
 gly glu gly lys met ser ser ser lys gly asn phe ile ala val asp asp ser pro glu glu ile arg ala lys ile lys lys ala tyr cys pro ala gly val val glu gly asn pro  
 721/241 751/251 781/261 811/271  
 ATA ATG GAG ATA GCT AAA TAC TTC CTT GAA TAT CCT TTA ACC ATA AAA AGG CCA GAA AAA TTT GGT GGA GAT TTG ACA GTT AAT AGC TAT GAG GAG TTA GAG AGT TTA TTT AAA AAT AAG  
 ile met glu ile ala lys tyr phe leu glu tyr pro leu thr ile lys arg pro glu lys phe gly gly asp leu thr val asn ser tyr glu glu leu leu glu ser leu phe lys asn lys  
 841/281 871/291 901/301  
 GAA TTG CAF CCA ATG GAT TTA AAA AAT GCT GTA GCT GAA GAA CTT ATA AAG ATT TTA GAG CCA ATT AGA ANG AGA TTA (SEQ ID NO:3)  
 glu leu his pro met asp leu lys asn ala val ala glu leu ile lys ile leu glu pro ile arg lys arg leu leu leu (SEQ ID NO:1)

Fig. 6