

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 456 331**

51 Int. Cl.:

C12P 21/06 (2006.01)

C12N 15/74 (2006.01)

C12N 1/21 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.09.2007 E 07861356 (9)**

97 Fecha y número de publicación de la concesión europea: **08.01.2014 EP 2052083**

54 Título: **Expresión genéticamente programada de proteínas sulfatadas selectivamente en eubacterias**

30 Prioridad:

21.09.2006 US 846519 P

28.10.2006 US 855210 P

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

22.04.2014

73 Titular/es:

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

72 Inventor/es:

**LIU, CHANG C. y
SCHULTZ, PETER G.**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 456 331 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

Expresión genéticamente programada de proteínas sulfatadas selectivamente en eubacterias

5 **Campo de la invención**

La invención pertenece al campo de la bioquímica de la traducción. La invención se refiere a composiciones y métodos para fabricar y utilizar ARNt ortogonales, aminoacil-ARNt sintetizadas ortogonales, y pares de los mismos, que incorporan aminoácidos no naturales en proteínas. La invención también se refiere a métodos para producir
10 proteínas en células utilizando dichos pares y proteínas fabricados por los métodos.

Antecedentes de la invención

La sulfatación de la tirosina es una modificación postraducciona común en proteínas segregadas y unidas a
15 membranas (Kehoe y Bertozzi, "Tyrosine sulfation: a modulator of extracellular protein-protein interactions," *Chem Biol* 7: R57-61 (2000)). Aunque estamos solo comenzando a comprender el grado de su función biológica, la sulfotirosina ya se ha identificado en diversos paradigmas de interacción proteína-proteína. Por ejemplo, la sulfatación de la tirosina desempeña una función determinante en la unión de quimiocinas a los receptores de quimiocinas CCR2 (Preobrazhensky *et al.*, "Monocyte chemotactic protein-1 receptor CCR2B is a glycoprotein that has tyrosine sulfation in a conserved extracellular N-terminal region" *J Immunol* 165: 5295-5303 (2000)), CCR5 (Farzan *et al.*, "Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry" *Cell* 96: 667-676 (1999)), CXCR4 (Farzan *et al.*, "The role of post-translational modifications of the CXCR4 amino terminus in stromal-derived factor 1 alpha association and HIV-1 entry," *J Biol Chem* 277: 29484-29489 (2002); Veldkamp *et al.*, "Recognition of a CXCR4 sulfotyrosine by the chemokine stromal cell-derived factor1alpha (SDF-1alpha/CXCL12)," *J Mol Biol* 359: 1400-1409 (2006)) y CX₃CR1 (Fong *et al.*, "CX₃CR1 tyrosine sulfation enhances fractalkine-induced cell adhesion," *J Biol Chem* 277: 19418-19423 (2002)). De manera similar, leucocitos rodantes bajo tensión cortante hidrodinámica requieren sulfatación de PSGL-1 para la correcta unión y adhesión (Somers *et al.*, "Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLe(X) and PSGL-1," *Cell* 103: 467-479 (2000)). La sulfatación de la tirosina está también implicada en la cascada de coagulación, habiéndose
20 identificado en diversos factores de coagulación así como en inhibidores de trombina naturales tales como la hirudina, un anticoagulante segregado por sanguijuelas (Dong *et al.*, "Tyrosine sulfation of the glycoprotein Ib-IX complex: identification of sulfated residues and effect on ligand binding," *Biochemistry* 33: 13946-13953 (1994); Bagdy *et al.*, "Hirudin," *Methods Enzymol* 45: 669-678 (1976)). Además, recientemente se descubrió que la sulfatación de la tirosina en una región bucle variable de anticuerpo era responsable de la actividad neutralizante de un subconjunto de anticuerpos del VHI-1 inducidos por CD4, demostrando de este modo la capacidad de la sulfotirosina para aumentar la afinidad antígeno anticuerpo (Choe *et al.*, "Tyrosine sulfation of human antibodies contributes to recognition of the CCR5 binding region of HIV-1 gp120," *Cell* 114: 161-170 (2003); Xiang *et al.*, "Functional mimicry of a human immunodeficiency virus type 1 coreceptor by a neutralizing monoclonal antibody," *J Virol* 79: 6068-6077 (2005)).
25

Un obstáculo principal para la determinación de las funciones de la sulfatación en más de las 60 proteínas conocidas y más de las 2100 proteínas predichas que contienen sulfotirosina (según un estudio de secuencias de proteínas de ratón) es la capacidad de sintetizar selectivamente proteínas sulfatadas (Moore, "The biology and enzymology of protein tyrosine O-sulfation," *J Biol Chem* 278: 24243-24246 (2003)). Los métodos actuales se basan en la síntesis
30 de péptidos convencional o en sulfatación enzimática *in vitro* (Veldkamp *et al.*, "Recognition of a CXCR4 sulfotyrosine by the chemokine stromal cell-derived factor-1alpha (SDF-1alpha/CXCL12)," *J Mol Biol* 359: 1400-1409 (2006); Kirano *et al.*, "Total synthesis of porcine cholecystokinin-33 (CCK-33)," *J. Chem. Soc., Chem. Commun.*, 323-325 (1987); Muramatsu *et al.*, "Enzymic O-sulfation of tyrosine residues in hirudins by sulfotransferase from *Eubacterium A-44*," *Eur J Biochem* 223: 243-248 (1994); Young y Kiessling, "A strategy for the synthesis of sulfated peptides," *Angew Chem Int Ed Engl* 41: 3449-3451 (2002)); sin embargo, ambos carecen de generalidades: el primero está limitado por restricciones de longitud y la tendencia hacia la desulfatación de sulfotirosina en condiciones ácidas; el último está limitado por la disponibilidad de sulfotransferasas accesorias y sus restricciones de secuencias de reconocimiento asociadas.
35

La incorporación directa de un aminoácido no natural de sulfotirosina codificado genéticamente en sitios definidos en proteínas directamente en organismos vivos superaría las limitaciones descritas anteriormente. La incorporación directa de sulfotirosina facilitaría enormemente el estudio de acontecimientos de sulfatación en la regulación de procesos biológicos y también permitiría la creación de bibliotecas de anticuerpos y péptidos sulfatados de diversidad significativa. Adicionalmente, la capacidad para producir una forma sulfatada en la proteína hirudina tiene
40 aplicación clínica inmediata para su uso como un anticoagulante mejorado (mejorado con respecto a la forma no sulfatada). Lo que se necesita en la técnica son nuevas estrategias para incorporar en proteínas el aminoácido no natural sulfotirosina.

Se ha desarrollado una metodología general para la incorporación *in vivo* específica de sitio de diversos aminoácidos no naturales en proteínas, en organismos tanto eucariotas como procariotas. Estos métodos se basan en componentes de traducción de proteínas ortogonales que reconocen un codón selector adecuado para insertar
45

un aminoácido no natural deseado en una posición definida durante la traducción de polipéptidos *in vivo*. Estos métodos utilizan un ARNt ortogonal (O-ARNt) que reconoce un codón selector, y en el que una aminoacil-ARNt sintetasa ortogonal (una O-RS) específica correspondiente carga el O-ARNt con el aminoácido no natural. Estos componentes no reaccionan en cruzado con ninguno de los ARNt, RSs, aminoácidos o codones endógenos en el organismo hospedador (es decir, debe ser ortogonal). El uso de dichos pares ARNt-RS ortogonales ha hecho que sea posible codificar genéticamente una gran cantidad de aminoácidos no naturales estructuralmente diversos.

La práctica de utilizar sistemas de traducción ortogonal que son adecuados para la fabricación de proteínas que comprende uno o más aminoácidos no naturales es generalmente conocida en la técnica, como lo son los métodos generales para producir sistemas de traducción ortogonal. Por ejemplo, véanse los Números de Publicación Internacional WO 2002/086075, titulada "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS"; WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE"; WO 2005/019415, presentada el 7 de julio de 2004; WO 2005/007870, presentada el 7 de julio del 2004; WO 2005/007624, presentada el 7 de julio de 2004 y WO 2006/110182, presentado el 27 de octubre de 2005, titulada "ORTHOGONAL TRANSLATION COMPONENTS FOR THE VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Para un análisis adicional de sistemas de traducción ortogonal que incorporan aminoácidos no naturales, y métodos para su producción y uso, véase también, Wang y Schultz, "Expanding the Genetic Code," Chem. Commun. (Camb.) 1: 1-11 (2002); Wang y Schultz "Expanding the Genetic Code," Angewandte Chemie Int. Ed., 44(1): 34-66 (2005); Xie y Schultz, "An Expanding Genetic Code," Methods 36(3): 227-238 (2005); Xie y Schultz, "Adding Amino Acids to the Genetic Repertoire," Curr. Opin. in Chemical Biology 9(6): 548-554 (2005); Wang *et al.*, "Expanding the Genetic Code," Annu. Rev. Biophys. Biomol. Struct., 35: 225-249 (2006; epub 13 de junio de 2006); y Xie y Schultz, "A chemical toolkit for proteins -an expanded genetic code," Nat. Rev. Mol. Cell Biol., 7(10): 775-782 (2006; publicación electrónica 23 de agosto de 2006). El documento WO 2004/035743 describe pares de ARNt/aminoacil ARNt sintetasa ortogonales que incorporan cetos aminoácidos en proteínas.

Existe necesidad en la técnica de desarrollar componentes de traducción ortogonal que incorporen el aminoácido no natural sulfotirosina en proteínas, en el que el aminoácido no natural puede incorporarse en cualquier posición definida. La invención descrita en el presente documento cumple estas y otras necesidades, como se apreciará después de la revisión de la siguiente divulgación.

Sumario de la invención

Aunque la sulfatación de la tirosina es una modificación postraduccional generalizada en eucariotas multicelulares (Moore, "The biology and enzymology of protein tyrosine O-sulfation," J Biol Chem 278: 24243-24246 (2003)), sus funciones biológicas siguen sin conocerse en gran medida. Esto se debe en parte a las dificultades asociadas con la síntesis de proteínas sulfatadas selectivamente. La invención proporciona la incorporación selectiva de sulfotirosina en proteínas en bacterias por codificación genética del aminoácido modificado en respuesta al codón sin sentido ámbar, TAG. Además, se demuestra que la sulfo-hirudina, previamente inaccesible a través de métodos recombinantes, puede expresarse directamente en *E. coli* utilizando esta estrategia. Como se describe en el presente documento, análisis cinéticos muestran una mejora en afinidad mayor de 10 veces frente a la trombina humana por sulfohirudina sobre la desulfohirudina, una observación que ofrece ventajas clínicas para la sulfohirudina en su uso como un anticoagulante (Di Nisio *et al.*, "Direct thrombin inhibitors," N Engl J Med 353: 1028-1040 (2005)). Esta estrategia general para la biosíntesis de proteínas sulfatadas facilita el estudio y la aplicación adicional de la modificación postraduccional emergente, la sulfatación de la tirosina.

La invención proporciona composiciones y métodos, como se define en las reivindicaciones, para la incorporación *in vivo* (por ejemplo, en una célula hospedadora) del aminoácido no natural sulfotirosina en una cadena polipeptídica en crecimiento en respuesta a un codón selector, por ejemplo, un codón de terminación ámbar. Estas composiciones incluyen pares de ARNt ortogonales (O-ARNt) y aminoacil-ARNt sintetasa ortogonales (O-RSs) que no interaccionan con la maquinaria de traducción de la célula hospedadora. Es decir, el O-ARNt no está cargado (o no está cargado a un nivel significativo) con un aminoácido (natural o no natural) por una aminoacil-ARNt sintetasa de la célula hospedadora endógena. De manera similar, la O-RSs proporcionada por la invención no carga ningún ARNt endógeno con un aminoácido (natural o no natural) a un nivel significativo o detectable. Estas nuevas composiciones permiten la producción de grandes cantidades de proteínas que tienen sulfotirosina incorporada traduccionalmente.

En un primer aspecto, la invención proporciona sistemas de traducción como se define en las reivindicaciones. Los sistemas de traducción comprenden: (a) un primer aminoácido no natural que es la sulfotirosina como se representa en la Figura 1; (b) una primera aminoacil ARNt sintetasa ortogonal (O-RS) que comprende una secuencia de aminoácidos que es una variante conservativa de SEC ID N°: 2, cuya variante conservativa es el menos el 90 % idéntica a la SEC ID N°: 2 y comprende: una leucina en una posición correspondiente a Tyr32; una prolina en una posición correspondiente a Leu65; una glicina en una posición correspondiente a Asp158, y una lisina a una posición correspondiente a Leu162, y; opcionalmente comprende un ácido glutámico en una posición correspondiente a Gln155, y/o una treonina o una cisteína en una posición correspondiente a Ile159, y; (c) un primer ARNt ortogonal (O-ARNt); en el que la primera O-RS aminoacila el primer O-ARNt con la sulfotirosina.

- 5 En determinadas realizaciones, la primera O-RS deriva de una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*, y/o deriva de una tirosil ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre y/o comprende una secuencia de aminoácidos expuesta en las SEC ID Nos: 4, 6, 8 o 10. El primer O-ARNt puede ser un ARNt supresor de ámbar y/o puede comprender o estar codificado por una secuencia de polinucleótidos expuesta en la SEC ID N^o: 1.
- 10 En determinadas realizaciones, el sistema de traducción comprende adicionalmente un ácido nucleico que codifica una proteína de interés. El ácido nucleico comprende al menos un codón selector que reconoce el primer O-ARNt. El sistema de traducción puede comprender adicionalmente una segunda O-RS y un segundo O-ARNt, en el que la segunda O-RS aminoacila el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y en el que el segundo O-ARNt reconoce un codón selector que es diferente del codón selector reconocido por el primer O-ARNt.
- 15 En determinadas realizaciones, el sistema de traducción comprende una célula hospedadora que comprende el primer aminoácido no natural, la primera O-RS y el primer O-ARNt. La célula hospedadora puede ser una célula de eubacteria, tal como una célula de *E. coli*. La célula hospedadora puede comprender un polinucleótido que codifica la primera O-RS y/o un polinucleótido que codifica el primer O-ARNt. El polinucleótido que codifica la primera O-RS puede comprender una secuencia de nucleótidos expuesta en las SEC ID Nos: 5, 7, 9 u 11.
- 20 En un segundo aspecto, la invención proporciona métodos para producir, en un sistema de traducción, una proteína que comprende un aminoácido no natural en una posición seleccionada como se define en las reivindicaciones. El método comprende: (a) proporcionar un sistema de traducción que comprende: (i) un primer aminoácido no natural que es la sulfotirosina como se representa en la Fig. 1; (ii) una primera O-RS que comprende la secuencia de aminoácidos que es una variante conservativa de la SEC ID N^o: 2, cuya variante conservativa es al menos el 90 % idéntica a la SEC ID N^o: 2 y comprende: una leucina en una posición correspondiente a Tyr32; una prolina en una posición correspondiente a Leu65; una glicina en una posición correspondiente a Asp158, y una lisina en una posición correspondiente a Leu162, y; opcionalmente comprende un ácido glutámico en una posición correspondiente a Gln155, y/o una treonina o una cisteína en una posición correspondiente a Ile159; (iii) un primer O-ARNt, en el que la primera O-RS aminoacila el primer O-ARNt con la sulfotirosina; y, (iv) un ácido nucleico que codifica la proteína, en el que el ácido nucleico comprende al menos un codón selector reconocido por el primer O-ARNt, y; (b) incorporar el aminoácido no natural en la posición seleccionada en la proteína durante la traducción de la proteína en respuesta al codón selector, produciendo de este modo la proteína que comprende el aminoácido no natural en la posición seleccionada.
- 30 En determinadas realizaciones de los métodos, la proteína que comprende un aminoácido no natural es la sulfohirudina.
- 35 En determinadas realizaciones de los métodos, el proporcionar un sistema de traducción puede comprender: (A) proporcionar un polinucleótido que codifique la O-RS o (B) proporcionar una O-RS derivada de una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*, o; (C) proporcionar una O-RS derivada de una tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre, o; (D) proporcionar una O-RS que comprenda una secuencia de aminoácidos expuesta en las SEC ID Nos: 4, 6, 8 o 10, o; (E) mutar un bolsillo de unión de aminoácidos de una aminoacil-ARNt sintetasa de tipo silvestre por mutagénesis de sitio dirigido y seleccionar una O-RS resultante que aminoacile dicho O-ARNt con dicho aminoácido no natural, en el que la etapa de selección comprende seleccionar positivamente y seleccionar negativamente la O-RS a partir de un conjunto que comprenda una pluralidad de moléculas de aminoacil-ARNt sintetasa resultantes después de la mutagénesis de sitio dirigido, o; (F) proporcionar un polinucleótido que codifique dicho O-ARNt, o; (G) proporcionar un O-ARNt que es un ARNt supresor de ámbar, o, (H) proporcionar un O-ARNt que comprenda o esté codificado por una secuencia de polinucleótidos expuesta en la SEC ID N^o: 1, o; (I) proporcionar un ácido nucleico que comprenda un codón selector ámbar, o; (J) proporcionar una célula hospedadora que comprenda el primer aminoácido no natural, la primera O-RS, el primer O-ARNt y el ácido nucleico, y la etapa de incorporación comprende cultivar la célula hospedadora, o; (K) proporcionar un extracto celular.
- 40 En determinadas realizaciones de (J), el proporcionar una célula hospedadora comprende proporcionar una célula hospedadora de eubacteria, que puede ser una célula hospedadora de *E. coli*, o proporcionar una célula hospedadora que comprenda un polinucleótido que codifique la O-RS, que puede comprender una secuencia de nucleótidos expuesta en las SEC ID Nos: 5, 7, 9 u 11.
- 50 En determinadas realizaciones de los métodos, la proteína comprende un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y el sistema de traducción comprende adicionalmente una segunda O-RS y un segundo O-ARNt. La segunda O-RS aminoacila el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y el segundo O-ARNt reconoce un codón selector en el ácido nucleico que es diferente del codón selector reconocido por el primer O-ARNt.
- 60 En un tercer aspecto, la invención proporciona composiciones que comprenden un polipéptido, que comprende una secuencia de aminoácidos que es una variante conservativa de la SEC ID N^o: 2, cuya variante conservativa es al
- 65

5 menos el 90 % idéntica a la SEC ID N°:2, y comprende: una leucina en una posición correspondiente a Tyr32; una prolina en una posición correspondiente a Leu65; una glicina en una posición correspondiente a Asp158 y una lisina en una posición correspondiente a Leu162, y; opcionalmente comprende un ácido glutámico en una posición correspondiente a Gln155, y/o una treonina o una cisteína en una posición correspondiente a Ile159; cuyo polipéptido aminoacila un O-ARNt con sulfotirosina como de representa en la Fig. 1.

10 En determinadas realizaciones de las composiciones, el polipéptido comprende la secuencia de aminoácidos de las SEC ID Nos: 4, 6, 8 o 10. En determinadas realizaciones, las composiciones comprenden una célula que comprende el polipéptido.

15 En un aspecto relacionado, la invención proporciona polinucleótidos que codifican los polipéptidos de las composiciones. En determinadas realizaciones, los polinucleótidos comprenden la secuencia de nucleótidos las SEC ID Nos: 5, 7, 9 u 11.

20 En un aspecto adicional relacionado, la invención proporciona vectores que comprenden el polinucleótido que codifica el polipéptido de las composiciones.

En un aspecto adicional relacionado, la invención proporciona células que comprenden los vectores.

25 En algunos aspectos, el sistema de traducción incorpora un segundo par ortogonal (es decir, una segunda O-RS y un segundo O-ARNt) que utiliza un segundo aminoácido no natural, de tal manera que el sistema puede ahora incorporar al menos dos aminoácidos no naturales en diferentes en sitios seleccionados en un polipéptido. En este sistema doble, la segunda O-RS aminoacila preferentemente el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y el segundo O-ARNt reconoce un codón selector que es diferente del codón selector reconocido por el primer O-ARNt.

30 En algunas realizaciones, el sistema de traducción reside en una célula hospedadora (e incluye la célula hospedadora). La célula hospedadora utilizada no está particularmente limitada, siempre que la O-RS y el O-ARNt conserven su ortogonalidad en su medio celular hospedador. La célula hospedadora puede ser una célula de eubacteria, tal como *E. coli*. La célula hospedadora puede comprender uno o más polinucleótidos que codifican componentes del sistema de traducción, incluyendo la O-RS o el O-ARNt. En algunas realizaciones, el polinucleótido que codifica la O-RS comprende una secuencia de nucleótidos de las SEC ID Nos: 5, 7, 9 u 11.

35 La invención también proporciona métodos para producir proteínas que tienen uno o más aminoácidos no naturales en posiciones seleccionadas. Estos métodos utilizan los sistemas de traducción descritos anteriormente. Generalmente, estos métodos comienzan con la etapa de proporcionar un sistema de traducción que comprende: (i) un primer aminoácido no natural que es el aminoácido no natural sulfotirosina; (ii) una primera aminoacil-ARNt sintetasa ortogonal (O-RS); (iii) un primer ARNt ortogonal (O-ARNt), en el que la O-RS aminoacila preferentemente el O-ARNt con el aminoácido no natural; y, (iv) un ácido nucleico que codifica la proteína, en el que el ácido nucleico comprende al menos un codón selector reconocido por el primer O-ARNt. Por tanto el método incorpora el aminoácido no natural en la posición seleccionada en la proteína durante la traducción de la proteína en respuesta al codón selector, produciendo de este modo la proteína que comprende el aminoácido no natural en la posición seleccionada. En algunos aspectos de estos métodos, la O-RS aminoacila preferentemente el O-ARNt con la sulfotirosina con una eficacia que es al menos el 50 % de la eficacia observada para un sistema de traducción que comprende el mismo O-ARNt, la sulfotirosina, y una aminoacil-ARNt sintetasa que comprende la secuencia de aminoácidos de las SEC ID Nos: 4, 6, 8 o 10. En algunos aspectos, los métodos se usan para producir la forma sulfatada de la hirudina.

45 Estos métodos pueden aplicarse ampliamente utilizando diversos reactivos y etapas. En algunas realizaciones, se proporciona un polinucleótido que codifica la O-RS. En algunas realizaciones, se proporciona una O-RS derivada de una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*, por ejemplo, puede proporcionarse una tirosil ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre. En algunas realizaciones, la etapa de proporcionar incluye proporcionar una O-RS que comprenda una secuencia de aminoácidos de las SEC ID Nos: 4, 6, 8 o 10, y variantes conservativas de las mismas.

55 En algunas realizaciones de estos métodos, el proporcionar una etapa del sistema de traducción comprende mutar un bolsillo de unión de aminoácido de una aminoacil-ARNt sintetasa de tipo silvestre mediante mutagénesis de sitio dirigido, y seleccionar una O-RS resultante que aminoacile preferentemente el O-ARNt con el aminoácido no natural. La etapa de selección puede comprender seleccionar de manera positiva y negativa la O-RS de un conjunto de moléculas de aminoacil-ARNt sintetasa resultante después de la mutagénesis de sitio dirigido. En algunas realizaciones, la etapa de proporcionar suministra un polinucleótido que codifica el O-ARNt, por ejemplo, un O-ARNt que es un ARNt supresor ámbar, o un O-ARNt que comprende, o está codificado por, un polinucleótido de la SEC ID N°: 1. En estos métodos, la etapa de proporcionar también puede suministrar un ácido nucleico que comprende un codón selector ámbar que es utilizado por el sistema de traducción.

65

Estos métodos también pueden modificarse para incorporar más de un aminoácido no natural en una proteína. En estos métodos, junto con el primer sistema de traducción se emplea un segundo sistema de traducción ortogonal, en el que el segundo sistema tiene diferentes especificidades de codones selectores y aminoácidos. Por ejemplo, la etapa de proporcionar pueden incluir proporcionar una segunda O-RS y un segundo O-ARNt, en la que la segunda O-RS aminoacila preferentemente el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y en la que el segundo O-ARNt reconoce un codón selector en el ácido nucleico que es diferente del codón selector reconocido por el primer O-ARNt.

Los métodos para producir una proteína con un aminoácido no natural también pueden realizarse en el contexto de una célula hospedadora. En estos casos, se proporciona una célula hospedadora, en la que la célula hospedadora comprende el aminoácido no natural, la O-RS, el O-ARNt y el ácido nucleico con al menos un codón selector que codifica la proteína, y en la que el cultivo de la célula hospedadora da como resultado la incorporación del aminoácido no natural. En algunas realizaciones, la etapa de proporcionar comprende proporcionar una célula hospedadora de eubacteria (por ejemplo, *E. coli*). En algunas realizaciones, la etapa de proporcionar incluye proporcionar una célula hospedadora que contiene un polinucleótido que codifica la O-RS. Por ejemplo, el polinucleótido que codifica la O-RS puede comprender un nucleótido.

Definiciones

Antes de describir la invención con 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 utilizada en el presente documento tiene la única finalidad de describir realizaciones particulares y no pretende ser limitante. Como se usa en esta memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un", "una", "el" y "la" incluyen los plurales, a menos que el contexto indique claramente lo contrario. Por tanto, la referencia, por ejemplo, a "una célula" incluye combinaciones de dos o más células; la referencia a "un polinucleótido" incluye, como cuestión práctica, muchas copias de ese polinucleótido.

A menos que se defina en el presente documento y más adelante en el resto de la memoria descriptiva, todos los términos científicos y técnicos utilizados en el presente documento tienen el mismo significado que el normalmente entendido por un experto habitual en la técnica a la cual pertenece la invención.

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 menor eficacia en comparación con una molécula correspondiente que es endógena a la célula o al sistema de traducción, o que fracasa al funcionar con componentes endógenos de la célula. En el contexto de los ARNt y las aminoacil-ARNt sintetetasas, ortogonal se refiere a la incapacidad o menor eficacia, por ejemplo, una eficacia reducida de un 20 %, menor de un 10 %, menor de un 5 %, o menor de un 1 %, de un ARNt ortogonal para funcionar con una ARNt sintetasa endógena en comparación con la capacidad de un ARNt endógeno para funcionar con la ARNt sintetasa endógena, o de una aminoacil ARNt sintetasa ortogonal para funcionar con un ARNt endógeno en comparación con la capacidad de una ARNt sintetasa endógena para funcionar con el ARNt endógeno. La molécula ortogonal carece de una molécula complementaria endógena funcionalmente normal en la célula. Por ejemplo, un ARNt ortogonal en una célula se aminoacila mediante cualquier RS endógena de la célula con una eficacia reducida o incluso nula en comparación con la aminoacilación de un ARNt endógeno por la RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno en una célula de interés con una eficacia reducida o incluso nula, en comparación con la aminoacilación del ARNt endógeno mediante una RS endógena. En la célula puede introducirse una segunda molécula ortogonal que funcione con la primera molécula ortogonal. Por ejemplo, un par de ARNt/RS ortogonal incluye componentes complementarios introducidos que funcionan juntos en la célula con una eficacia (por ejemplo, una eficacia del 45 %, 50 %, 60 %, 70 %, 75 %, 80 %, 90 %, 95 % o 99 % o mayor) en comparación con la de un control, por ejemplo, un par endógeno de ARNt/RS correspondiente, o un par ortogonal activo (por ejemplo, un par de tirosil ARNt/RS ortogonal).

Tirosil-ARNt ortogonal: como se usa en el presente documento, un tirosil-ARNt ortogonal (tirosil-O-ARNt) es un ARNt que es ortogonal a un sistema de traducción de interés, en el que el ARNt es: (1) idéntico o sustancialmente similar a un tirosil ARNt de origen natural, (2) derivado de un tirosil ARNt de origen natural por mutagénesis natural o artificial, (3) derivado por cualquier proceso que tiene en cuenta una secuencia de tirosil-ARNt de tipo silvestre o mutante de (1) o (2), (4) homólogo a un tirosil-ARNt de tipo silvestre o natural; (5) homólogo a cualquier ARNt de ejemplo que está diseñado como un sustrato para una tirosil ARNt sintetasa en la FIG. 7 o (6) una variante conservativa de cualquier ARNt de ejemplo que está diseñado como un sustrato para una tirosil-ARNt sintetasa en la FIG. 7. El tirosil-ARNt puede existir cargado con un aminoácido, o en un estado no cargado. También debe entenderse que un "tirosil-O-ARNt" está opcionalmente cargado (aminoacilado) por una sintetasa afín con un aminoácido distinto de tirosina, respectivamente, por ejemplo, con un aminoácido no natural. De hecho, se apreciará que un tirosil-O-ARNt de la presente descripción se usa ventajosamente para insertar esencialmente cualquier aminoácido, tanto si es natural como no natural, en el polipéptido en crecimiento, durante la traducción, en respuesta a un codón selector.

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

sistema de traducción de interés. El aminoácido que la tirosil-O-RS carga sobre tirosil-O-ARNt puede ser cualquier aminoácido, tanto si es natural como no natural o artificial y en el presente documento no está limitado. La sintetasa es opcionalmente la misma que u homóloga a una tirosil aminoácido sintetasa de origen natural, o la misma u homóloga a una sintetasa diseñada como O-RS en la **FIG. 7**. Por ejemplo, la O-RS puede ser una variante conservativa de una tirosil-O-RS de la **FIG. 7**, y/o puede tener una identidad de al menos 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o más con una O-RS de la **FIG. 7**.

Afín: el término “afín” se refiere a componentes que funcionan juntos, o que tienen algún aspecto de especificidad entre sí, es decir, un ARNt ortogonal y una aminoacil-ARNt sintetasa ortogonal. Los componentes también pueden denominarse complementarios.

Aminoacila preferentemente: como se usa en el presente documento en referencia a sistemas de traducción ortogonal, una O-RS “aminoacila preferentemente” un O-ARNt afín cuando la O-RS carga un O-ARNt con un aminoácido de forma más eficaz de lo que carga cualquier ARNt endógeno en un sistema de expresión. Es decir, cuando el O-ARNt y cualquier ARNt endógeno determinado están presente en un sistema de traducción aproximadamente en las mismas proporciones molares, la O-RS cargará el O-ARNt más frecuentemente de lo que lo cargará el ARNt endógeno. Preferentemente, la proporción relativa del O-ARNt cargado por la O-RS con respecto al endógeno ARNt cargado por la O-RS es alta, dando como resultado preferentemente la O-RS que carga el O-ARNt exclusivamente, o casi exclusivamente, cuando el O-ARNt y un ARNt endógeno están presentes en las mismas concentraciones molares en el sistema de traducción. La proporción relativa entre el O-ARNt y el ARNt endógeno que está cargado por la O-RS, cuando el O-ARNt y la O-RS están presentes a las mismas concentraciones molares, es mayor de 1:1, preferentemente al menos aproximadamente 2:1, más preferentemente 5:1, aún más preferentemente 10:1, incluso más preferentemente 20:1, aún más preferentemente 50:1, aún más preferentemente 75:1, aún más preferentemente 95:1, 98:1, 99:1, 100:1, 500:1, 1,000:1, 5,000:1 o mayor.

La O-RS “aminoacila preferentemente un O-ARNt con un aminoácido natural” cuando (a) la O-RS aminoacila preferentemente el O-ARNt en comparación con un ARNt endógeno y (b) cuando la aminoacilación es específica para el aminoácido no natural, en comparación con la aminoacilación del O-ARNt por la O-RS con un aminoácido natural. Es decir, cuando los aminoácidos no naturales y naturales están presentes en las mismas cantidades molares en un sistema de traducción que comprende la O-RS y un O-ARNt, la O-RS cargará el O-ARNt con el aminoácido no natural más frecuentemente que con el aminoácido natural. Preferentemente, la proporción relativa de O-ARNt cargado con el aminoácido no natural con respecto al O-ARNt cargado con el aminoácido natural es alta. Más preferentemente, la O-RS carga el O-ARNt exclusivamente, o casi exclusivamente, con el aminoácido no natural. La proporción relativa entre la carga del O-ARNt con el aminoácido no natural y la carga del O-ARNt con el aminoácido natural, cuando ambos aminoácidos, natural y no natural, están presentes en el sistema de traducción en las mismas concentraciones molares, es mayor que 1:1, preferentemente al menos aproximadamente 2:1, más preferentemente 5:1, aún más preferentemente 10:1, aún más preferentemente 20:1, aún más preferentemente 50:1, aún más preferentemente 75:1, aún más preferentemente 95:1, 98:1, 99:1, 100:1, 500:1, 1,000:1, 5,000:1 o más alta.

Codón selector: la expresión “codón selector” se refiere a codones reconocidos por el O-ARNt en el proceso de traducción y no reconocido por un ARNt endógeno. El bucle anticodón de O-ARNt reconoce el codón selector en el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido no natural, en este sitio en el polipéptido. Los codones selectores pueden incluir, por ejemplo, codones sin sentido, 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.

ARNt supresor: un ARNt supresor es un ARNt que altera la lectura de un ARN mensajero (ARNm) en un sistema de traducción determinado, normalmente permitiendo la incorporación de un aminoácido en respuesta a un codón de terminación (es decir, “lectura completa”) durante la traducción de un polipéptido. En algunos aspectos, un codón selector de la invención es un codón supresor, por ejemplo, un codón de terminación (por ejemplo, un codón ámbar, ocre u ópalo), un codón de cuatro bases, un codón raro, etc.

Actividad de supresión: como se usa en el presente documento, la expresión “actividad de supresión” se refiere, en general, a la capacidad de un ARNt (por ejemplo, un ARNt supresor) de permitir la lectura completa de traducción de un codón (por ejemplo, un codón selector que es un codón ámbar o un codón de cuatro o más bases) que de otra manera daría como resultado la terminación de la traducción o la traducción errónea (por ejemplo un desplazamiento de fase de lectura). La actividad de supresión de un ARNt supresor puede expresarse como un porcentaje de la actividad de lectura completa de traducción observada en comparación con un segundo ARNt supresor o en comparación con un sistema de control, por ejemplo, un sistema de control que carece de una O-RS.

La presente descripción proporciona diversos métodos mediante los cuales la actividad de supresión puede cuantificarse. El porcentaje de supresión de un O-ARNt particular y una O-RS contra un codón selector (por ejemplo, un codón ámbar) de interés se refiere al porcentaje de actividad de un marcador de ensayo expresado determinado (por ejemplo, LacZ), que incluye un codón selector, en un ácido nucleico que codifica el marcador de ensayo 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, en comparación con una construcción de control positivo, en el que el control positivo carece del O-ARNt, e la O-RS y del codón selector. Por tanto, por ejemplo, si una construcción de marcador de control positivo activo que carece de un codón de selector tiene una actividad observada de X en un sistema de traducción determinado, en unidades pertinentes al ensayo de marcador en cuestión, entonces el porcentaje de supresión de una construcción de ensayo que comprende el codón selector es el porcentaje de X que muestra la construcción de marcador de ensayo esencialmente en las mismas condiciones que el medio natural en las que se expresaba el marcador de control positivo, excepto porque la construcción de marcador de ensayo 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 ensayo también incluye un aminoácido que es reconocido por la O-RS y el O-ARNt. Opcionalmente, la medición del porcentaje de supresión puede refinarse por comparación del marcador de ensayo con una construcción de marcador de control de “fondo” o “negativo”, que incluye el mismo codón selector que el marcador de ensayo, pero en un sistema que no incluye el O-ARNt, la O-RS y/o el aminoácido pertinente reconocido por el O-ARNt y/o la O-RS. Este control negativo es útil para normalizar las mediciones del porcentaje de supresión para tener en cuenta los efectos de señal de fondo del marcador en el sistema de traducción de interés.

La eficacia de supresión puede determinarse mediante cualquiera de diversos ensayos conocidos en la técnica. Por ejemplo, puede utilizarse un ensayo indicador de β -galactosidasa, por ejemplo, un plásmido lacZ derivatizado (en el que la construcción tiene un codón selector en la secuencia de ácido nucleico lacZ) se introduce en células de un organismo apropiado (por ejemplo, un organismo en el que pueden utilizarse componentes ortogonales) junto con el plásmido que comprende un O-ARNt de la invención. También puede introducirse una sintetasa afín (bien como un polipéptido o como un polinucleótido que codifica la sintetasa afín cuando se expresa). Las células se cultivan en medio hasta una densidad deseada, por ejemplo, hasta una DO_{600} de aproximadamente 0,5, y se realizan ensayos de β -galactosidasa, utilizando, por ejemplo, el Kit de Ensayo de β -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 para la construcción lacZ derivatizada, en la que la construcción tiene un codón en sentido correspondiente en una posición deseada en vez de un codón selector.

Sistema de traducción: la expresión “sistema de traducción” se refiere a componentes que incorporan un aminoácido en una cadena polipeptídica en crecimiento (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetasa, ARNm y similares. El O-ARNt y/o las O-RS de la invención pueden añadirse o formar 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 vegetal, una célula de alga, una célula de hongo, una célula de insecto y/o similares.

Aminoácido no natural: como se usa en el presente documento, la expresión “aminoácido no natural” se refiere a cualquier aminoácido, aminoácido modificado y/o análogo de aminoácido, que no es uno de los 20 aminoácidos de origen natural comunes o los aminoácidos naturales raros selenocisteína o pirrolisina. Por ejemplo, el aminoácido no natural sulfotirosina; véase la **FIG. 1**) encuentra uso con la invención.

Derivado de: como se usa en el presente documento, la expresión “derivado de” se refiere a un componente que se aísla o se fabrica utilizando una molécula u organismo especificado, o información de la molécula o el organismo especificado. Por ejemplo, un polipéptido que deriva de un segundo polipéptido puede incluir una secuencia de aminoácidos que es idéntica o sustancialmente similar a la secuencia de aminoácidos del segundo polipéptido. En el caso de polipéptidos, las especies derivadas pueden obtenerse, por ejemplo, por mutagénesis de origen natural, mutagénesis de sitio dirigido artificial o mutagénesis al azar artificial. La mutagénesis usada para derivar polipéptidos puede ser intencionalmente dirigida o intencionalmente al azar, o una mezcla de cada una. La mutagénesis de un polipéptido para crear un polipéptido diferente derivado del primero puede ser un acontecimiento al azar (por ejemplo, producido por infidelidad de polimerasa) y la identificación del polipéptido derivado puede realizarse mediante métodos de exploración apropiados, por ejemplo, como se analiza en el presente documento. La mutagénesis de un polipéptido normalmente conlleva la manipulación del polinucleótido que codifica el polipéptido.

Marcador de selección o identificación positiva: como se usa en el presente documento, la expresión “marcador de selección o identificación positiva” se refiere a un marcador que, cuando está presente, por ejemplo, expresado, activado o similar, da como resultado la identificación de una célula que comprende el rasgo, por ejemplo, una célula con el marcador de selección positiva, entre aquellas que no tiene el rasgo.

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

Indicador: como se usa en el presente documento, el término “indicador” se refiere a un componente que puede utilizarse 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, β -lactamasa, cloranfenicol acetiltransferasa (CAT) y similar), un marcador de identificación 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 identificación basado en afinidad, o genes marcadores de selección positiva o negativa, tales como lacZ, β -gal/lacZ β -galactosidasa), ADH (alcohol deshidrogenasa), his3, ura3, leu2, lys2 o similares.

5 Eucariota: como se usa en el presente documento el término “eucariota” se refiere a organismos que pertenecen al Reino *Eucarya*. Los eucariotas se diferencian generalmente de los procariotas por su organización normalmente multicelular (pero no exclusivamente multicelular, por ejemplo, levaduras), la presencia de un núcleo delimitado por una membrana y otros orgánulos delimitados por una membrana, materiales genéticos lineales (por ejemplo, cromosomas lineales), la ausencia de operones, la presencia de intrones, ARNm con protección de mensaje y poli-
10 A y otras características bioquímicas, tal como una estructura ribosomal diferenciable. Los organismos eucariotas incluyen, por ejemplo, animales (por ejemplo, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (por ejemplo, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidios, protistas, etc.

15 Procariota: como se usa en el presente documento, el término “procariota” se refiere a organismos que pertenecen al Reino Monera (también denominado *Prokarya*). Los organismos procariotas se diferencian generalmente de los eucariotas por su organización unicelular, reproducción asexual por gemación o fisión, la ausencia de un núcleo delimitado por una membrana u otros orgánulos delimitados por una membrana, un cromosoma circular, la presencia de operones, la ausencia de intrones, protección de mensaje y poli-A ARNm y otras características bioquímicas tales como una estructura ribosomal diferenciable. Los Procariota incluyen los subreinos Eubacteria y
20 Archaea (a veces denominados “Arqueobacterias”). Las cianobacterias (algas verde azuladas) y los micoplasmas a veces se presentan en distintas clasificaciones en el Reino Monera.

Bacterias: como se usa en el presente documento, los términos “bacterias” y “eubacterias” se refieren a organismos procariotas que son diferenciables de los *Archaea*. De manera similar, Archaea se refiere a procariotas que son diferenciables de eubacterias. Las Eubacterias y Archaea pueden diferenciarse por diversos criterios morfológicos y bioquímicos. Por ejemplo, las diferencias en las secuencias de ARN ribosomal, estructura de la ARN polimerasa, la presencia o ausencia de intrones, sensibilidad a antibióticos, la presencia o ausencia de peptidoglucanos en la pared celular y otros componentes de pared celular, las estructuras ramificadas frente a no ramificadas de los lípidos de membrana y la presencia/ausencia de histonas y proteínas similares a histonas se usan para asignar a un
30 organismo en Eubacterias o en Archaea.

Como ejemplos de Eubacterias se incluyen *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis* y *Bacillus stearothermophilus*. Ejemplos de Arcaea incluyen *Methanococcus jannaschii* (Mj), *Methanosarcina mazei* (Mm), *Methanobacterium thermoautotrophicum* (Mt), *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Halobacterium*
35 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*, *Aeuropyrum pernix* (Ap), *Thermoplasma acidophilum* y *Thermoplasma volcanium*.

40 Variante conservativa: como se usa en el presente documento, la expresión “variante conservativa” en el contexto de un componente de traducción, se refiere a un componente de traducción, por ejemplo, un O-ARNt variante conservativo o una O-RS variante conservativa que funcionalmente se comporta de manera similar a un componente base al cual es similar la variante conservativa, por ejemplo, un O-ARNt o una O-RS, que tiene variaciones en la secuencia en comparación con un O-ARNt o una O-RS de referencia. Por ejemplo, una O-RS, o una variante conservativa de esta O-RS, aminoacilará un O-ARNt afín con un aminoácido natural, por ejemplo, sulfotirosina. En
45 este ejemplo, la O-RS y la O-RS variante conservativa no tienen las mismas secuencias de aminoácidos. La variante conservativa puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones o cinco o más variaciones en la secuencia, siempre que la variante conservativa sea complementaria (por ejemplo funcione con) al O-ARNt o a la O-RS afín correspondiente.

50 En algunas realizaciones, una O-RS variante conservativa comprende una o más sustituciones de aminoácidos conservativas en comparación con la O-RS de la que deriva. En algunas realizaciones, una O-RS variante conservativa comprende una o más sustituciones de aminoácidos conservativas en comparación con la O-RS de la que deriva, y adicionalmente conserva actividad biológica O-RS; por ejemplo, una O-RS variante conservativa que conserva al menos el 10 % de la actividad biológica de la molécula O-RS parental de la que deriva, o como
55 alternativa, al menos el 20 %, al menos el 30 % o al menos el 40 %. En algunas realizaciones preferidas, la O-RS variante conservativa conserva al menos el 50 % de la actividad biológica de la molécula O-RS parental de la que deriva. Las sustituciones de aminoácidos conservativas de una O-RS variante conservativa pueden producirse en cualquier dominio de la O-RS, incluyendo el bolsillo de unión de aminoácidos.

60 Agente de selección o identificación: como se usa en el presente documento, la expresión “agente de selección o identificación” se refiere a un agente que, cuando está presente, permite la selección/identificación de determinados componentes de una población. Por ejemplo, un agente de selección o identificación puede ser, pero sin limitación, por ejemplo, un nutriente, un antibiótico, una longitud de onda de luz, un anticuerpo, un polinucleótido expresado o similar. El agente de selección se puede variar, por ejemplo, por concentración, intensidad, etc.
65

En respuesta a: como se usa en el presente documento, la expresión “en respuesta a” se refiere al proceso en el que un O-ARNt de la invención reconoce un codón selector y media la incorporación del aminoácido no natural, que está acoplado al ARNt, en la cadena polipeptídica en crecimiento.

5 **Codifica:** como se usa en el presente documento, el término “codifica” se refiere a cualquier proceso mediante el cual la información en una macromolécula polimérica o cadena de secuencia se usa para dirigir la producción de una segunda molécula o cadena de secuencia que es diferente de la primera molécula o cadena de secuencia. Como se usa en el presente documento, el término se usa ampliamente, y puede tener diversas aplicaciones. En algunos aspectos, el término “codifica” describe el proceso de replicación de ADN semi-conservativa, donde una cadena de
10 una molécula de ADN bicatenaria se usa como un molde para codificar una cadena hermana complementaria recién sintetizada por una ADN polimerasa dependiente de ADN.

En otro aspecto, el término “codificar” se refiere a cualquier proceso mediante 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 incorporando una enzima ARN polimerasa dependiente de ADN). Además, una molécula de ARN puede codificar un polipéptido, como en el proceso de traducción. Cuando se usa para describir el proceso de traducción, el término “codificar” también se extiende al codón triplete que codifica un aminoácido. En algunos aspectos, una molécula de ARN puede codificar una molécula de ADN, por ejemplo, mediante el proceso de transcripción inversa incorporando una ADN polimerasa dependiente de ARN. En otro aspecto, una molécula de ADN puede codificar un polipéptido, donde se entiende que “codificar”, como se usa en este caso, incorpora los procesos tanto de traducción como los de transcripción.

Breve descripción de las figuras

25 La **FIG. 1** proporciona la estructura química del aminoácido no natural sulfotirosina.

La **FIG. 2** muestra un gel PAGE desnaturalizante teñido con azul de Coomassie que ilustra la migración de la desulfohirudina y de la sulfohirudina. El tamaño de la hirudina no puede juzgarse por patrones de peso molecular debido a la carga atípica de la hirudina.

Las **FIGS. 3A y 3B** proporcionan gráficas representativas de la inhibición de trombina con sus curvas de progreso ajustadas respectivas superpuestas en los puntos de datos no procesados. Se realizaron ensayos enzimáticos con sustrato fluorogénico 50 μ M, α -trombina humana 40 pM e hirudina expresada 100 pM en un tampón de solución salina Tris-HCl complementado con polietilenglicol 600 y HSA. La **FIG. 3A** muestra gráficas de intensidad de fluorescencia a lo largo del tiempo sin inhibición (*control*), con inhibición por desulfohirudina y con inhibición por sulfohirudina. La **FIG. 3B** muestra ampliación de gráficas de desulfohirudina y sulfohirudina para comparación.

Las **FIGS. 4A y 4B** ilustran la expresión dependiente de sulfotirosina del dominio Z. La **FIG. 4A** proporciona un gel PAGE desnaturalizante teñido con azul de Coomassie del lisado celular purificado con Ni-NTA (ácido níquel nitrilotriacético) de células que expresan el dominio Z con un codón ámbar en la posición 7. La expresión solo con medio complementado con sulfotirosina produce el dominio Z de longitud completa. La **FIG. 4B** proporciona un espectro MALDI-TOF en modo lineal de iones positivos (generado utilizando una matriz THAP) del lisado celular purificado con Ni-NTA (concentrado y dializado frente a agua) que muestra un pico correspondiente al dominio Z de longitud completa que contiene una sola sulfotirosina y que carece de metionina. También se observa un pico correspondiente a la pérdida de sulfato resultante de las condiciones del análisis espectral de masas.

Las **FIGS. 5A, 5B y 5C** muestran diversos espectros MALDI-TOF. La **FIG. 5A** muestra un espectro MALDI-TOF en modo lineal de iones positivos (generado utilizando una matriz THAP) de la sulfohirudina pura que muestra tanto el pico de sulfohirudina intacto [M+H] (7059 Da) como el pico correspondiente a la pérdida de sulfato durante el análisis espectral de masas (6979 Da). Obsérvese que los picos pequeños a la derecha de los principales son aductos de sodio. Se producen a intervalos adicionales de 22 Da. La **FIG. 5B** muestra un espectro MALDI-TOF (generado utilizando una matriz sinapínica) que documenta la pureza de la muestra. Para aumentar la detección de posibles impurezas, se utilizó una matriz sinapínica más rigurosa, que dio como resultado la predominancia del pico [M+H-80]. El pico a 13964 Da puede atribuirse a la dimerización de la sulfohirudina. No se observaron otras impurezas. La **FIG. 5C** muestra una ampliación de la región relevante que muestra la presencia del pico tanto de la sulfohirudina intacta y como de [M+H-80]. La sulfohirudina intacta es el pico menor debido al uso de la matriz sinapínica más rigurosa. Los picos pequeños a la derecha de los principales son aductos de sodio.

Las **FIGS. 6A y 6B** muestran diversos espectros MALDI-TOF. La **FIGURA 6A** muestra un espectro MALDI-TOF (generado utilizando una matriz sinapínica) de medios de expresión de la sulfohirudina no purificada correspondientes a la expresión en ausencia de sulfotirosina. Solo se encontró el pico de la hirudina truncada; no se observó proteína de longitud completa. La **FIG. 6B** muestra un espectro MALDI-TOF (generado utilizando una

matriz sinapínica) de medios de expresión de sulfohirudina no purificada correspondientes a la expresión en presencia de sulfotirosina demostrando la proporción de picos de la sulfohirudina truncada con respecto a la de longitud completa. Debido a las condiciones más rigurosas necesarias para una buena detección de las mezclas de muestras brutas, solo se observó claramente la forma ionizada de la sulfohirudina.

5

La **FIG. 7** proporciona secuencias de nucleótidos y de aminoácidos.

Descripción detallada de la invención

10 Aunque la sulfatación de la tirosina es una modificación postraduccional generalizada en eucariotas multicelulares (Moore, "The biology and enzymology of protein tyrosine O-sulfation," J Biol Chem 278: 24243-24246 (2003)), sus funciones biológicas siguen sin conocerse en gran medida. Esto se debe en parte a las dificultades asociadas con la síntesis de proteínas sulfatadas selectivamente. La invención proporciona la incorporación selectiva de sulfotirosina en proteínas en bacterias por codificación genética del aminoácido modificado en respuesta al codón sin sentido
15 ámbar, TAG. Además, se demuestra que la sulfohirudina, previamente inaccesible a través de métodos recombinantes, puede expresarse directamente en *E. coli* utilizando esta estrategia. Como se describe en el presente documento, análisis cinéticos muestran una mejora en afinidad mayor de 10 veces frente a la trombina humana por sulfohirudina sobre la desulfohirudina, una observación que ofrece ventajas clínicas para la sulfohirudina en su uso como un anticoagulante (Di Nisio *et al.*, "Direct thrombin inhibitors," N Engl J Med 353: 1028-
20 1040 (2005)). Esta estrategia general para la biosíntesis de proteínas sulfatadas facilita el estudio y la aplicación adicional de la modificación postraduccional emergente, la sulfatación de la tirosina.

Como un método general para la sulfatación específica de sitio de proteínas, la presente descripción describe el desarrollo de un par ARNt/aminoacil-ARNt sintetasa (aaRS) ortogonal que permite la incorporación eficaz y selectiva
25 de sulfotirosina en proteínas eucariotas tales como *E. coli* en respuesta al codón sin sentido ámbar. Utilizando este único par ARNt/aaRS supresor, la forma sulfatada nativa de la hirudina se expresa directamente y se muestra que tiene una afinidad de 10 veces mayor por la trombina humana que por la desulfohirudina, de acuerdo con publicaciones de bibliografía anterior (Stone y Hofsteenge, "Kinetics of the inhibition of thrombin by hirudin," Biochemistry 25: 4622-4628 (1986)).

30

La presente memoria descriptiva proporciona pares ARNt/aminoacil-ARNt sintetasa ortogonales que permiten la introducción selectiva *in vivo* de la sulfotirosina (véase la **FIG. 1**) en proteínas en *E. coli* en respuesta a un codón selector, por ejemplo, el codón de terminación ámbar TAG. La invención proporciona nuevos polipéptidos aminoacil-ARNt sintetasa (O-RS) ortogonales, que cargan específicamente un ARNt ortogonal (O-ARNt) afín con el
35 aminoácido no natural sulfotirosina.

En algunos aspectos, para demostrar (pero no limitar) la presente invención, la descripción del presente documento demuestra que el resto de aminoácido no natural puede incorporarse en diversas proteínas modelo. No se pretende limitar la incorporación del aminoácido no natural a ninguna proteína particular. A partir de la presente descripción,
40 será obvio que la incorporación del aminoácido no natural sulfotirosina en proteínas de interés particular es ventajosa para una amplia diversidad de finalidades.

La presente descripción describe el desarrollo de nuevos pares ARNt/aminoacil-ARNt sintetasa ortogonales que funcionan en eubacterias para incorporar específicamente en el sitio un aminoácido no natural de sulfotirosina
45 (proporcionado en la **FIG. 1**) en respuesta a codones selectores. En resumen, la invención proporciona nuevos mutantes de tirosil-ARNt sintetasa de *Methanococcus janaschii* como se define en las reivindicaciones, que cargan selectivamente un ARNt supresor con el aminoácido no natural sulfotirosina en células hospedadoras *E. coli*.

Estos pares de ARNt-sintetasa desarrollados pueden utilizarse para incorporar específicamente en el sitio el aminoácido no natural sulfotirosina en una proteína. La incorporación del aminoácido no natural en la proteína puede programarse para que se produzca en cualquier posición deseada por modificación con ingeniería genética del
50 polinucleótido que codifica la proteína de interés para contener un codón selector que señale la incorporación del aminoácido no natural.

La invención descrita en el presente documento proporciona pares ortogonales para la codificación e incorporación genética del aminoácido no natural sulfotirosina en proteínas en una eubacteria, por ejemplo, una célula de *E. coli*, en la que los componentes ortogonales no reaccionan en cruzado con componentes de *E. coli* endógenos de la maquinaria traduccional de la célula hospedadora, pero reconocen el aminoácido no natural deseado y lo incorporan en proteínas en respuesta a un codón selector (por ejemplo, un codón sin sentido ámbar, TAG). Los componentes ortogonales proporcionados por la invención incluyen aminoacil-ARNt sintetasa ortogonales derivadas de ARNt-sintetasa de *Methanococcus janaschii* y el supresor ámbar tirosil ARNt_{CUA}, mutante que funciona como un par
60 ortogonal en una célula hospedadora de eubacteria.

La invención proporciona composiciones de y métodos, como se define en las reivindicaciones, para identificar y producir pares adicionales de ARNt-aminoacil-ARNt sintetasa ortogonales, por ejemplo pares O-ARNt/O-RS que pueden utilizarse para incorporar sulfotirosina en proteínas. Un par O-ARNt/O-RS de la invención puede mediar la
65

incorporación de la sulfotirosina en una proteína que está codificada por un polinucleótido, en el que el polinucleótido comprende un codón selector que es reconocido por el O-ARNt. El bucle anticodón del O-ARNt reconoce el codón selector en un ARNm e incorpora el aminoácido no natural en este sitio en el polipéptido. Generalmente, una aminoacil-ARNt sintetasa ortogonal de la invención aminoacila (o carga) preferentemente su O-ARNt solo con un aminoácido no natural específico.

TECNOLOGÍA ARNt/AMINOACIL-ARNt SINTETASA ORTOGONAL

Un entendimiento de las nuevas composiciones y métodos de la presente invención requiere un entendimiento de las actividades asociadas con pares de ARNt ortogonal y aminoacil-ARNt sintetasa ortogonal. Para añadir aminoácidos no naturales adicionales al código genético, se requieren nuevos pares ortogonales que comprendan una aminoacil-ARNt sintetasa y un ARNt adecuado que puedan funcionar eficazmente en la maquinaria traduccional del hospedador, pero que sean "ortogonales" para el sistema de traducción en cuestión, lo que significa que funcionan independientemente de las sintetetas y de los ARNt endógenos al sistema traduccional. Características deseadas del par ortogonal incluyen ARNt que descodifica o reconoce solo un codón específico, por ejemplo, un codón selector, que no está descodificado por ningún ARNt endógeno, y aminoacil-ARNt sintetetas que aminoacilan (o "cargan") preferentemente su ARNt afín solo con un aminoácido específico no natural. Además el O-ARNt no está normalmente aminoacilado (o está mal aminoacilado, es decir, cargado) por sintetetas endógenas. Por ejemplo, en un sistema hospedador de *E. coli*, un par ortogonal incluirá una aminoacil-ARNt sintetasa que no reacciona en cruzado con ninguno de los ARNt endógenos, por ejemplo, en el que hay 40 en *E. coli*, y un ARNt ortogonal que no está aminoacilado por ningunas de las sintetetas endógenas, por ejemplo, de las que hay 21 en *E. coli*.

Los principios generales de los sistemas de traducción ortogonal que son adecuados para fabricar proteínas que comprendan uno o más aminoácidos naturales son conocidos en la técnica, así como en los métodos generales para la producción de sistemas de traducción ortogonal. Por ejemplo, véanse los Números de Publicación Internacionales WO 2002/086075, titulado "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS"; WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE"; WO 2005/019415, presentado el 7 de julio de 2004; WO 2005/007870, presentado el 7 de julio de 2004; WO 2005/007624, presentado el 7 de julio de 2004; WO 2006/110182, presentado el 27 de octubre de 2005, titulado "ORTHOGONAL TRANSLATION COMPONENTS FOR THE VIVO INCORPORATION OF UNNATURAL AMINO ACIDS" y WO 2007/103490, presentado el 7 de marzo de 2007, titulado "SYSTEMS FOR THE EXPRESSION OF ORTHOGONAL TRANSLATION COMPONENTS IN EUBACTERIAL HOST CELLS". Para un debate sobre los sistemas de traducción ortogonal que incorporan aminoácidos no naturales y métodos para su producción y uso, véase también Wang y Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1): 34-66 (2005), Xie y Schultz, "An Expanding Genetic Code," *Methods* 36(3): 227-238 (2005); Xie y Schultz, "Adding Amino Acids to the Genetic Repertoire," *Curr. Opin. in Chemical Biology* 9(6): 548-554 (2005); y Wang *et al.*, "Expanding the Genetic Code," *Annu. Rev. Biophys. Biomol. Struct.*, 35: 225-249 (2006).

Sistemas de traducción ortogonal

Los sistemas de traducción ortogonal generalmente comprenden células (que pueden ser células procariotas tales como *E. coli*) que incluyen un ARNt ortogonal (O-ARNt), una aminoacil ARNt sintetasa ortogonal (O-RS), y un aminoácido no natural, en el que la O-RS aminoacila el O-ARNt con el aminoácido no natural. Un par ortogonal de la invención puede incluir un O-ARNt, por ejemplo, un ARNt supresor, un ARNt de desplazamiento de fase de lectura, o similar, o una O-RS afín. Los sistemas ortogonales de la invención definidos en las reivindicaciones comprenden normalmente pares de O-ARNt/O-RS, en el contexto de una célula hospedadora o sin la célula hospedadora. Además de sistemas multicomponentes, la invención también proporciona nuevos componentes individuales, por ejemplo, nuevos polipéptidos aminoacil-ARNt sintetasa ortogonales (por ejemplo, SEC ID Nos: 4, 6, 8 o 10), y los polinucleótidos que codifican estos polipéptidos (por ejemplo, SEC ID Nos: 5, 7, 9 u 11).

En general, cuando un par ortogonal reconoce un codón selector y carga un aminoácido en respuesta al codón selector, se dice que el par ortogonal "suprime" el codón selector. Es decir, un codón selector que no está reconocido por la maquinaria endógena (por ejemplo, de la célula) del sistema de traducción, no está habitualmente cargado, lo que produce el bloqueo de la producción de un polipéptido que, de otra manera, se transcribiría a partir del ácido nucleico. En un sistema de pares ortogonales, la O-RS aminoacila el O-ARNt con un aminoácido no natural específico. El O-ARNt cargado reconoce el codón selector y suprime el bloqueo traduccional producido por el codón selector.

En algunos aspectos, un O-ARNt de la invención reconoce un codón selector y puede incluir una eficacia de supresión de al menos aproximadamente, por ejemplo, 45 %, 50 %, 60 %, 75 %, 80 % o 90 % o más en presencia de una sintetasa afín en respuesta a un codón selector en comparación con la eficacia de supresión de un O-ARNt que comprende o está codificado por una secuencia de polinucleótidos como se indica en la lista de secuencias del presente documento.

En algunas realizaciones, la eficacia de supresión de la O-RS y del O-ARNt en su conjunto puede ser aproximadamente, por ejemplo, 5 veces, 10 veces, 15 veces, 20 veces o 25 veces o más elevada que la eficacia de supresión del O-ARNt que carece de la O-RS. En algunos aspectos, la eficacia de supresión de la O-RS y del O-ARNt en su conjunto puede ser al menos aproximadamente, por ejemplo, 35 %, 40 %, 45 %, 50 %, 60 %, 75 %, 80 % o 90 % o más de la eficacia de supresión de un par sintetasa ortogonal como se indica en la lista de secuencias en el presente documento.

La célula hospedadora utiliza el par O-ARNt/O-RS que incorpora el aminoácido no natural en una cadena polipeptídica en crecimiento, 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 determinados aspectos preferidos, la célula puede incluir uno o más pares O-ARNt/O-RS adicionales, en el que el O-ARNt adicional se carga mediante la O-RS adicional con un aminoácido no natural diferente. Por ejemplo, uno de los ARNt puede reconocer un codón de cuatro bases y el otro O-ARNt puede reconocer un codón de terminación. Como alternativa, en el mismo ácido nucleico codificante pueden utilizarse codones de terminación diferentes múltiples o codones de cuatro bases diferentes múltiples.

Como se observa, en algunas realizaciones, existen pares O-ARNt/O-RS múltiples en una célula u otro sistema de traducción, que permiten la incorporación de más de un aminoácido no natural en un polipéptido. Por ejemplo, la célula puede incluir adicionalmente un par O-ARNt/O-RS diferente adicional y un segundo aminoácido no natural, en el que este O-ARNt reconoce un segundo codón selector y esta O-RS adicional aminoacila preferentemente el O-ARNt con el segundo aminoácido no natural. Por ejemplo, una célula que incluya un par O-ARNt/O-RS (en el que el O-ARNt reconoce, por ejemplo, un codón selector ámbar), puede comprender adicionalmente un segundo par ortogonal, en el que segundo O-ARNt reconoce un codón selector diferente, por ejemplo, un codón ópalo, un codón de cuatro bases o similar. Deseablemente, los pares ortogonales diferentes derivan de diferentes fuentes, que pueden facilitar el reconocimiento de diferentes codones selectores.

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

El O-ARNt y/o la O-RS pueden ser de origen natural o pueden derivar, por ejemplo, por mutación de un ARNt y/o una RS de origen natural, por ejemplo, generando bibliotecas de ARNt y/o bibliotecas de RS, a partir de cualquiera de diversos organismos y/o utilizando cualquiera de diversas estrategias de mutación disponibles. Por ejemplo, una estrategia para producir un par ARNt/aminoacil-ARNt sintetasa ortogonal implica importar un par ARNt/sintetasa (a la célula hospedadora) heterólogo a partir de, por ejemplo, una fuente distinta de la célula hospedadora, o múltiples fuentes, en la célula hospedadora. Las propiedades de la candidata sintetasa heteróloga incluyen, por ejemplo, que no carga ningún ARNt en la célula hospedadora, y las propiedades del candidato ARNt heterólogo incluyen, por ejemplo, que no está aminoacilado por ninguna sintetasa de la célula hospedadora. Además, el ARNt heterólogo es ortogonal con respecto a todas las sintetastas de la célula hospedadora. Una segunda estrategia para generar un par ortogonal implica generar bibliotecas mutantes a partir de las cuales se identifique y/o se seleccione un O-ARNt o una O-RS. Estas estrategias también pueden combinarse.

ARNt ortogonal (O-ARNt)

Un ARNt ortogonal (O-ARNt) de la invención media deseablemente la incorporación de un aminoácido no natural 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 determinadas realizaciones, un O-ARNt de la invención puede incluir una eficacia de supresión de al menos aproximadamente, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 % o un 90 % o más en presencia de una sintetasa afín en respuesta a un codón selector en comparación con un O-ARNt que comprende o está codificado por una secuencia de polinucleótidos como se indica en las secuencias de O-ARNt en la lista de secuencias del presente documento.

La eficacia de supresión puede determinarse por cualquiera de diversos ensayos conocidos en la técnica. Por ejemplo, puede utilizarse un ensayo indicador con β -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 utilizarse los componentes ortogonales) junto con plásmido que comprende un O-ARNt de la invención. También puede introducirse una sintetasa afín (como un polipéptido o un polinucleótido que codifica la sintetasa afín cuando se expresa). Las células se cultivan en medios a una densidad deseada, por ejemplo, a una DO₆₀₀ de aproximadamente 0,5, y se realizan ensayos con β -galactosidasa, por ejemplo, utilizando el Kit de Ensayo β -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 a partir de la construcción lacZ derivatizada, en el que la construcción tiene un

codón sentido correspondiente en la posición deseada en lugar de un codón selector.

En la lista de secuencias del presente documento, por ejemplo, véase la **FIG. 7** y la SEC ID N°: 1 se indican ejemplos de O-ARNt de la invención. La descripción del presente documento también proporciona orientaciones para el diseño de especies de O-ARNt adicionales equivalentes. En una molécula de ARN, tal como un ARNm de O-RS, o molécula de O-ARNt, la Timina (T) se reemplaza con Uracilo (U) con respecto a una secuencia determinada (o viceversa para un ADN codificante), o su complemento. También pueden estar presentes modificaciones adicionales para las bases para generar moléculas en gran parte funcionalmente equivalentes.

La invención también incluye variaciones conservativas de O-ARNt correspondientes a O-ARNt particulares del presente documento. Por ejemplo, variaciones conservativas de O-ARNt incluyen aquellas moléculas que funcionan como los O-ARNt particulares, por ejemplo, como en la lista de secuencias del presente documento y que conservan la estructura en forma de L del ARNt debido a auto-complementariedad apropiada, pero que no tienen una secuencia idéntica a las de, por ejemplo, la lista de secuencias de la **FIG. 7**, y deseablemente, son distintas de las moléculas de ARNt de tipo silvestre.

La composición comprende un O-ARNt que adicionalmente incluye una aminoacil-ARNt sintetasa ortogonal (O-RS) como se define en las reivindicaciones, en la que la O-RS aminoacila preferentemente el O-ARNt con un aminoácido no natural. En determinadas realizaciones, una composición que incluye un O-ARNt puede incluir adicionalmente un sistema de traducción (por ejemplo, *in vitro* o *in vivo*). En la célula también puede haber un ácido nucleico que comprenda un polinucleótido que codifique un polipéptido de interés, en el que el polinucleótido comprenda un codón selector reconocido por el O-ARNt, o una combinación de uno o más de estos.

Son también un aspecto de la presente invención métodos de producción de un ARNt ortogonal (O-ARNt). Un O-ARNt producido por el método se utiliza en las composiciones y métodos de la invención. En determinadas realizaciones de la presente descripción, los O-ARNt pueden producirse generando una biblioteca de mutantes. La biblioteca de ARNt mutantes puede generarse utilizando diversas técnicas de mutagénesis conocidas en el campo. Por ejemplo, los ARNt mutantes pueden generarse por mutaciones específicas de sitio, mutaciones puntuales aleatorias, recombinación homóloga, transposición de ADN u otros métodos de mutagénesis recursiva, construcción química o cualquier combinación de los mismos, por ejemplo, del O-ARNt de la SEC ID N°: 1.

Pueden introducirse mutaciones adicionales en una posición (o posiciones) específicas, por ejemplo, en una posición (o posiciones) no conservativa, o en una posición conservativa, en una posición (o posiciones) al azar, 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 TPC, otras regiones de la molécula de ARNt, o una combinación de los mismos. Normalmente, las mutaciones en un ARNt incluyen mutar el bucle anticodón de cada miembro de la biblioteca de ARNt mutantes para permitir el reconocimiento de un codón selector. El método puede incluir añadir adicionalmente secuencias adicionales al 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, conservando al mismo tiempo su afinidad hacia una RS deseada.

Los métodos incluyen analizar opcionalmente la similitud (y/u homología deducida) de secuencia de los ARNt y/o aminoacil-ARNt sintetasa 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 utilizarse para el análisis, por ejemplo, BLAST y pueden utilizarse programas de compilación. En un ejemplo, para seleccionar posibles componentes de traducción ortogonal para su uso en *E. coli*, se selecciona una sintetasa y/o un ARNt que no muestre similitud de secuencia próxima a organismos eubacterianos.

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.

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

Por ejemplo, para medir la actividad de ARNt supresores, se utiliza un sistema de selección que se basa en la supresión *in vivo* de codón selector, por ejemplo, mutaciones sin sentido (por ejemplo, de detección) o de desplazamiento de fase introducidas en un polinucleótido que codifica un marcador de selección negativa, por

ejemplo, un gen para β -lactamasa (*bla*). Por ejemplo, se construyen variantes de polinucleótido, por ejemplo variantes de *bla*, con un codón selector en una determinada posición (por ejemplo, A184). 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 no es ortogonal, o si una sintetasa heteróloga capaz de cargar el ARNt se coexpresa en el sistema, se observa un mayor nivel de antibiótico, por ejemplo, resistencia a ampicilina. Se seleccionan células, por ejemplo, bacterias, que no pueden cultivarse en placas de LB agar con concentraciones de antibiótico aproximadamente iguales a células transformadas sin plásmidos.

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

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

Por consiguiente, el mismo marcador puede ser un marcador positivo o negativo, dependiendo del contexto en el que se explore. Es decir, el mismo marcador es un marcador positivo si se explora para, pero es un marcador negativo si se explora contra.

La rigurosidad de la selección, por ejemplo, la selección positiva, la selección negativa o tanto la selección positiva como la negativa, en los métodos descritos anteriormente, incluye opcionalmente modificar la rigurosidad de la selección. Por ejemplo, dado 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 utilizando un promotor inducible. En otro ejemplo, se modifica la concentración del agente de selección o identificación (por ejemplo, concentración de ampicilina). En algunos aspectos, la rigurosidad se modifica debido a que la actividad deseada puede ser baja durante las primeras rondas. Por tanto, se aplican criterios de selección menos rigurosos en las primeras rondas y se aplican criterios más rigurosos en rondas de selección posteriores. En determinadas realizaciones, la selección negativa, la selección positiva o tanto la selección positiva como la negativa pueden repetirse múltiples veces. Se pueden utilizar múltiples marcadores diferentes de selección negativa, marcadores de selección positiva o tanto marcadores de selección positiva como negativa. En determinadas realizaciones, el marcador de selección positiva y negativa puede ser el mismo.

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

Métodos adicionales para producir un ARNt ortogonal recombinante pueden encontrarse, por ejemplo, en las Publicaciones de Solicitud Internacional WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOG-ONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS"; WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE" y WO 2005/019415, presentada el 7 de julio de 7, 2004. Véase también Forster *et al.*, (2003) Programming peptidomimetic synthetases by translating genetic codes designed de novo PNAS 100(11): 6353-6357; y, Feng *et al.*, (2003), Expanding tRNA recognition of a tRNA synthetase by a single amino acid change, PNAS 100(10): 5676-5681.

Aminoacil-ARNt sintetasa ortogonal (O-RS)

Una O-RS de la invención se define en las reivindicaciones y aminoacila preferentemente un O-ARNt con un aminoácido no natural, *in vitro* o *in vivo*. Una O-RS de la invención puede proporcionarse al sistema de traducción, por ejemplo, una célula, mediante un polipéptido que incluye una O-RS y/o un polinucleótido que codifica una O-RS o una parte de la misma. Por ejemplo, una O-RS ilustrativa comprende una secuencia de aminoácidos como se expone en las SEC ID Nos: 4, 6, 8 o 10 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 una secuencia que comprende aminoácidos en la lista de secuencias o los ejemplos del presente documento, o una secuencia de polinucleótidos complementaria de la misma. Véase, por ejemplo, el polinucleótido de las SEC ID Nos: 5, 7, 9 u 11.

Los métodos como se define en las reivindicaciones pueden comprender identificar una aminoacil-ARNt sintetasa ortogonal (O-RS), por ejemplo, una O-RS, para su uso con un O-ARNt. Por ejemplo, un método incluye someter a selección, por ejemplo, selección positiva, una población de células de una primera especie, en el que las células comprenden individualmente: 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 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. Se seleccionan o identifican las células que muestran un aumento en la eficacia de supresión en comparación con células que carecen de o con una cantidad reducida del miembro de la pluralidad de RS. La eficacia de supresión puede medirse por técnicas conocidas en el campo y descritas en el presente documento. Las células que tienen un aumento en la eficacia 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 mediante una sustancia detectable (por ejemplo, un aminoácido no natural marcado). Normalmente se selecciona la RS activa que aminoacila eficazmente el segundo conjunto de ARNt en comparación con el primer conjunto de ARNt, proporcionando de este modo una aminoacil ARNt sintetasa ortogonal eficaz (optimizada) para su uso con el O-ARNt. Una O-RS, identificada por el método, también es una característica de la invención.

Para determinar la aminoacilación pueden utilizarse cualquiera de diversos ensayos. Estos ensayos pueden realizarse *in vitro* o *in vivo*. Por ejemplo, se describen ensayos de aminoacilación *in vitro*, por ejemplo, en Hoben y Soll (1985) *Methods Enzymol.* 113: 55-59. La aminoacilación también puede determinarse utilizando un indicador junto con los componentes de traducción ortogonales y detectando el indicador en una célula que exprese un polinucleótido que comprende al menos un codón selector que codifica una proteína. Véase también, el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; y el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE".

La O-RS identificada puede manipularse adicionalmente para alterar la especificidad de sustrato de la sintetasa, de tal manera que solo se carga un aminoácido no natural deseado, pero ninguno de los 20 aminoácidos comunes, en el O-ARNt. Los métodos para generar un aminoacil-ARNt sintetasa ortogonal con una especificidad de sustrato para un aminoácido 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 sintetasa, o similares, y aplicar un proceso de selección. Se usa una estrategia que se basa en la combinación de una selección positiva seguida de una selección negativa. En la selección positiva, la supresión del codón selector introducido en una o más o posiciones no esenciales de un marcador positivo permite a las células sobrevivir bajo presión de selección positiva. En presencia de aminoácidos tanto naturales como no naturales, los supervivientes codifican así sintetasa activa que cargan el ARNt supresor ortogonal con un aminoácido natural o no natural. En la selección negativa, la supresión de un codón selector introducido en una posición o más posiciones no esenciales de un marcador negativo elimina las sintetasa con especificidades de aminoácidos naturales. Los supervivientes de la selección negativa y positiva codifican sintetasa que aminoacilan (cargan) el ARNt supresor ortogonal solo con aminoácidos no natural. Después, estas sintetasa pueden someterse a mutagénesis adicional, por ejemplo, transposición de ADN u otros métodos de mutagénesis recursiva.

Puede generarse una biblioteca de O-RS mutantes utilizando diversas técnicas de mutagénesis conocidas en la materia. Por ejemplo, pueden generarse RS mutantes mediante mutaciones con especificidad de sitio, mutaciones puntuales aleatorias, recombinación homóloga, transposición de ADN u otros métodos de mutagénesis recursiva, construcción quimérica o cualquier combinación de los mismos. Por ejemplo, puede producirse una biblioteca de RS mutantes a partir de dos o más, por ejemplo, "sub-bibliotecas" diferentes, más pequeñas, menos diversas. También se describen bibliotecas quiméricas de RS. Debe señalarse que las bibliotecas de ARNt sintetasa de diversos organismos (por ejemplo, microorganismos, tales como eubacterias o arqueobacterias) tales como bibliotecas que comprenden diversidad natural (véase, por ejemplo, la Patente de Estados Unidos N° 6.238.884 de Short *et al.*; la Patente de Estados Unidos N° 5.756.316 de Schallenberger *et al.*; la Patente de Estados Unidos N° 5.783.431 de Petersen *et al.*; la Patente de Estados Unidos N° 5.824.485 de Thompson *et al.*; la Patente de Estados Unidos N° 5.958.672 de Short *et al.*) se construyen y se exploran opcionalmente para pares ortogonales.

Una vez que las sintetetasas se han sometido a la estrategia de selección/identificación positiva y negativa, estas sintetetasas pueden someterse después a mutagénesis adicionales. Por ejemplo, un ácido nucleico que codifica la O-RS puede aislarse; un conjunto de polinucleótidos que codifican las O-RS mutadas (por ejemplo, por mutagénesis al azar, mutagénesis específica de sitio, recombinación o cualquier combinación de las mismas) puede generarse 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 aminoacila preferentemente el O-ARNt con el aminoácido no natural. En algunos aspectos de la invención, las etapas pueden realizarse múltiples veces, por ejemplo, al menos dos veces.

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

Detalles generales adicionales para la producción de O-RS y la modificación de la especificidad de la sintetasa por el sustrato pueden encontrarse en la Publicación Interna Número WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS"; y en el documento WO 2004/094593, titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". Véase también Wang y Schultz "Expanding the Genetic Code," *Ange-wandte Chemie Int. Ed.*, 44(1): 34-66 (2005).

ORGANISMOS FUENTE Y HOSPEDADORES

Los componentes de traducción ortogonal (O-ARNt y O-RS) de la invención definidos en las reivindicaciones pueden derivar de cualquier organismo (o una combinación de organismos) para su uso en un sistema de traducción hospedador de cualquier otra especie, con la advertencia de que los componentes O-ARNt/O-RS y el sistema del hospedador actúan de una manera ortogonal. No es necesario que el O-ARNt y la O-RS de un par ortogonal deriven del mismo organismo. En algunos aspectos, los componentes ortogonales derivan de genes *Archaea* (es decir, arqueobacterias) para su uso en un sistema hospedador eubacteriano.

Por ejemplo, el O-ARNt ortogonal puede derivar de un organismo *Archaea*, 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*, *Aeuryopyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (*Mm*), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (*Ss*), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o similar, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, *Bacillus stearothermophilus*, o similares, mientras que la O-RS ortogonal puede derivar de un organismo o 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*, *Aeuryopyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o similar, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, *Bacillus stearothermophilus*, o similar. En una realización, también pueden utilizarse fuentes eucariotas, por ejemplo, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos, artrópodos, etc.), o similar, como fuentes de O-ARNt y O-RS.

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

El O-ARNt, la O-RS o el par O-ARNt/O-RS puede seleccionarse o identificarse *in vivo* o *in vitro* y/o utilizarse en una célula, por ejemplo, una célula de eubacteria, para producir un polipéptido con un aminoácido natural. La célula de eubacteria utilizada no está limitada, por ejemplo, *Escherichia coli*, *Thermus thermophilus*, *Bacillus subtilis*, *Bacillus stearothermophilus*, o similar. Las composiciones de células de eubacterias que comprenden componentes de traducción de la invención son también una característica de la invención.

Véase también la Publicación de Solicitud Internacional Número WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE", presentada el 16 de abril de 2004, para identificar O-ARNt y/o O-RS en una especie para su uso en otra especie.

Aunque los sistemas de traducción ortogonal (por ejemplo, que comprenden una O-RS, un O-ARNt y un aminoácido no natural) pueden utilizar células hospedadoras cultivadas para producir proteínas que tengan aminoácidos no naturales, no se pretende que un sistema de traducción ortogonal de la invención requiera una célula hospedadora viable, intacta. Por ejemplo, un sistema de traducción ortogonal puede utilizar un sistema celular en presencia de un

extracto celular. De hecho, el uso de sistemas acelulares de transcripción/traducción *in vitro* para la producción de proteínas es una técnica bien establecida. La adaptación de estos sistemas *in vitro* para producir proteínas que tienen aminoácidos no naturales utilizando componentes de sistemas de traducción ortogonal descritos en el presente documento se encuentra bien dentro del ámbito de la invención.

5

CODONES SELECTORES

Los codones selectores de la invención expanden la fase de lectura del codón genético de la maquinaria biosintética de proteínas. Por ejemplo, un codón selector incluye, por ejemplo, un único codón de tres bases, un codón sin sentido, tal como un codón de terminación, por ejemplo, un codón ámbar (UAG), o un codón ópalo (UGA), un codón no natural, al menos un codón de cuatro bases, un codón raro, o similar. Pueden introducirse diversos codones selectores en un gen deseado, por ejemplo, uno o más, dos o más, más de tres, etc. Utilizando diferentes codones selectores, pueden utilizarse múltiples pares ortogonales de ARNt sintetasa para permitir la incorporación con especificidad de sitio simultánea de múltiples aminoácidos no naturales, por ejemplo, incluyendo al menos un aminoácido no natural, utilizando estos codones selectores diferentes.

10

15

En una realización, los métodos pueden implicar el uso de un codón selector que es un codón de terminación para la incorporación de un aminoácido no natural *in vivo* en una célula. Por ejemplo, se produce un O-ARNt que reconoce el codón de terminación y es aminoacilado por una O-RS con un aminoácido no natural. Este O-ARNt no es reconocido por las aminoacil-ARNt sintetasa del hospedador de origen natural. También puede utilizarse mutagénesis de sitio dirigido convencional para introducir el codón de terminación en el sitio de interés en un polinucleótido que codifica un polinucleótido de interés. Véase, por ejemplo, Sayers *et al.* (1988), 5', 3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*, 791-802. Cuando la O-RS, O-ARNt y el ácido nucleico que codifica un polipéptido de interés se combinan, por ejemplo, *in vivo*, el aminoácido no natural se incorpora en respuesta al codón de terminación para dar un polipéptido que contiene el aminoácido no natural en la posición especificada. En una realización de la invención, el codón de terminación utilizado como un codón selector es un codón ámbar, UAG, y/o puede ser un codón ópalo, UGA. En un ejemplo, un código genético en el que se utiliza tanto UAG como UGA como un codón selector puede codificar 22 aminoácidos conservando al mismo tiempo el codón sin sentido ocre UAA, que es la señal de terminación más abundante.

20

25

30

La incorporación de aminoácidos no naturales *in vivo* puede realizarse sin perturbación significativa de la célula hospedadora. Por ejemplo, en células no eucariotas, tales como *Escherichia coli*, como la eficacia de supresión para el codón UAG depende de la competitividad entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y el factor de liberación 1 (RF1) (que se une al codón UAG e inicia la liberación del péptido de crecimiento del ribosoma), la eficacia de supresión puede modularse, por ejemplo, aumentando el nivel de expresión de O-ARNt, por ejemplo, el ARNt supresor, o utilizando una cepa deficiente en RF1. En células eucariotas, dado que la eficacia de supresión de un codón UAG depende de la competitividad entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y el factor de liberación eucariota (por ejemplo, eRF) (que se une a un codón de terminación e inicia la liberación del péptido de crecimiento del ribosoma), la eficacia de supresión puede modularse, 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 ditiotreitil (DTT).

35

40

Los aminoácidos no naturales también pueden codificarse por codones raros. Por ejemplo, cuando la concentración de arginina en una región de síntesis de proteínas *in vitro* se reduce, el codón arginina raro AGG, se ha demostrado que es eficaz para la inserción de Ala mediante un ARNt sintético acilado con alanina. Véase, por ejemplo, Ma *et al.*, *Biochemistry*, 32: 7939 (1993). En este caso, el ARNt sintético compite con el ARNt^{Arg} de origen natural, que existe como una especie minoritaria en *Escherichia coli*. Además, algunos organismos no utilizan todos los codones de triplete. Se ha utilizado un codón AGA no asignado 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). Pueden generarse componentes de la invención para utilizar estos codones raros *in vivo*.

45

50

Los codones selectores también pueden comprender codones extendidos, por ejemplo, codones de cuatro o más bases, tales como, codones de cuatro, cinco, seis o más bases. Los ejemplos de codones de cuatro bases incluyen, por ejemplo, AGGA, CUAG, UAGA, CCCU, y similares. Los ejemplos de codones de cinco bases incluyen, por ejemplo, AGGAC, CCCCUC, CCCUC, CUAGA, CUACU, UAGGC y similares. Los métodos de la invención pueden incluir utilizar codones extendidos basados en supresión del desplazamiento de fase de lectura. Pueden insertarse codones de cuatro o más bases, por ejemplo, uno o múltiples aminoácidos no naturales, 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. Dado que hay 256 codones de cuatro bases posibles, pueden codificarse múltiples aminoácidos no naturales en la misma célula utilizando un codón de cuatro o más bases. Véase también, Anderson *et al.*, (2002) "Exploring the Limits of Codon and Anticodon Size," *Chemistry and Biology* 9: 237-244; y, Magliery, (2001) "Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in *Escherichia coli*," *J. Mol. Biol.*, 307: 755-769.

55

60

65

Por ejemplo, se han utilizado codones de cuatro bases para incorporar aminoácidos no naturales en proteínas utilizando métodos biosintéticos *in vitro*. Véase, por ejemplo, Ma *et al.*, (1993) *Biochemistry* 32: 7939; y Hohsaka *et al.*, (1999) *J. Am. Chem. Soc.*, 121:34. Se utilizaron 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 fase de lectura químicamente acilados. Véase, por ejemplo, Hohsaka *et al.*, (1999) *J. Am. Chem. Soc.*, 121: 12194. En un estudio *in vivo*, Moore *et al.* examinaron la capacidad de los derivados de ARNt^{Leu} con anticodones NCUA para suprimir codones UAGN (N puede ser U, A, G o C), y observaron que el cuadruplete UAGA podía ser descodificado por un ARNt^{Leu} con un anticodón UCUA con una eficacia del 13 al 26 % con poca descodificación en la fase 0 o -1. Véase Moore *et al.*, (2000) *J. Mol. Biol.*, 298: 195. En una realización, en la invención pueden utilizarse codones extendidos basados en codones raros o codones sin sentido, que pueden reducir la lectura completa sin sentido y la supresión de desplazamiento de fase de lectura en otros sitios no deseados. En diversos sistemas ortogonales se han utilizado codones de cuatro bases. Véanse, por ejemplo, los documentos WO 2005/019415; WO 2005/007870 y WO 2005/07624. Véase también, Wang y Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1): 34-66 (2005). Aunque los siguientes ejemplos utilizan un codón selector ámbar, también pueden utilizarse codones de cuatro o más bases, modificando los ejemplos del presente documento para incluir O-ARNt de cuatro bases y sintetasas modificadas para incluir mutaciones similares a las previamente descritas para diversas O-RS de aminoácidos no naturales.

Para un sistema determinado, un codón selector también puede incluir uno de los codones de tres bases naturales, donde el que el sistema endógeno no utiliza (o raramente utiliza) el codón de bases naturales. Por ejemplo, esto incluye un sistema que carece de un ARNt que reconoce el codón de tres bases natural, y/o un sistema en el que el codón de tres bases es un codón raro.

Los codones selectores incluyen opcionalmente pares de bases no naturales. Estos pares de bases no naturales expanden adicionalmente el alfabeto genético existente. Un par de bases adicional aumenta el número de codones de triplete de 64 a 125. Las propiedades de pares de tres bases incluyen un emparejamiento de bases estable y selectivo, una incorporación enzimática eficaz en ADN con alta fidelidad mediante una polimerasa, y la extensión por cebadores continuada eficaz después de la síntesis del par de bases no natural que se genera. Las descripciones de pares de bases no naturales que pueden adaptarse para métodos y composiciones incluyen, por ejemplo, Hiraio, *et al.*, (2002) "An unnatural base pair for incorporating amino acid analogues into protein," *Nature Biotechnology* 20: 177-182. Véase también Wu *et al.*, (2002) *J. Am. Chem. Soc.*, 124: 14626-14630. Más adelante se indican otras publicaciones relacionadas.

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

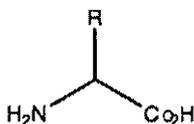
También puede utilizarse un sistema de derivación traduccional para incorporar un aminoácido no natural en un polipéptido no deseado. En un sistema de derivación traduccional, una secuencia grande se inserta en un gen pero no se traduce en una proteína. La secuencia contiene una estructura que sirve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

AMINOÁCIDOS NO NATURALES

Como se usa en el presente documento, un aminoácido no natural se refiere a cualquier aminoácido, aminoácido modificado o análogo de aminoácido diferente de selenocisteína y/o pirrolisina y los siguientes veinte alfa

aminoácidos codificados genéticamente: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. La estructura genérica de un alfa-aminoácido se ilustra en la Fórmula I:

I



5

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

10

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

15

Es de particular interés en el presente documento el aminoácido no natural sulfotirosina (véase la **FIG. 1**). Además del aminoácido no natural sulfotirosina, pueden incorporarse simultáneamente otros aminoácidos no naturales en un polipéptido de interés, por ejemplo, utilizando, un segundo par O-RS/O-ARNt apropiado junto con un par ortogonal proporcionado por la presente invención. Se conocen muchos de dichos aminoácidos no naturales adicionales y pares ortogonales adecuados. Véase, la presente descripción y las referencias citadas en su interior. Por ejemplo, véase Wang y Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1): 34-66 (2005); Xie y Schultz, "An Expanding Genetic Code," *Methods* 36(3): 227-238 (2005); Xie y Schultz, "Adding Amino Acids to the Genetic Repertoire," *Curr. Opinion in Chemical Biology* 9(6): 548-554 (2005); y Wang *et al.*, "Expanding the Genetic Code," *Annu. Rev. Biophys. Biomol. Struct.*, 35: 225-249 (2006).

20

25

En otros aminoácidos no naturales, por ejemplo, el grupo R en la Fórmula I comprende opcionalmente un alquilo, arilo, acilo, hidracina, ciano, halo, hidracida, alquenilo, éter, borato, boronato, fosfo, fosfona, fosfina, enona, imina, éster, hidroxilamina, amina y similares o cualquier combinación de los mismos. Otros aminoácidos no naturales de interés incluyen, pero sin limitación, aminoácidos que comprenden un agente de entrecruzamiento fotoactivable, aminoácidos marcados por espín, aminoácidos fluorescentes, aminoácidos de unión a metal, aminoácidos que contienen metal, aminoácidos radiactivos, aminoácidos con nuevos grupos funcionales, aminoácidos que interactúan de forma covalente o no covalente con otras moléculas, aminoácidos con grupos fotoprotectores y/o fotoisomerizables, aminoácidos que contienen biotina o análogos de biotina, aminoácidos que contienen cetona, aminoácidos glucosilados, un resto de sacárido unido a la cadena lateral del aminoácido, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles o fotoescindibles, aminoácidos con una cadena lateral alargada en comparación con aminoácidos naturales (por ejemplo, poliéteres o hidrocarburos de cadena larga, por ejemplo, de más de aproximadamente 5, de más de aproximadamente 10 carbonos, etc.), aminoácidos que contienen azúcar unido a carbono, aminoácidos que contienen aminoácido, y aminoácidos que contienen uno o más residuos tóxicos.

30

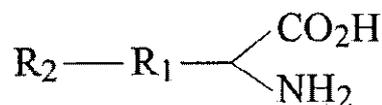
35

40

En otro aspecto, la presente descripción proporciona aminoácidos no naturales que tienen la estructura general ilustrada por la Fórmula IV siguiente:

45

IV

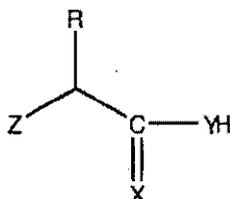


Un aminoácido no natural que tiene esta estructura es normalmente cualquier estructura en la que R₁ es un sustituyente utilizado en uno de los veinte aminoácidos naturales (por ejemplo, tirosina o fenilalanina) y R₂ es un sustituyente. Por tanto, este tipo de aminoácido no natural puede considerarse como un derivado de aminoácido natural.

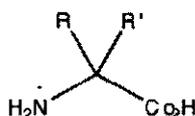
50

Además de aminoácidos no naturales que contienen la estructura sulfotirosina mostrada en la **FIG. 1**, un aminoácido no natural también puede comprender opcionalmente estructuras de cadena principal modificadas, por ejemplo, como se ilustra mediante las estructuras de las Fórmula II y III:

II



III



5

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

20 En algunos aspectos, la invención utiliza aminoácidos no naturales en la configuración L. Sin embargo, no se pretende que la invención se limite al uso de aminoácidos no naturales en la configuración L. Se contempla que en la invención también se utilicen los enantiómeros D de estos aminoácidos no naturales.

Un experto en la técnica reconocerá que puede obtenerse fácilmente una amplia diversidad de análogos no naturales de aminoácidos de origen natural. Por ejemplo, pero sin limitación, se producen fácilmente derivados no naturales de la tirosina. Los análogos de tirosina incluyen, por ejemplo, tirosinas para-sustituidas, tirosinas orto-sustituidas y tirosinas meta-sustituidas, en las que la tirosina sustituida comprende un grupo alquililo, un grupo acetilo, grupo benzoilo, un grupo amino, una hidrazina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo de una cadena lineal C₆-C₂₀ o ramificada, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, o similar. Además, también se contemplan anillos de arilo múltiples sustituidos. Los análogos de glutamina de la presente descripción incluyen, sin limitación, derivados α -hidroxi, derivados γ sustituidos, derivados cíclicos y derivados de glutamina sustituidos con amida. Ejemplos de análogos de fenilalanina incluyen, pero sin limitación, fenilalaninas para-sustituidas, fenilalaninas orto-sustituidas, y fenilalaninas meta-sustituidas, en las que el sustituyente comprende un grupo alquililo, un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído, un nitro, un grupo tiol, un grupo ceto o similar. Ejemplos específicos de aminoácidos naturales no incluyen, pero sin limitación, sulfotirosina, *p*-etil-tiocarbonil-L-fenilalanina, *p*-(3-oxobutanoil)-L-fenilalanina, 1,5-dansilalanina, aminoácido 7-amino-coumarina, aminoácido 7-hidroxi-coumarina, nitrobenzil-serina, O-(2-nitrobenzil)-L-tirosina, *p*-carboximetil-L-fenilalanina, *p*-ciano-L-fenilalanina, *m*-ciano-L-fenilalanina, bifenilalanina, 3-amino-L-tirosina, biperidil alanina, *p*-(2-amino-1-hidroxietil)-L-fenilalanina, *p*-isopropiltiocarbonil-L-fenilalanina, 3-nitro-L-tirosina y *p*-nitro-L-fenilalanina. También una *p*-propargiloxifenilalanina, una 3,4-dihidroxi-L-fenilalanina (DHP), una 3, 4, 6-trihidroxi-L-fenilalanina, una 3,4,5-tri-hidroxi-L-fenilalanina, 4-nitro-fenilalanina, una *p*-acetil-L-fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil) alanina, una 3-metilfenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una 3-nitro-tirosina, una 3-tiol-tirosina, una tri-O-acetil-GlcNAc β -serina, un 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 fosfotirosina, una *p*-yodo-fenilalanina, una *p*-bromofenilalanina, una *p*-amino-L-fenilalanina y una isopropil-L-fenilalanina y similares. Las estructuras de diversos

45

aminoácidos no naturales se describen en las referencias citadas en el presente documento. Véase también, el documento WO 2006/110182, presentado el 27 de octubre de 2005, titulado "ORTHOAGONAL TRANSLATION COMPONENTS FOR THE VIVO INCORPORATION OF UNNATURAL AMINO ACIDS".

5 Síntesis química de aminoácidos no naturales

Muchos de los aminoácidos no naturales proporcionados anteriormente se encuentran disponibles en el comercio, por ejemplo de Sigma (Estados Unidos) o Aldrich (Milwaukee, WI, Estados Unidos). Los que no se encuentran disponibles en el comercio se sintetizan opcionalmente según se proporciona en diversas publicaciones o utilizando métodos convencionales conocidos por los expertos en la materia. Para técnicas de síntesis orgánica, véase, por ejemplo, Organic Chemistry by Fessenden y Fessenden, (1982, Segunda Edición, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Tercera Edición, 1985, Wiley y Sons, Nueva York); y Advanced Organic Chemistry by Carey y Sundberg (tercera edición, partes A y B, 1990, Plenum Press, Nueva York). Publicaciones adicionales que describen la síntesis de aminoácidos no naturales incluyen, por ejemplo, el documento WO 2002/085923 titulado "In vivo incorporation of Unnatural Amino Acids;" Matsoukas *et al.*, (1995) J. Med. Chem., 38, 4660-4669; King y Kidd (1949) "A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates," J. Chem. Soc., 3315-3319; Friedman y Chatterji (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750-3752; Craig *et al.*, (1988) Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmon, M. y Frappier, F. (1991) Glutamine analogues as Potential Antimalarials., Eur. J. Med. Chem. 26, 201-5; Koskinen, 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. y 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 *et al.*, (1987) Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron Lett. 43: 4297-4308; y, Subasinghe *et al.*, (1992) Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35: 4602-7. Véase también, la Publicación Internacional WO 2004/058946, titulada "PROTEIN ARRAYS", presentada el 22 de diciembre del 2003.

Captación celular de aminoácidos no naturales

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

Biosíntesis de aminoácidos no naturales

Ya existen muchas rutas biosintéticas en las células para la producción de aminoácidos y otros compuestos. Aunque un método biosintético para un aminoácido no natural particular puede no existir en la naturaleza, por ejemplo, en una célula, la presente descripción proporciona dichos métodos. Por ejemplo, opcionalmente pueden generarse rutas biosintéticas para aminoácidos no naturales en células hospedadoras añadiendo nuevas enzimas o modificando rutas existentes de la célula hospedadora. Las nuevas enzimas adicionales son opcionalmente enzimas de origen natural o enzimas desarrolladas artificialmente. Por ejemplo, la biosíntesis de *p*-aminofenilalanina (como se presenta en un ejemplo del documento WO 2002/085923) se basa en la adición de una combinación de enzimas conocidas procedente de otros organismos. Los genes para estas enzimas pueden introducirse en una célula transformando la célula con un plásmido que comprenda los genes. Cuando los genes se expresan en la célula, proporcionan una ruta enzimática para sintetizar el compuesto deseado. En los ejemplos que se presentan más adelante se proporcionan ejemplos de los tipos de enzimas que se añaden opcionalmente. Se encuentra secuencias de enzimas adicionales, por ejemplo, en el Genbank. También se añaden opcionalmente enzimas desarrolladas artificialmente a una célula del mismo modo. De esta manera, la maquinaria celular y los recursos de una célula se manipulan para producir aminoácidos no naturales.

De hecho, puede utilizarse cualquiera de diversos métodos para producir nuevas enzimas para su uso en rutas biosintéticas, o para el desarrollo de rutas existentes, para la producción de aminoácidos no naturales *in vitro* o *in vivo*. Para producir aminoácidos no naturales (o, de hecho, para desarrollar sintetisas que tengan nuevas especificidades de sustrato u otras actividades de interés) pueden aplicarse muchos métodos disponibles para

desarrollar enzimas y otros componentes de ruta biosintética. Por ejemplo, la transposición de ADN se utiliza opcionalmente para desarrollar nuevas enzimas y/o rutas de dichas enzimas para la producción de aminoácidos no naturales (o para la 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. Una estrategia relacionada combina familias de genes relacionados (por ejemplo, homólogos) para desarrollar rápidamente enzimas con características deseadas. Un ejemplo de dichos métodos "de combinación de genes de familias" se encuentra en Cramer *et al.* (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature*, 391(6664): 288-291. También pueden generarse nuevas enzimas (componentes de rutas biosintéticas o sintetetasas) utilizando un procedimiento de recombinación de ADN conocido como "truncamiento incremental para la creación de enzimas híbridas" ("ITCHY"), por ejemplo, como se describe en Ostermeier *et al.* (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" *Nature Biotech* 17: 1205. Esta estrategia también puede utilizarse para generar una biblioteca de enzimas u otras variantes de ruta que pueden servir como sustratos para uno o más métodos de recombinación *in vitro* o *in vivo*: Véase también, Ostermeier *et al.* (1999) "Combinatorial Protein Engineering by Incremental Truncation," *Proc. Natl. Acad. Sci. Estados Unidos*, 96: 3562-67, y Ostermeier *et al.* (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts," *Biological and Medicinal Chemistry*, 7: 2139-44. Otra estrategia usa mutagénesis de conjunto exponencial para producir bibliotecas de enzimas u otras variantes de ruta que se seleccionan, por ejemplo, por la capacidad de catalizar una reacción biosintética pertinente para la producción de un aminoácido no natural (o una nueva sintetetasa). En esta estrategia se distribuyen al azar pequeños grupos de restos en una secuencia de interés en paralelo para identificar, en cada posición alterada, aminoácidos que conducen a proteínas funcionales. Pueden encontrarse ejemplos de dichos procedimientos, que pueden adaptarse a la presente invención para producir nuevas enzimas para la producción de aminoácidos no naturales (o nuevas sintetetasas), en Delegrave y Youvan (1993) *Biotechnology Research* 11: 1548-1552. En otra estrategia más, se puede utilizar mutagénesis aleatoria o semi-aleatoria utilizando oligonucleótidos dopados o generados para la producción por ingeniería genética de enzimas y/o componentes de ruta, por ejemplo, utilizando los métodos 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 *et al.* (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" *Methods Enzymol.* 208: 564-86. Otra estrategia más, denominada a menudo mutagénesis "no estocástica", que usa el reensamblaje de polinucleótidos y mutagénesis de saturación de sitio, puede utilizarse para producir enzimas y/o componentes de ruta que después se pueden explorar con respecto a una capacidad para realizar una o más funciones de sintetetasa o 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.

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

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

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

Componentes ortogonales para incorporar aminoácidos no naturales

- La invención proporciona composiciones y métodos como se define en las reivindicaciones para producir componentes ortogonales para incorporar el aminoácido no natural sulfotirosina (véase la **FIG. 1**) en una cadena polipeptídica en crecimiento en respuesta a un codón selector, por ejemplo, un codón de terminación ámbar, 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 utilizarse para incorporar un aminoácido no natural en cadenas polipeptídicas en crecimiento.
- Una composición de la invención incluye una aminoacil ARNt sintetasa ortogonal (O-RS), en la que la O-RS aminoacila preferentemente un O-ARNt con una sulfotirosina. En determinadas realizaciones, la O-RS comprende una secuencia de aminoácidos que comprende las SEC ID Nos: 4, 6, 8 o 10, y variaciones conservativas de la misma. En determinadas realizaciones de la invención, la O-RS aminoacila preferentemente el O-ARNt sobre cualquier ARNt endógeno con un aminoácido no natural particular, en el que la O-RS tiene una predilección por el O-ARNt, y en el que la proporción del O-ARNt cargado con un aminoácido no natural con respecto al ARNt endógeno cargado con el mismo aminoácido no natural es mayor de 1:1, y más preferentemente en el que la O-RS carga el O-ARNt exclusivamente o casi exclusivamente.
- Una composición que incluye una O-RS puede incluir opcionalmente de forma adicional un ARNt ortogonal (O-ARNt), en el que el O-ARNt reconoce un codón selector. Normalmente, un O-ARNt de la invención puede incluir al menos aproximadamente, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 % o un 90 % o más de eficacia de supresión en presencia de una sintetasa afín en respuesta a un codón selector en comparación con el O-ARNt que comprende o está codificado por una secuencia polinucleotídica como se indica en la lista de secuencias y los ejemplos del presente documento. En una realización, la eficacia de supresión de la O-RS y el O-ARNt conjuntamente es, por ejemplo, 5 veces, 10 veces, 15 veces, 20 veces o 25 veces o más veces mayor que la eficacia de supresión del O-ARNt que carece de la O-RS. En algunos aspectos, la eficacia de supresión de la O-RS y del O-ARNt conjuntamente es al menos un 45 % de la eficacia de supresión de un par ortogonal de tirosil ARNt sintetasa derivado de *Methanococcus jannaschii*.
- Una composición que incluye un O-ARNt puede incluir opcionalmente una célula (por ejemplo, una célula de eubacteria, tal como una célula *E. coli* y similar, o una célula eucariota tal como una célula de levadura) y/o un sistema de traducción.
- La invención también proporciona una célula (por ejemplo una célula de eubacteria o una célula de levadura) que comprende un sistema de traducción, como se define en las reivindicaciones, donde el sistema de traducción incluye un ARNt ortogonal (O-ARNt); una aminoacil-ARNt sintetasa ortogonal (O-RS); y un aminoácido no natural sulfotirosina. Normalmente, la O-RS aminoacila preferentemente el O-ARNt sobre cualquier ARNt endógeno con el aminoácido no natural, en el que la O-RS tiene una predilección por el O-ARNt, y en el que la proporción del O-ARNt cargado con el aminoácido no natural con respecto al ARNt endógeno cargado con el aminoácido no natural es mayor de 1:1, y más preferentemente en el que la O-RS carga el O-ARNt exclusivamente o casi exclusivamente. El O-ARNt reconoce el primer codón selector, y la O-RS aminoacila preferentemente el O-ARNt con un aminoácido no natural. En una realización, el O-ARNt comprende o está codificado por una secuencia de polinucleótidos como se indica en la SEC ID N°: 1, o una secuencia de polinucleótidos complementaria de la misma. En una realización, la O-RS comprende una secuencia de aminoácidos como se indica en las SEC ID Nos: 4, 6, 8 o 10, y variaciones conservativas de la misma.
- Una célula de la invención también puede comprender opcionalmente un par de O-ARNt/O-RS diferente adicional y un segundo aminoácido no natural, por ejemplo, donde este O-ARNt reconoce un segundo codón selector y esta O-RS aminoacila preferentemente el O-ARNt correspondiente con el segundo aminoácido no natural, donde el segundo aminoácido es diferente del primer aminoácido no natural. Opcionalmente, una célula de la invención incluye un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, donde el polinucleótido comprende un codón selector que es reconocido por el O-ARNt.
- En determinadas realizaciones, una célula de la invención es una célula de eubacteria (tal como *E. coli*), que incluye un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), un aminoácido no natural, y un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, donde el polinucleótido comprende el codón selector que es reconocido por el O-ARNt. En determinadas realizaciones de la invención, la O-RS aminoacila preferentemente el O-ARNt con el aminoácido no natural con una eficacia que es mayor que la eficacia con la que la O-RS aminoacila cualquier ARNt endógeno.
- En determinadas realizaciones de la invención, un O-ARNt de la invención comprende o está codificado por una secuencia de polinucleótidos como se indica en las listas de secuencias (por ejemplo SEC ID N°: 1) o en los ejemplos del presente documento, o una secuencia de polinucleótidos complementaria de la misma. En determinadas realizaciones de la invención, una O-RS comprende una secuencia de aminoácidos como se indica en la lista 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 como se indica en la lista

de secuencias o en los ejemplos en el presente documento, o una secuencia de polinucleótidos complementaria de la misma.

5 El O-ARNt y/o la O-RS de la invención pueden derivar de cualquiera de diversos organismos (por ejemplo, organismos eucariotas y/o no eucariotas).

10 Como se define en las reivindicaciones, los polinucleótidos son también una característica de la invención. Un polinucleótido de la invención (por ejemplo SEC ID Nos: 5, 7, 9 u 11) incluye un polinucleótido artificial (por ejemplo, fabricado por el hombre y no de origen natural) que comprende una secuencia de nucleótidos que codifica un polipéptido como se indica en la lista de secuencias en el presente documento y/o es complementario a esa secuencia de polinucleótidos. Un polinucleótido de la invención también puede incluir un ácido nucleico que hibrida con un polinucleótido descrito anteriormente, en condiciones de alta rigurosidad, a lo largo de sustancialmente toda la longitud del ácido nucleico. Un polinucleótido de la invención también incluye un polinucleótido que tiene una identidad de, por ejemplo, al menos 75 %, al menos 80 %, al menos 90 %, al menos 95 %, al menos 98 % o más con el de un ARNt de origen natural o un ácido nucleico codificante correspondiente (pero un polinucleótido de la invención es diferente de un ARNt de origen natural o ácido nucleico codificante correspondiente) donde el ARNt reconoce un codón selector, por ejemplo, un codón de cuatro bases. También se incluyen en los polinucleótidos de la invención polinucleótidos artificiales que tienen una identidad de al menos 80 %, al menos 90 %, al menos 95 %, al menos 98 % o más con cualquiera de los anteriores y/o un polinucleótido que comprende una variación conservativa de cualquiera de los anteriores.

25 También son una característica de la invención vectores que comprenden un polinucleótido de la invención, como se define en las reivindicaciones. Por ejemplo, un vector de la invención puede incluir un plásmido, un cósmido, un fago, un virus, un vector de expresión y/o similar. También es una característica de la invención una célula que comprende un vector de la invención.

30 También son características de la invención métodos para producir componentes de un par de O-ARNt/O-RS. Los componentes producidos mediante esos métodos son también una característica de la invención. Por ejemplo, los métodos para producir al menos un ARNt que es ortogonal 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, proporcionando de este modo una biblioteca de O-ARNt potenciales, y someter a selección negativa una primer población de células de una primera especie, donde las células comprenden un miembro de la biblioteca de ARNt potenciales. La selección negativa elimina las células que comprenden un miembro de la biblioteca de O-ARNt potenciales que se aminoacila 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, proporcionando de este modo al menos un O-ARNt. También se proporciona un O-ARNt producido mediante los métodos de la invención.

40 En determinadas realizaciones los métodos comprenden adicionalmente someter a selección positiva una segunda población de células de la primera especie, donde las células comprenden un miembro del conjunto de ARNt que son ortogonales a la célula de la primera especie, una aminoacil-ARNt sintetasa afín, y un marcador de selección positivo. Utilizando la selección positiva se seleccionan o se identifican las células que comprende un miembro del conjunto de ARNt que está aminoacilado por la aminoacil-ARNt sintetasa afín y que muestra una respuesta deseada en presencia del marcador de selección positiva, proporcionando de esta manera un O-ARNt. En determinadas realizaciones, la segunda población de células comprende células que no se han eliminado por la selección negativa.

50 También se proporcionan métodos para identificar una aminoacil-ARNt sintetasa ortogonal que carga un aminoácido natural. Por ejemplo, los métodos incluyen someter a selección una población de células de una primera especie, donde cada una de las células comprende: 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 o como RS procedentes de una especie diferente 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 que comprende al menos un codón selector.

55 Se seleccionan o identifican las células (por ejemplo una célula hospedadora) que muestran una potenciación en la eficacia de supresión en comparación con células que carecen o que tienen una cantidad reducida del miembro de la pluralidad de RS. Estas células seleccionadas/identificadas comprenden una RS activa que aminoacila el O-ARNt. También es una característica de la invención una aminoacil-ARNt sintetasa ortogonal identificada por el método de la invención.

60 También son una característica de la invención métodos para producir una proteína en una célula (por ejemplo, en una célula de eurobacteria, tal como una célula de *E. coli*, o similar, o en una célula de levadura) que tiene el aminoácido no natural en una posición seleccionada. Por ejemplo, un método incluye dejar crecer, en un medio apropiado, una célula, donde la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína, proporcionar el aminoácido no natural, e incorporar el aminoácido no natural en la posición

especificada en la proteína durante la traducción del ácido nucleico con dicho al menos un codón selector, produciendo de este modo la proteína. La célula comprende adicionalmente: un ARNt ortogonal (O-ARNt) que funciona en la célula y que reconoce el codón selector; y, una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente el O-ARNt con el aminoácido no natural. Una proteína producida mediante este método también es una característica de la invención. De particular interés son métodos para producir la forma sulfatada de hirudina, que se utiliza un anticoagulante.

La invención también proporciona composiciones que incluyen proteínas, donde las proteínas comprenden sulfotirosina. En determinadas realizaciones, la proteína puede comprender una secuencia de aminoácidos que tiene una identidad de al menos un 75 % con la de una proteína conocida, por ejemplo, hirudina, una proteína terapéutica, una proteína de diagnóstico, una enzima industrial o una parte de la misma. Opcionalmente, la composición comprende un vehículo farmacéuticamente aceptable.

SECUENCIAS DE ÁCIDOS NUCLEICOS Y DE POLIPÉPTIDOS Y VARIANTES

Como se describe en el presente documento, la invención proporciona secuencias de polinucleótidos como se define en las reivindicaciones que codifican, por ejemplo, los ARNt y las O-RS y secuencias de aminoácidos de polipéptidos, por ejemplo, O-RS y, por ejemplo, composiciones, sistemas y métodos que comprenden dichas secuencias de polinucleótidos o polipéptidos. En el presente documento se describen ejemplos de dichas secuencias, por ejemplo, O-ARNt y O-RS y secuencias de aminoácidos y nucleótidos (véase la **FIG. 7**, por ejemplo, SEC ID Nos: 1 y 4-11). Sin embargo, un experto en la técnica apreciará que la invención no se limita a estas secuencias descritas en el presente documento, por ejemplo, en los Ejemplos y en la lista de secuencias. Un experto apreciará que la invención también proporciona muchas secuencias relacionadas con las funciones descritas en el presente documento, por ejemplo, polinucleótidos y polipéptidos que codifican variantes conservativas de una O-RS descrita en este documento.

La construcción y el análisis de especies de sintetasa ortogonales (O-RS) que son capaces de aminoacilar un O-ARNt con una sulfotirosina se describen en el Ejemplo 1. Este Ejemplo describe la construcción y análisis de las especies de O-RS que son capaces de incorporar el aminoácido no natural sulfotirosina.

La invención proporciona polipéptidos (O-RS) y polinucleótidos, por ejemplo, O-ARNt, polinucleótidos que codifican O-RS o partes de las mismas, oligonucleótidos utilizados 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 como se indica en las SEC ID Nos: 5, 7, 9 u 11, y un polinucleótido que es complementario a, o que codifica, una secuencia polinucleotídica del mismo. Un polinucleótido de la invención también incluye cualquier polinucleótido que codifique una secuencia de aminoácidos O-RS que comprende las SEC ID Nos: 4, 6, 8 o 10. De manera similar, un ácido nucleico artificial que hibrida con un polinucleótido indicado anteriormente en condiciones de alta rigurosidad a lo largo de sustancialmente toda la longitud de un ácido nucleico (y es distinto de un polinucleótido de origen 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 antisueros específicamente inmunoreactivos con un polipéptido de la invención. Un polinucleótido artificial es un polinucleótido que está fabricado por el hombre y que no es de origen natural.

Un polinucleótido de la invención también incluye un polinucleótido artificial que tiene una identidad de, por ejemplo, al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 % o mayor con un ARNt de origen natural (pero es diferente de un ARNt de origen natural) o cualquier ARNt o ácido nucleico codificante del mismo en una lista o ejemplo del presente documento. Un polinucleótido también incluye un polinucleótido artificial que tiene una identidad de, por ejemplo, al menos 75 %, al menos 80 %, al menos 90 %, al menos 95 %, al menos 98 % o mayor (pero no 100 % de identidad) con la de un ARNt de origen natural.

En determinadas 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 de expresión puede ser un vector de expresión. En otra realización el vector de expresión puede incluir un promotor unido operativamente a uno o más polinucleótidos de la invención. En otra realización, una célula comprende un vector que incluye un polinucleótido de la invención.

Un experto en la técnica también apreciará que en la invención están incluidas muchas variantes de las secuencias descritas. Por ejemplo, en la invención están incluidas variaciones conservativas de las secuencias descritas que proporcionan una secuencia funcionalmente similar. Se considera que en la invención están incluidas variantes de las secuencias polinucleotídicas de ácido nucleico, en las que las variantes hibridan con al menos una secuencia descrita y reconocen un codón selector. En la invención también se incluyen sub-secuencias únicas de las secuencias descritas en el presente documento como se determina, por ejemplo, mediante técnicas de comparación de secuencias convencionales.

Variaciones conservativas

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

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

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

Sustituciones de aminoácidos conservativas

Cadenas Laterales No Polares y/o Alifáticas	Cadenas Laterales Polares, No Cargadas	Cadenas Laterales Aromáticas	Cadenas Laterales Cargadas Positivamente	Cargadas Laterales Cargadas Negativamente
Glicina Alanina Valina Leucina Isoleucina Prolina	Serina Treonina Cisteína Metionina Asparagina Glutamina	Fenilalanina Tirosina Triptófano	Lisina Arginina Histidina	Aspartato Glutamato

Hibridación de ácidos nucleicos

Para identificar ácidos nucleicos de la invención puede utilizarse hibridación comparativa, incluyendo variaciones conservativas de ácidos nucleicos de la invención, y este método de hibridación comparativo es un método preferido para diferenciar ácidos nucleicos de la invención. Además son una característica de la invención los ácidos nucleicos diana que hibridan con un ácido nucleico representado por las SEC ID Nos: 5, 7, 9 u 11, en condiciones de rigurosidad alta, ultra-alta y ultra ultra-alta. Los ejemplos de dichos ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones de ácido nucleico conservativas o silenciosas en comparación con una secuencia de ácido nucleico dada.

Se dice que un ácido nucleico de ensayo hibrida específicamente con un ácido nucleico sonda cuando hibrida al menos el 50 % de bien con la sonda como con la diana complementaria perfectamente emparejada, es decir, con una relación entre señal e interferencia de al menos la mitad de alta que la hibridación de la sonda con la diana en condiciones en las que la sonda perfectamente emparejada se une a la diana complementaria perfectamente coincidente con una relación entre señal interferencia que es al menos de aproximadamente 5 veces a 10 veces tan alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no emparejados.

Los ácidos nucleicos “hibridan” cuando se asocian, normalmente en solución. Los ácidos nucleicos hibridan debido a diversas fuerzas físicoquímicas bien caracterizadas, tales como formación de enlaces de hidrógeno, exclusión de disolvente, apilamiento de bases y similares. Una amplia guía para la hibridación de ácidos nucleicos se encuentra

en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, Parte I, Capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, Nueva York), así como en *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., Current Protocols, una sociedad conjunta entre Greene Publishing Associates, Inc. y John Wiley y Sons, Inc., (adscrita a lo largo del 2006); Hames y Higgins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, England; y Hames y Higgins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, England, proporcionan detalles sobre la síntesis, marcaje, detección y cuantificación de ADN y ARN, incluyendo oligonucleótidos.

Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 restos complementarios en un filtro en una transferencia de Southern o Northern es formalina al 50 % con 1 mg de heparina a 42 °C, realizándose la hibridación durante la noche. Un ejemplo de condiciones de lavado rigurosas es un lavado con SSC 0,2x a 65 °C durante 15 minutos (para una descripción de tampón SSC véase, Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3ª Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 2001). A menudo, el lavado de alta rigurosidad está precedido por un lavado de baja rigurosidad para eliminar la señal de fondo de la sonda. Un ejemplo del lavado de baja rigurosidad es SSC 2x a 40 °C durante 15 minutos. En general, una relación entre señal e interferencia 5 veces mayor (o superior) que la observada para una sonda no relacionada en el ensayo de hibridación particular indica detección de una hibridación específica.

Las "condiciones rigurosas de lavado de hibridación" en el contexto de los experimentos de hibridación de ácidos nucleicos tales como hibridaciones de Southern y Northern son dependientes de la secuencia y son diferentes con diferentes parámetros ambientales. Se encuentra una guía extensa sobre la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, Parte I, Capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, Nueva York); Hames y Higgins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, England; y Hames y Higgins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, England. Las condiciones de hibridación y lavado rigurosas pueden determinarse de forma sencilla empíricamente para cualquier ácido nucleico de ensayo. Por ejemplo, para determinar condiciones de hibridación y lavado rigurosas, las condiciones de hibridación y lavado se aumentan gradualmente (por ejemplo, aumentando la temperatura, disminuyendo la concentración de sal, aumentando la concentración de detergente y/o aumentando la concentración de disolventes orgánicos tales como formalina en la hibridación o el lavado), hasta que se cumpla un conjunto de criterios seleccionado. Por ejemplo, en condiciones de hibridación y lavado de alta rigurosidad, las condiciones de hibridación y lavado se aumentan gradualmente hasta que una sonda se une a una diana complementaria perfectamente emparejada con una relación entre señal e interferencia que es al menos 5 veces la observada para la hibridación de la sonda con una diana no emparejada.

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

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

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

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

Subsecuencias únicas

- En algunos aspectos, la presente descripción proporciona un ácido nucleico que comprende una subsecuencia única en un ácido nucleico seleccionado de las secuencias de O-ARNt y O-RS descritas en el presente documento. La subsecuencia única es única en comparación con un ácido nucleico correspondiente a cualquier secuencia de ácido nucleico de O-ARNt u O-RS conocida previamente. Se puede realizar el alineamiento utilizando, por ejemplo, un ajuste BLAST para parámetros por defecto. Cualquier subsecuencia única es útil, por ejemplo, como una sonda para identificar los ácidos nucleicos de la invención.
- De manera similar, la presente invención incluye un polipéptido que comprende una subsecuencia única en un polipéptido seleccionado de las secuencias de las O-RS descritas en el presente documento. En este caso, la subsecuencia única es única en comparación con un polipéptido correspondiente a cualquier secuencia de RS conocida previamente.
- La invención también proporciona ácidos nucleicos diana que hibridan en condiciones rigurosas con un único oligonucleótido codificante que codifica una subsecuencia única en un polipéptido seleccionado de las secuencias de O-RS donde la subsecuencia única es única en comparación con un polipéptido correspondiente a cualquiera de los polipéptidos de control (por ejemplo, secuencias parentales de las que se obtuvieron las sintetasas de la invención, por ejemplo mediante mutación). Las secuencias únicas se determinan como se ha señalado anteriormente.

Comparación, identidad y homología de secuencia

- Las expresiones “idéntico” o “porcentaje de identidad”, en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas, se refieren a dos o más secuencias o subsecuencias que son iguales o que tienen un porcentaje especificado de restos de aminoácidos o nucleótidos que son iguales cuando se comparan y alinean para la máxima correspondencia, como se mide utilizando uno de dos algoritmos de comparación de secuencia que se describen a continuación (u otros algoritmos disponibles para los expertos en la materia) o mediante inspección visual.
- La expresión “sustancialmente idéntico”, en el contexto de dos ácidos nucleicos o polipéptidos (por ejemplo, ADN que codifican un O-ARNt o una O-RS, o la secuencia de aminoácidos de una O-RS), se refiere a dos o más secuencias o subsecuencias que tienen una identidad de restos de aminoácidos o nucleótidos de al menos aproximadamente 60 %, aproximadamente 80 %, aproximadamente 90-95 %, aproximadamente 98 %, aproximadamente 99 % o más, cuando se comparan y se alinean para la máxima correspondencia, como se mide utilizando un algoritmo de comparación de secuencias o mediante inspección visual. Dichas secuencias “sustancialmente idénticas” se consideran normalmente “homólogas”, sin referencia a la ascendencia real. Preferentemente, la “identidad sustancial” existe a lo largo de una región de las secuencias que tiene una longitud de al menos aproximadamente 50 restos, más preferentemente a lo largo de una región de al menos aproximadamente 100 restos, y mucho más preferentemente, las secuencias son sustancialmente idénticas a lo largo de al menos aproximadamente 150 restos, o a lo largo de la longitud completa de las dos secuencias a comparar.

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

- Para la comparación de secuencias y determinación de homología, normalmente una secuencia actúa como una secuencia de referencia con la que se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencias, se introducen las secuencias de ensayo y de referencia en un ordenador, se diseñan las coordenadas de subsecuencia, si es necesario, y se diseñan parámetros de programa de algoritmo de secuencia. Después, el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencia para la secuencia (o secuencias) de ensayo con respecto a las secuencias de referencia basándose en los parámetros de programas diseñados.

65

Se puede realizar un alineamiento óptimo de secuencias para la comparación, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math* 2: 482 (1981), mediante el algoritmo de alineamiento de homología de Needleman y Wunsch, *J. Mol. Biol.*, 48: 443 (1970), mediante el método de búsqueda de similitud de Pearson y Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444 (1988), mediante aplicaciones informáticas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Paquete de Software de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante inspección visual (véase en general *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., *Current Protocols*, una sociedad conjunta entre Greene Publishing Associates, Inc. y John Wiley y Sons, Inc., adscrita a lo largo del 2006).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y similitud de secuencia es el algoritmo BLAST, que se describe en Altschul *et al.*, *J. Mol. Biol.*, 215: 403-410 (1990). El software para realizar análisis BLAST está disponible públicamente por el National Center for Biotechnology Information. Este algoritmo implica identificar en primer lugar pares de secuencias con alta puntuación (HSP, *High Scoring Sequence Pairs*) identificando palabras cortas de longitud *W* en la secuencia problema, que se ajustan o satisfacen alguna puntuación umbral *T* valorada positivamente cuando se alinean con una palabra de la misma longitud en una secuencia de bases de datos. Se denomina *T* al umbral de puntuación de la palabra vecina (Altschul *et al.*, *J. Mol. Biol.*, 215: 403-410 (1990). Estos aciertos de palabra vecina iniciales actúan como semillas para iniciar búsquedas para encontrar los HSP más largos que los contengan. Después, los aciertos de palabras se extienden en ambas direcciones a lo largo de cada secuencia hasta que se pueda aumentar la puntuación de alineamiento acumulativa. Las puntuaciones acumulativas se calculan utilizando, para secuencias de nucleótidos, los parámetros *M* (puntuación de recompensa para un par de restos coincidentes; siempre > 0) y *N* (puntuación de penalización para restos no coincidentes; siempre < 0). Para secuencias de aminoácidos se utiliza una matriz de puntuación para calcular la puntuación acumulativa. La extensión de los aciertos de palabra en cada dirección se detiene cuando: la puntuación del alineamiento acumulativa cae en la cantidad *X* desde su valor máximo conseguido; la puntuación acumulativa tiende a cero o a un valor inferior, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se alcanza el extremo de cualquier secuencia. Los parámetros *W*, *T* y *X* del algoritmo BLAST determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto una longitud de palabra (*W*) de 11, una esperanza (*E*) de 10, un límite de 100, *M*=5, *N*=-4 y una comparación de las dos cadenas. Para secuencias de aminoácidos, el programa BLASTP utiliza por defecto una longitud de palabra (*W*) de 3, una esperanza (*E*) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89: 10915).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST realiza un análisis estadístico de la similitud entre dos secuencias, (véase, por ejemplo, Karlin y Altschul, *Proc. Nat'l. Acad. Sci. USA* 90: 5873-5787 (1993)). Una medición de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (*P(N)*), que proporciona una indicación de la probabilidad con la que ocurriría por casualidad una coincidencia entre dos secuencias de nucleótidos o aminoácidos. Por ejemplo, se considera que un ácido nucleico es similar a una secuencia de referencia sin la probabilidad de suma más pequeña en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es menor de aproximadamente 0,1, más preferentemente menos de aproximadamente 0,01 y mucho más preferentemente menor de aproximadamente 0,001.

Mutagénesis y otras técnicas de biología molecular

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

En la invención se utilizan diversos tipos de mutagénesis, por ejemplo, para mutar moléculas de ARNt, para producir bibliotecas de ARNt, para producir bibliotecas de sintetasas, o para insertar codones selectores que codifican un aminoácido no natural en una proteína o un polipéptido de interés. Estos incluyen, pero sin limitación, mutagénesis puntual aleatoria dirigida, recombinación homóloga, transposición de ADN u otros métodos de mutagénesis recursiva, construcción quimérica, mutagénesis utilizando moldes que contienen uracilo, mutagénesis de sitio dirigido por oligonucleótido, mutagénesis de ADN modificado por fósforotioato, mutagénesis utilizando ADN dúplex discontinuo o similares o cualquier combinación de los mismos. Otros métodos adecuados incluyen reparación puntual de emparejamiento erróneo, mutagénesis utilizando cepas de hospedador deficientes en reparación, selección por restricción y purificación por restricción, mutagénesis por delección, mutagénesis por síntesis génica total, reparación de roturas de cadena doble o similar. También se incluye en la presente invención, por ejemplo, la mutagénesis que implica construcciones quiméricas. En una realización, la mutagénesis puede dirigirse por una información conocida de la molécula de origen natural o de la molécula natural alterada o mutada, por ejemplo, por

información de la secuencia, de comparaciones de secuencias, de propiedades físicas, de la estructura cristalina o similar.

5 Las células hospedadoras se crean por ingeniería genética (por ejemplo, se transforman, transducen o transfectan) con los polinucleótidos de la invención o con construcciones que incluyen un polinucleótido de la invención, por ejemplo, un vector de la invención, que puede ser, por ejemplo, un vector de clonación o un vector de expresión. Por ejemplo, las regiones codificantes para el ARNt ortogonal, la ARNt sintetasa ortogonal y la proteína que se tiene que derivatizar están unidas operativamente a elementos de control de la expresión génica que son funcionales en la célula hospedadora deseada. Los vectores típicos contienen terminadores de la transcripción y traducción, secuencias de inicio de la transcripción y 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 procariontes o en ambos (por ejemplo, vectores lanzadera) y marcadores de selección tanto para sistemas procariontes como para sistemas eucariotas. Los vectores son adecuados para replicación y/o integración en procariontes, eucariotas o, preferentemente, en ambos. Véase Gilman y Smith, *Gene* 8: 81 (1979); Roberts, *et al.*, *Nature*, 328: 731 (1987); Schneider *et al.*, *Protein Expr. Purif.*, 6435: 10 (1995); Berger y Kimmel, "Guide to Molecular Cloning Techniques," *Methods in Enzymology*, volume 152 Academic Press, Inc., San Diego, CA; Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3^a ed.), vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 2001, and *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., *Current Protocols*, una sociedad conjunta entre Greene Publishing Associates, Inc. y John Wiley y Sons, Inc., (adscrita a lo largo del 2006). 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 por métodos convencionales incluyendo electroporación (From *et al.*, *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infección por vectores virales, penetración balística a gran velocidad con partículas pequeñas con el ácido nucleico en el interior de la matriz de pequeñas perlas o partículas o sobre la superficie (Klein *et al.*, *Nature* 327, 70-73 (1987)), y/o similares.

30 Para la incorporación específica de sitio de aminoácidos naturales en proteínas sin respuesta al codón de terminación ámbar (UAG) en *E. coli*, se desarrolló un sistema de un solo plásmido versátil y altamente eficaz. En el nuevo sistema, el par ARN^{tyr}(CUA) y tirosil-ARNt sintetasa supresor de *M. jannaschii* se codifica en un solo plásmido, que es compatible con la mayoría de los vectores de expresión de *E. coli*. Se construyó un operón de ARNt monocistrónico bajo el control del promotor proK y el terminador para una estructura secundaria óptima y procesamiento de ARNt. La introducción de una forma mutada del promotor glnS para la sintetasa dio como resultado un aumento significativo de la eficacia de supresión y fidelidad. También se obtuvieron aumentos en la eficacia de supresión por copias múltiples de gen de ARNt así como por mutación específica (D286R) en la sintetasa (Kobayashi *et al.*, "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion," *Nat. Struct. Biol.*, 10(6): 425-432 [2003]). La generalidad del sistema optimizado también se demostró mediante una incorporación altamente eficaz y exacta de diversos aminoácidos no naturales diferentes, cuyas utilidades exclusivas en el estudio de la función y estructura de las proteínas se aprobaron previamente.

40 La ATCC proporciona un catálogo de Bacterias y Bacteriófagos útiles para la clonación, por ejemplo, el Catálogo de la ATCC de Bacterias y Bacteriófagos (1996) Ghema *et al.* (eds) publicado por la ATCC. También se encuentran procedimientos básicos adicionales para la secuenciación, la clonación y otros aspectos de biología molecular y consideraciones teóricas fundamentales en Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3^a ed.), vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 2001; *Current Protocols in Molecular Biology*, Ausubel *et al.*, eds., *Current Protocols*, una sociedad conjunta entre Greene Publishing Associates, Inc. y John Wiley y Sons, Inc., (adscrita a lo largo del 2006), y en Watson *et al.* (1992) *Recombinant DNA*, Segunda Ed., Scientific American Books, NY. Además, esencialmente cualquier ácido nucleico (y prácticamente cualquier ácido nucleico marcado, convencional o no convencional) puede encargarse a petición del cliente o de forma convencional en cualquiera de diversas fuentes comerciales, tal como la Midland Certified Reagent Company, The Great American Gene Company (Ramona, CA), ExpressGen Inc. (Chicago, IL), Operon Technologies Inc. (Alameda, CA) y muchas otras.

55 Las células hospedadoras creadas por ingeniería genética se pueden cultivar en medios nutrientes convencionales modificados de forma apropiada para actividades tales como, por ejemplo, etapas de identificación, activación de promotores o selección de transformantes. Estas células pueden cultivarse opcionalmente en organismos transgénicos. Otras referencias útiles, por ejemplo, para el aislamiento y cultivo de células (por ejemplo, para el aislamiento posterior de ácidos nucleicos) incluyen Freshney (1994) *Culture of Animal Cells*, a *Manual of Basic Technique*, tercera edición, Wiley Liss, Nueva York y las referencias citadas en este documento; Payne *et al.* (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley y 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

65 Como se define en las reivindicaciones también constituyen una característica de la invención los métodos para

producir una proteína en una célula con un aminoácido no natural en una posición específica. Por ejemplo, un método 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 no natural; en el que la célula comprende adicionalmente; un ARNt ortogonal (O-ARNt) que funciona en la célula y conoce el codón selector; y, una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente el O-ARNt con el aminoácido no natural. Es también una característica de la invención una proteína producida por este método.

En determinadas realizaciones, la O-RS comprende una predilección por la aminoacilación del O-ARNt afín sobre cualquier ARNt endógeno en un sistema de expresión. La proporción relativa entre O-ARNt y el ARNt endógeno que está cargado por la O-RS, cuando el O-ARNt y la O-RS están presentes a concentraciones equimolares, puede ser mayor de 1:1, preferentemente al menos aproximadamente de 2:1, más preferentemente de 5:1, aún más preferentemente de 10:1, aún más preferentemente de 20:1, aún más preferentemente de 50:1, aún más preferentemente de 75:1, aún más preferentemente de 95:1, 98:1, 99:1, 100:1, 500:1, 1.000:1, 5.000:1 o una proporción mayor.

La invención también proporciona composiciones que incluyen proteínas, en la que las proteínas comprenden un aminoácido no natural. En determinadas realizaciones, la proteína puede comprender una secuencia de aminoácidos que tiene una identidad de al menos 75 % con la de una proteína terapéutica, una proteína de diagnóstico, una enzima industrial o una parte de las mismas.

Las composiciones de la invención y las composiciones fabricadas mediante los métodos de la invención están presentes opcionalmente en una célula. Los pares O-ARNt/O-RS o los componentes individuales de la invención pueden por tanto utilizarse en una maquinaria de traducción de un sistema hospedador, lo que da como resultado la incorporación un aminoácido no natural en una proteína. Los Números de Publicación Internacional WO 2004/094593, presentada el 16 de abril de 2004, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE", y WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", describen este proceso. Por ejemplo, cuando un par O-ARNt/O-RS se introduce en un hospedador, por ejemplo, una célula de *Escherichia coli*, el par conduce a la incorporación *in vivo* de un aminoácido no natural, tal como sulfotirosina, en una proteína en respuesta a un codón selector. El aminoácido no natural que se añade al sistema puede ser un aminoácido sintético, tal como un derivado de una fenilalanina o tirosina, que puede añadirse exógenamente al medio de cultivo. Opcionalmente, las composiciones de la presente invención pueden estar en un sistema de traducción *in vitro* o en un sistema (o sistemas) *in vivo*.

Una célula de la invención proporciona la capacidad de sintetizar proteínas que comprenden aminoácidos no naturales en grandes cantidades útiles. En algunos aspectos, 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 no natural o una cantidad que puede conseguirse con los métodos de producción de proteína *in vivo* (en el presente documento se proporcionan detalles de la producción y purificación de proteínas recombinantes). 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, por ejemplo, en un lisado celular, un tampón, un tampón farmacéutico u otra suspensión líquida (por ejemplo, en un volumen de, por ejemplo, cualquiera de aproximadamente 1 nl a aproximadamente 100 l). Una característica de la presente descripción es la producción de grandes cantidades (por ejemplo, superiores a las normalmente posibles con otros métodos, por ejemplo, traducción *in vitro*) de una proteína en una célula que incluye al menos un aminoácido no natural.

La incorporación de un aminoácido no natural puede realizarse, por ejemplo, para ajustar cambios en la estructura y/o función de la proteína, por ejemplo, para cambiar el tamaño, la acidez, la nucleofilia, la formación de enlaces de hidrógeno, la hidrofobia, la accesibilidad de sitios diana de proteasa, el direccionamiento a un resto (por ejemplo, para una matriz de proteínas), la incorporación de marcadores o grupos reactivos, etc. Las proteínas que incluyen un aminoácido no natural pueden tener propiedades catalítica o físicas mejoradas o incluso completamente nuevas. Por ejemplo, mediante la inclusión de un aminoácido no natural en una proteína se modifican opcionalmente las siguientes propiedades: toxicidad, biodistribución, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, capacidad catalítica, semivida (por ejemplo, semivida en suero), capacidad de reaccionar con otras moléculas, por ejemplo, de forma covalente o no covalente y similares. Las composiciones que incluyen proteínas que incluyen al menos un aminoácido no natural son útiles para, por ejemplo, nuevas enzimas terapéuticas, de diagnóstico, catalíticas, industriales, proteínas de unión (por ejemplo, anticuerpos), y por ejemplo, para el estudio de la estructura y función de las proteínas. Véase, por ejemplo, Dougherty, (2000) Unnatural Amino Acids as Probes of Protein Structure and Function, *Current Opinion in Chemical Biology*, 4: 645-652.

En algunos aspectos de la presente descripción, una composición incluye al menos una proteína con al menos una, por ejemplo, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos

ocho, al menos nueve o al menos diez o más aminoácidos no naturales. Los aminoácidos no naturales pueden ser iguales o diferentes, por ejemplo, puede haber 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más sitios diferentes en la proteína que comprenden 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más aminoácidos no naturales diferentes. En otro aspecto, una composición incluye una proteína en la que al menos uno, pero no todos, los aminoácidos particulares presentes en la proteína es un aminoácido no natural. Para una proteína determinada 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 de los mismos aminoácidos no naturales). Para una proteína determinada con más de dos aminoácidos no naturales, los aminoácidos no naturales pueden ser iguales, diferentes o una combinación de un aminoácido no natural múltiple del mismo tipo con al menos un aminoácido no natural diferente.

Utilizando las composiciones y los métodos del presente documento puede producirse esencialmente cualquier proteína (o parte de la misma) que incluya un aminoácido no natural (y cualquier ácido nucleico codificante correspondiente, que incluya, por ejemplo, uno o más codones selectores). No se ha intentado identificar los cientos de miles de proteínas conocidas que pueden modificarse para incluir uno o más aminoácidos no naturales, por ejemplo, adaptando cualquiera de los métodos de mutación disponibles para incluir uno o más codones selectores apropiados en un sistema de traducción pertinente. Los archivos depositarios de secuencias comunes para proteínas conocidas incluyen el GenBank EMBL, el DDBJ y el NCBI. Buscando en internet pueden identificarse fácilmente otros archivos depositarios.

Normalmente, las proteínas tienen una identidad de, por ejemplo, al menos 60 %, al menos 70 %, al menos 75 %, al menos 80 %, al menos 90 %, al menos 95 % o al menos 99 % o más con cualquiera de las proteínas disponibles (por ejemplo, una proteína terapéutica, una proteína de diagnóstico, una enzima industrial o una parte de la misma, y similares), y comprenden uno o más aminoácidos no naturales. Ejemplos de proteínas terapéuticas, de diagnóstico y otras proteínas que pueden modificarse para que comprendan uno o más aminoácidos no naturales pueden encontrarse, pero sin limitación, en las Publicaciones Internacionales WO 2004/094593, presentada el 16 de abril de 2004, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE"; y WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Ejemplos de proteínas terapéuticas, de diagnóstico y otras proteínas que pueden modificarse para que comprendan uno o más aminoácidos naturales incluyen, pero sin limitación, por ejemplo, hirudina, alfa-1 antitripsina, 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 de Complemento, inhibidor del complemento, receptor 1 del complemento, citocinas, (por ejemplo, péptido 78 activador de neutrófilos epiteliales, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-18, MCP-1), Factor de Crecimiento Epidérmico (EGF), Eritropoyetina ("EPO"), toxinas exfoliantes A y B, Factor IX, Factor VII, Factor VIII, Factor X, Factor de Crecimiento de Fibroblastos (FGF), Fibrinógeno, Fibronectina, G-CSF, GM-CSF, glucocerebrosidasa, gonadotropina, factores de crecimiento, proteínas erizo (por ejemplo, Sonic, Indian, Desert), hemoglobina, Factor de Crecimiento de Hepatocitos (HGF), hirudina, albúmina de suero humano, insulina, factor de crecimiento similar a insulina (IGF), interferones (por ejemplo, IFN- α , IFN- β , IFN- γ), 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, neurturina, factor inhibidor de neutrófilos (NIF), oncostatina M, proteína osteogénica, hormona paratiroidea, PD-ECSF, PDGF, hormonas peptídicas (por ejemplo, Hormona de Crecimiento Humano), pleiotropina, Proteína A, Proteína G, exotoxinas Pirogénicas A, B y C, relaxina, renina, SCF, receptor I de complemento Soluble, I-CAM 1 Soluble, receptores solubles de interleucina (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), receptor soluble de TNF, somatomedina, somatostatina, somatotropina, estreptoquinasa, superantígenos, es decir, enterotoxinas de estafilococos (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutasa (SOD), toxina de síndrome de choque tóxico (TSST-1), timosina alfa 1, activador de plasminógeno tisular, factor beta de necrosis tumoral (TNF beta), receptor del factor de necrosis tumoral (TNFR), factor alfa de necrosis tumoral (TNF alfa), Factor de Crecimiento del Endotelio Vascular (VEGEF), uroquinasa y muchas otras.

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

- Una clase de proteínas de la presente descripción (por ejemplo, proteínas con uno o más aminoácidos no naturales) incluyen proteínas biológicamente activas tales como hirudina, 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- α , TGF- β , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, e hialurina/CD44; moléculas de transducción de señales y productos oncogénicos correspondientes, por ejemplo, Mos, Ras, Raf y Met; y activadores y supresores de la transcripción, por ejemplo, p53, Tat, Fos, Myc, Jun, Myb, Rel y receptores de hormonas esteroideas tales como los receptores de estrógeno, progesterona, testosterona, aldosterona, el ligando del receptor de LDL y corticosterona.
- 10 La descripción también proporciona enzimas (por ejemplo enzimas industriales) o partes de las mismas con al menos un aminoácido no natural. Los ejemplos de enzimas incluyen, pero sin limitación, por ejemplo, amidasas, aminoácido racemasas, acilasas, deshalogenasas, dioxigenasas, diarilpropano peroxidasas, epimerasas, epóxido hidrolasas, estererasas, isomerasas, quinasas, glucosa isomerasas, glucosidasas, glucosil transferasas, haloperoxidasas, monooxigenasas (por ejemplo p450s), lipasas, lignina peroxidasas, nitrilo hidratasas, nitrilasas, proteasas, fosfatasas, subtilisinas, transaminasas y nucleasas.
- 15 Muchas de estas proteínas están disponibles en el mercado (Véase, por ejemplo, el catálogo y la lista de precios de Sigma BioSciences 2002) y las secuencias de proteínas y genes correspondientes y, normalmente, muchas variantes de las mismas, son muy conocidas (véase, por ejemplo, Genbank). Cualquiera de ellas pueden modificarse mediante la inserción de uno o más aminoácidos no naturales de acuerdo con la invención, por ejemplo, para alterar la proteína con respecto a una o más propiedades terapéuticas, de diagnóstico o enzimáticas de interés. Los ejemplos de propiedades terapéuticamente pertinentes incluyen semivida en suero, semivida de conservación, estabilidad, inmunogenicidad, actividad terapéutica, detectabilidad (por ejemplo, mediante la inclusión de grupos indicadores (por ejemplo, marcadores o sitios de unión a marcadores) en los aminoácidos no naturales), reducción de DL₅₀ u otros efectos secundarios, capacidad de entrar en el organismo a través del tracto gástrico (por ejemplo, disponibilidad oral) o similares. Los ejemplos de las propiedades de diagnóstico incluyen semivida de conservación, estabilidad, actividad enzimática, capacidad de producción o similares.
- 20 Utilizando las composiciones y los métodos de la invención también puede modificarse diversas proteínas diferentes para incluir uno o más aminoácidos no naturales. Por ejemplo, la invención puede incluir sustituir uno o más aminoácidos naturales en una o más proteínas de vacuna con un aminoácido no natural, por ejemplo, en proteínas de hongos infecciosos, por ejemplo *Aspergillus*, especies de *Candida*; bacterias, particularmente *E. coli*, que sirve como modelo de bacterias patógenas, así como bacterias médicamente importantes tales como *Staphylococci* (por ejemplo *aureus*), o *Streptococci* (por ejemplo *pneumoniae*); protozoos tales como esporozoos (por ejemplo *Plasmodia*), rizópodos (por ejemplo *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); virus tales como virus de ARN (+) (los ejemplos incluyen Poxvirus, por ejemplo, *vaccinia*; Picornavirus, por ejemplo, *polio*; Togavirus, por ejemplo, *rubeola*; Flavivirus, por ejemplo, VHC; y Coronavirus), virus de ARN (-) (por ejemplo, Rabdovirus, por ejemplo, VSV; Paramixovirus, por ejemplo, VSR; 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 VLTV y determinados virus de ADN a ARN tales como el virus de la Hepatitis B.
- 30 Las proteínas relacionadas con la agricultura, tales como las proteínas de resistencia a insectos (por ejemplo las proteínas Cry), enzimas de producción de almidón y lípidos, toxinas de plantas e insectos, proteínas de resistencia a toxinas, proteínas desintoxicantes de micotoxina, enzimas de crecimiento de plantas (por ejemplo, Ribulosa 1,5-Bifosfato Carboxilasa/Oxigenasa, "RUBISCO"), lipooxigenasa (LOX) y Fosfoenolpiruvato (PEP) carboxilasa son también dianas apropiadas para la modificación con aminoácidos no naturales.
- 35 En determinadas realizaciones, la proteína o el polipéptido de interés (o partes de los mismos) en los métodos y/o composiciones de la invención están codificados por un ácido nucleico. Normalmente, el ácido nucleico comprende al menos un codón selector, y puede comprender 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.
- 40 Los genes que codifican proteínas o polipéptidos de interés pueden mutagenizarse utilizando métodos bien conocidos por los expertos en la técnica y descritos en el presente documento en "mutagenesis y otras técnicas de biología molecular" para incluir, por ejemplo, uno o más codones selectores para la incorporación de un aminoácido no natural. Por ejemplo, se mutageniza un ácido nucleico para una proteína de interés para incluir uno o más codones selectores, proporcionando la inserción de uno o más aminoácidos no naturales. La invención incluye cualquiera de dichas variantes, por ejemplo, mutantes, versiones de cualquier proteína, por ejemplo, que incluye al menos un aminoácido no natural. De forma similar, la invención también incluye ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifique uno o más aminoácidos no naturales.
- 45 Para fabricar una proteína que incluya un aminoácido no natural, pueden utilizarse células y organismos hospedadores que estén adaptados para la incorporación *in vivo* del ácido nucleico no natural mediante pares

ARNt/RS ortogonales. Las células hospedadoras se modifican por ingeniería genética (por ejemplo, se transforman, transducen o transfectan) con uno o más vectores que expresan el ARNt ortogonal, la ARNt sintetasa ortogonal y un vector que codifica la proteína que va a derivatizarse. Cada uno de estos componentes pueden estar en el mismo vector o cada uno de ellos puede estar en un vector distinto, o los dos componentes pueden estar en un vector y el tercer componente en un segundo vector. El vector puede estar, por ejemplo, en forma de plásmido, bacteria, virus, polinucleótido desnudo o polinucleótido conjugado.

Definición de polipéptidos por inmunoreactividad

Como los polipéptidos de la invención proporcionan diversas secuencias polipeptídicas nuevas (por ejemplo, polipéptidos que comprenden aminoácidos no naturales en el caso de proteínas sintetizadas en los sistemas de traducción del presente documento o, por ejemplo, en el caso de nuevas sintetasas, nuevas secuencias de aminoácidos convencionales), los polipéptidos también proporcionan nuevas características estructurales que pueden reconocerse, por ejemplo, en ensayos inmunológicos. La generación de antisueros, que se unen específicamente a los polipéptidos de la invención, así como los polipéptidos que se unen por dichos antisueros, constituye una característica de la invención. El término "anticuerpo", como se utiliza en el presente documento, incluye, pero sin limitación, un polipéptido codificado sustancialmente por uno o más genes de inmunoglobulina, o fragmentos de los mismos que se unen específicamente y reconocen un analito (antígeno). Los ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos, monocatenarios y similares. Los fragmentos de inmunoglobulinas, incluyendo fragmentos Fab y fragmentos producidos mediante una biblioteca de expresión, incluyendo presentación en fagos, también están incluidos en el término "anticuerpo" como se usa en el presente documento. Véase, por ejemplo, Paul, *Fundamental Immunology*, 4^a ed., 1999, Raven Press, Nueva York, para estructura y terminología de anticuerpos.

Con el fin de producir los antisueros para su uso en un inmunoensayo, se producen uno o más polipéptidos inmunogénicos y se purifican según se describe en el presente documento. Por ejemplo, la proteína recombinante puede producirse en una célula recombinante. Una cepa endogámica de ratón (utilizada en este ensayo ya que los resultados son más reproducibles debido a la identidad genética implícita de los ratones), se inmuniza con la proteína (o proteínas) inmunogénica en combinación con un adyuvante convencional, tal como adyuvante de Freund, y un protocolo de inmunización de ratón 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 utilizarse para determinar la inmunorreactividad específica). Pueden encontrarse detalles adicionales sobre proteínas, anticuerpos, antisueros, etc. en las Publicaciones Internacionales Números WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE"; WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; WO 2004/035605, titulada "GLYCOPROTEIN SYNTHESIS"; y WO 2004/058946, titulada "PROTEIN ARRAYS".

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

Las composiciones de la invención y las composiciones preparadas mediante los métodos de la invención como se define en las reivindicaciones están opcionalmente en una célula. Los pares O-ARNt/O-RS o componentes individuales de la invención que pueden utilizarse después en una maquinaria de traducción de un sistema hospedador, dando como resultado la incorporación de un aminoácido no natural en una proteína. La Publicación Internacional Número WO 2002/085923 de Schultz, *et al.*, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACID", describe este proceso. Por ejemplo, cuando se introduce un par O-ARNt/O-RS en un hospedador, por ejemplo, *Escherichia coli*, el par conduce a la incorporación *in vivo* de un aminoácido no natural, que puede añadirse de forma exógena al medio de cultivo, en una proteína, por ejemplo, una proteína de ensayo de mioglobina o una proteína terapéutica, en respuesta a un codón selector, por ejemplo, un codón sin sentido ámbar. Opcionalmente, las composiciones de la invención pueden estar en un sistema de traducción *in vitro*, o en uno o más sistemas *in vivo*. Se pueden utilizar proteínas con el aminoácido no natural en cualquiera de una gran diversidad de aplicaciones. Por ejemplo, el resto no natural incorporado en la proteína puede servir como una diana para cualquiera de una gran variedad de modificaciones, por ejemplo, entrecruzamientos con otras proteínas, con moléculas pequeñas tales como marcadores o colorantes y/o biomoléculas. Con estas modificaciones, la incorporación del aminoácido no natural puede dar como resultado proteínas terapéuticas mejoradas y pueden utilizarse para alterar o mejorar la función catalítica de enzimas. En algunos aspectos, la incorporación y posterior modificación de un aminoácido no natural en una proteína puede facilitar estudios sobre la estructura de las proteínas, interacciones con otras proteínas y similares.

Ejemplos

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

EJEMPLO 1

Selección genética de sintetasas mutantes específicas de sulfotirosina

Las metodologías que permiten la adición sistemática de aminoácidos no naturales a los códigos genéticos de células de *E. coli* (Wang *et al.*, “Expanding the genetic code of Escherichia coli”, Science 292: 498-500 (2001)), levadura (Chin *et al.*, “An expanded eukaryotic genetic code”, Science 301: 964-967 (2003)) y de mamífero (Zhang *et al.*, “Selective incorporation of 5-hydroxytryptophan into proteins in mammalian cells”, Proc Natl Acad Sci USA 101: 8882-8887 (2004)) se han descrito anteriormente. Dichos métodos se basan en el desarrollo de un par ARNt/aaRS supresor antisentido que tiene la propiedad de la ortogonalidad, definida como la capacidad de incorporar selectivamente un aminoácido determinado en respuesta a un único codón sin reaccionar en cruzado con los ARNt, las aminoacil-ARNt sintetasas o los aminoácidos engendros del hospedador.

Para generar un par ARNt/aaRS ortogonal que inserta únicamente sulfotirosina (**FIG. 1**), se utilizó una biblioteca de mutantes de sitio activo de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* (MjTyrRS), que carga específicamente un supresor sinsentido de *M. jannaschii* (MjARNt^{Tyr_{CUA}}) modificado por ingeniería genética no reconocido por sintetasas de *E. coli* (Wang *et al.*, “Expanding the genetic code of Escherichia coli”, Science 292: 498-500 (2001)). Esta biblioteca, cuyo diseño y generación se describe en cualquier parte (Bose *et al.*, “The incorporation of a photoisomerizable amino acid into proteins in *E. coli*”, J Am Chem Soc 128: 388-389 (2006)), se sometió a una serie de selecciones positivas y negativas (3 positivas y 2 negativas). La supervivencia en la selección positiva depende de la supresión de una mutación ámbar en el gen cloranfenicol acetiltransferasa (CAT) en presencia de sulfotirosina 2 mM; la supervivencia en la selección negativa depende de la supresión inadecuada de tres mutaciones ámbar en un gen que codifica la proteína barnasa tóxica en ausencia de sulfotirosina (Wang *et al.*, “Expanding the genetic code of Escherichia coli”, Science 292: 498-500 (2001)). Los clones sobreviven a través de rondas de selección tanto positiva como negativa solo si estos incorporan únicamente sulfotirosina en respuesta al codón ámbar.

Después de estas selecciones, se identificaron numerosos clones que permitían a las células, que albergaban el gen CAT con una mutación ámbar en el sitio permisivo 112, sobrevivir en cloranfenicol 130 µg/ml en presencia de sulfotirosina 2 mM. En ausencia de sulfotirosina, las mismas células no crecieron en cloranfenicol 20 µg/ml, coherente con la incorporación eficiente de sulfotirosina con poco a ningún fondo de incorporación de aminoácidos endógenos. La secuenciación de los clones sintetasa mutante candidatos (denominados STyrRS) reveló cuatro clones sintetasa diferentes, cada uno de los cuales cumplía los criterios de un sistema de traducción ortogonal. El clon 1 fue predominantemente (Tyr32Leu, Leu65Pro, Asp158Gly, Ile159Cys, Leu162Lys). En la **FIG. 7** se proporcionan las secuencias de nucleótidos y de aminoácidos de cada uno de estos clones y las especies de tipo silvestre.

	Aminoácido de Mjtirosil-ARNt sintetasa (y codón correspondiente)						
	32	65	155	158	159	162	SEC ID Nº:
tipo silvestre	Tyr (TAC)	Leu (TTG)	Gln (CAG)	Asp (GAT)	Ile (ATT)	Leu (TTA)	2 (3)
Clon 1	Leu (CTG)	Pro (CCT)	Gln (CAG)	Gly (GGT)	Cys (TGT)	Lys (AAG)	4 (5)
Clon 2	Leu (CTG)	Pro (CCG)	Gln (CAG)	Gly (GGT)	Thr (ACT)	Lys (AAG)	6 (7)
Clon 3	Leu (CTG)	Pro (CCT)	Glu (GAG)	Gly (GGT)	Cys (TGT)	Lys (AAG)	8 (9)
Clon 4	Leu (CTG)	Pro (CCG)	Gln (CAG)	Gly (GGT)	Ile (ATT)	Lys (AAG)	10 (11)

Es posible asignar funciones posibles de estas mutaciones, particularmente Lys162, que probablemente forma una interacción de puente salino con la sulfotirosina SO₃⁻. Leu32 y Gly158 pueden alojar el grupo SO₃⁻ más grande y eliminar la afinidad de la tirosina endógena (Tyr32 y Asp158 están implicadas en la formación de enlaces de hidrógeno con el grupo fenólico de la tirosina en la enzima de tipo silvestre). El reemplazo de Asp158 aniónico por Gly posiblemente obvia interacciones electrostáticas desfavorables con la sulfotirosina. Sin embargo, para realizar o utilizar la invención, no se requiere un entendimiento del mecanismo o funciones de las diversas posiciones sustituidas.

METODOLOGÍA DETALLADA PARA LA SELECCIÓN DE SULFOTIROSINA AMINOACIL ARNt SINTETASA

Para seleccionar la STyrRS, se utilizó una biblioteca de sitio activo de MjTyrRS contenida en el vector pBK (pBK-lib) (Bose *et al.*, “The incorporation of a photoisomerizable amino acid into proteins in *E. coli*”, J Am Chem Soc 128: 388-389 (2006)). Células DH10B que incluían pRep, un plásmido de selección positiva que contenía un MjARNt^{Tyr_{CUA}}

modificado por ingeniería genética, un gen de cloranfenicol acetiltransferasa con un codón ámbar introducido en la posición 112 (un sitio permisivo) y un marcador de resistencia a tetraciclina, se transformaron con pBK-lib y se sembraron en placas de agar con GMML complementado con sulfotirosina 2 mM (Senn Chemicals) y cloranfenicol 68 µg/ml. Después de 72 horas a 37 °C, las placas rasparon y se extrajeron los vectores pBK-lib.

5 Esta colección de plásmidos de biblioteca se utilizó después para transformar células DH10B que contenían pNeg, un plásmido de selección negativa, que contenía un MjARN^{Tyr}_{CUA} modificado por ingeniería genética, un gen de barnasa tóxico con tres codones ámbar introducidos, y un marcador de resistencia a cloranfenicol. Las células se sembraron en placas de LB agar que no contenía sulfotirosina y se cultivaron a 37 °C durante 12 horas después de las cuales los vectores pBK-lib se extrajeron de las células supervivientes. Este ciclo de selección positiva y negativa se repitió una vez, y los vectores pBK-lib seleccionados se transformaron posteriormente en células DH10B que contenían pRep y se replicaron sembrando en placas de agar con GMML con y sin sulfotirosina. Las células que crecieron en placas que contenían cloranfenicol 130 µg/ml en presencia de sulfotirosina pero que no crecieron en placas que contenían cloranfenicol 20 µg/ml en ausencia de sulfotirosina se consideraron buenos aciertos.

15 Estos aciertos se clasificaron y la ortogonalidad de sus sintetetas correspondientes se confirmó expresando la proteína de dominio Z que contenía un codón ámbar en la posición 7 en presencia y en ausencia de sulfotirosina. Las sintetetas ortogonales fueron aquellas que permitieron la expresión del dominio Z de longitud completa solo o en presencia de sulfotirosina. Se utilizó MALDI-TOF para confirmar que la sulfotirosina se incorporaba realmente en el dominio Z de longitud completa.

EJEMPLO 2

Expresión y caracterización de una proteína modelo mutante que contiene sulfotirosina

25 Para verificar la incorporación única de la sulfotirosina por la sinteteta STyrRS seleccionada, un mutante ámbar (resto 7) de una proteína de dominio Z marcada con His₆ en el extremo C, se expresó en *E. coli* que incluía plásmidos para el dominio Z mutante ámbar, MjARN^{Tyr}_{CUA} y STyrRS (clon 1). El análisis de electroforesis en gel de poliacrilamida (PAGE) después de purificación con Ni-NTA mostró una fuerte banda para el dominio Z solo cuando la proteína se expresaba en medios que contenían sulfotirosina 2 mM - no se observó banda en ausencia de sulfotirosina, lo que confirmaba la dependencia de la supresión ámbar en la sulfotirosina (**FIG. 4A**).

30 Para una caracterización adicional, se realizó análisis MALDI-TOF en el dominio Z mutante purificado. Debe indicarse que los análisis MALDI-TOF y ESI de proteínas de tirosina sulfatadas dan como resultado la pérdida parcial de sulfato, cuyo grado depende la rigurosidad de las condiciones (22, 23). Por lo tanto, se utilizaron condiciones suaves en modo ión positivo con una matriz a pH moderado (2,4,6-trihidroxi-acetofenona), en las que apareció un pico predominante [M+H] de 7876 Da ($M_{teórico} = 7877,5$ Da) correspondiente al dominio Z que contenía una sola sulfotirosina y que carecía de metionina. También se observó un pico pequeño (<10 %) [M+H] de 7798 Da ($M_{teórico} = 7797,5$ Da) que es el resultado de la pérdida de sulfato durante el MALDI-TOF, saliendo de la tirosina (**FIG. 4B**). Aunque por si solos estos datos de espectrometría de masas no descartan la incorporación de tirosina del fondo mediante STyrRS, puede hacerse basándose en el análisis en gel PAGE. Por tanto la STyrRS incorpora únicamente sulfotirosina, lo que permite la expresión recombinante de proteínas sulfatadas en bacterias.

EJEMPLO 3

Expresión de una proteína modelo sulfatada (hirudina) derivada de un organismo superior

45 Se examinó si este sistema ortogonal para la producción de proteínas sulfatadas podría utilizarse para generar una proteína nativa selectivamente sulfatada normalmente biosintetizada solo en organismos superiores. Para ello, se seleccionó la proteína hirudina, que está sulfatada en la posición 63 de tirosina. La hirudina, segregada por la sanguijuela medicinal *Hirudo medicinalis*, es el inhibidor natural más fuerte de la trombina, y su forma recombinante se administra clínicamente como un anticoagulante. Sin embargo, la expresión recombinante de la hirudina en *E. coli* y levadura utilizada para la producción comercial del fármaco produce la forma no sulfatada (desulfohirudina) debido a la ausencia de sulfotransferasas necesarias en estos organismos (Markwardt, "Hirudin as alternative anticoagulant-a historical review," *Semin Thromb Hemost* 28, 405-414 (2002)). Aunque a pesar de que la desulfohirudina sea un inhibidor eficaz de trombina, su afinidad por la trombina humana es al menos un orden de magnitud inferior que el de la sulfohirudina, que tiene una K_i de aproximadamente 20 fM (Braun *et al.*, "Use of site-directed mutagenesis to investigate the basis for the specificity of hirudin," *Biochemistry* 27, 6517-6522 (1988)).

60 Para expresar la sulfohirudina, se clonó el gen STyrRS (clon 1) en la estructura vectorial pSup que contenía seis copias de MjARN^{Tyr}_{CUA} con promotores optimizados (Ryu y Schultz, "Efficient incorporation of unnatural amino acids into proteins in *Escherichia coli*," *Nat Methods* 3: 263-265 (2006)). El gen de hirudina con un codón ámbar en la posición 63 y una secuencia señal periplásmica gIII se sintetizó y se insertó en el vector pBAD. Después de la cotransformación de células *E. coli* DH10B con ambos plásmidos, se realizó la expresión en un matraz agitador en medio mínimo líquido con glicerol (GMML) complementado con sulfotirosina 10 mM. Dado que la hirudina es pequeña, la dirección dentro del periplasma da como resultado la secreción de modo eficaz; por lo tanto, la

sulfohirudina se purificó directamente del medio concentrado por FPLC utilizando una columna de intercambio aniónico Q Sepharose seguido de cromatografía de exclusión por tamaño para producir un rendimiento de 5 mg/l. Para comparar, la desulfohirudina con la tirosina codificada en la posición 63 se expresó de manera similar y se purificó con un rendimiento de 12 mg/l.

5 METODOLOGÍA DETALLADA PARA LA CLONACIÓN, EXPRESIÓN Y PURIFICACIÓN DE SULFOHIRUDINA Y DESULFOHIRUDINA

10 El gen correspondiente a [Leu¹, Thr²]-63-desulfohirudina (comercialmente conocido como Lepirudina (Refludan[®])) fusionado con una secuencia señal periplásmica gIII para la secreción se sintetizó por BlueHeron[®] con Optimización de Expresión. Este gen se insertó en el vector pBAD (Invitrogen) para producir pBAD-Hirudin bajo el control del promotor *araBAD*. Se utilizó mutagénesis de sitio dirigido Quickchange (Stratagene) para introducir TAG en la posición 63 del gen Lepirudina para producir pBAD-HirudinTAG para la expresión de sulfohirudina.

15 El gen correspondiente a la STirRS seleccionada (clon 1) se insertó en el vector pSup entre los sitios *PstI* y *NdeI* bajo el control del promotor *glnS* para producir pSup-STyrRS. El pSup-STyrRS también contenía seis copias de MjARN^{1yr}_{CUA} modificado por ingeniería genética bajo el control del promotor *proK*.

20 Células DH10B electrocompetentes cotransformadas con pSup-STyrRS y pBAD-HirudinTAG se cultivaron en medio GMMML con ampicilina 50 µg/ml, 20 µg/ml de cloranfenicol y sulfotirosina 10 mM a 37 °C. Cuando las células alcanzaron una DO₆₀₀ de 0,6, se añadió L-arabinosa a una concentración final de 0,2 % para inducir la expresión de proteínas. Las células se cultivaron durante 24 horas más a 37 °C. Las células se sedimentaron y los medios se concentraron utilizando un dispositivo agitador celular.

25 Los medios concentrados se dializaron frente a agua y se aplicaron a una columna de intercambio aniónico (HiLoad 26/10 Q Sepharose, GE Healthcare) previamente equilibrada con Tris-HCl 50 mM, EDTA 1 mM y β-mercaptoetanol 10 mM, pH 7,4. Las proteínas se eluyeron con un gradiente lineal de NaCl de 0,025 a 1 M. Se analizaron las fracciones pico por PAGE. Las fracciones procedentes de un pico principal que eluyeron a NaCl 0,3 M se agruparon entre sí, se concentraron, se dializaron frente a agua y se aplicaron a filtración en gel (Superdex 200 10/300 GL, GE Healthcare). Las proteínas se eluyeron con solución salina tamponada con Tris (Tris-HCl 25 mM, NaCl 125 mM y KCl 2 mM, pH 7,6). La concentración final de sulfohirudina se determinó por titulación frente a α-trombina humana 1 nM (Diapharma) utilizando 50 µM de sustrato fluorogénico de trombina Boc-Asp(OBzl)-Pro-Arg-MCA (Peptides International, Inc.) para medir la actividad trombina. Esto supone una inhibición estequiométrica de 1:1 de trombina por hirudina, que es válida a las concentraciones utilizadas tal y como estipula la cinética de unión estrecha (Szedlacsek y Duggleby, "Kinetics of slow and tight-binding inhibitors," *Methods Enzymol* 249: 144-180 (1995)). Para expresar, purificar y cuantificar [Leu¹, Thr²]-63-desulfohirudina se utilizaron procedimientos similares.

EJEMPLO 4 DE REFERENCIA

40 **Caracterización de una hirudina sulfatada codificada genéticamente**

Las hirudinas resultantes descritas en el ejemplo anterior se caracterizaron por análisis PAGE y cada una de ellas se presentó como una sola banda. La sulfohirudina podría diferenciarse de la desulfohirudina ya que la primera migra más lejos que la última para ofrecer un desplazamiento en gel (**FIG. 2**). El análisis MALDI-TOF mostró las masas correctas [M+H] tanto de la sulfohirudina (7059 Da; M_{teórico} = 7059,5 Da) como de la desulfohirudina (6979 Da; M_{teórico} = 6979,5 Da) con dos picos en el caso de la sulfohirudina por la pérdida de sulfato dando como resultado una señal secundaria [M+H-80] (véase la **FIG. 5**).

50 Para verificar adicionalmente que este segundo pico resultaba exclusivamente del análisis espectral de masas, se realizaron dos experimentos. En primer lugar, se aprovechó el hecho de que la elución de la sulfohirudina de la columna de intercambio aniónico se produce a una fuerza iónica mayor del 10 % en comparación con la elución de la desulfohirudina en las mismas condiciones de gradiente, lo que permitiría la separación completa de las dos hirudinas si hubieran estado presentes de manera simultánea. (Esto se confirmó añadiendo sulfohirudina con desulfohirudina). Dado que no se observó ningún pico de desulfohirudina en la purificación por intercambio aniónico de la sulfohirudina, determinado por la ausencia de un pico de desulfohirudina en los espectros de masas de las correspondientes fracciones eluidas, se llegó a la conclusión de que cuando se expresaba la sulfohirudina no se producía desulfohirudina.

60 En segundo lugar, se realizó una expresión de control en la que no se añadió sulfotirosina. Un análisis MALDI-TOF posterior de los medios concentrados en bruto que contenían una mezcla de todas las proteína segregadas mostró solo un pico [M+H] de 6578 Da correspondiente a la proteína truncada resultante de un comportamiento alternativo de TAG como un codón de terminación (M_{teórico} = 6575 Da); no se observó ningún pico correspondiente a la proteína de longitud completa (véase la **FIG. 6A**). Esto difiere de la expresión en presencia de sulfotirosina en la que en los espectros de masas se encuentran picos de proteína tanto truncada como de longitud completa a aproximadamente las mismas intensidades (véase **FIG. 6B**), lo que sugiere una estricta dependencia de supresión ámbra sobre la presencia de sulfotirosina. A partir de estos dos experimentos, se llega a la conclusión de que la señal [M+H-80] en

el MALDI-TOF de la sulfohirudina es exclusivamente atribuible a la escisión de SO_3^- durante la espectrometría de masas, lo que confirma que STyrRS carga su ARNt afín exclusivamente con sulfotirosina sin aminoacilación observable de tirosina

- 5 Cabe señalar que las intensidades similares de los picos de la proteína truncada y de longitud completa en los espectros de masas de los medios de expresión de la sulfohirudina en bruto combinado con el hecho de que la expresión de la desulfohirudina produce aproximadamente el doble de proteínas que la expresión de la sulfohirudina en las mismas condiciones sugiere la supresión de aproximadamente la mitad de los acontecimientos de traducción durante la expresión de la sulfohirudina. Puede por tanto deducirse que la doble supresión en nuestro sistema producirá aproximadamente el 75 % de proteína truncada y el 25 % de longitud completa, suponiendo la ausencia de efectos en el contexto de supresión ámbar. Se contempla que la presencia de la proteína truncada se debe a una baja permeabilidad de la sulfotirosina aniónica en células *E. coli*, dando como resultado una población disminuida de MjARNt^{Tyr}_{CUA} cargada con aminoácido. De hecho, la expresión de hirudina utilizando el mismo sistema, pero con la *p*-acetil fenilalanina altamente permeable y su correspondiente sintetasa mutante, produce la incorporación de *p*-acetil fenilalanina con proteína truncada indetectable (datos no mostrados). Una estrategia profármaco para administrar sulfotirosina puede por lo tanto eliminar la presencia de proteína truncada y aumentar el rendimiento.

EJEMPLO 5 DE REFERENCIA

20 Caracterización de la actividad biológica de sulfohirudina codificada genéticamente

Para caracterizar la eficacia de la sulfohirudina expresada como un anticoagulante, se determinó la cinética de la inhibición de trombina utilizando un ensayo enzimático fluorogénico basado en el método de curva de progreso sencillo previamente descrito en la bibliografía (Cha, "Tight-binding inhibitors--III. A new approach for the determination of competition between tight-binding inhibitors and substrates--inhibition of adenosine deaminase by coformycin," *Biochem Pharmacol* 25: 2695-2702 (1976); Komatsu *et al.*, "CX-397, a novel recombinant hirudin analog having a hybrid sequence of hirudin variants-1 and -3," *Biochem Biophys Res Commun* 196: 773-779 (1993)). En este ensayo, se mezclaron 100 pM de sulfohirudina o desulfohirudina con sustrato fluorogénico 50 μM al que se añadió α -trombina humana para iniciar la reacción. La escisión de sustrato fluorogénico por trombina, cuya actividad inhibe a diferentes grados la sulfohirudina y la desulfohirudina, da como resultado una gráfica de intensidad de fluorescencia a lo largo del tiempo (FIG. 3).

Las concentraciones exactas de la hirudina y sulfohirudina se determinaron por titulación frente a trombina en un intervalo de concentración en el que podía suponerse una unión de 1:1. Como resultado de la cinética de unión estrecha apropiada para la hirudina (Stone y Hofsteenge, "Kinetics of the inhibition of thrombin by hirudin," *Biochemistry* 25: 4622-4628 (1986)), estos datos experimentales se ajustaron a la ecuación 1, produciendo valores de K_i , k_{on} y k_{off} después del tratamiento de las constantes extraídas. Este análisis ofreció valores K_i para la sulfohirudina y desulfohirudina de 26 fM y 307 fM respectivamente, de acuerdo con las publicaciones bibliográficas (17). Como se esperaba, el valor de k_{on} para la sulfohirudina ($0,95 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$) fue mayor que el de la desulfohirudina ($0,38 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$), mientras que el valor de k_{off} para la sulfohirudina fue menor ($0,22 \times 10^{-5} \text{ s}^{-1}$) que el de la desulfohirudina ($1,18 \times 10^{-5} \text{ s}^{-1}$). En la siguiente tabla se muestra un promedio de estas constantes cinéticas de la inhibición de trombina derivadas de un ajuste de curvas de progreso no lineal a lo largo de al menos 3 lecturas con desviaciones típicas.

	K_i	$k_{on} \times 10^8 (\text{M}^{-1} \text{s}^{-1})$	$k_{off} \times 10^5 (\text{s}^{-1})$
Sulfohirudina	26 \pm 9,8	0,95 \pm 0,56	0,22 \pm 0,06
Desulfohirudina	307 \pm 72	0,38 \pm 0,07	1,18 \pm 0,45

45 La ventaja de la mayor afinidad de la sulfohirudina sobre la desulfohirudina debe ser especialmente pronunciada en el intervalo de concentración de la trombina débilmente unida por sus K_i respectivas (Szedlacsek y Duggleby, "Kinetics of slow and tight-binding inhibitors," *Methods Enzymol* 249: 144-180 (1995)). Por lo tanto es interesante que la concentración basal fisiológica del estado de equilibrio estacionario de la trombina humana activa se encuentre dentro de este intervalo (Velan y Chandler, "Effects of surgical trauma and cardiopulmonary bypass on active thrombin concentrations and the rate of thrombin inhibition in vivo," *Pathophysiol Haemost Thromb* 33: 144-156 (2003)), lo que sugiere un posible impulso evolutivo para la sulfatación en la hirudina de sanguijuela nativa. Esta observación debe servir como una orientación para posibles aplicaciones terapéuticas para la desulfohirudina codificada genéticamente (descrita en el presente documento) sobre la forma recombinante no sulfatada predominante.

60 La incorporación cotraduccional de la sulfotirosina en proteínas debe hacer posible la expresión eficaz de muchas proteínas más selectivamente sulfatadas en *E. coli* incluyendo anticuerpos, motivos de receptores de quimiocina y factores de coagulación, facilitando de esta manera estudios estructurales-funcionales así como la aplicación terapéutica práctica de proteínas sulfatadas. Además, esta estrategia *in vivo* puede aplicarse a la construcción de

bibliotecas de anticuerpos sulfatados y presentación de fagos de proteínas sulfatadas, caminos prometedoros inaccesibles mediante métodos disponibles de síntesis peptídica, ligamiento químico nativo y ligamiento de proteínas expresadas. Como alternativa, debe ser posible ampliar esta estrategia a la expresión directa de proteínas tirosina sulfatadas en organismos eucariotas.

5 METODOLOGÍA DETALLADA DE LA CARACTERIZACIÓN CINÉTICA DE ESPECIES DE HIRUDINA EXPRESADAS

10 La liberación de 7-amino-4-metilcoumarina de Boc-Asp(OBzl)-Pro-Arg-MCA 50 μM como resultado de la actividad de la trombina se controló midiendo la intensidad de fluorescencia (longitud de onda de excitación = 365 nm; longitud de onda de emisión = 450 nm) con un lector de placa fluorescente (Molecular Devices SpectraMax Gemini). La reacción enzimática se realizó por triplicado y se repitió tres veces en placas de 96 pocillos a 37 °C en tampón Tris-HCl 50 mM, pH 7,8, que contenía Polietilenglicol 6000 al 0,1 % (Fluka), NaCl 100 mM, y HSA 250 μg/ml (Calbiochem). La constante de Michaelis del sustrato en estas condiciones es 11,6 μM (Komatsu *et al.*, "CX-397, a novel recombinant hirudin analog having a hybrid sequence of hirudin variants-1 and -3," Biochem Biophys Res Commun 196: 773-779 (1993)).

15 Los parámetros cinéticos de la inhibición de trombina por la sulfohirudina y desulfohirudina expresada se extrajo de ajuste no lineal de curvas de progreso obtenido a 40 pM de α-trombina y 100 pM de sulfohirudina o desulfohirudina utilizando el método de curva de progreso sencillo, como se ha descrito previamente (Komatsu *et al.*, "CX-397, a novel recombinant hirudin analog having a hybrid sequence of hirudin variants-1 and -3," Biochem Biophys Res Commun 196: 773-779 (1993)). De acuerdo con el mecanismo de inhibición competitiva de unión estrecha, lenta de las hirudinas, la formación de productos puede describirse mediante la ecuación 1 (Stone y Hofsteenge, "Kinetics of the inhibition of thrombin by hirudin," Biochemistry 25: 4622-4628 (1986); Cha, "Tight-binding inhibitors--III. A new approach for the determination of competition between tight-binding inhibitors and substrates--inhibition of adenosine deaminase by coformycin," Biochem Pharmacol 25: 2695-2702 (1976)):

$$P = v_s t + \frac{(1-\gamma)(v_0 - v_s)}{\lambda \gamma} \ln \left(\frac{1 - \gamma e^{-\lambda t}}{1 - \gamma} \right)$$

20 en la que P es la cantidad de producto formado en el tiempo t y v_0 y v_s son las velocidades inicial y de estado de equilibrio estacionario de la reacción. En la ecuación 1, v_s , γ y λ pueden describirse mediante las siguientes expresiones:

$$v_s = v_0 \left(\frac{E_t - I_t - K_i' + Q}{2E_t} \right)$$

35
$$\gamma = \frac{K_i' + E_t + I_t - Q}{K_i' + E_t + I_t + Q}$$

$$\lambda = k_{on} Q,$$

en las que

40
$$K_i' = K_i \left(1 + \frac{S}{K_m} \right)$$

y

$$Q = \sqrt{(K_i' + E_t + I_t)^2 - 4E_t I_t}$$

45 Utilizando estas ecuaciones, se determinaron los valores de K_i y k_{on} . El valor de k_{off} es el producto de k_{on} y K_i . Los ajustes de regresión no lineal se calcularon utilizando el programa GraphPad Prism.

REIVINDICACIONES

1. Un sistema de traducción que comprende:

- 5 (a) un primer aminoácido no natural que es la sulfotirosina tal como se representa en la Figura 1;
 (b) una primera aminoacil ARNt sintetasa ortogonal (O-RS) en donde la O-RS comprende una secuencia de aminoácidos que es una variante conservativa de la SEC ID N^o: 2, la cual variante conservativa tiene una identidad de al menos el 90 % con la SEC ID N^o: 2 y comprende: una leucina en una posición correspondiente a Tyr32; una prolina en una posición correspondiente a Leu65; una glicina en una posición correspondiente a Asp158 y una lisina en una posición correspondiente a Leu162, y;
 10 opcionalmente comprende un ácido glutámico en una posición correspondiente a Gln155 y/o una treonina o una cisteína en una posición correspondiente a Ile159, y;
 (c) un primer ARNt ortogonal (O-ARNt); en el que dicha primera O-RS aminoacila dicho primer O-ARNt con dicha sulfotirosina.

15 2. El sistema de traducción de la reivindicación 1, en el que:

- (i) dicha primera O-RS deriva de una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; y/o,
 (ii) dicha primera O-RS deriva de una tirosil ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre; y/o
 20 (iii) dicha primera O-RS comprende una secuencia de aminoácidos expuesta en las SEC ID Nos: 4, 6, 8 o 10; y/o,
 (iv) dicho primer O-ARNt es un ARNt supresor de ámbar; y/o
 (v) dicho primer O-ARNt comprende o está codificado por una secuencia de polinucleótidos expuesta en la SEC ID N^o: 1.

25 3. El sistema de traducción de la reivindicación 1, que adicionalmente comprende un ácido nucleico que codifica una proteína de interés, comprendiendo dicho ácido nucleico al menos un codón selector, en el que dicho codón selector es reconocido por dicho primer O-ARNt.

30 4. El sistema de traducción de la reivindicación 3, que adicionalmente comprende una segunda O-RS y un segundo O-ARNt, en el que la segunda O-RS aminoacila el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y en el que el segundo O-ARNt reconoce un codón selector que es diferente del codón selector reconocido por el primer O-ARNt.

35 5. El sistema de traducción de la reivindicación 1, en el que dicho sistema comprende una célula hospedadora que comprende dicho primer aminoácido no natural, dicha primera O-RS y dicho primer O-ARNt.

40 6. El sistema de traducción de la reivindicación 5, en el que dicha célula hospedadora es una célula de eubacteria, tal como una célula de *E. coli*.

7. El sistema de traducción de la reivindicación 5, en el que dicha célula hospedadora comprende un polinucleótido que codifica dicha primera O-RS y/o un polinucleótido que codifica dicho primer O-ARNt.

45 8. El sistema de traducción de la reivindicación 7, en el que dicho primer polinucleótido comprende una secuencia de nucleótidos expuesta en las SEC ID Nos: 5, 7, 9 u 11.

9. Un método para producir, en un sistema de traducción, una proteína que comprende un aminoácido no natural en una posición seleccionada, comprendiendo el método:

- 50 (a) proporcionar un sistema de traducción que comprende:
 (i) un primer aminoácido no natural que es la sulfotirosina tal como se representa en la Figura 1;
 (ii) una primera aminoacil ARNt sintetasa ortogonal (O-RS) en la que la O-RS comprende una secuencia de aminoácidos que es una variante conservativa de la SEC ID N^o: 2, la cual dicha variante conservativa tiene una identidad de al menos el 90 % con la SEC ID N^o: 2 y comprende: una leucina en una posición correspondiente a Tyr32; una prolina en una posición correspondiente a Leu65; una glicina en una posición correspondiente a Asp158, y una lisina en una posición correspondiente a Leu162, y;
 55 opcionalmente comprende un ácido glutámico en una posición correspondiente a Gln155, y/o una treonina o una cisteína en una posición correspondiente a Ile159;
 (iii) un primer ARNt ortogonal (O-ARNt), en el que dicha primera O-RS aminoacila dicho primer O-ARNt con dicha sulfotirosina; y
 (iv) un ácido nucleico que codifica dicha proteína, en donde dicho ácido nucleico comprende al menos un codón selector que es reconocido por dicho primer O-ARNt; e
 60
 65 (b) incorporar dicho aminoácido natural en dicha posición seleccionada en dicha proteína durante la traducción de dicha proteína en respuesta a dicho codón selector, produciendo de este modo dicha proteína que comprende

dicho aminoácido no natural en la posición seleccionada.

10. El método de la reivindicación 9, en el que dicha proteína que comprende un aminoácido no natural es la sulfohirudina.

5

11. El método de la reivindicación 9, en el que:

(A) dicha disposición de un sistema de traducción comprende proporcionar un polinucleótido que codifica dicha O-RS; o

10 (B) dicha disposición de un sistema de traducción comprende proporcionar una O-RS derivada de una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o,

(C) dicha disposición de un sistema de traducción comprende proporcionar una O-RS derivada de una tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre; o,

15 (D) dicha disposición de un sistema de traducción comprende proporcionar una O-RS que comprende una secuencia de aminoácidos expuesta en las SEC ID Nos: 4, 6, 8 o 10; o,

(E) dicha disposición de un sistema de traducción comprende mutar un bolsillo de unión de aminoácidos de una aminoacil-ARNt sintetasa de tipo silvestre por mutagénesis de sitio dirigido y seleccionar una O-RS resultante que aminoacila dicho O-ARNt con dicho aminoácido no natural, en donde dicha etapa de selección comprende realizar una selección positiva y una selección negativa de dicha O-RS a partir de un conjunto que comprende una pluralidad de moléculas de aminoacil-ARNt sintetasa resultantes después de la mutagénesis de sitio dirigido; o,

20 (F) dicha disposición de un sistema de traducción comprende proporcionar un polinucleótido que codifica dicho O-ARNt; o,

25 (G) dicha disposición de un sistema de traducción comprende proporcionar un O-ARNt que es un ARNt supresor ámbar; o,

(H) dicha disposición de un sistema de traducción comprende proporcionar un O-ARNt que comprende o está codificado por una secuencia de polinucleótidos expuesta en la SEC ID N°: 1; o,

30 (I) dicha disposición de un sistema de traducción comprende proporcionar un ácido nucleico que comprende un codón selector ámbar; o

(J) dicha disposición de un sistema de traducción comprende proporcionar una célula hospedadora, en donde dicha célula hospedadora comprende dicho primer aminoácido no natural, dicha primera O-RS, dicho primer O-ARNt y dicho ácido nucleico, y en el que dicha etapa de incorporación comprende cultivar dicha célula hospedadora; o,

35 (K) dicha disposición de un sistema de traducción comprende proporcionar un extracto celular.

12. El método de la reivindicación 11 (J), en el que:

(i) dicha disposición de una célula hospedadora comprende proporcionar una célula hospedadora de eubacteria; o

40 (ii) dicha disposición de una célula hospedadora de eubacteria comprende proporcionar una célula hospedadora de *E. coli*; o,

(iii) dicha disposición de una célula hospedadora comprende proporcionar una célula hospedadora que comprende un polinucleótido que codifica dicha O-RS; o,

45 (iv) dicha disposición de una célula hospedadora que comprende un polinucleótido que codifica dicha O-RS comprende proporcionar una célula hospedadora que comprende un polinucleótido que comprende una secuencia de nucleótidos expuesta en las SEC ID Nos: 5, 7, 9 u 11.

13. El método de la reivindicación 9, en el que adicionalmente dicha proteína comprende un segundo aminoácido no natural que es diferente de dicho primer aminoácido no natural, y en el que dicho sistema de traducción comprende adicionalmente una segunda O-RS y un segundo O-ARNt, en donde la segunda O-RS aminoacila el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y en el que el segundo O-ARNt reconoce un codón selector en el ácido nucleico que es diferente del codón selector reconocido por el primer O-ARNt.

55 14. Una composición que comprende un polipéptido que comprende una secuencia de aminoácidos que es una variante conservativa de la SEC ID N°: 2, la cual dicha variante conservativa tiene una identidad de al menos el 90 % con la SEC ID N°: 2 y comprende: una leucina en una posición correspondiente a Tyr32; una prolina en una posición correspondiente a Leu65; una glicina en una posición correspondiente a Asp158, y una lisina en una posición correspondiente a Leu162, y;

60 opcionalmente comprende un ácido glutámico en una posición correspondiente a Gln155 y/o una treonina o una cisteína en una posición correspondiente a Ile159, y; el cual polipéptido aminoacila un O-ARNt con sulfotirosina tal como se representa en la Figura 1.

65 15. La composición de la reivindicación 14, en la que el polipéptido comprende la secuencia de aminoácidos de las SEC ID Nos: 4, 6, 8 o 10.

16. Un polinucleótido que codifica el polipéptido de las reivindicaciones 14 o 15.

17. El polinucleótido de la reivindicación 16, en donde dicho polinucleótido comprende la secuencia de nucleótidos de las SEC ID N°: 5, 7, 9 u 11.

5 18. La composición de las reivindicaciones 14 o 15, en donde dicha composición comprende una célula que comprende el polipéptido.

19. Un vector que comprende el polinucleótido de acuerdo con la reivindicación 16.

10 20. Una célula que comprende un vector, comprendiendo el vector el polinucleótido de la reivindicación 16.

Fig. 1

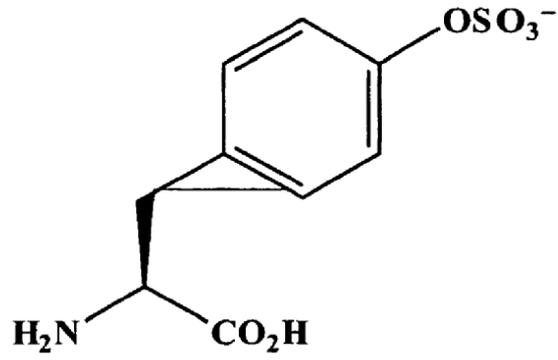
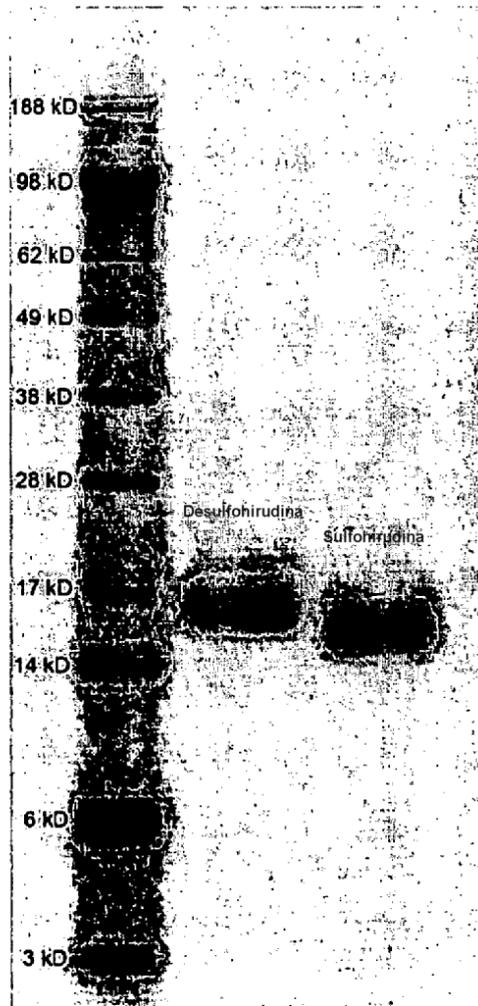


Fig. 2



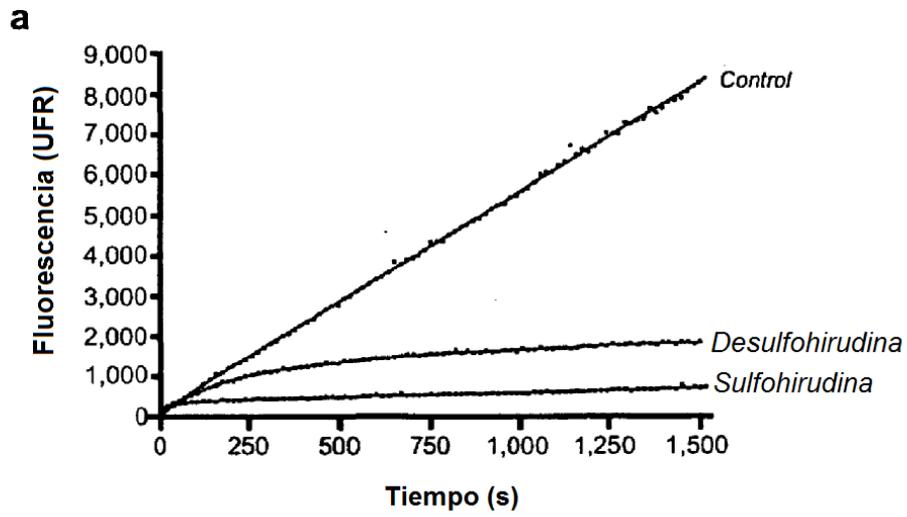


Fig. 3A

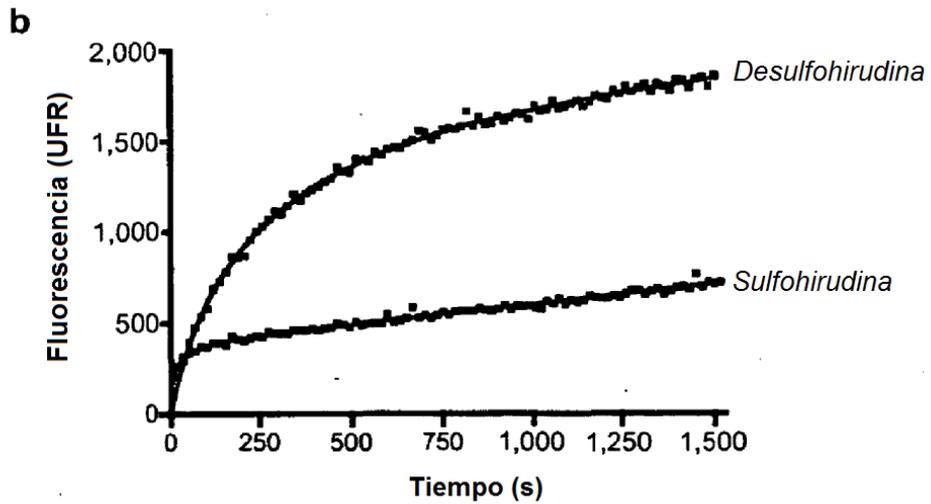


Fig. 3B

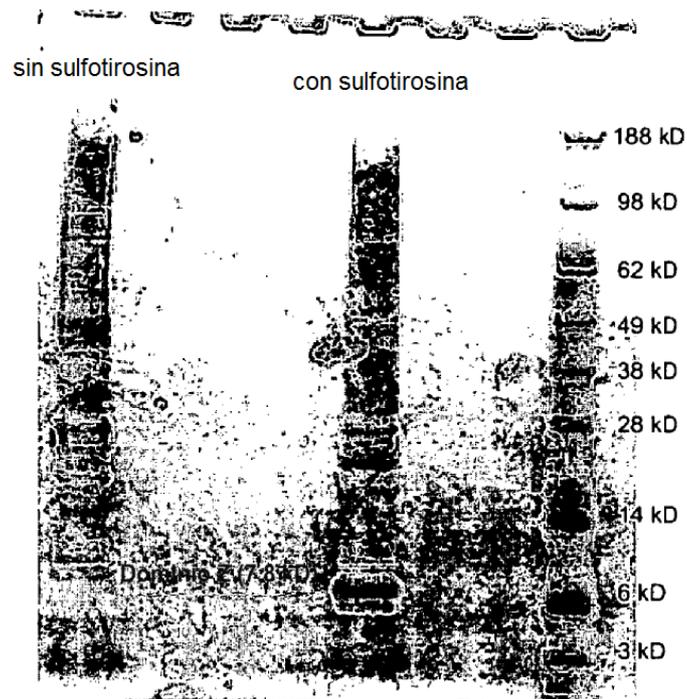


Fig. 4A

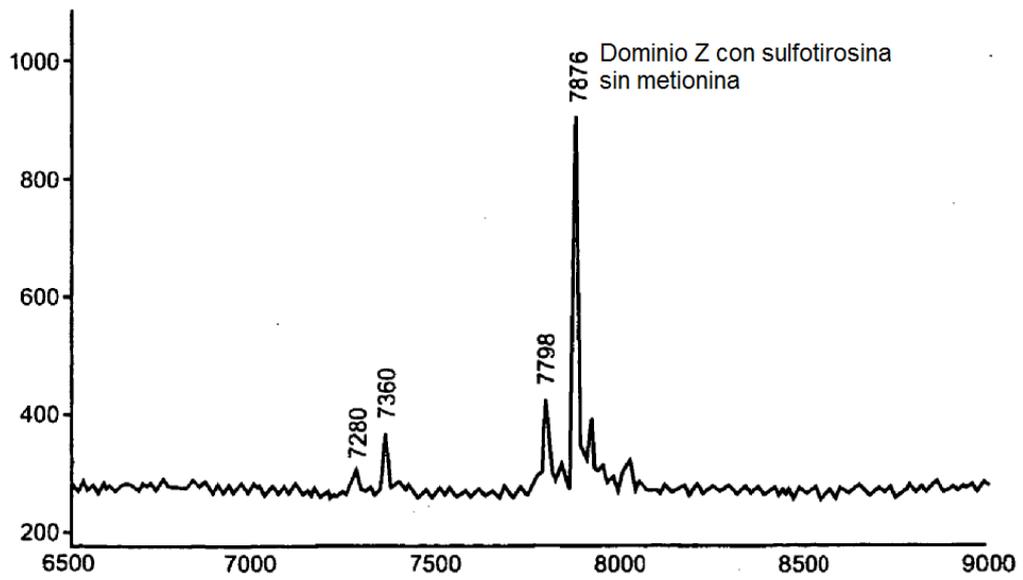


Fig. 4B

Fig. 5A

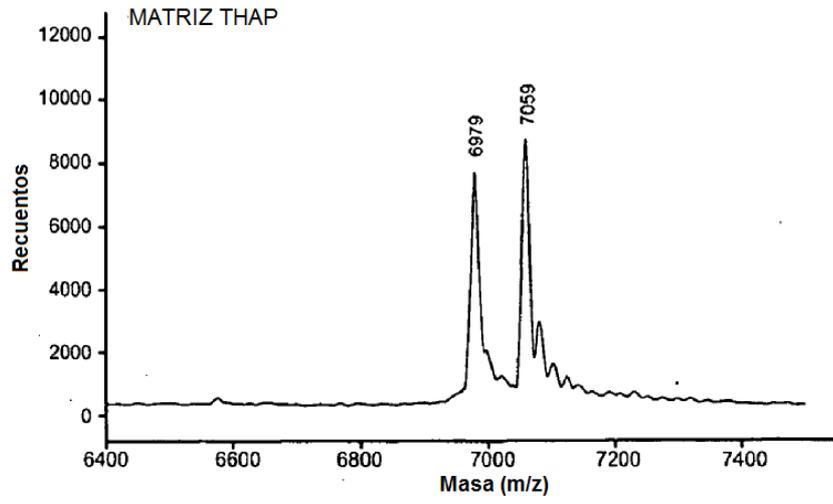


Fig. 5B

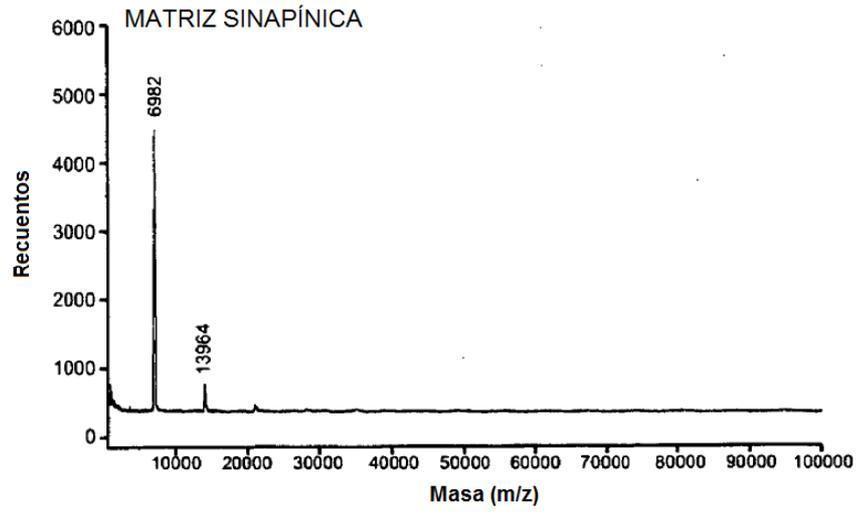
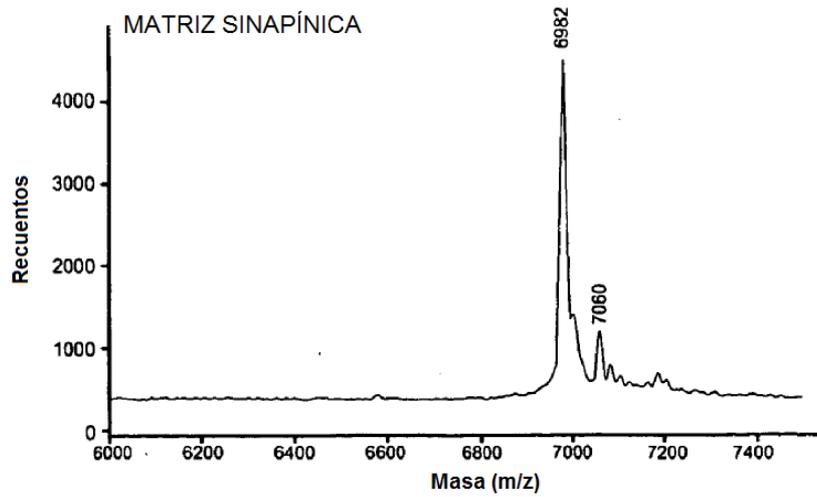


Fig. 5C



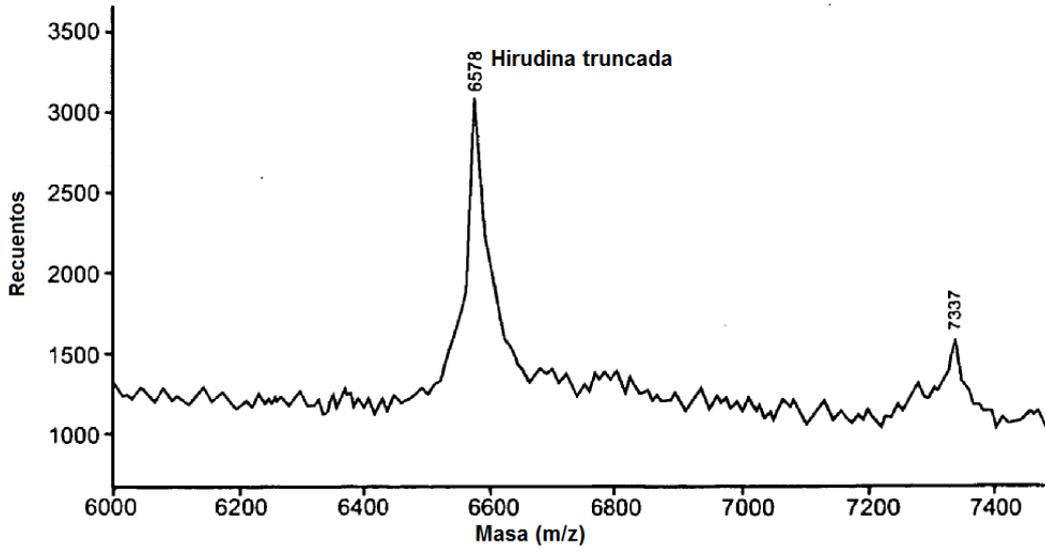


Fig. 6A

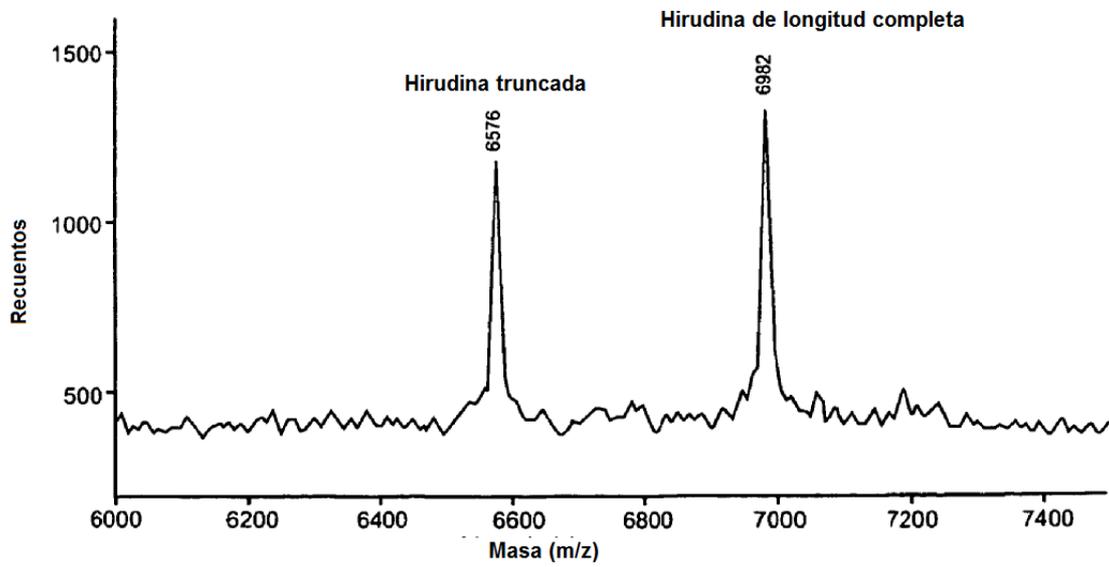


Fig. 6B

Secuencias de nucleótidos y aminoácidos

SEC ID N°:	Descripción	SECUENCIA
1	tirosil- ARN _t ^{CUA} supresor de <i>Methanococcus jannaschii</i> tcc MjARNt-Tyr(CUA) o muARNt ^{Tyr} _{CUA}	CCGGCGGUAGUUCAGCAGGGCAGAACGGCGGACUCUAAAUCCG CAUGGCGCUGGUUCAAAUCCGGCCCGCCGGACCA
2	Secuencia de aminoácidos de la tirosil-ARNt sintetasa (MjTyrRS) de <i>Methanococcus jannaschii</i> de tipo silvestre	MDEFEMIKRNTSEI ISEEELREVLKKDEKSAYIGFEPGKIHL GHYLQIKKMIDLQAGFDI I ILLADLHAYLNQKGELDEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNDIHYLGVDVAVGGME QRKIHMLARELLPKKVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP EKFGDLTVNSYEELESFLFNKELHPMDLKNVAEELIKILEP IRKRL
3	Secuencia de nucleótidos de la tirosil-ARNt sintetasa (MjTyrRS) de <i>Methanococcus jannaschii</i> de tipo silvestre	ATGGACGAATTTGAAATGATAAAGAGAAAACACATCTGAAATATCAGC GAGGAAGAGTTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCTTAC ATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGGCATATCTCCAA ATAAAAAAGATGATGATTACAAAAATGCTGGATTTGATATAAATTATA TTGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTGTGAAGCAATG GGTTAAAGGCAAAATATGTTTATGGAAGTGAATCCAGCTTGATAAG GATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTAAAA AGAGCAAGAAGGAGTATGGAAGTTATAGCAAGAGAGGATGAAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGATATTCAT TATTTAGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTAC AACCTGTCTTAACGGGTTGGATGGAGAAGGAAGATGAGTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAA AGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTAAAA AATGCTGTAGCTGAAGAAGTTATAAAGATTTTAGAGCCAATTAGAAAG AGATTA
4	CLON 1 Secuencia de aminoácidos de la sulfotirosina aminoacil-ARNt sintetasa (derivada de la tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre)	MDEFEMIKRNTSEI ISEEELREVLKKDEKSALIGFEPGKIHL GHYLQIKKMIDLQAGFDI I IPLADLHAYLNQKGELDEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNGCHYKGVVDVAVGGM EQRKIHMLARELLPKKVCIHNPVLTGLDGEKMSSSKGNFIA VDDSPPEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRP PEKFGDLTVNSYEELESFLFNKELHPMDLKNVAEELIKILEP PIRKRL

Fig. 7

SEC ID Nº:	Descripción	SECUENCIA
5	<p style="text-align: center;">CLON 1</p> <p style="text-align: center;">Secuencia de nucleótidos de la sulfotirosina aminoacil-ARNt sintetasa</p>	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCCTCG ATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGGCATTATCTCCAA ATAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATA CTTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTTGAAGCAATG GGGTTAAAGGCAAAAATATGTTTATGGAAGTGAATTCACGCTTGATAAG GATTATACACTGAATGCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGTTGTCAT TATAAAGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAAGGTTGTTTGTATTAC AACCTGTCTTAAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTAAAA AATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAAG AGATTA</p>
6	<p style="text-align: center;">CLON 2</p> <p style="text-align: center;">Secuencia de aminoácidos de la sulfotirosina aminoacil-ARNt sintetasa (derivada de la tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre)</p>	<p>MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSPGKIHL GHYLQIKKMIDLQONAGFDIIIPLADLHAYLNQKGELDEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMQVNGTHYKGV DVAVGGM EQRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIA VDDSP EEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKR PEKFGGDLTVNSYEELESFLKKNKELHPMDLKNVAEELIKILE PIRKRL</p>
7	<p style="text-align: center;">CLON 2</p> <p style="text-align: center;">Secuencia de nucleótidos de la sulfotirosina aminoacil-ARNt sintetasa</p>	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTAAAAAAGATGAAAAATCTGCCTCG ATAGGTTTGAACCAAGTGGTAAAAATACATTTAGGGCATTATCTCCAA ATAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATA CCGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAAGTTTTGAAGCAATG GGGTTAAAGGCAAAAATATGTTTATGGAAGTGAATTCACGCTTGATAAG GATTATACACTGAATGCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGTACTCAT TATAAGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAAGGTTGTTTGTATTAC AACCTGTCTTAAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACTTCCTTGAATATCCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTTAAAA AATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAAG AGATTA</p>
8	<p style="text-align: center;">CLON 3</p> <p style="text-align: center;">Secuencia de aminoácidos de la sulfotirosina aminoacil-ARNt sintetasa (derivada de la tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> de tipo silvestre)</p>	<p>MDEFEMIKRNTSEIISEEELREVLKKDEKSALIGFEPSPGKIHL GHYLQIKKMIDLQONAGFDIIIPLADLHAYLNQKGELDEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIMEVNGCHYKGV DVAVGGM EQRKIHMLARELLPKKVVCIHNPVLTGLDGEKMSSSKGNFIA VDDSP EEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKR PEKFGGDLTVNSYEELESFLKKNKELHPMDLKNVAEELIKILE PIRKRL</p>

Fig. 7 (Cont.)

SEC ID Nº:	Descripción	SECUENCIA
9	<p>CLON 3</p> <p>Secuencia de nucleótidos de la sulfotirosina aminoacil-ARNt sintetasa</p>	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATCTGCCTCG ATAGGTTTTGAACCAAGTGGTAAAAATACATTTAGGGCATTATCTCCAA ATAAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATA CCTTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAGTTTTGAAGCAATG GGTTAAAGGCAAAATATGTTTTATGGAAGTGAATTCACGCTTGATAAG GATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGGAGGTTAATGGTTGTCAT TATAAGGGCGTTGATGTTGCAGTTGGAGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTTCAC AACCTGTCTTAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACCTCCTTGAATATCCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTAAAA AATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAA AGATTA</p>
10	<p>CLON 4</p> <p>Secuencia de aminoácidos de la sulfotirosina aminoacil-ARNt sintetasa (derivada de la tirosil ARNt-sintetasa de <i>Methanococcus jannaschii</i> tipo silvestre)</p>	<p>MDEFEMIKRNTSEII SEELREVLKKDEKSALIGFEPGKIHL GHYLQIKKMIDLQNAFDII IPLADLHAYLNQKGELEIRKIG DYNKKVFEAMGLKAKYVYGSEFQLDKDYTLNVYRLALKTTLKR ARRSMELIAREDENPKVAEVIYPIQVNGIHYKGVDSVAVGGM EQRKIHMLARELLPKKVCIHNPVLTGLDGEKMFSSSKNFI A VDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEPLTIKR PEKFGDLTVNSYEELESFLKKNELHPMDLKNVAEELIKILE PIRKRL</p>
11	<p>CLON 4</p> <p>Secuencia de nucleótidos de la sulfotirosina aminoacil-ARNt sintetasa</p>	<p>ATGGACGAATTTGAAATGATAAAGAGAAACACATCTGAAATTATCAGC GAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAATCTGCCTCG ATAGGTTTTGAACCAAGTGGTAAAAATACATTTAGGGCATTATCTCCAA ATAAAAAAGATGATTGATTTACAAAATGCTGGATTTGATATAATTATA CCGTTGGCTGATTTACACGCCTATTTAAACCAGAAAGGAGAGTTGGAT GAGATTAGAAAAATAGGAGATTATAACAAAAAGTTTTGAAGCAATG GGTTAAAGGCAAAATATGTTTTATGGAAGTGAATTCACGCTTGATAAG GATTATACACTGAATGTCTATAGATTGGCTTTAAAAACTACCTTAAAA AGAGCAAGAAGGAGTATGGAACCTATAGCAAGAGAGGATGAAAAATCCA AAGGTTGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGTATTTCAT TATAAGGGCGTTGATGTTGCAGTTGGAGGATGGAGCAGAGAAAAATA CACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATTTCAC AACCTGTCTTAACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCA AAAGGGAATTTTATAGCTGTTGATGACTCTCCAGAAGAGATTAGGGCT AAGATAAAGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAATCCA ATAATGGAGATAGCTAAATACCTCCTTGAATATCCTTTAACCATAAAA AGGCCAGAAAAATTTGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCAATGGATTAAAA AATGCTGTAGCTGAAGAACTTATAAAGATTTTAGAGCCAATTAGAAA AGATTA</p>

Fig. 7 (Cont.)