

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 408 581**

51 Int. Cl.:

C12N 9/00 (2006.01)
C12N 15/11 (2006.01)
C12P 21/02 (2006.01)
C12N 1/21 (2006.01)
C12N 5/10 (2006.01)
C12N 15/18 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **01.12.2005 E 05853071 (8)**

97 Fecha y número de publicación de la concesión europea: **13.03.2013 EP 1937824**

54 Título: **Composiciones de ARNt y usos de las mismas**

30 Prioridad:

18.08.2005 US 709364 P

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

21.06.2013

73 Titular/es:

**AMBRX, INC. (100.0%)
10975 NORTH TORREY PINES ROAD, SUITE 100
LA JOLLA CA 92037, US**

72 Inventor/es:

TIAN, FENG

74 Agente/Representante:

IZQUIERDO FACES, José

ES 2 408 581 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

Composiciones de ARNt y usos de las mismas

5 CAMPO DE LA PRESENTE INVENCION

La invención concierne al campo de la bioquímica de la traducción. La invención se refiere a métodos de producción y composiciones de ARNt y usos de las mismas. La invención también se refiere a métodos para producir proteínas en células utilizando tal ARNt y composiciones relacionadas.

10

ANTECEDENTES DE LA PRESENTE INVENCION

15

El código genético de cada organismo conocido, desde las bacterias hasta los seres humanos, codifica los mismos veinte aminoácidos comunes. Diferentes combinaciones de los mismos veinte aminoácidos naturales forman las proteínas que llevan a cabo prácticamente todos los complejos procesos de la vida, desde la fotosíntesis a la transducción de señales y la respuesta inmunitaria. Con el fin de estudiar y modificar la estructura y función de las proteínas, los científicos han tratado de manipular tanto el código genético como la secuencia de aminoácidos de la proteína. Sin embargo, ha sido difícil eliminar las limitaciones impuestas por el código genético que limitan las proteínas a veinte componentes básicos convencionales codificados genéticamente (con la rara excepción de la selenocisteína (véase, por ejemplo, A. Bock *et al.*, (1991), *Molecular Microbiology* 5:515-20) y la pirrolisina (véase, por ejemplo, Srinivasan, G, *et al.*, (2002), *Science* 296:1459-1462).

20

25

Se han realizado algunos avances para eliminar estas limitaciones, aunque estos avances han sido limitados y la capacidad para controlar racionalmente la estructura y la función de las proteínas está todavía en ciernes. Por ejemplo, los químicos han desarrollado métodos y estrategias para sintetizar y manipular las estructuras de moléculas pequeñas (véase, por ejemplo, E.J. Corey y X.M. Cheng, *The Logic of Chemical Synthesis* (Wiley-Interscience, Nueva York, 1995)). La síntesis total (véase, por ejemplo, B. Merrifield, (1986), *Science* 232:341-7 (1986)), y las metodologías de síntesis parcial (véase, por ejemplo, D.Y. Jackson *et al.*, (1994) *Science* 266:243-7; y P.E. Dawson, y S.B. Kent, (2000), *Annual Review of Biochemistry* 69:923-60), han hecho posible sintetizar péptidos y proteínas pequeñas, pero estas metodologías tienen una utilidad limitada con proteínas de más de 10 kilo Daltons (kDa). Los métodos de mutagénesis, aunque potentes, están restringidos a un número limitado de cambios estructurales. En varios casos, ha sido posible incorporar competitivamente análogos estructurales cercanos de los aminoácidos comunes en las proteínas. Véase, por ejemplo, R. Furter, (1998), *Protein Science* 7:419-26; K. Kirshenbaum, *et al.*, (2002), *ChemBioChem* 3:235-7; y V. Doring *et al.*, (2001), *Science* 292:501-4. La ligación química de péptidos y la ligación química nativa se describen en la patente de EE.UU. N° 6.184.344, la publicación de patente de EE.UU. N° 2004/0138412, la publicación de patente de EE.UU. N° 2003/0208046, el documento WO 02/098902 y el documento WO 03/042235. Lu *et al.* en *Mol. Cell.* 2001 Oct; 8(4):759-69 describen un método en el que una proteína se liga químicamente a un péptido sintético que contiene aminoácidos no naturales (ligación de proteína expresada).

30

35

40

45

Los primeros trabajos demostraron que la maquinaria de traducción de *E. coli* se adapta a aminoácidos de estructura similar a los veinte comunes. Véase Hortin, G., y Boime, I. (1983) *Methods Enzymol.* 96:777-784. Este trabajo se amplió adicionalmente relajando la especificidad de las sintetetasas endógenas de *E. coli* para que activaran aminoácidos no naturales, así como su aminoácido natural afín. Por otra parte, se demostró que las mutaciones en los dominios de corrección de errores también podían utilizarse para ampliar el campo de los sustratos de la sintetasa endógena. Véase Doring, V., *et al.*, (2001) *Science* 292:501-504. Sin embargo, estas estrategias se limitan a *recodificar* el código genético más que *ampliar* el código genético y conducir a grados variables de sustitución de uno de los veinte aminoácidos comunes con un aminoácido no natural.

50

55

Más tarde se demostró que pueden incorporarse aminoácidos no naturales de manera específica para el sitio en proteínas *in vitro* añadiendo ARNt supresores de ámbar ortogonales aminoacilados químicamente a una reacción de transcripción/traducción *in vitro*. Véase, por ejemplo, Noren, C. J., *et al.* (1989) *Science* 244:182-188; Bain, J.D., *et al.*, (1989) *J. Am. Chem. Soc.* 111:8013-8014; Dougherty D. A. (2000) *Curr. Opin. Chem. Biol.* 4, 645-652; Cornish, V. W., *et al.* (1995) *Angew. Chem., Int. Ed.* 34:621-633; J.A. Ellman, *et al.*, (1992), *Science* 255:197-200; y D. Mendel, *et al.*, (1995), *Annual Review of Biophysics and Biomolecular Structure* 24:435-462. Estos estudios demuestran que el ribosoma y los factores de traducción son compatibles con un gran número de aminoácidos no naturales, incluso aquellos con estructuras inusuales. Lamentablemente, la aminoacilación química de los ARNt resulta difícil, y la naturaleza estequiométrica de este proceso limita seriamente la cantidad de proteína que podría generarse.

60

65

Se han microinyectado aminoácidos no naturales en células. Por ejemplo, se introdujeron aminoácidos no naturales en el receptor nicotínico de acetilcolina en ovocitos de *Xenopus* (por ejemplo, M.W. Nowak, *et al.* (1998), *In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system*, *Method Enzymol* 293:504-529) mediante microinyección de un ARNt de *Tetrahymena thermophila* acilado químicamente de manera incorrecta (por ejemplo, M.E. Saks, *et al.* (1996), *An engineered Tetrahymena tRNA^{Gln} for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression*, *J. Biol. Chem.* 271:23169-23175), y el ARNm

5 pertinente. Véase, también, D. A. Dougherty (2000), Unnatural amino acids as probes of protein structure and function, *Curr. Opin. Chem. Biol.* 4:645-652 y M.W. Nowak, P.C. Kearney J.R. Sampson, M.E. Saks, C.G. Labarca, S.K. Silverman, W.G. Zhong, J. Thorson, J.N. Abelson, N. Davidson, P.G. Schultz, D.A. Dougherty y H.A. Lester, *Science*, 268:439 (1995). Se inyecta conjuntamente un ovocito de *Xenopus* con dos especies de ARN producidas *in vitro*: un ARNm que codifica la proteína diana con un codón de terminación UAG en la posición de aminoácido de interés y un ARNt supresor de ámbar aminoacilado con el aminoácido no natural deseado. A continuación la maquinaria de traducción del ovocito inserta el aminoácido no natural en la posición determinada por el UAG. Lamentablemente, esta metodología está limitada a las proteínas de células que pueden ser microinyectadas, y debido a que el ARNt pertinente se acila químicamente *in vitro*, y no se puede volver a acilar, los rendimientos de proteína son muy bajos.

10 Para superar estas limitaciones, se añadieron nuevos componentes, por ejemplo, ARNt ortogonales, aminoacil-ARNt sintetasas ortogonales y pares de los mismos, a la maquinaria de biosíntesis de proteínas del procarionta *Escherichia coli* (*E. coli*) (véase, por ejemplo, L. Wang, *et al.*, (2001), *Science* 292:498-500) y el eucariota *Saccharomyces cerevisiae* (*S. cerevisiae*) (por ejemplo, J. Chin *et al.*, *Science* 301:964-7 (2003)) lo que ha permitido incorporar *in vivo* aminoácidos codificados no genéticamente a proteínas. Mediante esta metodología, varios aminoácidos nuevos con propiedades químicas, físicas o biológicas novedosas, que incluyen marcadores de fotoafinidad y aminoácidos fotoisomerizables, aminoácidos de fotoentrecruzamiento (véase, por ejemplo, J.W. Chin, *et al.* (2002) *Proc. Natl. Acad. Sci. EE.UU.* 99:11020-11024; y Chin, J.W., *et al.*, (2002) *J. Am. Chem. Soc.* 124:9026-9027), cet aminoácidos (véase, por ejemplo, Wang, L., *et al.*, (2003) *Proc. Natl. Acad. Sci. EE.UU.* 100:56-61 y Zhang, Z. *et al.*, *Biochem.* 42(22):6735-6746 (2003)), aminoácidos que contienen un átomo pesado, y aminoácidos glicosilados, han sido incorporados a proteínas de manera eficaz y con elevada fidelidad en *E. coli* y en levadura en respuesta a, por ejemplo, el codón ámbar (TAG). Véase, por ejemplo, J.W. Chin, y P.G. Schultz, (2002), *ChemBioChem* 3(11):1135-1137 y L. Wang, y P.G. Schultz, (2002), *Chem. Comm*, 1:1-11.

25 Se han presentado otros varios pares ortogonales. Se han descrito sistemas glutaminilo (véase, por ejemplo, Liu, D.R., y Schultz, P.G. (1999) *Proc. Natl. Acad. Sci. EE.UU.* 96:4780-4785), aspartilo (véase, por ejemplo, Pastrnak, M., *et al.*, (2000) *Helv. Chim. Acta* 83:2277-2286), y tirosilo (véase, por ejemplo, Ohno, S., *et al.*, (1998) *J. Biochem. (Tokio, Japón)* 124:1065-1068, y Kowal, A.K., *et al.*, (2001) *Proc. Natl. Acad. Sci. EE.UU.* 98:2268-2273) provenientes de ARNt y sintetasas de *S. cerevisiae* para la potencial incorporación de aminoácidos no naturales en *E. coli*. Se han descrito sistemas provenientes de glutaminil sintetasa (véase, por ejemplo, Kowal, A.K., *et al.*, (2001) *Proc. Natl. Acad. Sci. EE.UU.* 98:2268-2273) y de tirosil sintetasa (véase, por ejemplo, Edwards, H., y Schimmel, P. (1990) *Mol. Cell. Biol.* 10:1633-1641) de *E. coli* para su uso en *S. cerevisiae*. Se ha utilizado el sistema tirosilo de *E. coli* para incorporar 3-yodo-L-tirosina *in vivo*, en células de mamífero. Véase Sakamoto, K, *et al.*, (2002) *Nucleic Acids Res.* 30:4692-4699. Por lo general, estos sistemas han utilizado el codón de terminación ámbar. Para ampliar adicionalmente el código genético, existe la necesidad de desarrollar componentes mejorados y/o adicionales de la maquinaria de biosíntesis, por ejemplo, los ARNt. La presente invención satisface estas y otras necesidades, como resultará evidente tras la revisión de la siguiente divulgación.

40 RESUMEN DE LA PRESENTE INVENCION

45 Para ampliar el código genético, la invención proporciona composiciones de y métodos de producción de ARNt ortogonales. Las aminoacil-ARNt sintetasas aminoacilan los ARNt de la presente invención con un aminoácido codificado de forma no natural. Pueden utilizarse estos componentes de traducción para incorporar un aminoácido seleccionado en una posición específica en una cadena polipeptídica en crecimiento (durante la traducción del ácido nucleico) en respuesta a un codón selector que es reconocido por el ARNt. En concreto, la invención proporciona una composición que comprende un ARNt codificado por la secuencia de polinucleótido de la SEQ ID NO:1; la SEQ ID NO:2; la SEQ ID NO:3; una secuencia de polinucleótido complementaria de la SEQ ID NO:1; una secuencia de polinucleótido complementaria de la SEQ ID NO:2; o una secuencia de polinucleótido complementaria de la SEQ ID NO:3. La invención también proporciona un vector que comprende un polinucleótido que codifica un ARNt, en el que dicho polinucleótido tiene la secuencia de ácido nucleico de la: SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; una secuencia de polinucleótido complementaria de la SEQ ID NO:1; una secuencia de polinucleótido complementaria de la SEQ ID NO:2; o una secuencia de polinucleótido complementaria de la SEQ ID NO:3.

55 Son también una característica de la presente invención métodos para producir una proteína en una célula con un aminoácido seleccionado en una posición determinada. Por ejemplo, un método incluye cultivar una célula en un medio apropiado, en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína; y proporcionar el aminoácido seleccionado. La célula comprende adicionalmente: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector; y una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente el O-ARNt con el aminoácido seleccionado. Por lo general, el O-ARNt comprende una actividad supresora en presencia de una sintetasa afín. También es una característica de la presente invención una proteína producida mediante este método. En concreto, la invención proporciona un método para producir un polipéptido en una célula con un aminoácido seleccionado en una posición determinada, comprendiendo el método: cultivar la célula en un medio apropiado, en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica un polipéptido; y proporcionar el aminoácido seleccionado; en el que la célula comprende adicionalmente: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el

codón selector, en el que dicho O-ARNt es codificado por el secuencia de polinucleótido de la: SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; una secuencia de polinucleótido complementaria de la SEQ ID NO:1; una secuencia de polinucleótido complementaria de la SEQ ID NO:2; o una secuencia de polinucleótido complementaria de SEQ ID NO:3; una aminoacil-ARNt sintetasa ortogonal (O-RS), en el que dicha O-RS aminoacila el O-ARNt con el aminoácido seleccionado.

BREVE DESCRIPCIÓN DE LOS DIBUJOS

Figura 1 - Se muestra la estructura de hoja de trébol del ARNt de J17 con los sitios de mutación en el tallo de TΨC.

Figura 2 - Se muestra la supresión de una mutación ámbar en la hormona del crecimiento humana utilizando J17 o mutantes de J17 (F12, F13, F14). El lisado de células total de cada muestra se analizó mediante SDS-PAGE.

Figura 3 - Se muestra la supresión de una mutación ámbar en la hormona del crecimiento humana en diferentes líneas celulares utilizando F13.

DEFINICIONES

Antes de describir la invención detalladamente, debe entenderse que la presente invención no está limitada a sistemas biológicos concretos, que pueden, por supuesto, variar. También debe entenderse que la terminología utilizada en el presente documento tiene la finalidad de describir las formas de realización concretas solamente, y no pretende limitar el alcance de la presente invención, que será limitado sólo por las reivindicaciones adjuntas. Tal como se utilizan en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un", "uno", "una", "el" y "la" incluyen los referentes plurales a menos que el contenido dicte claramente lo contrario. Por lo tanto, por ejemplo, la referencia a "una célula" incluye una combinación de dos o más células e incluye equivalentes de las mismas conocidos por los expertos en la técnica, y así sucesivamente. La referencia a "bacteria" incluye mezclas de bacterias y similares.

A menos que se defina en el presente documento y más adelante en el resto de la memoria descriptiva, todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado como entiende comúnmente un experto en la materia a la que pertenece la invención.

Las publicaciones analizadas en el presente documento se proporcionan únicamente por su divulgación antes de la fecha de presentación de la presente solicitud. Nada en el presente documento debe interpretarse como una admisión de que los inventores no tienen derecho a antedatar tal divulgación en virtud de una invención anterior o por cualquier otra razón.

Homólogo: Las proteínas y/o secuencias de proteínas son "homólogas" cuando provienen, natural o artificialmente, de una proteína o secuencia de proteína ancestral común. Del mismo modo, los ácidos nucleicos y/o secuencias de ácido nucleico son homólogos cuando provienen, natural o artificialmente, de un ácido nucleico o secuencia de ácido nucleico ancestral común. Por ejemplo, puede modificarse cualquier ácido nucleico de origen natural mediante cualquier método de mutagénesis disponible para que incluya uno o más codones selectores. Cuando se expresa, este ácido nucleico mutagenizado codifica un polipéptido que comprende uno o más aminoácidos seleccionados, por ejemplo, un aminoácido no natural. El proceso de mutación puede, por supuesto, modificar adicionalmente uno o más codones convencionales, cambiando de este modo también uno o más aminoácidos convencionales en la proteína mutante resultante. El uno o más aminoácidos convencionales pueden cambiarse a un aminoácido no natural o un aminoácido natural. La homología se deduce generalmente a partir de la similitud de secuencia entre dos o más ácidos nucleicos o proteínas (o secuencias de los mismos). El porcentaje preciso de similitud entre secuencias que resulta útil para establecer la homología varía con el ácido nucleico y la proteína en cuestión, pero para establecer la homología se utiliza de forma rutinaria una similitud de secuencia tan reducida como el 25%. Para establecer la homología también pueden utilizarse niveles de similitud de secuencia más elevados, por ejemplo, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, ó 99% o más. Los métodos para determinar los porcentajes de similitud de secuencia (por ejemplo, BLASTN y BLASTP utilizando los parámetros por defecto) se describen en el presente documento y se encuentran disponibles generalmente.

Ortogonal: Tal como se utiliza 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 es utilizada con una eficacia reducida por un sistema de interés (por ejemplo, un sistema de traducción, por ejemplo, una célula). Ortogonal se refiere a la incapacidad o a la eficacia reducida, por ejemplo, una eficacia inferior a un 20%, una eficacia inferior a un 10%, una eficacia inferior a un 5%, o, por ejemplo, una eficacia inferior a un 1%, de un ARNt ortogonal y/o una RS ortogonal para funcionar en el sistema de traducción de interés. Por ejemplo, un ARNt ortogonal en un sistema de traducción de interés es aminoacilado por cualquier RS endógena de un sistema de traducción de interés con una eficacia reducida o incluso nula, en comparación con la aminoacilación de un ARNt

endógeno por una RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno en el sistema de traducción de interés con una eficacia reducida o incluso nula, en comparación con la aminoacilación del ARNt endógeno por una RS endógena. Puede introducirse en la célula 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 conjuntamente en la célula con una eficacia (por ejemplo, una eficacia de aproximadamente un 50%, una eficacia de aproximadamente un 60%, una eficacia de aproximadamente un 70%, una eficacia de aproximadamente un 75%, una eficacia de aproximadamente un 80%, una eficacia de aproximadamente un 85%, una eficacia de aproximadamente un 90%, una eficacia de aproximadamente un 95%, o una eficacia de aproximadamente un 99% o más) con respecto a la de un par ARNt/RS endógeno correspondiente. "Mejora de la ortogonalidad" se refiere a una ortogonalidad potenciada en comparación con un material de partida o un ARNt o una RS de origen natural.

Afín: El término "afín" se refiere a componentes que funcionan conjuntamente, por ejemplo, un ARNt y una aminoacil-ARNt sintetasa. Los componentes pueden denominarse también complementarios.

Aminoacila preferentemente: La expresión "aminoacila preferentemente" se refiere a una eficacia de, por ejemplo, aproximadamente un 70%, una eficacia de aproximadamente un 75%, una eficacia de aproximadamente un 80%, una eficacia de aproximadamente un 85%, una eficacia de aproximadamente un 90%, una eficacia de aproximadamente un 95%, o una eficacia de aproximadamente un 99% o más, con la que una O-RS aminoacila un O-ARNt con un aminoácido seleccionado, por ejemplo, un aminoácido no natural, en comparación con la O-RS que aminoacila un material de partida o un ARNt de origen natural utilizado para generar el O-ARNt. A continuación, se incorpora el aminoácido no natural a una cadena polipeptídica en crecimiento con elevada fidelidad, por ejemplo, con una eficacia de más de aproximadamente un 70% para un codón selector determinado, una eficacia de más de aproximadamente un 75% para un codón selector determinado, una eficacia de más de aproximadamente un 80% para un codón selector determinado, una eficacia de más de aproximadamente un 85% para un codón selector determinado, una eficacia de más de aproximadamente un 90% para un codón selector determinado, una eficacia de más de aproximadamente un 95% para un codón selector determinado, o una eficacia de más de aproximadamente un 99% para un codón selector determinado.

Codón selector: La expresión "codón selector" se refiere a los codones reconocidos por el O-ARNt en el proceso de traducción y no reconocidos por un ARNt endógeno. El bucle del anticodón del O-ARNt reconoce el codón selector en el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido seleccionado, tal como un aminoácido no natural, en este sitio en el polipéptido. Los codones selectores pueden incluir, pero no se limitan a, por ejemplo, codones sin sentido, tales como, codones de terminación, que incluyen, pero no se limitan a, los codones ámbar, ocre y ópalo; codones de cuatro bases o más; codones raros; codones provenientes de pares de bases naturales o no naturales y/o similares. Para un sistema determinado, un codón selector puede incluir también uno de los codones naturales de tres bases, en el que el sistema endógeno no utiliza (o utiliza raramente) dicho codón natural de tres bases. Por ejemplo, esto incluye un sistema que carece de un ARNt que reconoce el codón natural de tres bases, y/o un sistema en el que el codón natural de tres bases es un codón raro.

ARNt supresor: Un ARNt supresor es un ARNt que modifica la lectura de un ARN mensajero (ARNm) en un sistema de traducción determinado, por ejemplo, proporcionando un mecanismo para incorporar un aminoácido en una cadena polipeptídica en respuesta a un codón selector. Por ejemplo, un ARNt supresor puede ultraleer un codón que incluye, pero no se limita a, un codón de terminación, un codón de cuatro bases o un codón raro.

Actividad supresora: La expresión "actividad supresora" se refiere a la capacidad de un ARNt, por ejemplo, un ARNt supresor, para ultraleer un codón selector. La actividad puede expresarse como un porcentaje de la actividad observada en comparación con un testigo (por ejemplo, que carece de una sintetasa afín).

Sistema de traducción: La expresión "sistema de traducción" se refiere a los componentes necesarios para incorporar un aminoácido de origen natural a una cadena polipeptídica en crecimiento (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetasa, ARNm y similares. Los componentes de la presente invención pueden añadirse a un sistema de traducción *in vivo* o *in vitro*. Los ejemplos de sistemas de traducción incluyen, pero no se limitan a, una célula no eucariota, por ejemplo, una bacteria (tal como *E. coli*), 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, un sistema de traducción libre de células, por ejemplo, un lisado de células, y/o similares.

Los sistemas de traducción pueden ser celulares o libres de células, y pueden ser de procariotas o de eucariotas. Los sistemas de traducción celulares incluyen, pero no se limitan a, preparaciones de células completas tales como células permeabilizadas o cultivos celulares en los que una secuencia de ácido nucleico deseado puede transcribirse a ARNm y puede traducirse el ARNm. Se encuentran disponibles comercialmente sistemas de traducción libres de células y se conocen bien muchos sistemas y tipos diferentes. Los ejemplos de sistemas libres de células incluyen, pero no se limitan a, lisados de procariotas tales como lisados de *Escherichia coli*, y lisados de eucariotas tales como extractos de germen de trigo, lisados de células de insectos, lisados de reticulocitos de conejo, lisados de ovocitos de conejo y los lisados de células humanas. Los extractos o lisados de eucariotas

pueden resultar preferentes cuando la proteína resultante está glicosilada, fosforilada o modificada de otro modo, porque muchas de estas modificaciones sólo son posibles en los sistemas de eucariotas. Algunos de estos extractos y lisados se encuentran disponibles comercialmente (Promega; Madison, Wis.; Stratagene; La Jolla, Calif.; Amersham, Arlington Heights, IL; GIBCO/BRL; Grand Island, N.Y.). También se encuentran disponibles extractos membranosos, tales como extractos pancreáticos caninos que contienen membranas microsomales, que resultan útiles para traducir proteínas de secreción.

También pueden utilizarse sistemas de traducción reconstituidos. También se han utilizado con éxito mezclas de factores de traducción purificados para traducir el ARNm a proteínas, así como combinaciones de lisados o lisados complementados con factores de traducción purificados, tales como factor de iniciación 1 (IF-1), IF-2, IF-3 (α o β), factor de elongación T (EF-Tu), o factores de terminación. Los sistemas libres de células también pueden ser sistemas de transcripción/traducción acoplados en los que el ADN se introduce en el sistema, se transcribe a ARNm y el ARNm se traduce como se describe en Current Protocols in Molecular Biology (F.M. Ausubel *et al.* editores, Wiley Interscience, 1993). El ARN transcrito en el sistema de transcripción eucariota puede estar en forma de ARN heteronuclear (ARNhn) o protecciones terminales en el extremo 5' (7-metil guanosina) y ARNm maduro con cola de poli A en el extremo 3', que puede ser una ventaja en determinados sistemas de traducción. Por ejemplo, los ARNm con protección terminal se traducen con una alta eficacia en el sistema de lisado de reticulocitos.

Aminoácido seleccionado: La expresión "aminoácido seleccionado" se refiere a cualquier aminoácido de origen natural o aminoácido no natural deseado. Tal como se utiliza en el presente documento, la expresión "aminoácido no natural" o "aminoácido codificado de forma 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 selenocisteína o pirrolisina. Otras expresiones que pueden utilizarse como sinónimos de la expresión "aminoácido codificado de forma no natural" y "aminoácido no natural" son "aminoácido de origen no natural" y diversas versiones con guiones y sin guiones de la misma. La expresión "aminoácido codificado de forma no natural" también incluye, pero no se limita a, los aminoácidos que se producen por modificación (por ejemplo, modificaciones post-traduccionales) de un aminoácido codificado de forma natural (que incluye, pero no se limita a, los 20 aminoácidos comunes o pirrolisina y selenocisteína) pero ellos mismos no son incorporados de forma natural a una cadena polipeptídica en crecimiento por el complejo de traducción. Los ejemplos de tales aminoácidos de origen no natural incluyen, pero no se limitan a, *N*-acetilglucosaminil-L-serina, *N*-acetilglucosaminil-L-treonina, y O-fosfotirosina.

Proveniente de: Tal como se utiliza en el presente documento, la expresión "proveniente de" se refiere a un componente que se aísla a partir de o se produce utilizando información de una molécula u organismo determinado.

Marcador de cribado o de selección positiva: Tal como se utiliza en el presente documento, la expresión "marcador de cribado o de selección positiva" se refiere a un marcador que cuando se encuentra presente, por ejemplo, expresado, activado o similares, da como resultado la identificación de una célula con el marcador de selección positiva frente a aquellas sin el marcador de selección positiva.

Marcador de cribado o de selección negativa: Tal como se utiliza en el presente documento, la expresión "marcador de cribado o de selección negativa" se refiere a un marcador que cuando se encuentra presente, por ejemplo, expresado, activado o similares, permite la identificación de una célula que no posee la propiedad deseada (por ejemplo, en comparación con una célula que sí posee la propiedad deseada).

Indicador: Tal como se utiliza en el presente documento, el término "indicador" se refiere a un componente que puede utilizarse para seleccionar componentes diana de un sistema de interés. Por ejemplo, un indicador puede incluir una proteína, por ejemplo, una enzima, que confiere sensibilidad o resistencia a antibióticos (que incluye, pero no se limita a, β -lactamasa, cloranfenicol acetiltransferasa (CAT), y similares), un marcador de cribado fluorescente (que incluye, pero no se limita a, proteína fluorescente verde (por ejemplo, GFP), YFP, EGFP, RFP, un marcador luminiscente (que incluye, pero no se limita a, una proteína de luciferasa de luciérnaga), un marcador de cribado 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).

Eucariota: Tal como se utiliza en el presente documento, el término "eucariota" se refiere a organismos pertenecientes al dominio filogenético *Eucarya*, tales como animales (que incluyen, pero no se limitan a, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (que incluyen, pero no se limitan a, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidios, protistas, etc.

No eucariota: Tal como se utiliza en el presente documento, la expresión "no eucariota" se refiere a los organismos no son eucariotas. Por ejemplo, un organismo no eucariota puede pertenecer al dominio filogenético *Eubacteria* (que incluye, pero no se limita a, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), o al dominio filogenético *Archaea* (por ejemplo, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloflex volcanii* y las especies de *Halobacterium* *NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, etc.).

Variante conservadora: La expresión " variante conservadora" se refiere a un componente de traducción, por ejemplo, un O-ARNt variante conservadora o una O-RS variante conservadora, que se comporta funcionalmente como el componente en el que se basa la variante conservadora, por ejemplo, un O-ARNt o una O-RS, pero que tiene variaciones en la secuencia. Por ejemplo, una O-RS aminoacilará un O-ARNt complementario o un O-ARNt variante conservadora con un aminoácido seleccionado, por ejemplo, un aminoácido no natural, aunque el O-ARNt y el O-ARNt variante conservadora no tengan la misma secuencia. Del mismo modo, un ARNt será aminoacilado con un aminoácido seleccionado, por ejemplo, un aminoácido no natural, por una O-RS complementaria o una O-RS variante conservadora, aunque la O-RS y la O-RS variante conservadora no tengan la misma secuencia. La variante conservadora puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones o cinco o más variaciones en la secuencia, siempre que la variante conservadora sea complementaria al O-ARNt o a la O-RS correspondiente.

Agente de selección o cribado: Tal como se utiliza en el presente documento, la expresión "agente de selección o cribado" se refiere a un agente que, cuando se encuentra presente, permite la selección/cribado de determinados componentes de una población. Por ejemplo, un agente de selección o cribado incluye, pero no se limita a, por ejemplo, un nutriente, un antibiótico, una longitud de onda de luz, un anticuerpo, un polinucleótido expresado, o similares. El agente de selección puede modificarse, por ejemplo, en concentración, intensidad, etc.

La expresión "no reconocido de manera eficaz" se refiere a una eficacia, por ejemplo, inferior a aproximadamente un 10%, inferior a aproximadamente un 5%, o inferior a aproximadamente un 1%, con la que una RS de un organismo aminoacila el O-ARNt.

DESCRIPCIÓN DETALLADA

Los sistemas de traducción que resultan adecuados para producir proteínas que incluyen uno o más aminoácidos seleccionados, por ejemplo, un aminoácido no natural, se describen en las solicitudes de patente de EE.UU 10/126.931, titulada "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL tRNA SYNTHETASE PAIRS" y 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Además, véase el documento USSN 10/825.867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". Tales sistemas de traducción comprenden generalmente células que incluyen un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), y un aminoácido seleccionado, por ejemplo, un aminoácido no natural, en los que la O-RS aminoacila el O-ARNt con el aminoácido seleccionado. Un par ortogonal de la presente invención se compone de un O-ARNt, por ejemplo, un ARNt supresor, un ARNt de desplazamiento de marco de lectura, o similares, y una O-RS. El O-ARNt reconoce un primer codón selector y tiene actividad supresora en presencia de una sintetasa afín en respuesta a un codón selector. La célula utiliza los componentes para incorporar el aminoácido seleccionado a una cadena polipeptídica en crecimiento. Por ejemplo, también puede estar presente 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. El sistema de traducción también puede ser un sistema *in vitro*. Las moléculas de ARNt de la presente invención resultan útiles en cualquier sistema de traducción, incluyendo los sistemas que utilizan ribosomas en la traducción.

El sistema de traducción también puede ser un sistema de traducción libre de células (*in vitro*). En estos sistemas, que pueden incluir ARNm como molde (traducción *in vitro*) o ADN como molde (transcripción y traducción *in vitro* combinadas), la síntesis *in vitro* es dirigida por los ribosomas. Se ha realizado un esfuerzo considerable para el desarrollo de sistemas de expresión de proteínas libres de células. Véase, por ejemplo, Kim, D.M. y J.R. Swartz, *Biotechnology and Bioengineering*, 74:309-316 (2001), Kim, D.M. y J.R. Swartz, *Biotechnology Letters*, 22, 1537-1542, (2000), Kim, D.M. y J.R. Swartz, *Biotechnology Progress*, 16, 385-390, (2000), Kim, D.M. y J.R. Swartz, *Biotechnology and Bioengineering*, 66, 180-188, (1999); y Patnaik, R. y J.R. Swartz, *Biotechniques* 24, 862-868, (1998), la patente de EE.UU. N° 6.337.191; la publicación de patente de EE.UU. N° 2002/0081660, el documento WO 00/55353; el documento WO 90/05785. Otro enfoque que puede aplicarse incluye la técnica de fusión de ARNm-péptido. Véase, por ejemplo, R. Roberts y J. Szostak, *Proc. Natl Acad. Sci. (EE.UU)* 94:12297-12302 (1997); A. Frankel, *et al.*, *Chemistry & Biology* 10:1043-1050 (2003). En este enfoque, un molde de ARNm unido a puromicina se traduce a péptido en el ribosoma. Si se han modificado una o más moléculas de ARNt, también pueden incorporarse aminoácidos no naturales al péptido. Después de haber sido leído el último codón del ARNm, la puromicina captura el extremo C-terminal del péptido. Si se descubre que el conjugado de ARNm-péptido resultante tiene propiedades interesantes en un ensayo *in vitro*, su identidad puede ponerse de manifiesto fácilmente a partir de la secuencia de ARNm. De esta manera, podrían cribarse bibliotecas de polipéptidos que comprenden uno o más aminoácidos codificados de forma no natural para identificar los polipéptidos que tienen las propiedades deseadas. Más recientemente, se han presentado traducciones en ribosomas *in vitro* con componentes purificados que permiten la síntesis de péptidos sustituidos con aminoácidos codificados de forma no natural. Véase, por ejemplo, A. Forster *et al.*, *Proc. Natl Acad. Sci. (EE.UU)* 100:6353 (2003).

En determinadas formas de realización, una célula de *E. coli* que comprende el ARNt de la presente invención incluye un sistema de traducción de este tipo. Por ejemplo, la célula de *E. coli* de la presente invención incluye un ARNt ortogonal (O-ARNt), en el que el O-ARNt comprende actividad supresora en presencia de una sintetasa afín en respuesta a un codón selector; una aminoacil-ARNt sintetasa ortogonal (O-RS); un aminoácido

seleccionado; y 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.

5 La invención también presenta pares O-ARNt/O-RS múltiples en una célula, lo que permite incorporar más de un aminoácido seleccionado. En determinadas formas de realización, la célula puede incluir adicionalmente un par O-ARNt/O-RS diferente adicional y un segundo aminoácido seleccionado, en el que el O-ARNt reconoce un segundo codón selector y la O-RS aminoacila preferentemente el O-ARNt con el segundo aminoácido seleccionado. Por ejemplo, una célula puede comprender adicionalmente, por ejemplo, un par ARNt supresor de ámbar/aminoacil-ARNt sintetasa proveniente de la tirosil-ARNt sintetasa de *Methanococcus jannaschii*.

10 El O-ARNt y/o la O-RS pueden ser de origen natural o pueden obtenerse por mutación de un ARNt y/o RS de origen natural, por ejemplo, que genera bibliotecas de ARNt y/o bibliotecas de RS, a partir de diversos organismos. Por ejemplo, una estrategia de producción de un par ARNt ortogonal/aminoacil-ARNt sintetasa implica importar a la célula hospedadora un par ARNt/sintetasa heterólogo de, por ejemplo, una fuente distinta de la célula hospedadora, o de fuentes múltiples. Las propiedades del candidato sintetasa heteróloga incluyen, por ejemplo, que no cargue ningún ARNt de la célula hospedadora, y las propiedades del candidato ARNt heterólogo incluyen, por ejemplo, que no sea aminoacilado por ninguna sintetasa de la célula hospedadora. Además, el ARNt heterólogo es ortogonal a todas las sintetasas de la célula hospedadora.

20 Una segunda estrategia para generar un par ortogonal implica la generación de bibliotecas de mutantes a partir de las cuales cribar y/o seleccionar un O-ARNt o una O-RS. Estas estrategias también pueden combinarse.

25 En diversas formas de realización, el O-ARNt y la O-RS provienen de al menos un organismo. En otra forma de realización, el O-ARNt proviene de un ARNt de origen natural o de origen natural mutado de un primer organismo y la O-RS proviene de una RS de origen natural o de origen natural mutada de un segundo organismo. En una forma de realización, el primer y el segundo organismo son diferentes. Por ejemplo, un par ortogonal puede incluir una ARNt sintetasa proveniente de *Methanobacterium thermoautotrophicum*, y un ARNt proveniente de un ARNt de *Archaea* (por ejemplo, de *Halobacterium sp. NRC-1*). Como alternativa, el primer y el segundo organismo son iguales. Para obtener información adicional véase la sección titulada "Organismos hospedadores y de partida" en el presente documento.

30 En determinadas formas de realización de la presente invención, un O-ARNt de la presente invención es codificado por la secuencia de polinucleótido de la: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, una secuencia de polinucleótido complementaria de la SEQ ID NO:1, una secuencia de polinucleótido complementaria de la SEQ ID NO:2 o una secuencia de polinucleótido complementaria de la SEQ ID NO:3. Véase también la sección titulada "Secuencia de polipéptido y ácido nucleico y variantes" en el presente documento.

ARNt ortogonal (O-ARNt)

40 Un ARNt ortogonal (O-ARNt) interviene en la incorporación de un aminoácido seleccionado a una proteína que es codificada por un polinucleótido que comprende un codón selector que es reconocido por el O-ARNt, por ejemplo, *in vivo* o *in vitro*. Un O-ARNt de la presente invención puede ser aminoacilado con un aminoácido deseado mediante cualquier método o técnica, que incluye, pero no se limita a, aminoacilación química o enzimática. El O-ARNt aminoacilado de la presente invención puede añadirse directamente a un sistema de traducción. Un O-ARNt de la presente invención puede ser aminoacilado por una RS con un aminoácido seleccionado *in vivo* o *in vitro*. Además, la RS puede ser una O-RS. Puede proporcionarse un O-ARNt de la presente invención al sistema de traducción (por ejemplo, componentes de traducción *in vitro*, o una célula) directamente, o proporcionando un polinucleótido que codifica un O-ARNt o una porción del mismo. Por ejemplo, un O-ARNt, o una porción del mismo, es codificado por la secuencia de polinucleótido de la: SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, una secuencia de polinucleótido complementaria de la SEQ ID NO:1, una secuencia de polinucleótido complementaria de la SEQ ID NO:2 o una secuencia de polinucleótido complementaria de la SEQ ID NO:3.

55 Un O-ARNt de la presente invención comprende actividad supresora en presencia de una sintetasa afín en respuesta a un codón selector. La actividad supresora puede determinarse mediante cualquiera de varios ensayos conocidos en la técnica. Por ejemplo, puede utilizarse un ensayo de indicador β -galactosidasa. Se utiliza un derivado de un plásmido que expresa el gen *lacZ* bajo el control del promotor, por ejemplo, en el que la Leu-25 del péptido VVLQRRDWEN de *lacZ* es sustituido por un codón selector, por ejemplo, los codones TAG, TGA, AGGA, etc., o codones codificantes (como un testigo) para la tirosina, serina, leucina, etc. Se introduce el plásmido *lacZ* derivado en las 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 presente invención. También puede introducirse una sintetasa afín (como polipéptido o como polinucleótido que codifica la sintetasa afín cuando se expresa). Las células se cultivan en medio a una densidad deseada, por ejemplo, a una DO_{600} de aproximadamente 0,5, y se realizan ensayos de β -galactosidasa, por ejemplo, utilizando el kit de ensayo de β -galactosidasa BetaFluor™ (Novagen). La supresión porcentual se calcula como el porcentaje de actividad para una muestra con respecto a un testigo comparable, por ejemplo, el valor observado a partir del constructo *lacZ* derivado, en el que el constructo tiene un codón codificante correspondiente en la posición deseada más que un codón selector.

En la molécula de ARNt, la Timina (T) se sustituye con Uracilo (U). Además, pueden estar presentes modificaciones adicionales a las bases. La divulgación también incluye variaciones conservadoras de O-ARNt. Por ejemplo, las variaciones conservadoras de O-ARNt incluyen aquellas moléculas que funcionan como el O-ARNt y mantienen la estructura en forma de L del ARNt, pero no tienen la misma secuencia (y son distintas de las moléculas de ARNt de tipo silvestre). Véase también la sección del presente documento titulada "Secuencia de polipéptido y ácido nucleico y variantes".

La composición que comprende un O-ARNt puede incluir adicionalmente una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que la O-RS aminoacila preferentemente el O-ARNt con un aminoácido seleccionado (por ejemplo, un aminoácido no natural). En determinadas formas de realización, una composición que incluye un O-ARNt puede incluir adicionalmente un sistema de traducción (por ejemplo, un sistema de traducción *in vivo* o *in vitro*). Un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende uno o más codones selectores reconocidos por el O-ARNt, o una combinación de uno o más de estos, también puede estar presente en la célula o en otro sistema de traducción. Véase también la sección del presente documento titulada "Aminoacil-ARNt sintetasa ortogonales (O-RS)".

Son también una característica de la presente invención métodos para producir un ARNt ortogonal (O-ARNt), por ejemplo, un O-ARNt. Es también una característica de la presente invención un ARNt, por ejemplo, un O-ARNt, producido mediante el método.

Los métodos para producir un ARNt ortogonal incluyen mutar el bucle del anticodón de cada uno de una combinación de ARNt para permitir el reconocimiento de un codón selector (por ejemplo, un codón ámbar, un codón ópalo, un codón de cuatro bases, etc.), proporcionando de este modo una pluralidad de O-ARNt potenciales; y analizar la estructura secundaria de un miembro de una pluralidad de O-ARNt potenciales para identificar pares de bases no canónicos en la estructura secundaria, y opcionalmente mutar los pares de bases no canónicos (por ejemplo, los pares de bases no canónicos son mutados a pares de bases canónicos). Los pares de bases no canónicos pueden estar ubicados en la región del tallo de la estructura secundaria. Un O-ARNt puede poseer una mejora de una o más características o actividades, tales como la mejora de la ortogonalidad para un organismo deseado en comparación con el material de partida, por ejemplo, la pluralidad de secuencias de ARNt, al tiempo que conserva su afinidad hacia una RS deseada.

Como alternativa, puede desarrollarse un O-ARNt mutando un ARNt conocido para modular su interacción con o afinidad de unión a una o más moléculas que influyen en la traducción o son componentes de la maquinaria de traducción. Tales componentes incluyen, pero no se limitan a, factores de elongación. El factor de elongación EF-Tu bacteriano juega un papel clave en la etapa de elongación en la síntesis de proteínas. Después de la aminoacilación del ARNt por la ARNt sintetasa, el EF-Tu se une al ARNt aminoacilado y lo lleva al sitio A del ribosoma. El enlace éster entre el aminoácido cargado y el ARNt está protegido de la hidrólisis espontánea debido a la unión entre el EF-Tu y el ARNt aminoacilado. Stortchevoi *et al.* investigaron mutantes del par de bases oscilante U50:G64 del ARNt^{fMet} iniciador de *E. coli* en el tallo de TΨC, ya que se descubrió que este par de bases era un determinante negativo secundario que bloqueaba la actividad de la ARNt en la elongación, presumiblemente debido a un debilitamiento de la interacción entre el EF-Tu.GTP y el ARNt aminoacilado (JBC 2003 278(20):17672-17679). Además, LaRiviere *et al.* describieron en Science del 5 de octubre de 2001; 294(5540):165-8 las contribuciones termodinámicas del cuerpo del ARNt y el aminoácido a la afinidad de unión global a EF-Tu. Indicaron que las contribuciones del cuerpo del ARNt y el aminoácido son independientes entre sí y que se compensan entre sí cuando los ARNt están correctamente acilados. Las modificaciones en la interacción entre el EF-Tu.GTP y el ARNt aminoacilado con el aminoácido no natural pueden influir en la eficacia de la carga del ARNt al sitio A del ribosoma. También pueden encontrarse sitios de mutación potenciales analizando las estructuras cristalinas de los complejos entre el ARNt y otros componentes de la maquinaria de traducción, tales como EF-Tu. Por ejemplo, Nissen *et al.* han indicado que EF-Tu.GTP se une directamente a la cadena principal de fosfato del tallo de TΨC del ARN de transferencia de la fenilalanina (ARNtPhe) de levadura (Science 1995 270(5241):1464-1472).

Los métodos incluyen opcionalmente analizar la homología de secuencia de los ARNt y/o las aminoacil-ARNt sintetasa para determinar candidatos potenciales para un O-ARNt, una O-RS y/o pares de los mismos, que parecen ser ortogonales para un organismo específico. Para el análisis pueden utilizarse los programas informáticos conocidos en la técnica y descritos en el presente documento. En un ejemplo, para elegir los posibles componentes de traducción ortogonales para su uso en un organismo procarionta, se elige una sintetasa y/o un ARNt que no presenten homología inusual para los organismos procariontas.

También puede producirse una combinación de ARNt mediante una estrategia de consenso. Por ejemplo, la combinación de ARNt se produce alineando una pluralidad de secuencias de ARNt; determinando una secuencia de consenso; y generando una biblioteca de ARNt utilizando al menos una parte, la mayor parte de, o toda la secuencia de consenso. Por ejemplo, puede compilarse una secuencia de consenso con un programa informático, por ejemplo, el programa GCG *pileup*. Opcionalmente, las posiciones degeneradas determinadas por el programa se cambian a la base más frecuente en esas posiciones. Se sintetiza una biblioteca mediante técnicas conocidas en la técnica utilizando la secuencia de consenso. Por ejemplo, puede utilizarse la extensión de solapamiento de los oligonucleótidos en los que cada sitio del gen del ARNt puede sintetizarse como una mezcla dopada de un 90% de

la secuencia de consenso y un 10% de una mezcla de las otras 3 bases, para proporcionar la biblioteca en base a la secuencia de consenso. También pueden utilizarse otras mezclas, por ejemplo, un 75% de la secuencia de consenso y un 25% de una mezcla de las otras 3 bases, un 80% de la secuencia de consenso y un 20% de una mezcla de las otras 3 bases, un 95% de la secuencia de consenso y un 5% mezcla de las otras 3 bases, etc.

Pueden generarse bibliotecas de ARNt mutante utilizando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, puede generarse el ARNt mutante mediante mutaciones específicas de sitio, mutaciones puntuales aleatorias, recombinación homóloga, barajado de ADN u otros métodos de mutagénesis recursivos, construcción quimérica o cualquier combinación de los mismos.

Pueden introducirse mutaciones adicionales en una posición o posiciones específicas, por ejemplo, en una posición o posiciones no conservadoras, o en una posición o posiciones conservadoras, en una posición o posiciones aleatorizadas, o una combinación de las mismas en una región o bucle deseado de un ARNt, por ejemplo, un bucle del anticodón, el tallo del aceptor, brazo o bucle D, bucle variable, brazo o bucle TΨC, otras regiones de la molécula de ARNt, o una combinación de los mismos. Las mutaciones pueden incluir pares de bases apareados de la región del tallo.

Por lo general, un O-ARNt se obtiene sometiendo a selección negativa una población de células de una primera especie, en la que las células comprenden un miembro de la pluralidad de O-ARNt potenciales. La selección negativa elimina las células que comprenden un miembro de la pluralidad de O-ARNt potenciales que es aminoacilado por una aminoacil-ARNt sintetasa (RS) que es endógena a las células. Esto proporciona una combinación de ARNt que son ortogonales a la célula de la primera especie.

En determinadas formas de realización de la selección negativa, se introduce un codón o 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, etc. El cribado/selección puede realizarse cultivando la población de células en presencia de un agente de selección (por ejemplo, un antibiótico, tal como ampicilina). En una forma de realización, se modifica la concentración del agente de selección.

Por ejemplo, para medir la actividad de los ARNt supresores, se utiliza un sistema de selección que se basa en la supresión *in vivo* del codón selector, por ejemplo, mutaciones terminadoras o del marco de lectura introducidas en un polinucleótido que codifica un marcador de selección negativa, por ejemplo, un gen para β-lactamasa (*bla*). Por ejemplo, se construyen variantes de polinucleótidos, por ejemplo, las variantes *bla*, con, por ejemplo, TAG, AGGA, y TGA, en una posición determinada. Las células, por ejemplo, bacterias, se transforman con estos polinucleótidos. En el caso de un ARNt ortogonal, que no puede ser cargado de manera eficaz por las sintetasas endógenas de *E. coli*, la resistencia a antibióticos, por ejemplo, resistencia a la ampicilina, debería ser aproximadamente, o inferior a, la de una bacteria transformada sin plásmidos. Si el ARNt no es ortogonal, o si se co-expresa en el sistema una sintetasa heteróloga capaz de cargar el ARNt, se observa un mayor nivel de resistencia a antibiótico, por ejemplo, ampicilina. Se eligen células, por ejemplo, bacterias, que son incapaces de crecer en placas de agar LB con concentraciones de antibiótico aproximadamente iguales a las de las células transformadas sin plásmidos.

En el caso de un producto tóxico (por ejemplo, ribonucleasa barnasa), cuando un miembro de la pluralidad de ARNt potenciales es aminoacilado por las sintetasas del hospedador endógeno, por ejemplo, *Escherichia coli* (es decir, no es ortogonal a las sintetasas del hospedador, por ejemplo, *Escherichia coli*), el codón selector es suprimido y el producto polinucleótido tóxico producido conduce a la muerte celular. Las células que albergan ARNt ortogonales o ARNt no funcionales sobreviven.

En una forma de realización, se somete a continuación la combinación de ARNt que son ortogonales a un organismo deseado a una selección positiva en la que se coloca un codón selector en un marcador de selección positiva, por ejemplo, codificado por un gen de resistencia a fármacos, tal como un gen de la β-lactamasa. La selección positiva se realiza en una célula que comprende un polinucleótido que codifica o comprende un miembro de la combinación de ARNt, un polinucleótido que codifica un marcador de selección positiva, y un polinucleótido que codifica una RS afín. Estos polinucleótidos se expresan en la célula y se cultiva la célula en presencia de un agente de selección, por ejemplo, ampicilina. A continuación se seleccionan los ARNt por su capacidad para ser aminoacilados por la sintetasa afín co-expresada y para insertar un aminoácido en respuesta a este codón selector. Por lo general, estas células muestran una potenciación de la eficacia de supresión en comparación con las células que albergan ARNt no funcionales, o ARNt que no pueden ser reconocidos de manera eficaz por la sintetasa de interés. Las células que albergan los ARNt no funcionales o que no son reconocidos de manera eficaz por la sintetasa de interés, son sensibles al antibiótico. Por lo tanto, los ARNt que: (i) no son sustratos para las sintetasas endógenas del hospedador, por ejemplo, *Escherichia coli*; (ii) pueden ser aminoacilados por la sintetasa de interés; y (iii) son funcionales en la traducción, sobreviven a ambas selecciones.

En los métodos anteriormente descritos, puede modificarse opcionalmente la rigurosidad de la selección,

por ejemplo, la selección positiva, la selección negativa o la selección positiva y la negativa. Por ejemplo, debido a que la barnasa es una proteína extremadamente tóxica, la rigurosidad de la selección negativa puede controlarse introduciendo diferentes números de codones selectores en el gen de la barnasa y/o utilizando un promotor inducible. En otro ejemplo, se modifica la concentración del agente de selección o cribado (por ejemplo, ampicilina).
 5 En un aspecto, se modifica la rigurosidad debido a que la actividad deseada puede ser baja durante las primeras rondas. Por lo tanto, se aplican criterios de selección menos estrictos en las primeras rondas y criterios más estrictos en las rondas de selección posteriores. En determinadas formas de realización, la selección negativa, la selección positiva, o la selección positiva y la negativa pueden repetirse múltiples veces. Pueden utilizarse múltiples marcadores de selección negativa, marcadores de selección positiva o marcadores de selección positiva y negativa
 10 diferentes. En determinadas formas de realización, el marcador de selección positiva y negativa puede ser el mismo.

Pueden utilizarse otros tipos de selecciones/cribados en la invención para producir componentes de traducción ortogonales, por ejemplo, un O-ARNt, una O-RS, y un par O-ARNt/O-RS. Por ejemplo, el marcador de selección negativa, el marcador de selección positiva o los marcadores de selección positiva y negativa pueden
 15 incluir un marcador que emite fluorescencia o cataliza una reacción luminiscente en presencia de un reactivo adecuado. En otra forma de realización, se detecta un producto del marcador mediante clasificación de células activadas por fluorescencia (FACS) o mediante luminiscencia. Opcionalmente, el marcador incluye un marcador de cribado basado en afinidad. Véase 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.
 20 EE.UU. 90:10444-8.

Pueden encontrarse métodos adicionales para producir un ARNt ortogonal recombinante, por ejemplo, en las solicitudes de patente de EE.UU. 10/126.931, titulada "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs" y 10/126.127, titulada "In vivo Incorporation of Unnatural Amino
 25 Acids", y el documento USSN 10/825.867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". 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.

Puede aminoacilarse un ARNt de la presente invención con un aminoácido deseado mediante cualquier
 30 método o técnica, que incluye, pero no se limita a, aminoacilación química o enzimática.

La aminoacilación puede ser llevada a cabo por aminoacil-ARNt sintetasas o por otras moléculas enzimáticas, que incluyen, pero no se limitan a, ribozimas. El término "ribozima" es intercambiable con "ARN
 35 catalítico". Cech y colaboradores (Cech, 1987, Science, 236:1532-1539; McCorkle *et al.*, 1987, Concepts Biochem. 64:221-226) demostraron la presencia de ARN de origen natural que pueden actuar como catalizadores (ribozimas). Sin embargo, aunque estos catalizadores de ARN naturales sólo han demostrado actuar en sustratos de ácido ribonucleico para la escisión y empalme, el reciente desarrollo de la evolución artificial de ribozimas ha ampliado el repertorio de catálisis a diversas reacciones químicas. Los estudios han identificado moléculas de ARN que pueden
 40 catalizar enlaces aminoacil-ARN en sus propios extremos (2')3'-terminales (Illangakere *et al.*, 1995 Science 267:643-647), y una molécula de ARN que puede transferir un aminoácido de una molécula de ARN a otra (Lohse *et al.*, 1996, Nature 381:442-444).

La publicación de solicitud de patente de EE.UU. 2003/0228593, describe métodos para construir ribozimas y su uso en la aminoacilación de ARNt con aminoácidos codificados de forma natural y codificados de forma no
 45 natural. Las formas inmovilizadas sobre sustrato de moléculas enzimáticas que pueden aminoacilar los ARNt, que incluyen, pero no se limitan a, ribozimas, pueden permitir la purificación por afinidad eficaz de los productos aminoacilados. Los ejemplos de sustratos adecuados incluyen agarosa, sefarosa, y perlas magnéticas. Se describe la producción y el uso de una forma de ribozima inmovilizada sobre sustrato para la aminoacilación en Chemistry
 50 and Biology 2003, 10:1077-1084 y en la publicación de solicitud de patente de EE.UU. 2003/0228593.

Los métodos químicos de aminoacilación incluyen, pero no se limitan a, los introducidos por Hecht y colaboradores (Hecht, S. M. Acc. Chem. Res. 1992, 25, 545; Heckler, T. G.; Roesser, J. R.; Xu, C.; Chang, P.;
 55 Hecht, S. M. Biochemistry 1988, 27, 7254; Hecht, S. M.; Alford, B. L.; Kuroda, Y.; Kitano, S. J. Biol. Chem. 1978, 253, 4517) y por Schultz, Chamberlin, Dougherty y otros (Cornish, V. W.; Mendel, D.; Schultz, P. G. Angew. Chem. Int. Ed. Engl. 1995, 34, 621; Robertson, S. A.; Ellman, J. A.; Schultz, P. G. J. Am. Chem. Soc. 1991, 113, 2722; Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G. Science 1989, 244, 182; Bain, J. D.; Glabe, C. G.; Dix, T. A.; Chamberlin, A. R. J. Am. Chem. Soc. 1989, 111, 8013; Bain, J. D. *et al.* Nature 1992, 356, 537; Gallivan, J. P.; Lester, H. A.; Dougherty, D. A. Chem. Biol. 1997, 4, 740; Turcatti, *et al.* J. Biol. Chem. 1996, 271, 19991; Nowak, M.
 60 W. *et al.* Science, 1995, 268, 439; Saks, M. E. *et al.* J. Biol. Chem. 1996, 271, 23169; Hoshaka, T. *et al.* J. Am. Chem. Soc. 1999, 121, 34), para evitar el uso de sintetasas en la aminoacilación. Pueden utilizarse tales métodos u otros métodos químicos de aminoacilación para aminoacilar las moléculas de ARNt de la invención.

Se han utilizado métodos de biosíntesis que emplean aminoacil-ARNt químicamente modificados para
 65 incorporar varias sondas biofísicas en las proteínas sintetizadas *in vitro*. Véanse las siguientes publicaciones y referencias citadas en: Brunner, J. New Photolabeling and crosslinking methods, Annu. Rev Biochem, 62:483-514

(1993); y, Krieg, U.C., Walter, P., Hohnson, A.E. Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle, Proc. Natl. Acad. Sci, 83(22):8604-8608 (1986).

Anteriormente, se ha demostrado que pueden incorporarse aminoácidos no naturales de manera específica para el sitio en proteínas *in vitro* añadiendo ARNt supresores químicamente aminoacilados a las reacciones de síntesis de proteínas programadas con un gen que contiene una mutación terminadora ámbar deseada. Utilizando estos enfoques, pueden sustituirse varios de los veinte aminoácidos comunes con homólogos estructurales cercanos, por ejemplo, fluorofenilalanina por fenilalanina, utilizando cepas auxotróficas para un aminoácido concreto. Véase, por ejemplo, Noren, C.J., Anthony-Cahill, Griffith, M.C., Schultz, P.G. A general method for site-specific incorporation of unnatural amino acids into proteins, Science, 244: 182-188 (1989); M.W. Nowak, *et al.*, Science 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., Diala, E.S. Biosynthetic sitespecific incorporation of a non-natural amino acid into a polypeptide, J. Am Chem Soc, 111:8013-8014 (1989); N. Budisa *et al.*, FASEB J. 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins, Methods in Enz., vol. 202, 301-336 (1992); y, Mendel, D., Cornish, V.W. & Schultz, P.G. Site-Directed Mutagenesis with an Expanded Genetic Code, Annu Rev Biophys. Biomol Struct. 24, 435-62 (1995).

Por ejemplo, se preparó un ARNt supresor que reconocía el codón de terminación UAG y se aminoaciló químicamente con un aminoácido no natural. Se utilizó la mutagénesis convencional dirigida al sitio para introducir el codón de terminación TAG, en el sitio de interés en el gen de la proteína. Véase, por ejemplo, Sayers, J.R., Schmidt, W. Eckstein, F. 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, Nucleic Acids Res, 16(3):791-802 (1988). Cuando el ARNt supresor acilado y el gen mutante se combinaron en un sistema de transcripción/traducción *in vitro*, el aminoácido no natural se incorporó en respuesta al codón UAG que dio una proteína que contenía ese aminoácido en la posición determinada. Los experimentos utilizando [³H]-Phe y los experimentos con α -hidroxiácidos demostraron que sólo el aminoácido deseado se incorpora a la posición determinada por el codón UAG y que este aminoácido no se incorpora a ningún otro sitio en la proteína. Véase, por ejemplo, Noren, *et al.*, *supra*; Kobayashi *et al.*, (2003) Nature Structural Biology 10(6):425-432; y, Ellman, J.A., Mendel, D., Schultz, P.G. Site-specific incorporation of novel backbone structures into proteins, Science, 255(5041):197-200 (1992).

Los métodos para generar ARN catalítico pueden implicar generar combinaciones independientes de secuencias de ribozima aleatorizadas, realizar la evolución dirigida en las combinaciones, cribar las combinaciones en busca de la actividad aminoacilación deseada, y seleccionar las secuencias de las ribozimas que presentan la actividad de aminoacilación deseada.

Las ribozimas pueden comprender motivos y/o regiones que facilitan la actividad de acilación, tales como un motivo GGU y una región rica en U. Por ejemplo, se ha notificado que las regiones ricas en U pueden facilitar el reconocimiento de un sustrato de aminoácido, y un motivo GGU puede formar pares de bases con los extremos terminales 3' de un ARNt. En combinación, el GGU y el motivo y la región rica en U facilitan el reconocimiento simultáneo del aminoácido y del ARNt simultáneamente, y de ese modo facilitan la aminoacilación del extremo terminal 3' del ARNt.

Pueden generarse ribozimas mediante selección *in vitro* utilizando un r24mini parcialmente aleatorizado conjugado con ARNt^{Asm}_{CCCG}, seguido de obtención por ingeniería genética sistemática de una secuencia de consenso que se encuentra en los clones activos. Una ribozima ejemplar obtenida mediante este método se denomina "ribozima Fx3" y se describe en la solicitud de publicación de EE.UU. N° 2003/0228593, actúa como un catalizador versátil para la síntesis de diversos aminoacil-ARNt cargados con aminoácidos no naturales afines.

Puede utilizarse la inmovilización sobre un sustrato para permitir la purificación eficaz por afinidad de los ARNt aminoacilados. Los ejemplos de sustratos adecuados incluyen, pero no se limitan a, agarosa, sefrosa, y perlas magnéticas. Las ribozimas pueden inmovilizarse en resinas aprovechando la estructura química del ARN, tal como el 3'-cis-diol en la ribosa del ARN puede oxidarse con peryodato para producir el dialdehído correspondiente para facilitar la inmovilización del ARN sobre la resina. Pueden utilizarse diversos tipos de resinas incluyendo resinas de hidrazida de bajo costo en las que la aminación reductora hace de la interacción entre la resina y la ribozima un enlace irreversible. La síntesis de aminoacil-ARNt puede ser facilitada de manera significativa por esta técnica de aminoacilación "on-column". Kourouklis *et al.* Methods 2005; 36:239-4 describe un sistema de aminoacilación basado en columna.

El aislamiento de los ARNt aminoacilados puede llevarse a cabo de diversas maneras. Un método adecuado es eluir los ARNt aminoacilados de una columna con un tampón tal como una solución de acetato de sodio con EDTA 10 mM, un tampón que contiene N-(2-hidroxiethyl)piperazina-N'-(3-ácido propanosulfónico) 50 mM, KCl 12,5 mM, pH 7,0, EDTA 10 mM, o simplemente agua tamponada con EDTA (pH 7,0).

Los ARNt aminoacilados de la presente invención pueden añadirse a reacciones de traducción con el fin de incorporar el aminoácido con el que se aciló el ARNt en una posición de elección en un polipéptido producido por la reacción de traducción. Los ejemplos de sistemas de traducción en los que pueden utilizarse los ARNt

aminoacilados de la presente invención incluyen, pero no se limitan a, lisados de células. Los lisados de células proporcionan componentes de reacción necesarios para la traducción *in vitro* de un polipéptido a partir de un ARNm de entrada. Los ejemplos de tales componentes de reacción incluyen, pero no se limitan a, proteínas ribosómicas, ARNr, aminoácidos, ARNt, GTP, ATP, factores de iniciación y elongación de la traducción y factores adicionales asociados con la traducción. Además, los sistemas de traducción pueden ser de traducción discontinua o de traducción compartimentada. Los sistemas de traducción discontinua combinan componentes de reacción en un solo compartimento, mientras que los sistemas de traducción compartimentada separan los componentes de reacción de la traducción de los productos de reacción que pueden inhibir la eficacia de la traducción. Tales sistemas de traducción se encuentran disponibles comercialmente.

Además, puede utilizarse un sistema de transcripción/traducción acoplada. Los sistemas de transcripción/traducción acoplada permiten la transcripción de un ADN de entrada a un ARNm correspondiente, que a su vez es traducido por los componentes de reacción. Un ejemplo de transcripción/traducción acoplada disponible comercialmente es el Sistema de Traducción Rápida (RTS, Roche Inc). El sistema incluye una mezcla que contiene lisado de *E. coli* para proporcionar componentes de traducción tales como ribosomas y factores de traducción. Además, se incluye una ARN polimerasa para la transcripción del ADN de entrada a un molde de ARNm para su uso en la traducción. El RTS puede utilizar la compartimentación de los componentes de reacción por medio de una membrana interpuesta entre los compartimentos de reacción, incluyendo un compartimento de suministro/residuos y un compartimento de transcripción/traducción.

La aminoacilación del ARNt puede ser realizada por otros agentes, que incluyen, pero no se limitan a, transferasas, polimerasas, anticuerpos catalíticos, proteínas multifuncionales, y similares.

Aminoacil-ARNt sintetasa ortogonales (O-RS)

Una O-RS aminoacila preferentemente un O-ARNt de la presente invención con un aminoácido seleccionado *in vivo* o *in vitro*. Una O-RS de la presente invención puede ser proporcionada al sistema de traducción (por ejemplo, componentes de traducción *in vitro*, o una célula) por un polipéptido que incluya una O-RS y/o por un polinucleótido que codifique una O-RS o una parte de la misma. Una O-RS, o una parte de la misma, es codificada por una secuencia de polinucleótido o una secuencia de polinucleótido complementaria de la misma, o una variación conservadora de la misma. Un O-ARNt de la presente invención puede ser aminoacilado por varias moléculas diferentes de O-RS, que incluyen, pero no se limitan a, las descritas en el presente documento.

Son también una característica de la presente invención métodos para identificar una aminoacil-ARNt sintetasa ortogonal (O-RS), por ejemplo, una O-RS, para su uso con un O-ARNt, por ejemplo, un O-ARNt. Por ejemplo, un método incluye someter a selección positiva una población de células de una primera especie, en el que cada una de las células comprende: 1) un miembro de una pluralidad de aminoacil-ARNt sintetasa (RS), en el que la pluralidad de RS comprende RS mutantes, RS provenientes de una especie distinta de la primera especie o RS mutantes y RS provenientes de una especie distinta de la primera especie; 2) el ARNt ortogonal (O-ARNt) de una segunda especie; y 3) un polinucleótido que codifica un marcador de selección positiva y que comprende al menos un codón selector. Las células se seleccionan o se criban para detectar las que muestran una potenciación de la eficacia de supresión en comparación con las células que carecen de o con una cantidad reducida del miembro de la pluralidad de RS. Las células que tienen una potenciación de la eficacia de supresión comprenden una RS activa que aminoacila el O-ARNt. Se compara un nivel de aminoacilación (*in vivo* o *in vitro*) por la RS activa de un primer conjunto de ARNt de la primera especie con el nivel de aminoacilación (*in vivo* o *in vitro*) 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 marcado o un aminoácido no natural). Se selecciona la RS activa que aminoacila de manera más eficaz el segundo conjunto de ARNt en comparación con el primer conjunto de ARNt, proporcionando de este modo la aminoacil-ARNt sintetasa ortogonal para su uso con el O-ARNt. Es también una característica de la presente invención una O-RS, por ejemplo, una O-RS, identificada mediante el método.

Puede utilizarse cualquiera de varios ensayos para determinar la aminoacilación. Estos ensayos pueden realizarse *in vivo* o *in vitro*. Por ejemplo, se describen ensayos de aminoacilación *in vitro*, por ejemplo, en Hoben, P., y Soll, D. (1985) *Methods Enzymol.* 113:55-59 y en la publicación de solicitud de patente de EE.UU. Nº 2003/0228593. La aminoacilación también puede determinarse utilizando un indicador junto con componentes de traducción ortogonales y detectando el indicador en una célula que expresa un polinucleótido que comprende al menos un codón selector que codifica una proteína. Véase también, la solicitud de patente de EE.UU. 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; y el documento USSN 10/825,867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE".

Puede manipularse adicionalmente una O-RS identificada para modificar la especificidad de sustrato de la sintetasa de modo que sólo un aminoácido no natural deseado, pero no cualquiera de los 20 aminoácidos comunes, sea cargado al O-ARNt. Los métodos para generar una aminoacil-ARNt sintetasa ortogonal con especificidad de sustrato para un aminoácido no natural incluyen mutar la sintetasa, por ejemplo, en el sitio activo de la sintetasa, en el sitio del mecanismo de corrección de errores de la sintetasa, en diferentes sitios combinando diferentes dominios de sintetasa, o similares, y aplicar un proceso de selección. Se utiliza una estrategia que se basa en la combinación

de una selección positiva seguida de una selección negativa. En la selección positiva, la supresión del codón selector introducido en una posición o posiciones no esenciales de un marcador positivo permite que las células sobrevivan bajo presión de selección positiva. En presencia de aminoácidos naturales y no naturales, los supervivientes codifican por lo tanto sintetasas activas que cargan el ARNt supresor ortogonal con un aminoácido natural o no natural. En la selección negativa, la supresión de un codón selector introducido en una posición o posiciones no esenciales de un marcador negativo elimina las sintetasas con especificidades de aminoácido natural. Los supervivientes de la selección positiva y negativa codifican sintetasas que aminoacilan (cargan) el ARNt supresor ortogonal sólo con aminoácidos no naturales. Estas sintetasas pueden someterse a mutagénesis adicional, por ejemplo, barajado de ADN u otros métodos de mutagénesis recursivos.

Pueden generarse biblioteca de O-RS mutantes utilizando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, pueden generarse RS mutantes mediante mutaciones específicas de sitio, mutaciones puntuales aleatorias, recombinación homóloga, barajado de ADN u otros métodos de mutagénesis recursivos, construcción quimérica o cualquier combinación de los mismos. Por ejemplo, puede producirse una biblioteca de RS mutantes a partir de dos o más de otras, por ejemplo, "sub-bibliotecas" pequeñas menos diversas. También se incluyen en la invención bibliotecas químicas de RS. Cabe señalar que, para la detección de pares ortogonales, se construyen y se criban opcionalmente bibliotecas de ARNt sintetasas de diversos organismos (por ejemplo, microorganismos tales como eubacterias o arqueobacterias) tales como bibliotecas que comprenden diversidad natural (véase, por ejemplo, la patente de EE.UU. N° 6.238.884 de Short *et al.*; la patente de EE.UU. N° 5.756.316 de Schallenberger *et al.*; la patente de EE.UU. N° 5.783.431 de Petersen *et al.*; la patente de EE.UU. N° 5.824.485 de Thompson *et al.*; la patente de EE.UU. N° 5.958.672 de Short *et al.*).

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

También pueden utilizarse niveles adicionales de rigurosidad de selección/cribado en los métodos de la presente invención, para producir O-ARNt, O-RS o pares de los mismos. La rigurosidad de selección o de cribado puede modificarse en una o ambas etapas del método para producir una O-RS. Esto podría incluir, por ejemplo, modificar la cantidad de agente de selección/cribado que se utiliza, etc. También pueden realizarse rondas adicionales de selecciones positivas y/o negativas. La selección o el cribado también puede comprender una o más selecciones o cribados positivos o negativos o que incluyen, por ejemplo, un cambio en la permeabilidad del aminoácido, un cambio en la eficacia de traducción, un cambio en la fidelidad de traducción, etc. Por lo general, el uno o más cambios se basan en una mutación en uno o más genes en un organismo en el que se utiliza un par ARNt ortogonal-ARNt sintetasa para producir la proteína.

Pueden utilizarse otros tipos de selecciones en la presente invención para, por ejemplo, la O-RS, el O-ARNt, y el par O-ARNt/O-RS. El marcador de selección positiva puede ser cualquiera de una diversidad de moléculas, que incluyen, pero no se limitan a, un producto que proporciona un complemento nutricional para el crecimiento, y la selección se realiza en un medio que carece del complemento nutricional. Los ejemplos de polinucleótidos que codifican marcadores de selección positiva incluyen, pero no se limitan a, por ejemplo, un gen informador basado en la complementación de la auxotrofia para aminoácidos de una célula, un gen *his3* (por ejemplo, en el que el gen *his3* codifica una imidazol glicerol fosfato deshidratasa, detectada proporcionando 3-aminotriazol (3-AT)), gen *ura3*, gen *leu2*, gen *lys2*, gen *lacZ*, gen *adh*, etc. Véase, por ejemplo, G.M. Kishore, y D.M. Shah, (1988), Amino acid biosynthesis inhibitors as herbicides, *Annual Review of Biochemistry* 57:627-663. En una forma de realización, la producción de *lacZ* se detecta mediante hidrólisis de orto-nitrofenil- β -D-galactopiranosido (ONPG). Véase, por ejemplo, I.G. Serebriiskii, y E.A. Golemis, (2000), Uses of *lacZ* to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, *Analytical Biochemistry* 285:1-15. Los marcadores de selección positiva adicionales incluyen, por ejemplo, luciferasa, proteína fluorescente verde (GFP), YFP, EGFP, RFP, el producto de un gen resistente a los antibióticos (por ejemplo, cloranfenicol acetiltransferasa (CAT)), una proteína moduladora de la transcripción (por ejemplo, GAL4), etc. Opcionalmente, un polinucleótido que codifica un marcador de selección positiva comprende un codón selector.

Un polinucleótido que codifica el marcador de selección positiva puede estar unido operativamente a un elemento de respuesta. También puede estar presente un polinucleótido adicional que codifica una proteína moduladora de la transcripción que modula la transcripción a partir del elemento de respuesta, y comprende al menos un codón selector. La incorporación del aminoácido no natural a la proteína moduladora de la transcripción por el O-ARNt aminoacilado con los aminoácidos no naturales da como resultado la transcripción del polinucleótido (por ejemplo, gen indicador) que codifica el marcador de selección positiva. Opcionalmente, el codón selector se encuentra en o substancialmente cerca de una porción del polinucleótido que codifica un dominio de unión a ADN de la proteína moduladora de la transcripción.

También puede unirse operativamente un polinucleótido que codifica el marcador de selección negativa a un elemento de respuesta del que la transcripción es mediada por la proteína moduladora de la transcripción. Véase, por ejemplo, A.J. DeMaggio, *et al.*, (2000), The yeast split-hybrid system, *Method Enzymol.* 328:128-137; H.M. Shih, *et al.*, (1996), A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP, *Proc. Natl. Acad. Science. EE.UU.* 93:13896-13901; M. Vidal, *et al.*, (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Science. EE.UU.* 93:10321-10326; y M. Vidal, *et al.*, (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions (*Proc. Natl. Acad. Sci. EE.UU.* 93:10315-10320). La incorporación de un aminoácido natural en la proteína moduladora de la transcripción por el O-ARNt aminoacilado con un aminoácido natural da como resultado la transcripción del marcador de selección negativa. Opcionalmente, el marcador de selección negativa comprende un codón selector. El marcador de selección positiva y/o el marcador de selección negativa de la invención pueden comprender al menos dos codones selectores, pudiendo comprender cada uno o ambos al menos dos codones selectores diferentes o al menos dos codones selectores iguales.

La proteína moduladora de la transcripción es una molécula que se une (directa o indirectamente) a una secuencia de ácido nucleico (por ejemplo, un elemento de respuesta) y modula la transcripción de una secuencia que está unida operativamente al elemento de respuesta. Una proteína moduladora de la transcripción puede ser una proteína activadora de la transcripción (por ejemplo, GAL4, receptores de hormonas nucleares, AP1, CREB, miembros de la familia LEF/tcf, SMAD, VP16, SP1, etc.), una proteína represora de la transcripción (por ejemplo, receptores de hormonas nucleares, familia Groucho/tle, familia Engrailed, etc.), o una proteína que puede tener actividades que dependen del entorno (por ejemplo, LEF/tcf, proteínas homobox, etc.) Un elemento de respuesta es por lo general una secuencia de ácido nucleico que es reconocida por la proteína moduladora de la transcripción o un agente adicional que actúa concertadamente con la proteína moduladora de la transcripción.

Otro ejemplo de una proteína moduladora de la transcripción es la proteína activadora de la transcripción, GAL4. Véase, por ejemplo, A. Laughon, *et al.* (1984), Identification of two proteins encoded by the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:268-275; A. Laughon, y R.F. Gesteland, (1984), Primary structure of the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:260-267; L. Keegan, *et al.*, (1986), Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein, *Science* 231:699-704; y M. Ptashne, (1988), How eukaryotic transcriptional activators work, *Nature* 335:683-689. Los 147 aminoácidos del extremo N-terminal de esta proteína de 881 aminoácidos forman un dominio de unión a ADN (DBD) que se une específicamente a la secuencia de ADN. Véase por ejemplo, M. Carey, *et al.*, (1989), An amino-terminal fragment of GAL4 binds DNA as a dimer, *J. Mol. Biol.* 209:423-432; y E. Giniger, *et al.*, (1985), Specific DNA binding of GAL4, a positive regulatory protein of yeast, *Cell* 40: 767-774. El DBD se une, mediante una interacción, a un dominio de activación (AD) de 113 aminoácidos C-terminal que puede activar la transcripción cuando se une al ADN. Véase, por ejemplo, J. Ma, y M. Ptashne, (1987), Deletion analysis of GAL4 defines two transcriptional activating segments, *Cell* 48:847-853; y Ma J, y M. Ptashne, (1987), The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell* 50:137-142. Colocando los codones ámbar hacia, por ejemplo, el DBD N-terminal de un solo polipéptido que contiene el DBD N-terminal de GAL4 y su AD C-terminal, puede vincularse la supresión ámbar por el par O-ARNt/O-RS a la activación de la transcripción por GAL4. Pueden utilizarse genes indicadores activados por GAL4 para realizar las selecciones positiva y negativa con el gen.

El medio utilizado para la selección negativa puede comprender un agente de selección o cribado que es convertido en una sustancia detectable por el marcador de selección negativa. En un aspecto de la invención, la sustancia detectable es una sustancia tóxica. Un polinucleótido que codifica un marcador de selección negativa puede ser, por ejemplo, un gen *ura3*. Por ejemplo, puede colocarse el indicador *ura3* bajo el control de un promotor que contenga los sitios de unión a ADN de GAL4. Cuando se produce el marcador de selección negativa, por ejemplo, por traducción de un polinucleótido que codifica el GAL4 con codones selectores, GAL4 activa la transcripción de *ura3*. La selección negativa se lleva a cabo en un medio que comprende ácido 5-fluoroorótico (5-FOA), que es convertido en una sustancia detectable (por ejemplo, una sustancia tóxica que mata a la célula) por el producto génico del gen *ura3*. Véase, por ejemplo, J.D. Boeke, *et al.*, (1984), A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoroorotic acid resistance, *Molecular & General Genetics* 197:345-346; M. Vidal, *et al.*, (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse twohybrid system, *Proc. Natl. Acad. Science. EE.UU.* 93:10321-10326; y M. Vidal, *et al.*, (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions, *Proc. Natl. Acad. Science. EE.UU.* 93:10315-10320.

Al igual que con el marcador de selección positiva, el marcador de selección negativa también puede ser cualquiera de una diversidad de moléculas. El marcador de selección positiva y/o el marcador de selección negativa puede ser un polipéptido que emite fluorescencia o cataliza una reacción luminiscente en presencia de un reactivo adecuado. Por ejemplo, los marcadores de selección negativa incluyen, pero no se limitan a, por ejemplo, luciferasa, proteína fluorescente verde (GFP), YFP, EGFP, RFP, el producto de un gen resistente a los antibióticos (por ejemplo, cloranfenicol acetiltransferasa (CAT)), el producto de un gen *lacZ*, una proteína moduladora de la transcripción, etc. El marcador de selección positiva y/o el marcador de selección negativa puede ser detectado mediante clasificación de células activadas por fluorescencia (FACS) o mediante luminiscencia. El marcador de

selección positiva y/o el marcador de selección negativa puede comprender un marcador de cribado basado en afinidad. El mismo polinucleótido puede codificar el marcador de selección positiva y el marcador de selección negativa. Por ejemplo, la etapa de selección positiva, la etapa de selección negativa o ambas etapas de selección positiva y negativa pueden incluir el uso de un indicador, en el que el indicador se detecta mediante clasificación de células activadas por fluorescencia (FACS). Por ejemplo, puede realizarse primero una selección positiva con un marcador de selección positiva, por ejemplo, el gen de la cloranfenicol acetiltransferasa (CAT), en el que el gen CAT comprende un codón selector, por ejemplo, un codón de terminación ámbar, en el gen CAT, que va seguido de un cribado de selección negativa, que se basa en la incapacidad para suprimir un codón o codones selectores, por ejemplo, dos o más, en las posiciones dentro de un marcador negativo, por ejemplo, el gen de la ARN polimerasa de T7. El marcador de selección positiva y el marcador de selección negativa pueden encontrarse en el mismo vector, por ejemplo, plásmido. La expresión del marcador negativo activa la expresión del indicador, por ejemplo, la proteína fluorescente verde (GFP). Puede modificarse la rigurosidad de la selección y el cribado, por ejemplo, puede modificarse la intensidad de luz necesaria para hacer fluorescente el indicador. Puede realizarse una selección positiva con un indicador como un marcador de selección positiva, que se criba mediante FACS, seguido de un cribado por selección negativa, que se basa en la incapacidad para suprimir un codón o codones selectores, por ejemplo, dos o más, en las posiciones dentro de un marcador negativo, por ejemplo, el gen de la barnasa.

Opcionalmente, el indicador se presenta en la superficie de una célula, por ejemplo, en una presentación en fagos o similar. La presentación en la superficie celular, por ejemplo, el sistema de presentación en la superficie celular basado en OmpA, depende de la expresión de un epítipo en concreto, por ejemplo, un péptido de poliovirus C3 fusionado a una porina de membrana externa OmpA, en la superficie de la célula de *Escherichia coli*. El epítipo se presenta en la superficie celular sólo cuando se suprime un codón selector en el mensaje de proteína durante la traducción. El péptido presentado contiene entonces el aminoácido reconocido por una de las aminoacil-ARNt sintetasas mutantes de la biblioteca, y puede aislarse la célula que contiene el gen de la sintetasa correspondiente con anticuerpos generados contra péptidos que contienen aminoácidos no naturales específicos. El sistema de presentación en la superficie celular basado en OmpA fue desarrollado y optimizado por Georgiou *et al.* como una alternativa a la presentación en fagos. Véase Francisco, J.A., Campbell, R., Iverson, B.L. y Georgiou, G. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. Proc. Natl. Acad. Science. EE.UU. 90:10444-8 (1993).

Otras formas de realización de la presente invención incluyen llevar a cabo una o más de las etapas de selección *in vitro*. A continuación puede introducirse en una célula el componente seleccionado, por ejemplo, sintetasa y/o ARNt, para su uso en la incorporación *in vivo* de un aminoácido no natural.

Pueden encontrarse detalles adicionales para producir O-RS, y la modificación de la especificidad de sustrato de la sintetasa en la solicitud de patente de EE.UU. 10/126.931 titulada "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs"; y el documento USSN 10/825.867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". Pueden encontrarse detalles adicionales para producir O-RS en Hamano-Takaku *et al.*, (2000), A mutant *Escherichia coli* Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrine More Efficiently than Tyrosine, Journal of Biological Chemistry 275(51):40324-40328; Kiga *et al.* (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cellfree system, PNAS 99(15):9715-9723, y Francklyn. *et al.*, (2002), Aminoacyl-tRNA synthetases: Versatile players in the changing theater of translation; RNA, 8:1363-1372.

ORGANISMOS HOSPEDADORES Y DE PARTIDA

Los componentes de traducción de la presente invención provienen por lo general de organismos no eucariotas. Por ejemplo, el O-ARNt ortogonal puede provenir de un organismo no eucariota, por ejemplo, una arqueobacteria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tales como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, o similares, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares, mientras que la O-RS ortogonal puede provenir de un organismo no eucariota, por ejemplo, *Methanobacterium thermoautotrophicum*, *Halobacterium* tales como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, o similares, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares. En una forma de realización, también pueden utilizarse fuentes eucariotas, que incluyen, pero no se limitan a, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos, artrópodos, etc.), o similares.

Los componentes individuales de un par O-ARNt/O-RS pueden provenir del mismo organismo o de organismos diferentes. En una forma de realización, el par O-ARNt/O-RS es del mismo organismo. Como alternativa, el O-ARNt y la ORS del par O-ARNt/O-RS son de diferentes organismos. Por ejemplo, el O-ARNt puede provenir, por ejemplo, de un *Halobacterium sp* NRC-1, y la O-RS puede provenir, por ejemplo, de un *Methanobacterium thermoautotrophicum*.

El O-ARNt, la O-RS o el par O-ARNt/O-RS pueden seleccionarse o cribarse *in vivo* o *in vitro* y/o utilizarse en una célula, por ejemplo, una célula no eucariota (tal como una célula de *E. coli*), o una célula eucariota, para producir un polipéptido con un aminoácido seleccionado (por ejemplo, un aminoácido no natural). Una célula no eucariota puede ser de diversas fuentes, tal como el dominio filogenético *Archaea*, que incluye, pero no se limita a, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tales como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum pernix*, o similares, o puede pertenecer al dominio filogenético *Eubacteria* (que incluye, pero no se limita a, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), o similar. Una célula eucariota puede ser de diversas fuentes, que incluyen, pero no se limitan a, una planta (por ejemplo, una planta compleja tal como monocotiledóneas o dicotiledóneas), un alga, un protista, un hongo, una levadura (que incluye, pero no se limita a, *Saccharomyces cerevisiae*), un animal (que incluye, pero no se limita a, un mamífero, un insecto, un artrópodo, etc.), o similares. Son también una característica de la presente invención las composiciones de células con componentes de traducción de la presente invención. Véase también el documento USSN 10/825.867 titulado "Expanding the Eukaryotic Genetic Code" para el cribado de O-ARNt y/o O-RS en una especie para su uso en otra especie.

Para expresar un polipéptido de interés con un aminoácido seleccionado en una célula hospedadora, pueden subclonarse polinucleótidos que codifican un polipéptido de interés en un vector de expresión que contenga un promotor para dirigir la transcripción, un terminador de la transcripción/traducción, y en el caso de un ácido nucleico que codifica una proteína, un sitio de unión al ribosoma para la iniciación de la traducción. Los promotores bacterianos adecuados son bien conocidos en la técnica y se describen, por ejemplo, en Sambrook *et al.* y Ausubel *et al.*

Los sistemas de expresión bacterianos para expresar un polipéptido de interés se encuentran disponibles en, incluyendo pero sin limitarse a, *E. coli*, *Bacillus sp.*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, y *Salmonella* (Palva *et al.*, Gene 22:229-235 (1983); Mosbach *et al.*, Nature 302:543-545 (1983)). Los kits para tales sistemas de expresión se encuentran disponibles comercialmente. Los sistemas de expresión eucariotas para las células de mamífero, levaduras, y células de insecto son bien conocidos en la técnica y también se encuentran disponibles comercialmente.

Puede utilizarse y/o expresarse un ARNt y/o una RS de la presente invención y/o un polipéptido de interés en multitud de sistemas de expresión adecuados incluyendo, por ejemplo, levadura, células de insecto, células de mamífero, y bacterias. Más adelante se proporciona una descripción de sistemas de expresión ejemplares.

Levadura: Tal como se utiliza en el presente documento, el término "levadura" incluye cualquiera de las diversas levaduras capaces de expresar un polipéptido de interés. Tales levaduras incluyen, pero no se limitan a, levaduras ascospórogenas (Endomycetales), levaduras basidiosporógenas y levaduras que pertenecen al grupo de Hongos Imperfectos (Blastomycetes). Las levaduras ascospórogenas se dividen en dos familias, *Spermophthoraceae* y *Saccharomycetaceae*. Esta última está compuesta por cuatro subfamilias, *Schizosaccharomycoidae* (por ejemplo, el género *Schizosaccharomyces*), *Nadsonioideae*, *Lipomycoideae* y *Saccharomycoidae* (por ejemplo, los géneros *Pichia*, *Kluyveromyces* y *Saccharomyces*). Las levaduras basidiosporógenas incluyen los géneros *Leucosporidium*, *Rhodospodium*, *Sporidiobolus*, *Filobasidium*, y *Filobasidiella*. Las levaduras pertenecientes al grupo de Hongos Imperfectos (Blastomycetes) se dividen en dos familias, *Sporobolomycetaceae* (por ejemplo, los géneros *Sporobolomyces* y *Bullera*) y *Cryptococcaceae* (por ejemplo, el género *Candida*).

Resultan de especial interés para su uso con la presente invención las especies dentro de los géneros *Pichia*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Torulopsis* y *Candida*, que incluyen, pero no se limitan a, *P. pastoris*, *P. guilliermondii*, *S. cerevisiae*, *S. carlsbergensis*, *S. diastaticus*, *S. douglasii*, *S. kluyveri*, *S. norbensis*, *S. oviformis*, *K. lactis*, *K. fragilis*, *C. albicans*, *C. maltosa* y *H. polymorpha*. Las levaduras se encuentran disponibles generalmente en diversas fuentes, que incluyen, pero no se limitan a, Yeast Genetic Stock Center, Departamento de Biofísica y Física Médica, Universidad de California (Berkeley, CA), y la colección americana de cultivos tipo ("ATCC") (Manassas, VA).

La expresión "levadura hospedadora" o "célula hospedadora de levadura" incluye levadura que puede ser, o ha sido, utilizada como receptor para los vectores recombinantes u otro ADN de transferencia. La expresión incluye la progenie de la célula hospedadora de levadura original que ha recibido los vectores recombinantes u otro ADN de transferencia. Se entiende que la progenie de una sola célula parental no tiene por qué ser completamente idéntica en su morfología o en su genómica o en su complemento de ADN total a la parental original, debido a mutación accidental o deliberada. La progenie de la célula parental que es suficientemente similar a la parental que se caracteriza por la propiedad pertinente, tal como la presencia de una secuencia de nucleótidos que codifica un polipéptido de interés, se incluye en la progenie contemplada en esta definición.

Se han desarrollado vectores de expresión y transformación, que incluyen replicones extracromosómicos o vectores de integración, para la transformación en muchas levaduras hospedadoras. Por ejemplo, se han desarrollado vectores de expresión para *S. cerevisiae* (Sikorski *et al.*, GENETICS (1989) 122:19; Ito *et al.*, J.

5 BACTERIOL (1983) 153:163; Hinnen *et al.*, PROC NATL. ACAD. SCI. USA (1978) 75:1929); *C. albicans* (Kurtz *et al.*,
MOL. CELL. BIOL. (1986) 6:142); *C. maltosa* (Kunze *et al.*, J. BASIC MICROBIOL. (1985) 25:141); *H. polymorpha*
(Gleeson *et al.*, J. GEN. MICROBIOL. (1986) 132:3459; Roggenkamp *et al.*, MOL. GENETICS AND GENOMICS
(1986) 202:302); *K. fragilis* (Das *et al.*, J. BACTERIOL. (1984) 158:1165); *K. lactis* (De Louvencourt *et al.*, J.
10 BACTERIOL. (1983) 154:737; Van den Berg *et al.*, BIOTECHNOLOGY (NY) (1990) 8:135); *P. guilliermondii* (Kunze
et al., J. BASIC MICROBIOL. (1985) 25:141); *P. pastoris* (patentes de EE.UU. N^{os} 5.324.639; 4.929.555; y
4.837.148; Cregg *et al.*, MOL. CELL. BIOL. (1985) 5:3376); *Schizosaccharomyces pombe* (Beach *et al.*, NATURE
(1982) 300:706); e *Y. lipolytica*; *A. nidulans* (Ballance *et al.*, BIOCHEM. BIOPHYS. RES. COMMUN. (1983)
112:284-89; Tilburn *et al.*, GENE (1983) 26:205-221; y Yelton *et al.*, PROC. NATL. ACAD. SCI. EE.UU. (1984)
81:1470-74); *A. niger* (Kelly and Hynes, EMBO J. (1985) 4:475-479); *T. reesia* (EP 0 244 234); y hongos
filamentosos tales como, por ejemplo, *Neurospora*, *Penicillium*, *Tolyopocladium* (WO 91/00357).

15 Los expertos en la técnica conocen secuencias testigo para vectores de levaduras e incluyen, pero no se
limitan a, las regiones promotoras de genes tales como alcohol deshidrogenasa (ADH) (documento EP 0 284 044),
enolasa; glucoquinasa; glucosa-6-fosfato isomerasa; gliceraldehído-3-fosfato deshidrogenasa (GAP o GAPDH);
hexoquinasa; fosfofructoquinasa; 3-fosfoglicerato mutasa; y piruvato quinasa (PyK) (documento EP 0 329 203). El
gen de levadura PHO5, que codifica la fosfatasa ácida, también puede proporcionar secuencias promotoras útiles
(Miyano-hara *et al.*, PROC. NATL. ACAD. SCI. EE.UU. (1983) 80:1). Otras secuencias promotoras adecuadas para
20 su uso con levaduras hospedadoras pueden incluir los promotores para la 3-fosfoglicerato quinasa (Hitzeman *et al.*,
J. BIOL. CHEM. (1980) 255:12073); y otras enzimas glicolíticas, tales como piruvato descarboxilasa, trisfosfato
isomerasa, y fosfoglucosa isomerasa (Holland *et al.*, BIOCHEMISTRY (1978) 17:4900; Hess *et al.*, J. ADV ENZYME
REG. (1969) 7:149). Los promotores de levadura inducibles que tienen la ventaja adicional de una transcripción
controlada por las condiciones de crecimiento pueden incluir las regiones promotoras para la alcohol deshidrogenasa
25 2; isocitocromo C; fosfatasa ácida; metalotioneína; gliceraldehído-3-fosfato deshidrogenasa; enzimas degradantes
asociadas con el metabolismo del nitrógeno; y enzimas responsables de la utilización de la maltosa y la galactosa.
En el documento EP 0 073 657 se describen adicionalmente vectores y promotores adecuados para su uso en la
expresión de levaduras.

30 También pueden utilizarse potenciadores de levadura con los promotores de levadura. Además, los
promotores sintéticos también pueden funcionar como promotores de levadura. Por ejemplo, las secuencias de
activación cadena arriba (UAS) de un promotor de levadura pueden unirse con la región de activación de la
transcripción de otro promotor de levadura, creando un promotor híbrido sintético. Los ejemplos de tales promotores
híbridos incluyen la secuencia reguladora de la ADH unida a la región de activación de la transcripción de la GAP.
Véanse las patentes de EE.UU. N^{os} 4.880.734 y 4.876.197. Otros ejemplos de promotores híbridos incluyen
35 promotores que consisten en las secuencias reguladoras de los genes ADH2, GAL4, GAL10, o PRO5, combinados
con la región de activación de la transcripción de un gen de enzima glicolítica tal como GAP o PyK. Véase el
documento EP 0 164 556. Además, un promotor de levadura puede incluir promotores de origen natural de origen no
levadura que tienen la capacidad de unirse a la ARN polimerasa de levadura e iniciar la transcripción.

40 Otros elementos testigo que pueden comprender parte de los vectores de expresión de levadura incluyen
terminadores, por ejemplo, de GAPDH o los genes de enolasa (Holland *et al.*, J. Biol. CHEM. (1981) 256:1385).
Además, el origen de replicación del origen del plásmido 2 μ resulta adecuado para las levaduras. Un gen de
selección adecuado para su uso en levaduras es el gen *trp1* presente en el plásmido de levadura. Véase Tschumper
et al., GENE (1980) 10:157; Kingsman *et al.*, GENE (1979) 7:141. El gen *trp1* proporciona un marcador de selección
45 para una cepa mutante de levadura que carece de la capacidad de crecer en triptófano. Del mismo modo, las cepas
de levadura deficientes en Leu2 (ATCC 20.622 ó 38.626) se complementan mediante plásmidos conocidos que
portan el gen *Leu2*.

50 Los expertos en la técnica conocen métodos para introducir ADN exógeno en levaduras hospedadoras, y
por lo general incluyen, pero no se limitan a, la transformación de esferoplastos o de células hospedadoras de
levadura intactas tratadas con cationes alcalinos. Por ejemplo, la transformación de la levadura puede llevarse a
cabo de acuerdo con el método descrito en Hsiao *et al.*, PROC. NATL. ACAD. SCI. EE.UU. (1979) 76:3829 y Van
Solingen *et al.*, J. BACT. (1977) 130:946. Sin embargo, también pueden utilizarse otros métodos para introducir ADN
55 en células, tales como mediante inyección nuclear, electroporación, o fusión de protoplastos como se describe en
general en SAMBROOK *ET AL.*, MOLECULAR CLONING: A LAB. MANUAL (2001). A continuación, pueden
cultivarse las células hospedadoras de levadura utilizando técnicas convencionales conocidas por los expertos en la
técnica.

60 Los expertos en la técnica conocen otros métodos para la expresión de proteínas heterólogas en células
hospedadoras de levadura. Véanse en general la publicación de patente de EE.UU. N^o 20020055169, las patentes
de EE.UU. N^{os} 6.361.969; 6.312.923; 6.183.985; 6.083.723; 6.017.731; 5.674.706; 5.629.203; 5.602.034; y
5.089.398; las patentes reexaminadas de EE.UU. N^{os} RE37.343 y RE35.749; las solicitudes de patente publicadas
PCT WO 99/07862, WO 98/37208; y WO 98/26080; las solicitudes de patente europea EP 0 946 736; EP 0 732 403;
EP 0 480 480; WO 90/10277; EP 0 349 986; EP 0 329 203; EP 0 324 274; y EP 0 164 556. Véase también Gellissen
65 *et al.*, ANTONIE VAN LEEUWENHOEK (1992) 62(1-2):79-93; Romanos *et al.*, YEAST (1992) 8(6):423-488; Goeddel,
METHODS IN ENZYMOLOGY (1990) 185:3-7.

Las cepas hospedadoras de levadura pueden cultivarse en fermentadores durante la etapa de amplificación utilizando métodos convencionales de fermentación en cultivo discontinuo conocidos por los expertos en la técnica. Los métodos de fermentación pueden adaptarse para tener en cuenta las diferencias en una ruta concreta de utilización del carbono de una levadura hospedadora o el modo de control de la expresión. Por ejemplo, la fermentación de una levadura hospedadora *Saccharomyces* puede requerir una sola alimentación de glucosa, fuente de nitrógeno compleja (por ejemplo, hidrolizados de caseína), y complementación con múltiples vitaminas. Por el contrario, la levadura metilotrófica *P. pastoris* puede requerir suministros de glicerol, metanol, y minerales traza, pero sólo sales de amonio simples (nitrógeno) para el crecimiento y la expresión óptimos. Véase, por ejemplo, la patente de EE.UU. Nº 5.324.639; Elliott *et al.*, J. PROTEIN CHEM. (1990) 9:95; y Fieschko *et al.*, BIOTECH. BIOENG. (1987) 29:1113.

Sin embargo, tales métodos de fermentación pueden tener determinadas características comunes independientes de la cepa hospedadora de levadura empleada. Por ejemplo, puede añadirse al fermentador durante la fase de amplificación un nutriente limitante del crecimiento, por lo general de carbono, para permitir el máximo crecimiento. Además, los métodos de fermentación emplean generalmente un medio de fermentación diseñado para contener cantidades adecuadas de carbono, nitrógeno, sales basales, fósforo, y otros nutrientes minoritarios (vitaminas, minerales traza y sales, etc.). En las patentes de EE.UU. N^{os} 5.324.639 y 5.231.178 se describen ejemplos de medios de fermentación adecuados para su uso con *Pichia*.

Células de insecto infectadas con baculovirus: La expresión "insecto hospedador" o "célula hospedadora de insecto" se refiere a un insecto que puede ser, o ha sido, utilizado como receptor para los vectores recombinantes u otro ADN de transferencia. La expresión incluye la progenie de la célula hospedadora de insecto original que ha sido transfectada. Se entiende que la progenie de una sola célula parental no tiene por qué ser completamente idéntica en su morfología o en su genómica o en su complemento de ADN total a la parental original, debido a mutación accidental o deliberada. La progenie de la célula parental que es suficientemente similar a la parental que se caracteriza por la propiedad pertinente, tal como la presencia de una secuencia de nucleótidos que codifica un polipéptido de interés, se incluye en la progenie contemplada en esta definición.

Los expertos en la técnica conocen la selección de células de insecto adecuadas para la expresión de un polipéptido de interés. Varias especies de insectos están bien descritas en la técnica y se encuentran disponibles comercialmente, incluidas, *Aedes aegypti*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, y *Trichoplusia ni*. En la selección de insectos hospedadores para la expresión, los hospedadores adecuados pueden incluir los que demuestran tener, entre otras, buena capacidad de secreción, baja actividad proteolítica, y robustez global. Los insectos están generalmente disponibles en diversas fuentes, que incluyen, pero no se limitan al, Insect Genetic Stock Center, Departamento de Biofísica y Física Médica, Universidad de California (Berkeley CA), y la colección americana de cultivos tipo ("ATCC") (Manassas, VA).

En general, los componentes de un sistema de expresión de insecto infectado con baculovirus incluyen un vector de transferencia, normalmente un plásmido bacteriano, que contiene un fragmento del genoma de baculovirus, y un sitio de restricción conveniente para la inserción del gen heterólogo a expresar; un baculovirus de tipo silvestre con secuencias homólogas al fragmento específico de baculovirus en el vector de transferencia (esto permite la recombinación homóloga del gen heterólogo en el genoma del baculovirus), y células hospedadoras de insecto y medios de cultivo apropiados. Los materiales, métodos y técnicas utilizadas en la construcción de vectores, la transfección de células, selección de placas, crecimiento de células en cultivo, y similares, son conocidos en la técnica y hay manuales disponibles que describen estas técnicas.

Después de la inserción del gen heterólogo en el vector de transferencia, el vector y el genoma viral de tipo silvestre se transfectan a una célula hospedadora de insecto en la que se recombinan el vector y el genoma viral. Se expresa el virus recombinante empaquetado y se identifican y purifican las placas recombinantes. Los materiales y métodos para los sistemas de expresión de células de baculovirus/insecto se encuentran disponibles comercialmente en forma de kit, por ejemplo, en Invitrogen Corp. (Carlsbad, CA). Los expertos en la técnica conocen en general estas técnicas y se describen completamente en SUMMERS AND SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN Nº 1555 (1987). Véase también, RICHARDSON, 39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS (1995); AUSUBEL *ET AL.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 16.9-16.11 (1994); KING AND POSSEE, THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE (1992); y O'REILLY *ET AL.*, BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

De hecho, los expertos en la técnica conocen la producción de diversas proteínas heterólogas utilizando sistemas de expresión de células de baculovirus/insecto. Véanse, por ejemplo, las patentes de EE.UU. N^{os} 6.368.825; 6.342.216; 6.338.846; 6.261.805; 6.245.528; 6.225.060; 6.183.987; 6.168.932; 6.126.944; 6.096.304; 6.013.433; 5.965.393; 5.939.285; 5.891.676; 5.871.986; 5.861.279; 5.858.368; 5.843.733; 5.762.939; 5.753.220; 5.605.827; 5.583.023; 5.571.709; 5.516.657; 5.290.686; los documentos WO 02/06305; WO 01190390; WO 01/27301; WO 01/05956; WO 00/55345; WO 00/20032; WO 99/51721; WO 99/45130; WO 99/31257; WO 99/10515; WO 99/09193; WO 97/26332; WO 96/29400; WO 96/25496; WO 96/06161; WO 95/20672;

WO 93/03173; WO 92/16619; WO 92/02628; WO 92/01801; WO 90/14428; WO 90/10078; WO 90/02566; WO 90/02186; WO 90/01556; WO 89/01038; WO 89/01037; WO 88/07082.

Los vectores que resultan útiles en los sistemas de expresión en células de insectos/baculovirus son conocidos en la técnica e incluyen, por ejemplo, vectores de expresión y de transferencia de insectos provenientes del baculovirus virus de la poliedrosis nuclear de *Autographa californica* (AcNPV), que es un vector de expresión viral independiente de auxiliares. Los vectores de expresión viral provenientes de este sistema utilizan normalmente el promotor fuerte del gen de la polihedrina viral para activar la expresión de genes heterólogos. Véase, en general, O'Reilly *et al.*, BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

Antes de insertar el gen extraño en el genoma de baculovirus, los componentes anteriormente descritos, que comprenden un promotor, un líder (si se desea), la secuencia codificante de interés, y la secuencia de terminación de la transcripción, se ensamblan por lo general en un constructo de transposición intermedio (vector de transferencia). Los constructos de transposición intermedios se mantienen a menudo en un replicón, tal como un elemento extracromosómico (por ejemplo, plásmidos) capaz de mantenerse estable en un hospedador, tal como bacterias. El replicón tendrá un sistema de replicación, lo que le permite mantenerse en un hospedador adecuado para la clonación y la amplificación. Más concretamente, el plásmido puede contener la señal de poliadenilación de la polihedrina (Miller, ANN. REV. MICROBIOL. (1988) 42:177) y un origen de replicación y gen de resistencia a la ampicilina (*amp*) procarionta para la selección y la propagación en *E. coli*.

Un vector de transferencia utilizado comúnmente para introducir genes extraños en AcNPV es pAc373. También se han diseñado muchos otros vectores, conocidos por los expertos en la materia, incluyendo, por ejemplo, pVL985, que modifica el codón de inicio de polihedrina del ATG al ATT, y que introduce un sitio de clonación BamHI 32 pares de bases cadena abajo del ATT. Véase Luckow y Summers, VIROLOGY 170:31 (1989). Otros vectores disponibles comercialmente incluyen, por ejemplo, PBlueBac4.5/V5-His; pBlueBacHis2; pMelBac; pBlueBac4.5 (Invitrogen Corp, Carlsbad, CA).

Después de la inserción del gen heterólogo, el vector de transferencia y el genoma de baculovirus de tipo silvestre son co-transfectados a una célula hospedadora de insecto. Los métodos para introducir ADN heterólogo en el sitio deseado en el virus baculovirus son conocidos en la técnica. Véase SUMMERS Y SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN NO. 1555 (1987); Smith *et al.*, MOL. CELL. BIOL. (1983) 3:2156; Luckow y Summers, VIROLOGY (1989) 170:31. Por ejemplo, la inserción puede ser en un gen tal como el gen de la polihedrina, mediante recombinación homóloga de doble entrecruzamiento; la inserción también puede ser en un sitio de la enzima de restricción por ingeniería genética en el gen de baculovirus deseado. Véase Miller *et al.*, BIOESSAYS (1989) 11(4):91.

La transfección puede lograrse mediante electroporación. Véase TROTTER Y WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Mann y King, J. GEN. VIROL. (1989) 70:3501. Como alternativa, pueden utilizarse liposomas para transfectar las células de insecto con el vector de expresión recombinante y el baculovirus. Véase, por ejemplo, Liebman *et al.*, BIOTECHNIQUES (1999) 26(1):36; Graves *et al.*, BIOCHEMISTRY (1998) 37:6050; Nomura *et al.*, J. BIOL. CHEM. (1998) 273(22):13570; Schmidt *et al.*, PROTEIN EXPRESSION AND PURIFICATION (1998) 12:323; Siffert *et al.*, NATURE GENETICS (1998) 18:45; TILKINS ET AL., CELL BIOLOGY: A LABORATORY HANDBOOK 145-154 (1998); Cai *et al.*, PROTEIN EXPRESSION AND PURIFICATION (1997) 10:263; Dolphin *et al.*, NATURE GENETICS (1997) 17:491; Kost *et al.*, GENE (1997) 190:139; Jakobsson *et al.*, J. BIOL. CHEM. (1996) 271:22203; Rowles *et al.*, J. BIOL. CHEM. (1996) 271(37):22376; Reverey *et al.*, J. BIOL. CHEM. (1996) 271(39):23607-10; Stanley *et al.*, J. BIOL. CHEM. (1995) 270:4121; Sisk *et al.*, J. VIROL. (1994) 68(2):766; y Peng *et al.*, BIOTECHNIQUES (1993) 14(2):274. Los liposomas disponibles comercialmente incluyen, por ejemplo, Cellfectin® y Lipofectina® (Invitrogen, Corp, Carlsbad, CA). Además, puede utilizarse la transfección con fosfato de calcio. Véase TROTTER AND WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Kitts, NAR (1990) 18(19):5667; y Mann y King, J. GEN. VIROL. (1989) 70:3501.

Los vectores de expresión de baculovirus contienen normalmente un promotor de baculovirus. Un promotor de baculovirus es cualquier secuencia de ADN capaz de unirse a una ARN polimerasa de baculovirus e iniciar la transcripción cadena abajo (3') de una secuencia codificante (por ejemplo, un gen estructural) a ARNm. Un promotor tendrá una región de iniciación de la transcripción que se coloca normalmente proximal al extremo 5' de la secuencia codificante. Esta región de iniciación de la transcripción incluye por lo general un sitio de unión a la ARN polimerasa y un sitio de iniciación de la transcripción. Un promotor de baculovirus también puede tener un segundo dominio denominado potenciador, que, si se encuentra presente, es normalmente distal al gen estructural. Por otra parte, la expresión puede ser regulada o constitutiva.

Los genes estructurales, transcritos abundantemente en los últimos momentos del ciclo de infección, proporcionan secuencias promotoras especialmente útiles. Los ejemplos incluyen secuencias provenientes del gen que codifica la proteína poliédrica viral (FRIESEN ET AL., The Regulation of Baculovirus Gene Expression in THE MOLECULAR BIOLOGY OF BACULOVIRUSES (1986); EP 0 127 839 y 0 155 476) y el gen que codifica la proteína p10 (Vlak *et al.*, J. GEN. VIROL. (1988) 69:765).

El vector de expresión de baculovirus recién formado se empaqueta en un baculovirus recombinante infeccioso y posteriormente las placas cultivadas pueden purificarse mediante técnicas conocidas por los expertos en la técnica. Véase Miller *et al.*, BIOESSAYS (1989) 11(4):91; SUMMERS Y SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN N° 1555 (1987).

Se han desarrollado vectores de expresión de baculovirus recombinantes para la infección en varias células de insecto. Por ejemplo, se han desarrollado baculovirus recombinantes para, entre otros, *Aedes aegypti* (ATCC N° CCL-125), *Bombyx mori* (ATCC N° CRL-8910), *Drosophila melanogaster* (ATCC N° 1963), *Spodoptera frugiperda*, y *Trichoplusia ni*. Véase Wright, NATURE (1986) 321:718; Carbonell *et al.*, J. VIROL. (1985) 56:153; Smith *et al.*, MOL. CELL. BIOL. (1983) 3:2156. Véase en general, Fraser *et al.*, IN VITRO CELL. DEV. BIOL. (1989) 25:225. Más concretamente, las líneas celulares utilizadas para los sistemas de vectores de expresión de baculovirus incluyen comúnmente, pero no se limitan a, Sf9 (*Spodoptera frugiperda*) (ATCC N° CRL-1711), Sf21 (*Spodoptera frugiperda*) (Invitrogen Corp, Cat. N° 11497-013 (Carlsbad, CA)), Tri-368 (*Trichoplusia ni*) y High-Five™ BTI-TN-5B1-4 (*Trichoplusia ni*).

Se encuentran disponibles comercialmente células y medios de cultivo para la expresión directa y de fusión de polipéptidos heterólogos en un baculovirus/expresión, y los expertos en la técnica conocen en general la tecnología de cultivo celular.

E. coli, especies de *Pseudomonas*, y otros procariontas: Los expertos en la técnica conocen las técnicas de expresión en bacterias. Se encuentra disponible una gran variedad de vectores para su uso en hospedadores bacterianos. Los vectores pueden ser vectores de copia única o de alta o baja multiplicidad de copias. Los vectores pueden servir para la clonación y/o la expresión. En vista de la amplia literatura con respecto a los vectores, la disponibilidad comercial de muchos vectores, e incluso manuales que describen los vectores y sus características y mapas de restricción, no se requiere en el presente documento un análisis extenso. Como es bien conocido, los vectores normalmente implican marcadores que permiten la selección, marcadores que pueden proporcionar resistencia a agentes citotóxicos, prototrofia o inmunidad. Con frecuencia, se encuentra presente una pluralidad de marcadores, que proporcionan diferentes características.

Un promotor bacteriano es cualquier secuencia de ADN capaz de unirse a la ARN polimerasa bacteriana e iniciar la transcripción cadena abajo (3') de una secuencia codificante (por ejemplo, un gen estructural) a ARNm. Un promotor tendrá una región de iniciación de la transcripción que se encuentra normalmente proximal al extremo 5' de la secuencia codificante. Esta región de iniciación de la transcripción incluye por lo general un sitio de unión a la ARN polimerasa y un sitio de iniciación de la transcripción. Un promotor bacteriano también puede tener un segundo dominio denominado operador que puede solapar un sitio de unión a la ARN polimerasa adyacente en el que comienza la síntesis del ARN. El operador permite la transcripción regulada negativa (inducible), ya que una proteína represora del gen puede unirse al operador y por lo tanto inhibir la transcripción de un gen específico. La expresión constitutiva puede ocurrir en ausencia de elementos de regulación negativa, tales como el operador. Además, puede conseguirse la regulación positiva mediante una secuencia de unión a la proteína activadora del gen, que, si se encuentra presente es normalmente proximal (5') a la secuencia de unión a la ARN polimerasa. Un ejemplo de proteína activadora del gen es la proteína activadora de catabolitos (CAP), que ayuda a iniciar la transcripción del operón lac en *Escherichia coli* (*E. coli*) [Raibaud *et al.*, ANNU. REV. GENET. (1984) 18:173]. Por lo tanto, la expresión regulada puede ser positiva o negativa, potenciando o reduciendo de ese modo la transcripción.

Las secuencias que codifican enzimas de la ruta metabólica proporcionan secuencias promotoras especialmente útiles. Los ejemplos incluyen secuencias promotoras proveniente de enzimas que metabolizan azúcares, tales como galactosa, lactosa (lac) [Chang *et al.*, NATURE (1977) 198:1056], y maltosa. Los ejemplos adicionales incluyen secuencias promotoras provenientes de enzimas biosintéticas tales como triptófano (trp) [Goeddel *et al.*, NUC. ACIDS RES. (1980) 8:4057; Yelverton *et al.*, NUCL. ACIDS RES. (1981) 9:731; patente de EE.UU. N° 4.738.921; publicaciones EP N°s 036 776 y 121 775]. El sistema promotor de la β -galactosidasa (bla) [Weissmann (1981) "The cloning of interferon and other mistakes" en Interferon 3 (Ed. I. Gresser)]. Los sistemas promotores del bacteriófago lambda PL [Shimatake *et al.*, NATURE (1981) 292:128] y T5 [patente de EE.UU N° 4.689.406] también proporcionan secuencias promotoras útiles. Pueden utilizarse promotores fuertes, tales como el promotor de T7 para inducir el polipéptido de interés a niveles elevados. Los expertos en la técnica conocen ejemplos de tales vectores e incluyen la serie pET29 de Novagen, y los vectores pPOP descritos en el documento WO99/05297. Tales sistemas de expresión producen altos niveles de polipéptido en el hospedador sin comprometer los parámetros de crecimiento o la viabilidad de la célula hospedadora. Otro vector conocido en la técnica es pET19 (Novagen).

Además, los promotores sintéticos que no se producen en la naturaleza también funcionan como promotores bacterianos. Por ejemplo, pueden unirse secuencias de activación de la transcripción de un promotor bacteriano o de bacteriófago con las secuencias operón de otro promotor bacteriano o de bacteriófago, creando un promotor híbrido sintético [patente de EE.UU N° 4.551.433]. Por ejemplo, el promotor tac es un promotor híbrido trp-lac que está compuesto por las secuencias del promotor trp y del operón lac que es regulado por el represor de lac [Amann *et al.*, GENE (1983) 25:167; de Boer *et al.*, PROC. NATL. ACAD. SCI. (1983) 80:21]. Además, un promotor bacteriano puede incluir promotores de origen natural de origen no bacteriano que tengan la capacidad de

unirse a la ARN polimerasa bacteriana e iniciar la transcripción. Puede acoplarse también un promotor de origen natural de origen no bacteriano a una ARN polimerasa compatible para producir altos niveles de expresión de algunos genes en procariotas. El sistema ARN polimerasa/promotor del bacteriófago T7 es un ejemplo de sistema promotor acoplado [Studier *et al.*, J. MOL. BIOL. (1986) 189:113; Tabor *et al.*, Proc Natl. Acad. Sci. (1985) 82:1074]. Además, un promotor híbrido también puede estar compuesto por un promotor de bacteriófago y una región operadora de *E. coli* (publicación EP N° 267 851).

Además de una secuencia promotora operativa, también resulta útil un sitio de unión al ribosoma eficaz para la expresión de genes extraños en procariotas. En *E. coli*, el sitio de unión al ribosoma se denomina secuencia de Shine-Dalgarno (SD) e incluye un codón de iniciación (ATG) y una secuencia de 3-9 nucleótidos de longitud situada 3-11 nucleótidos cadena arriba del codón de iniciación [Shine *et al.*, NATURE (1975) 254:34]. Se piensa que la secuencia SD promueve la unión del ARNm al ribosoma mediante el apareamiento de bases entre la secuencia SD y el extremo 3' del ARNr 16S de *E. coli* [Steitz *et al.* "Genetic signals and nucleotide sequences in messenger RNA", En Biological Regulation and Development: Gene Expression (Ed. R F. Goldberger, 1979)]. Para expresar genes eucariotas y genes procariotas con sitio de unión al ribosoma débil [Sambrook *et al.* "Expression of cloned genes in Escherichia coli", Molecular Cloning: A Laboratory Manual, 1989].

La expresión "hospedador bacteriano" o "célula hospedadora bacteriana" se refiere a una bacteria que puede ser, o ha sido, utilizada como receptor para vectores recombinantes u otro ADN de transferencia. La expresión incluye la progenie de la célula hospedadora bacteriana original que ha sido transfectada. Se entiende que la progenie de una sola célula parental no tiene por qué ser completamente idéntica en su morfología o en su genómica o en su complemento de ADN total a la parental original, debido a mutación accidental o deliberada. La progenie de la célula parental que es suficientemente similar a la parental que se caracteriza por la propiedad pertinente, tal como la presencia de una secuencia de nucleótidos que codifica un polipéptido de interés, se incluye en la progenie contemplada en esta definición.

Los expertos en la técnica conocen la selección de bacterias hospedadoras adecuadas para la expresión de polipéptidos. En la selección de hospedadores bacterianos para la expresión, los hospedadores adecuados pueden incluir aquellos que se demuestren tener, entre otras, buena capacidad de formación de cuerpos de inclusión, baja actividad proteolítica, y robustez global. Los hospedadores bacterianos están generalmente disponibles en diversas fuentes, que incluyen, pero no se limitan a, Bacterial Genetic Stock Center, Departamento de Biofísica y Física Médica, Universidad de California (Berkeley CA), y la colección americana de cultivos tipo ("ATCC") (Manassas, VA). La fermentación industrial/farmacéutica utiliza generalmente bacterias provenientes de cepas K (por ejemplo, W3110) o bacterias provenientes de cepas B (por ejemplo, BL21). Estas cepas son especialmente útiles debido a que sus parámetros de crecimiento son extremadamente robustos y bien conocidos. Además, estas cepas no son patógenas, lo cual es comercialmente importante por razones ambientales y de seguridad. Otros ejemplos de hospedadores *E. coli* adecuados incluyen, pero no se limitan a, cepas de BL21, DH10B, o derivados de las mismas. En otra forma de realización de los métodos de la presente invención, el hospedador *E. coli* es una cepa sin proteasas que incluye, pero no se limita a, OMP- y LON-. La cepa de célula hospedadora puede ser una especie de *Pseudomonas*, que incluye, pero no se limita a, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, y *Pseudomonas putida*. Se sabe que *Pseudomonas fluorescens* biovar 1, denominada cepa MB101, resulta útil para la producción recombinante y se encuentra disponible para los procesos de producción de proteínas terapéuticas. Los ejemplos de un sistema de expresión de *Pseudomonas* incluyen el sistema disponible en The Dow Chemical Company como cepa hospedadora (Mideland, MI disponible en la web dow.com). Las patentes de EE.UU. N°s 4.755.465 y 4.859.600, describen el uso de cepas de *Pseudomonas* como célula hospedadora para la producción de hGH.

Una vez que se ha establecido una cepa de célula hospedadora recombinante (es decir, se ha introducido el constructo de expresión en la célula hospedadora y se aíslan células hospedadoras con el constructo de expresión adecuado), se cultiva la cepa de célula hospedadora recombinante en condiciones apropiadas para la producción del polipéptido de interés. Como resultará evidente para un experto en la materia, el método de cultivo de la cepa de célula hospedadora recombinante dependerá de la naturaleza del constructo de expresión utilizado y de la identidad de la célula hospedadora. Las cepas hospedadoras recombinantes se cultivan normalmente utilizando métodos que son bien conocidos en la técnica. Las células hospedadoras recombinantes se cultivan por lo general en medio líquido que contiene fuentes asimilables de carbono, nitrógeno y sales inorgánicas y, opcionalmente, que contiene vitaminas, aminoácidos, factores de crecimiento y otros complementos proteicos de cultivo conocidos por los expertos en la técnica. Los medios líquidos para el cultivo de células hospedadoras pueden contener opcionalmente antibióticos o antifúngicos para evitar el crecimiento de compuestos y/o microorganismos indeseables, que incluyen, pero no se limitan a, antibióticos para la selección de las células hospedadoras que contienen el vector de expresión.

Las células hospedadoras recombinantes pueden cultivarse de forma continua o discontinua, con recolección de células (en caso de que el polipéptido de interés se acumule intracelularmente) o recolección de sobrenadante del cultivo en las formas continua o discontinua. Para la producción en células hospedadoras de procariotas, resultan preferentes el cultivo discontinuo y la recolección de células.

CODONES SELECTORES

Los codones selectores de la presente invención amplían la estructura de los codones del genoma de la maquinaria de biosíntesis 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, que incluye, pero no se limita a, un codón ámbar (UAG), un codón ocre, o un codón ópalo (UGA), un codón no natural, un codón de cuatro bases (o más), un codón raro, o similares. Pueden introducirse varios codones selectores en un polinucleótido o gen deseado, por ejemplo, uno o más, dos o más, tres o más, etc.

En una forma de realización, los métodos implican el uso de un codón selector que es un codón de terminación para la incorporación de un aminoácido seleccionado, por ejemplo, un aminoácido no natural, *in vivo*. Por ejemplo, se produce un O-ARNt que reconoce el codón de terminación y es aminoacilado por una O-RS con un aminoácido seleccionado. Este O-ARNt no es reconocido por las aminoacil-ARNt sintetasas de origen natural del hospedador. Puede utilizarse la mutagénesis dirigida al sitio convencional para introducir el codón de terminación en el sitio de interés en un polipéptido de interés. Véase, por ejemplo, Sayers, J.R., *et al.* (1988), 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*, 16:791-802. Cuando se combinan la O-RS, el O-ARNt y el ácido nucleico que codifica un polipéptido de interés, por ejemplo, *in vivo*, el aminoácido seleccionado se incorpora en respuesta al codón de terminación para dar un polipéptido que contiene el aminoácido seleccionado, por ejemplo, un aminoácido no natural, en la posición determinada. En una forma de realización de la presente invención, un codón de terminación utilizado como codón selector es un codón ámbar, UAG, y/o un codón ópalo, UGA. Por ejemplo, véase la SEQ ID NO: 6 para un ejemplo de O-ARNt que reconoce un codón ámbar, y véase la SEQ ID NO: 7 para un ejemplo de O-ARNt que reconoce un codón ópalo. Un código genético en el que se utilizan UAG y UGA como codón selector puede codificar 22 aminoácidos al tiempo que conserva el codón de terminación ocre, UAA, que es la señal de terminación más abundante.

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

Los aminoácidos no naturales también pueden codificarse con codones raros. Por ejemplo, cuando se reduce la concentración de arginina en una reacción de síntesis de proteínas *in vitro*, el codón de arginina raro, AGG, ha demostrado ser 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 ARNtArg de origen natural, que existe como una especie minoritaria en *Escherichia coli*. Algunos organismos no utilizan todos los codones triplete. Se ha utilizado un codón no asignado AGA en *Micrococcus luteus* para la inserción de aminoácidos en un extracto de transcripción/traducción *in vitro*. Véase, por ejemplo, Kowal y Oliver, *Nucl. Acid. Res*, 25:4685 (1997). Pueden generarse los componentes de la presente invención para utilizar estos codones raros *in vivo*.

Los codones selectores también comprenden codones prolongados, por ejemplo, codones de cuatro o más bases, tal como, codones de cuatro, cinco, seis o más bases. Los ejemplos de codones de cuatro bases incluyen, pero no se limitan a, AGGA, CUAG, UAGA, CCCU y similares. Los ejemplos de codones de cinco bases incluyen, pero no se limitan a, AGGAC, CCCCUC, CCCUC, CUAGA, CUACU, UAGGC y similares. Una característica puede incluir el uso de codones prolongados en base a la supresión del desplazamiento del marco de lectura. Los codones de cuatro o más bases pueden insertar, por ejemplo, uno o múltiples aminoácidos seleccionados, que incluyen, pero no se limitan a, aminoácidos no naturales, en la misma proteína. Por ejemplo, en presencia de O-ARNt mutados, por ejemplo, un ARNt supresor del desplazamiento del marco de lectura especial, con bucles anticodón, por ejemplo, con una secuencia CU(X)_n XXXAA (en la que n=1), el codón de cuatro o más bases se lee como un solo aminoácido. Por ejemplo, véase las SEQ ID NOS: 6, 12 del PCT/US04/22061 para los O-ARNt que reconocen un codón de cuatro bases. En otras formas de realización, los bucles anticodón pueden decodificar, por ejemplo, al menos un codón de cuatro bases, al menos un codón de cinco de base, o al menos un codón de seis bases o más. Puesto que hay 256 posibles codones de cuatro bases, pueden codificarse múltiples aminoácidos no naturales en la misma célula utilizando un codón de cuatro bases o más. Véase, Anderson *et al.*, (2002) Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Cordons and Identification of "Shifty" Four-base Cordons with a Library Approach in *Escherichia coli*, *J. Mol. Biol.* 307:755-769.

Por ejemplo, se han utilizado codones de cuatro bases para incorporar aminoácidos no naturales a proteínas utilizando métodos de biosíntesis *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 NBD de lisina a estreptavidina *in vitro* con dos ARNt supresores del desplazamiento del marco 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 ARNtLeu con anticodones NCUA para suprimir los codones UAGN (N puede ser U, A, G, o C), y se descubrió que el cuadruplete UAGA puede ser decodificado por un ARNtLeu con un anticodón UCUA con una eficacia del 13% al 26% con poca decodificación en el marco 0 ó -1. Véase Moore *et al.*, (2000) J. Mol. Biol., 298:195. En una forma de realización, pueden utilizarse en la invención codones prolongados basados en codones raros o codones sin sentido, que pueden reducir la ultralectura por mutación de aminoácido y la supresión del desplazamiento del marco de lectura en otros sitios no deseados.

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

Los codones selectores incluyen opcionalmente pares de bases no naturales. Estos pares de bases no naturales amplían adicionalmente el alfabeto genético existente. Un par de bases adicional aumenta el número de codones triplete de 64 a 125. Las propiedades de los terceros pares de bases incluyen el apareamiento de bases estable y selectivo, la incorporación enzimática eficaz en el ADN con elevada fidelidad mediante una polimerasa, y la prolongación del cebador continuada eficaz después de la síntesis del par de bases no natural naciente. Las descripciones de pares de bases no naturales que pueden adaptarse para los métodos y las composiciones incluyen, por ejemplo, Hirao, *et al.*, (2002) An unnatural base pair for incorporating amino acid analogues into protein, Nature Biotechnology, 20:177-182. Véase también, Wu, Y *et al.*, (2002) J. Am. Chem. Soc. 124:14626-14630. Más adelante se enumeran otras publicaciones pertinentes.

Para el uso *in vivo*, el nucleósido no natural es permeable a la membrana y está fosforilado para formar el trifosfato correspondiente. Además, el aumento de la información genética es estable y no es destruida por las enzimas celulares. Los esfuerzos anteriores de Benner y otros aprovechaban los patrones de unión de hidrógeno que son diferentes de aquellos en los pares de Watson-Crick canónicos, cuyo ejemplo más notable es el par iso-C:iso-G. Véase por ejemplo, Switzer *et al.*, (1989) J. Am. Chem. Soc, 111:8322; y Piccirilli *et al.*, (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. Estas bases en general se aparean incorrectamente en cierta medida con las bases naturales y no pueden replicarse enzimáticamente. Kool y colaboradores demostraron que las interacciones de empaquetamiento hidrófobas entre bases pueden reemplazar los enlaces de hidrógeno para activar 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 cumpla todos los requisitos anteriormente indicados, Schultz, Romesberg y colaboradores han sintetizado y estudiado de forma sistemática una serie de bases hidrófobas no naturales. Un auto-par PICS:PICS resulta ser más estable que los pares de bases naturales, y puede ser incorporado de manera eficaz en el ADN por el fragmento Klenow de la ADN polimerasa I (KF) de *Escherichia coli*. Véase por ejemplo, McMinin *et al.*, (1999) J. Am. Chem. Soc, 121:11585-6; y Ogawa *et al.*, (2000) J. Am. Chem. Soc, 122:3274. Un auto-par 3MN:3MN puede ser sintetizado por KF con eficacia y selectividad suficiente para la función biológica. Véase por ejemplo, Ogawa *et al.*, (2000) J. Am. Chem. Soc., 122:8803. Sin embargo, ambas bases actúan como terminador de cadena para la replicación adicional. Se ha desarrollado recientemente una ADN polimerasa mutante que puede utilizarse para replicar el auto-par PICS. Además, puede replicarse un auto-par 7AI. Véase por ejemplo, Tae *et al.*, (2001) J. Am. Chem. Soc, 123:7439. También se ha desarrollado un par de bases metálico novedoso, Dipic:Py, que forma un par estable tras su unión a Cu(II). Véase Meggers *et al.*, (2000) J. Am. Chem. Soc, 122:10714. Debido a que los codones prolongados y los codones no naturales son intrínsecamente ortogonales a los codones naturales, los métodos de la presente invención pueden aprovechar esta propiedad para generar ARNt ortogonales para ellos.

También puede utilizarse un sistema de desviación de la traducción para incorporar un aminoácido seleccionado, por ejemplo, un aminoácido no natural, a un polipéptido deseado. En un sistema de desviación de la traducción, se inserta una secuencia grande en un gen pero no se traduce a proteína. La secuencia contiene una estructura que sirve como entrada para inducir al ribosoma a saltar sobre la secuencia y reanudar la traducción cadena abajo de la inserción.

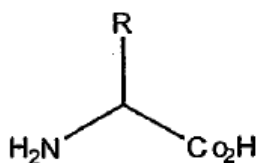
AMINOÁCIDOS SELECCIONADOS Y NO NATURALES

Tal como se utiliza en el presente documento, un aminoácido seleccionado se refiere a cualquier aminoácido no natural o aminoácido de origen natural deseado. Un aminoácido de origen natural incluye cualquiera de los 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. En una forma de realización, el aminoácido seleccionado se incorpora a una cadena polipeptídica en crecimiento con elevada fidelidad, por ejemplo, con una eficacia superior a aproximadamente un 70% para un codón selector determinado, con una eficacia superior a un 75% para un codón selector determinado, con una eficacia superior a aproximadamente un 80% para un codón selector determinado, con una eficacia superior a aproximadamente un 85% para un codón selector determinado, con una eficacia superior

a aproximadamente un 90% para un codón selector determinado, con una eficacia superior a aproximadamente un 95% para un codón selector determinado, o con una eficacia superior a aproximadamente un 99% o más para un codón selector determinado.

Tal como se utiliza 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 mediante la Fórmula I:

I

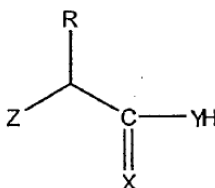


Un aminoácido no natural es por lo general cualquier estructura que tenga la Fórmula I en la que el grupo R es cualquier sustituyente distinto de uno utilizado en los veinte aminoácidos naturales. Véase, por ejemplo, Biochemistry de L. Stryer, 3ª ed. 1988, Freeman and Company, Nueva York, para las estructuras de los veinte aminoácidos naturales. Adviértase que los aminoácidos no naturales de la presente invención pueden ser compuestos de origen natural distintos de los veinte alfa-aminoácidos anteriormente indicados.

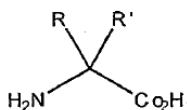
Debido a que los aminoácidos no naturales de la presente invención difieren por lo general de los aminoácidos naturales solamente en la estructura de la cadena lateral, los aminoácidos no naturales forman enlaces amida con otros aminoácidos, que incluyen, pero no se limitan a, naturales o no naturales, de la misma manera en que se forman en las proteínas de origen natural. Sin embargo, los aminoácidos no naturales tienen grupos de cadena lateral que los distinguen de los aminoácidos naturales. Por ejemplo, R en la Fórmula I puede comprender un grupo alquil-, aril-, acil-, ceto-, azido-, hidroxil-, hidracina, ciano-, halo-, hidracida, alquenilo, alquinilo, éter, tiol, seleno-, sulfonil-, borato, boronato, fosfo, fosfono, fosfina, heterociclo, enona, imina, aldehído, éster, tioácido, hidroxilamina, amina, y similares, o cualquier combinación de los mismos. Otros aminoácidos de origen no natural incluyen, pero no se limitan a, aminoácidos que comprenden un entrecruzador fotoactivable, aminoácidos marcados con espín, aminoácidos fluorescentes, aminoácidos de unión a metal, aminoácidos que contienen metales, aminoácidos radiactivos, aminoácidos con grupos funcionales novedosos, aminoácidos que interaccionan de forma covalente o no covalente con otras moléculas, aminoácidos fotoenjaulados ("photocaged") y/o fotoisomerizables, aminoácidos que comprenden biotina o un análogo de biotina, aminoácidos glicosilados, tales como una serina sustituida por azúcar, otros aminoácidos modificados con hidratos de carbono, aminoácidos que contienen un grupo ceto, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles y/o fotoescindibles, aminoácidos con cadenas laterales alargadas en comparación con los aminoácidos naturales, que incluyen, pero no se limitan a, poliéteres o hidrocarburos de cadena larga, que incluyen, pero no se limitan a, superior a aproximadamente 5 o superior a aproximadamente 10 átomos de carbono, aminoácidos que contiene azúcares unidos a carbono, aminoácidos con actividad redox, aminoácidos que contiene tioaminoácidos, y aminoácidos que comprenden uno o más restos tóxicos. Véanse, también, las publicaciones de las solicitudes de patente de EE.UU 2003/0082575 y 2003/0108885. Los aminoácidos no naturales pueden tener un entrecruzador fotoactivable que se utiliza, por ejemplo, para unir una proteína a un soporte sólido. Los aminoácidos no naturales pueden tener un resto de sacárido fijado a la cadena lateral de aminoácido.

Además de los aminoácidos no naturales que contienen cadenas laterales novedosas, los aminoácidos no naturales también comprenden opcionalmente estructuras de cadena principal modificadas, por ejemplo, como se ilustra mediante las estructuras de las Fórmula II y III:

II



III



5
10
15
20

en las que Z comprende por lo general OH, NH₂, SH, NH-R', o S-R'; X e Y, que pueden ser iguales o diferentes, comprenden por lo general S u O, y R y R', que son opcionalmente iguales o diferentes, están seleccionadas por lo general 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 pueden comprender sustituciones en el grupo amino o carboxilo como se ilustra mediante las Fórmulas II y III. Los aminoácidos no naturales de este tipo incluyen, pero no se limitan a, α-hidroxiácidos, α-tioácidos, α-aminotiocarboxilatos, por ejemplo, con cadenas laterales que corresponden a los veinte aminoácidos naturales comunes o cadenas laterales no naturales. Además, las sustituciones en el α-carbono 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, aminoácidos β y γ tales como ácido γ-aminobutírico y β-alanina sustituida.

25
30
35
40

Muchos aminoácidos no naturales se basan en aminoácidos naturales, tales como tirosina, glutamina, fenilalanina, y similares. Los análogos de tirosina incluyen tirosinas para-sustituidas, tirosinas orto-sustituidas, y tirosinas meta-sustituidas, en las que la tirosina sustituida comprende un grupo ceto (que incluye, pero no se limita a, un grupo acetilo), un grupo benzoilo, un grupo amino, una hidracina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo C₆-C₂₀ de cadena lineal o ramificada, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, o similares. Además, también se contemplan anillos de arilo múltiplemente sustituidos. Los análogos de glutamina incluyen, pero no se limitan a, derivados α-hidroxi, derivados γ-sustituidos, derivados cíclicos y derivados de glutamina sustituidos con amida. Los ejemplos de análogos de fenilalanina incluyen, pero no se limitan a, fenilalaninas para-sustituidas, fenilalaninas orto-sustituidas, y fenilalaninas meta-sustituidas, en las que el sustituyente comprende un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído, un azido, un yodo, un bromo, un grupo ceto (que incluye, pero no se limita a, un grupo acetilo), o similares. Los ejemplos específicos de aminoácidos no naturales incluyen, pero no se limitan a, una p-acetil-L-fenilalanina, una p-propargil-fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metil-fenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una tri-O-acetil-GlcNAcβ-serina, una L-Dopa, una fenilalanina fluorada, una isopropil-L-fenilalanina, una p-azido-L-fenilalanina, una p-acil-L-fenilalanina, una p-benzoil-L-fenilalanina, una L-fosfoserina, una fosfoserina, una fosfotirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina, una isopropil-L-fenilalanina, y una p-propargiloxi-fenilalanina, y similares. Se proporcionan ejemplos de estructuras de diversos aminoácidos no naturales, por ejemplo, en el documento WO 2002/085923 titulado "In vivo incorporation of unnatural amino acids". Véase también Kiick *et al.*, (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, PNAS 99:19-24, para análogos de metionina adicionales.

45

Un aminoácido no natural incorporado a un polipéptido en el extremo terminal amino puede estar compuesto por un grupo R que es cualquier sustituyente distinto del que se utiliza en los veinte aminoácidos naturales y un segundo grupo reactivo diferente del grupo NH₂ presente normalmente en los α-aminoácidos (véase la Fórmula I). Puede incorporarse un aminoácido no natural similar al extremo terminal carboxilo con un segundo grupo reactivo diferente del grupo COOH normalmente presente en los α-aminoácidos (véase la Fórmula I).

50
55

Los aminoácidos no naturales de la invención pueden seleccionarse o diseñarse para que proporcionen características adicionales no disponibles en los veinte aminoácidos naturales. Por ejemplo, pueden diseñarse o seleccionarse opcionalmente aminoácidos no naturales para que modifiquen las propiedades biológicas de una proteína, por ejemplo, a la que se incorporan. Por ejemplo, pueden modificarse opcionalmente las siguientes propiedades incluyendo un aminoácido no natural en una proteína: toxicidad, biodistribución, solubilidad, estabilidad, por ejemplo, térmica, hidrolítica, oxidativa, resistencia a la degradación enzimática, y similares, facilidad de purificación y procesamiento, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, actividad catalítica, potencial redox, semivida, la capacidad de reaccionar con otras moléculas, por ejemplo, covalentemente o no covalentemente, y similares.

60

Se proporcionan estructuras de diversos aminoácidos no naturales, por ejemplo, en las Figuras 16, 17, 18, 19, 26, y 29 del documento WO 2002/085923 titulado "In vivo incorporation of unnatural amino acids". Los ejemplos no pretenden en modo alguno limitar los aminoácidos que pueden fijarse a un ARNt de la presente invención.

65

Una de las ventajas de un aminoácido no natural es que presenta restos químicos adicionales que pueden utilizarse para añadir moléculas adicionales. Estas modificaciones pueden realizarse *in vivo* en una célula eucariota o no eucariota, o *in vitro*. Por lo tanto, en determinadas formas de realización, la modificación post-traduccional es a

través del aminoácido no natural. Puede utilizarse un aminoácido no natural en un polipéptido para fijar otra molécula al polipéptido, que incluye, pero no se limita a, un marcador, un colorante, un polímero, un polímero soluble en agua, un derivado de polietilenglicol, un fotoentrecruzador, un radionúclido, un compuesto citotóxico, un fármaco, un marcador de afinidad, un marcador de fotoafinidad, un compuesto reactivo, una resina, una segunda proteína o polipéptido o análogo de polipéptido, un anticuerpo o fragmento de anticuerpo, un quelante de metales, un cofactor, un ácido graso, un hidrato de carbono, un polinucleótido, un ADN, un ARN, un polinucleótido antisentido, un sacárido, un dendrímero soluble en agua, una ciclodextrina, un ácido ribonucleico inhibidor, un biomaterial, una nanopartícula, un marcador de espín, un fluoróforo, un resto que contiene metal, un resto radiactivo, un grupo funcional novedoso, un grupo que interactúa de forma covalente o no covalente con otras moléculas, un resto fotoenjaulado, un resto excitable por radiación actínica, un resto fotoisomerizable, biotina, un derivado de biotina, un análogo de biotina, un resto que incorpora un átomo pesado, un grupo químicamente escindible, un grupo fotoescindible, una cadena lateral alargada, un azúcar unido a carbono, un agente con actividad redox, un tioaminoácido, un resto tóxico, un resto marcado isotópicamente, una sonda biofísica, un grupo fosforescente, un grupo quimioluminiscente, un grupo electrodenso, un grupo magnético, un grupo intercalante, un cromóforo, un agente de transferencia de energía, un agente biológicamente activo, un marcador detectable, una molécula pequeña, un punto cuántico, un nanoemisor, o cualquier combinación de los anteriores o mediante cualquier otro compuesto o sustancia deseable, que comprenda un segundo grupo reactivo para al menos un aminoácido no natural que comprende un primer grupo reactivo que utiliza la metodología química que un experto en la técnica sabe que resulta adecuada para los grupos de reactivos concretos.

Por ejemplo, la modificación post-traducciona puede ser a través de una reacción nucleófila-electrófila. La mayoría de las reacciones que actualmente se utilizan para la modificación selectiva de proteínas implica la formación de enlaces covalentes entre parejas de reacción nucleófila y electrófila, que incluye, pero no se limita a, la reacción de α -halocetonas con cadenas laterales de cisteína o histidina. La selectividad en estos casos se determina por el número y la accesibilidad de los residuos nucleófilos en la proteína. En las proteínas de la invención, pueden utilizarse otras reacciones más selectivas tales como la reacción de un ceto-aminoácido no natural con compuestos aminooxi o hidrazidas, *in vitro* e *in vivo*. Véase por ejemplo, Cornish, *et al.*, (1996) J. Am. Chem. Soc. 118:8150-8151; Mahal, *et al.*, (1997) Science, 276:1125-1128; Wang, *et al.*, (2001) Science 292:498-500; Chin, *et al.*, (2002) J. Am. Chem. Soc. 124:9026-9027; Chin, *et al.*, (2002) Proc. Natl. Acad. Sci. 99:11020-11024; Wang, *et al.*, (2003) Proc. Natl. Acad. Sci. 100: 56-61; Zhang, *et al.*, (2003) Biochemistry 42:6735-6746; y Chin, *et al.*, (2003) Science, 301:964-7. Esto permite el marcaje selectivo de prácticamente cualquier proteína con varios reactivos que incluyen fluoróforos, agentes de entrecruzamiento, derivados de sacáridos y moléculas citotóxicas. Véase también, la patente de EE.UU. Nº 6.927.042 titulada "Glycoprotein synthesis". También pueden realizarse modificaciones post-traduccionales, que incluyen, pero no se limitan a, a través de un aminoácido azido, a través de la ligación de Staudinger (que incluye, pero no se limita a, con reactivos de triaril-fosfina). Véase, por ejemplo, Kiick *et al.*, (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, PNAS 99:19-24.

Síntesis química de aminoácidos no naturales

Muchos aminoácidos no naturales se encuentran disponibles comercialmente, por ejemplo, en Sigma-Aldrich (St. Louis, MO, EE.UU), Novabiochem (una división de EMD Biosciences, Darmstadt, Alemania), o Peptech (Burlington, MA, EE.UU). Aquellos que no se encuentran disponibles comercialmente se sintetizan opcionalmente conforme a lo dispuesto en el presente documento o utilizando los métodos convencionales conocidos por los expertos en la técnica. Para las técnicas de síntesis orgánica, véase, por ejemplo, Organic Chemistry de Fessenden y Fessenden, (1982, segunda edición, Willard Grant Press, Boston, MA); Advanced Organic Chemistry de March (tercera edición, 1985, Wiley and Sons, Nueva York); y Advanced Organic Chemistry de Carey y Sundberg (Tercera edición, partes A y B, 1990, Plenum Press, Nueva York). Las 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, F.E. y Kidd, D.A.A. (1949) A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319; Friedman, O.M. y Chatterji, R. (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750-3752; Craig, J. C. *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, Vilmont, M. y Frappier, F. (1991) Glutamine analogues as Potential Antimalarials, Eur. J. Med. Chem. 26, 201-5; Koskinen, A.M.P. y 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 Decarbonylation and Iminium Ion Cyclization. J. Org. Chem. 50:1239-1246; 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 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 de patente de EE.UU., Nº US 2004/0198637 titulada "Protein Arrays".

Captación celular de aminoácidos no naturales

La captación de aminoácidos no naturales por una célula es una cuestión que por lo general se tiene en cuenta cuando se diseñan y seleccionan aminoácidos no naturales, por ejemplo, para su incorporación en una proteína. Por ejemplo, la alta densidad de carga de los α -aminoácidos sugiere que es poco probable que estos compuestos sean permeables a las células. Los aminoácidos naturales son incorporados a la célula por medio de varios sistemas de transporte basados en proteínas. Puede realizarse un cribado rápido que evalúe qué aminoácidos no naturales, de haberlos, son incorporados por las células. Véanse, por ejemplo, los ensayos de toxicidad, por ejemplo, en la publicación de patente de EE.UU. N° US 2004/0198637 titulada "Protein Arrays", y Liu, D.R. y Schultz, P.G. (1999) Progress toward the evolution of an organism with an expanded genetic code. PNAS EE.UU. 96:4780-4785. Aunque la captación se analiza fácilmente con diversos ensayos, una alternativa para diseñar aminoácidos no naturales que son susceptibles a las rutas de captación celular es proporcionar rutas biosintéticas para crear aminoácidos *in vivo*.

Biosíntesis de aminoácidos no naturales

En las células ya existen muchas rutas biosintéticas para la producción de aminoácidos y otros compuestos. Aunque puede que no exista en la naturaleza un método de biosíntesis para un aminoácido no natural concreto, que incluye, pero no se limita a, en una célula, la invención proporciona tales métodos. Por ejemplo, las rutas biosintéticas para aminoácidos no naturales se generan opcionalmente en la célula hospedadora añadiendo nuevas enzimas o modificando las 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 titulado "In vivo incorporation of unnatural amino acids") se basa en la adición de una combinación de enzimas conocidas de otros organismos. Los genes para estas enzimas pueden introducirse en una célula transformando la célula con un plásmido que comprenda los genes. Los genes, cuando se expresan en la célula, proporcionan una ruta enzimática para sintetizar el compuesto deseado. Se proporcionan ejemplos de los tipos de enzimas que se añaden opcionalmente en los ejemplos que se presentan más adelante. Se encuentran secuencias de enzimas adicionales, por ejemplo, en el Genbank. Las enzimas desarrolladas artificialmente también se añaden opcionalmente a una célula de la misma manera. De esta manera, se manipulan los recursos y la maquinaria celular de una célula para producir aminoácidos no naturales.

Se encuentran disponibles diversos métodos para producir enzimas novedosas para su uso en rutas biosintéticas o para la evolución de las rutas existentes. Por ejemplo, se utiliza opcionalmente la recombinación recursiva, por ejemplo, como la desarrollada por Maxygen, Inc. (disponible en la web en maxygen.com), para desarrollar enzimas y rutas novedosas. 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. Science. EE.UU, 91:10747-10751. Del mismo modo se utiliza opcionalmente DesignPath™, desarrollado por Genencor (disponible en la web en genencor.com) para la obtención de rutas metabólicas por ingeniería genética, por ejemplo, para obtener por ingeniería genética una ruta para crear O-metil-L-tirosina en una célula. Esta tecnología reconstruye las rutas existentes en los organismos hospedadores utilizando una combinación de nuevos genes, que incluyen, pero no se limitan a, aquellos identificados a través de la genómica funcional, y el diseño y la evolución molecular. Diversa Corporation (disponible en la web en diversa.com) también proporciona tecnología para el cribado rápido de genotecas y rutas génicas, que incluye, pero no se limita a, para crear nuevas rutas.

Por lo general, el aminoácido no natural producido con una ruta biosintética obtenida por ingeniería genética de la presente invención se produce a una concentración suficiente para la biosíntesis eficaz de proteínas, por ejemplo, una cantidad celular natural, pero no hasta el punto de influir en la concentración de los demás aminoácidos o agotar los recursos celulares. Las concentraciones típicas producidas *in vivo* de esta manera son de aproximadamente 10 mM a aproximadamente 0,05 mM. Una vez que una célula es transformada con un plásmido que comprende los genes utilizados para producir las enzimas deseadas para una ruta específica y se genera un aminoácido no natural, se utilizan opcionalmente selecciones *in vivo* para optimizar adicionalmente la producción del aminoácido no natural para la síntesis de la proteína ribosómica y el crecimiento celular.

SECUENCIA DE POLIPÉPTIDO Y ÁCIDO NUCLEICO Y VARIANTES

Como se ha descrito anteriormente y como se hace más adelante, la invención proporciona secuencias de polinucleótido de ácidos nucleicos y secuencias de aminoácidos de polipéptidos, por ejemplo, ARNt y RS, y por ejemplo, composiciones y métodos que comprenden dichas secuencias. En el presente documento se describen ejemplos de dichas secuencias, por ejemplo, ARNt y RS. Sin embargo, un experto en la materia entenderá que la invención no se limita a las secuencias descritas en el presente documento, por ejemplo, los Ejemplos. Un experto entenderá que la invención también proporciona muchas secuencias relacionadas y no relacionadas con las funciones descritas en el presente documento, por ejemplo, que codifican un O-ARNt o una O-RS.

La invención proporciona polipéptidos (O-RS) y polinucleótidos, por ejemplo, O-ARNt, polinucleótidos que

codifican O-RS o porciones de los mismos, oligonucleótidos utilizados para aislar clones de aminoacil-ARNt sintetasa, etc. Los polinucleótidos de la presente invención incluyen aquellos que codifican proteínas o polipéptidos de interés de la presente invención con uno o más codones selectores. Además, los polinucleótidos de la presente invención incluyen, por ejemplo, un polinucleótido que comprende una secuencia de nucleótidos como se expone en cualquiera de las SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, un polinucleótido que es complementario a la SEQ ID NO:1, un polinucleótido que es complementario a la SEQ ID NO:2, o un polinucleótido que es complementario a la SEQ ID NO:3. Del mismo modo, un ácido nucleico que hibrida con un polinucleótido indicado anteriormente en condiciones muy rigurosas sobre sustancialmente toda la longitud del ácido nucleico es un polinucleótido de la presente invención.

En determinadas formas de realización, un vector (por ejemplo, un plásmido, un cósmido, un fago, una bacteria, un virus, un polinucleótido desnudo, un polinucleótido conjugado, etc.) comprende un polinucleótido de la presente invención. En una forma de realización, el vector es un vector de expresión. En otra forma de realización, el vector de expresión incluye un promotor unido operativamente a uno o más de los polinucleótidos de la presente invención. En otra forma de realización, una célula comprende un vector que incluye un polinucleótido de la presente invención.

Un experto también entenderá que se incluyen muchas variantes de las secuencias descritas en la divulgación. Por ejemplo, en la divulgación se incluyen las variaciones conservadoras de las secuencias descritas que producen una secuencia funcionalmente idéntica. Se consideran incluidas en la divulgación las variantes de las secuencias de polinucleótido de ácidos nucleicos, en las que las variantes hibridan con al menos una secuencia descrita. También están incluidas en la divulgación las subsecuencias únicas de las secuencias descritas en el presente documento, tal como se determinan mediante, por ejemplo, técnicas convencionales de comparación de secuencias.

Variaciones conservadoras

Debido a la degeneración del código genético, las "sustituciones silenciosas" (es decir, sustituciones en una secuencia de ácido nucleico que no dan como resultado una modificación en un polipéptido codificado) son una característica implícita de cada secuencia de ácido nucleico que codifica un aminoácido. Del mismo modo, las "sustituciones conservadoras de aminoácidos", en uno o unos pocos aminoácidos en una secuencia de aminoácidos se sustituyen con diferentes aminoácidos con propiedades muy similares, también se identifican fácilmente como muy similares a un constructo descrito. Tales variaciones conservadoras de cada secuencia descrita son una característica de la presente divulgación.

"Variaciones conservadoras" de una secuencia concreta de ácidos nucleicos se refiere a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o, en las que el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias esencialmente idénticas. Un experto en la técnica reconocerá que las deleciones, adiciones o sustituciones individuales que modifican, añaden o eliminan un solo aminoácido o un pequeño porcentaje de aminoácidos en una secuencia codificada son "variaciones modificadas de manera conservadora" o "variantes modificadas de manera conservadora" en las que las modificaciones dan como resultado la deleción de un aminoácido, la adición de un aminoácido, o la sustitución de un aminoácido con un aminoácido químicamente similar. Por lo tanto, las "variaciones conservadoras" de una secuencia de polipéptido indicada de la presente invención incluyen sustituciones de un pequeño porcentaje, por lo general inferior a un 5%, más generalmente inferior a un 4%, 2% ó 1%, de los aminoácidos de la secuencia de polipéptido, con un aminoácido seleccionado de manera conservadora del mismo grupo de sustitución conservador. La adición de secuencias que no modifican la actividad codificada de una molécula de ácido nucleico, tal como la adición de una secuencia no funcional, es una variación conservadora del ácido nucleico básico.

Los expertos en la técnica conocen tablas de sustituciones conservadoras que proporcionan aminoácidos funcionalmente similares. Cada uno de los siguientes ocho grupos contiene aminoácidos que son sustituciones conservadoras recíprocas:

- 1) Alanina (A), Glicina (G);
- 2) Ácido aspártico (D), Ácido glutámico (E);
- 3) Asparagina (N), Glutamina (Q);
- 4) Arginina (R), Lisina (K);
- 5) Isoleucina (I), Leucina (L), Metionina (M), Valina (V);
- 6) Fenilalanina (F), Tirosina (Y), Triptófano (W);
- 7) Serina (S), Treonina (T), y
- 8) Cisteína (C), Metionina (M)

(véase, por ejemplo, Creighton, Proteins: Structures and Molecular Properties (WH Freeman & Co.; segunda edición (diciembre de 1993).

Hibridación de ácidos nucleicos

Puede utilizarse la hibridación comparativa para identificar los ácidos nucleicos de la presente divulgación, tales como las SEQ ID NO:1-3, que incluyen las variaciones conservadoras de ácidos nucleicos de la presente divulgación, y este método de hibridación comparativo es un método preferente para distinguir los ácidos nucleicos de la presente invención. Además, los ácidos nucleicos diana que hibridan con los ácidos nucleicos representados por las SEQ ID NO:1-3 en condiciones de rigurosidad alta, ultra-alta, y/o ultra-ultra alta son una característica de la presente invención. Los ejemplos de tales ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones silenciosas o conservadoras de ácidos nucleicos en comparación con una determinada secuencia de ácido nucleico.

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

Los ácidos nucleicos "hibridan" cuando se asocian, por lo general en solución. Los ácidos nucleicos hibridan debido a diversas fuerzas físico-químicas bien caracterizadas, tales como enlaces de hidrógeno, exclusión de disolvente, apilamiento de bases y similares. La expresión "condiciones de hibridación rigurosas" se refiere a condiciones de baja fuerza iónica y alta temperatura tal como se conocen en la técnica. Por lo general, en condiciones rigurosas una sonda hibridará con su subsecuencia diana en una mezcla compleja de ácidos nucleicos (que incluyen, pero no se limitan a, ADN o ARN de biblioteca o celular total), pero no hibrida con otras secuencias en la mezcla compleja. Se encuentra una extensa guía para la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, parte I, capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays", (Elsevier, Nueva York), así como en Ausubel *et al.*, *Current Protocols in Molecular Biology* (1995). Hames y Higgins (1995) *Gene Probes 1* IRL Press en Oxford University Press, Oxford, Inglaterra, (Hames y Higgins 1) y Hames y Higgins (1995) *Gene Probes 2* IRL Press en Oxford University Press, Oxford, Inglaterra (Hames y Higgins 2) proporcionan detalles sobre síntesis, marcaje, detección y cuantificación de ADN y ARN, incluyendo oligonucleótidos. Generalmente, las condiciones rigurosas se seleccionan para que sean aproximadamente 5°C-10°C más bajas que el punto de fusión térmico (T_m) para la secuencia específica a un pH de fuerza iónica definida. La T_m es la temperatura (bajo una fuerza iónica, pH, y concentración de ácido nucleico definidos) a la que el 50% de las sondas complementarias a la diana hibridan con la secuencia diana en equilibrio (ya que las secuencias diana están presentes en exceso, a la T_m , el 50% de las sondas están ocupadas en equilibrio). Las condiciones rigurosas pueden ser aquellas en las que la concentración de sales es inferior a aproximadamente 1,0 M de ion sódico, por lo general una concentración de ion sódico (u otras sales) de aproximadamente 0,01 M a 1,0 M a un pH de 7,0 a 8,3 y la temperatura es al menos aproximadamente 30°C para las sondas cortas (que incluyen, pero no se limitan a, 10 a 50 nucleótidos) y de al menos aproximadamente 60°C para las sondas largas (que incluyen, pero no se limitan a, más de 50 nucleótidos). Las condiciones rigurosas también pueden conseguirse con la adición de agentes desestabilizantes tales como formamida. Para la hibridación selectiva o específica, una señal positiva puede ser al menos dos veces la de fondo, opcionalmente 10 veces la hibridación de fondo. Las condiciones de hibridación rigurosas ejemplares pueden ser de la siguiente manera: formamida al 50%, 5X SSC, y SDS al 1%, incubando a 42°C, o 5X SSC, SDS al 1%, incubando a 65°C, con lavado en 0,2X SSC y SDS al 0,1% a 65°C. Estos lavados pueden realizarse durante 5, 15, 30, 60, 120, o más minutos.

Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 residuos complementarios en un filtro en una transferencia de Southern o Northern es formalina al 50% con 1 mg de heparina a 42°C, llevándose a cabo la hibridación durante toda la noche. Un ejemplo de condiciones de lavado rigurosas es un lavado con 0,2 x SSC a 65°C durante 15 minutos (véase, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, (3ª ed. 2001) para una descripción de tampón SSC). A menudo, el lavado de alta rigurosidad va precedido de un lavado de baja rigurosidad para eliminar la señal de la sonda de fondo. Un ejemplo de lavado de baja rigurosidad es 2x SSC a 40°C durante 15 minutos. En general, una relación señal-ruido de 5 veces (o más) la observada para una sonda no relacionada en el ensayo de hibridación concreto indica la detección de una hibridación específica.

Las "condiciones de lavado de hibridación rigurosas" en el contexto de los experimentos de hibridación de ácidos nucleicos, tales como las hibridaciones de Southern y Northern, son dependientes de la secuencia y son diferentes bajo diferentes parámetros ambientales. Las secuencias más largas hibridan específicamente a temperaturas más altas. Se encuentra una extensa guía para la hibridación de ácidos nucleicos en Tijssen (1993), *supra*, y en Hames y Higgins, 1 y 2, *supra*. Las condiciones de hibridación y lavado rigurosas pueden determinarse fácilmente de manera empírica para cualquier ácido nucleico de ensayo. Por ejemplo, en la determinación de las condiciones de hibridación y lavado muy rigurosas, las condiciones de hibridación y de lavado se aumentan gradualmente (por ejemplo, aumentando la temperatura, disminuyendo la concentración de sales, aumentando la concentración de detergente y/o aumentando la concentración de disolventes orgánicos tales como formalina en la hibridación o el lavado), hasta que se cumplan varios criterios seleccionados. Por ejemplo, las condiciones de hibridación y lavado se aumentan gradualmente hasta que una sonda se une a una diana complementaria perfectamente coincidente con una relación señal-ruido que es al menos 5 veces tan alta como la observada para la

hibridación de la sonda con una diana no coincidente.

Las condiciones "muy rigurosas" se seleccionan para que sean iguales al punto de fusión térmico (T_m) para una sonda concreta. La T_m es la temperatura (bajo una fuerza iónica y un pH definidos) a la que el 50% de la secuencia de ensayo hibrida con una sonda perfectamente coincidente. Para los fines de la presente invención, generalmente, las condiciones de hibridación y lavado "muy rigurosas" se seleccionan para que sean aproximadamente 5°C inferiores a 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 señal-ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente coincidente sea al menos 10 veces tan alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no coincidentes. Se dice que un ácido nucleico diana que hibrida con una sonda en tales condiciones, con una relación señal-ruido de al menos la mitad de la del ácido nucleico diana complementario perfectamente coincidente, se unen a la sonda en condiciones de rigurosidad ultra-alta.

Del mismo modo, puede determinarse niveles de rigurosidad incluso más altos aumentando gradualmente las condiciones de hibridación y/o lavado del ensayo de hibridación pertinente. Por ejemplo, aquellos en los que la rigurosidad de las condiciones de hibridación y lavado se aumentan hasta que la relación señal-ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente coincidente sea al menos 10 veces, 20 veces, 50 veces, 100 veces, 500 veces o más tan alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no coincidentes. Se dice que un ácido nucleico diana que hibrida con una sonda en tales condiciones, con una relación señal-ruido de al menos la mitad de la del ácido nucleico diana complementario perfectamente coincidente, se une a la sonda en condiciones de rigurosidad ultra-ultra-alta.

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

Subsecuencias únicas

En un aspecto, la invención proporciona un ácido nucleico que comprende una subsecuencia única en un ácido nucleico seleccionado de entre 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 o de O-RS conocida. El alineamiento puede realizarse utilizando, por ejemplo, BLAST ajustado a los parámetros por defecto. Cualquier subsecuencia única resulta útil, por ejemplo, como sonda para identificar los ácidos nucleicos de la presente invención.

Del mismo modo, la invención incluye un polipéptido que comprende una subsecuencia única en un polipéptido seleccionado de las secuencias de O-RS descritas en el presente documento. En el presente documento, la subsecuencia única es única en comparación con un polipéptido correspondiente a cualquier secuencia de polipéptido conocida.

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

Homología, identidad y comparación de secuencias

Los términos "idéntico" o porcentaje de "identidad", en el contexto de dos o más secuencias de polipéptido o ácidos nucleicos, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje determinado de residuos de aminoácidos o nucleótidos que son iguales, cuando se comparan y se alinean para una correspondencia máxima en una ventana de comparación, o región asignada, tal como se mide utilizando uno de los algoritmos de comparación de secuencias que se describen más adelante (u otros algoritmos disponibles para las personas expertas en la técnica) o mediante alineamiento manual e inspección visual.

La expresión "sustancialmente idénticos", en el contexto de dos ácidos nucleicos o polipéptidos (por ejemplo, los 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 al menos aproximadamente un 60%, aproximadamente un 80%, aproximadamente un 90%-95%, aproximadamente un 98%, aproximadamente un 99% o más de identidad en una ventana de comparación, o región asignada, tal como se mide utilizando un algoritmo de comparación de secuencias (u otros algoritmos disponibles para las personas expertas en la técnica) o mediante alineamiento manual e

inspección visual. Tales secuencias "sustancialmente idénticas" se consideran por lo general "homólogas", sin hacer referencia a la ascendencia real. Puede existir "identidad sustancial" en una región de las secuencias que sea de al menos aproximadamente 50 residuos de longitud, una región de al menos aproximadamente 100 residuos, o en una región de al menos aproximadamente 150 residuos, o por toda la longitud de las dos secuencias a comparar.

Para la comparación de secuencias y la determinación de la homología, por lo general una secuencia actúa como secuencia de referencia con la que se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de ensayo y de referencia se introducen en un ordenador, se asignan las coordenadas de subsecuencia, en caso necesario, y se asignan los parámetros de programa de algoritmo de secuencia. A continuación, el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencia para la(s) secuencia(s) de ensayo con respecto a la secuencia de referencia, en base a los parámetros de programa asignados.

Los expertos en la técnica conocen los métodos de alineamiento de secuencias para la comparación. El alineamiento óptimo de secuencias para la comparación puede llevarse a cabo, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, *Adv. Appl. Math.* 2:482c (1981), mediante el algoritmo de alineamiento por 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. Science. EE.UU.* 85:2444 (1988), mediante aplicaciones informatizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA en el paquete de software Wisconsin Genetics, Genetics Computer Group, 575 Science Dr. Madison, WI), o mediante el alineamiento manual y la inspección visual (véase, por ejemplo, Ausubel *et al.*, *Current Protocols in Molecular Biology* (suplemento de 1995)).

Es un ejemplo de algoritmo que resulta adecuado para determinar el porcentaje de identidad de secuencia y la similitud de secuencia el algoritmo BLAST y los algoritmos BLAST 2.0, que se describen en Altschul *et al.* (1997) *Nuc. Acids Res.* 25:3389-3402, y Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). El software para realizar los análisis BLAST está a disposición del público a través del Centro Nacional de Información sobre Biotecnología disponible en la web en ncbi.nlm.nih.gov. Este algoritmo implica identificar en primer lugar los pares de secuencias de alta puntuación (HSP) identificando palabras cortas de longitud W en la secuencia de consulta, que coinciden o satisfacen alguna puntuación umbral de valor positivo T cuando se alinean con una palabra de la misma longitud en una secuencia de base de datos. T se denomina umbral de puntuación de palabra vecina (Altschul *et al.*, *supra*). Estas coincidencias ("hits") de palabras vecinas actúan como semillas para iniciar las búsquedas para encontrar HSP más largos que los contienen. A continuación, las coincidencias de palabras se extienden en ambas direcciones a lo largo de cada secuencia tanto como pueda aumentarse la puntuación de alineamiento acumulada. Las puntuaciones acumuladas se calculan utilizando, para las secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y N (puntuación de penalización para los residuos no coincidentes, siempre < 0). Para las secuencias de aminoácidos, se utiliza una matriz de puntuación para calcular la puntuación acumulada. La extensión de las coincidencias de palabras en cada dirección se detiene cuando: la puntuación de alineamiento acumulativo disminuye la cantidad X desde su valor máximo alcanzado; la puntuación acumulada llega a cero o menos, debido a la acumulación de uno o más alineamientos de residuos de puntuación negativa; o se alcanza el final de cualquiera de las secuencias. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un punto de corte de 100, M=5, N=-4, y una comparación de ambas hebras. Para las secuencias de aminoácidos, el programa BLASTP utiliza por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff (1992) *Proc. Natl. Acad. Sci. EE.UU.* 89:10915) alineamientos (B) de 50, expectativa (E) de 10, M=5, N=-4, y una comparación de ambas hebras. El algoritmo BLAST se realiza por lo general con el filtro de "baja complejidad" desactivado.

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, *Proc. Natl. Acad. Sci. EE.UU.* 90:5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona un indicio de la probabilidad con la que se produciría por casualidad una coincidencia entre dos secuencias de aminoácidos o nucleótidos. Por ejemplo, un ácido nucleico puede considerarse similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es inferior a aproximadamente 0,2; inferior a aproximadamente 0,01; o inferior a aproximadamente 0,001.

Mutagénesis y otras técnicas de biología molecular

Los polinucleótidos y polipéptidos de la presente invención y utilizados en la invención pueden manipularse utilizando técnicas de biología molecular. Puede modificarse convenientemente una secuencia de nucleótidos mediante mutagénesis dirigida al sitio de acuerdo con los métodos convencionales. Como alternativa, puede prepararse la secuencia de nucleótidos mediante síntesis química, que incluye, pero no se limita a, el uso de un sintetizador de oligonucleótidos, en el que los oligonucleótidos se diseñan en base a la secuencia de aminoácidos del polipéptido deseado, y preferentemente seleccionando aquellos codones que se vean favorecidos en la célula hospedadora en la que se producirá el polipéptido recombinante. Por ejemplo, pueden sintetizarse y ensamblarse

varios oligonucleótidos pequeños que codifican para porciones del polipéptido deseado mediante PCR, ligación o reacción en cadena de ligación. Véase, por ejemplo, Barany *et al.*, Proc. Natl. Acad. Sci. 88:189-193 (1991); la patente de EE.UU. 6.521.427.

5 La presente invención utiliza técnicas de rutina en el campo de la genética recombinante. Los textos básicos que describen los métodos generales de uso en la presente invención incluyen Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (3ª ed. 2001); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); y Current Protocols in Molecular Biology (Ausubel *et al.*, eds., 1994)).

10 Los textos generales que describen las técnicas biológicas moleculares incluyen Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology, volumen 152 Academic Press, Inc, San Diego, CA (Berger); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (2ª ed.). Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989 ("Sambrook") y Current Protocols in Molecular Biology F.M. Ausubel *et al.*, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc, (complementado a lo largo de 1999) ("Ausubel"). Estos textos describen mutagénesis, el uso de vectores, promotores y muchos otros temas pertinentes relacionados con, por ejemplo, la generación de genes o polinucleótidos que incluyen codones selectores para la producción de proteínas que incluyen aminoácidos seleccionados (por ejemplo, aminoácidos no naturales), ARNt ortogonales, sintetisas ortogonales y pares de los mismos.

20 En la invención se utilizan diversos tipos de mutagénesis para diversos fines, que incluyen pero no se limitan a, para producir nuevas sintetisas o ARNt, para mutar moléculas de ARNt, para producir bibliotecas de ARNt, para mutar moléculas de RS, para producir bibliotecas de sintetisas, para producir codones selectores, para insertar codones selectores que codifican un aminoácido seleccionado en una proteína o polipéptido de interés. Estos incluyen, pero no se limitan a mutagénesis puntual aleatoria dirigida al sitio, recombinación homóloga, barajado de ADN u otros métodos de mutagénesis recursivos, construcción quimérica, mutagénesis utilizando moldes que contienen uracilo, mutagénesis dirigida por oligonucleótidos, mutagénesis de ADN modificado con fosforotioato, mutagénesis utilizando ADN dúplex interrumpido o similares, o cualquier combinación de los mismos. Los métodos adecuados adicionales incluyen reparación de desapareamientos puntuales, mutagénesis utilizando cepas hospedadoras defectuosas para la reparación, selección por restricción y purificación por restricción, mutagénesis por delección, mutagénesis mediante síntesis génica total, reparación de rotura de la doble hebra, y similares. La mutagénesis, que incluye, pero no se limita a, que implica constructos quiméricos, también está incluida en la presente invención. En una forma de realización, la mutagénesis puede ser guiada por la información conocida de la molécula de origen natural o la molécula de origen natural modificada o mutada, que incluye, pero no se limita a, secuencia, comparaciones de secuencias, propiedades físicas, estructura secundaria, terciaria o cuaternaria, estructura cristalina, o similares.

35 Los textos y ejemplos encontrados en el presente documento describen estos procedimientos. Se encuentra información adicional en las siguientes publicaciones y referencias citadas en: Ling *et al.*, Approaches to DNA mutagenesis: an overview, Anal Biochem. 254 (2):157-178 (1997); Dale *et al.*, Oligonucleotide-directed random mutagenesis using the phosphorothioate method, Methods Mol. Biol. 57:369-374 (1996); Smith, In vitro mutagenesis, Ann. Rev. Genet. 19:423-462 (1985); Botstein y Shortle, Strategies and applications of in vitro mutagenesis, Science 229:1193-1201 (1985); Carter, Site-directed mutagenesis, Biochem. J. 237:1-7 (1986); Kunkel, The efficiency of oligonucleotide directed mutagenesis, in Nucleic Acids & Molecular Biology (Eckstein, F. y Lilley, D.M.J. eds., Springer Verlag, Berlin) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, Proc. Natl. Acad. Sci. USA 82:488-492 (1985); Kunkel *et al.*, Rapid and efficient site-specific mutagenesis without phenotypic selection, Methods in Enzymol. 154, 367-382 (1987); Bass *et al.*, Mutant Trp repressors with new DNA-binding specificities, Science 242:240-245 (1988); Zoller y Smith, Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the Production of point mutations in any DNA fragment, Nucleic Acids Res. 10:6487-6500 (1982); Zoller y Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, Methods in Enzymol. 100:468-500 (1983); Zoller y Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, Methods in Enzymol. 154:329-350 (1987); Taylor *et al.*, The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, Nucl. Acids Res. 13: 8749-8764 (1985); Taylor *et al.*, The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, Nucl. Acids Res. 13: 8765-8785 (1985); Nakamaye y Eckstein, Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, Nucl. Acids Res. 14: 9679-9698 (1986); Sayers *et al.*, 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, Nucl. Acids Res. 16:791-802 (1988); Sayers *et al.*, Strand specific cleavage of phosphorothioate-containing DNA by reactions with restriction endonucleases in the presence of ethidium bromide, (1988) Nucl. Acids Res. 16: 803-814; Kramer *et al.*, The gapped duplex DNA approach to oligonucleotide-directed mutation construction, Nucl. Acids Res. 12: 9441-9456 (1984); Kramer y Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, Methods in Enzymol. 154:350-367 (1987); Kramer *et al.*, Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, Nucl. Acids Res. 16:7207 (1988); Fritz *et al.*, Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, Nucl. Acids Res. 16:6987-6999 (1988); Kramer *et al.*, Different base/base mismatches are corrected with

different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*, *Cell* 38:879-887 (1984); Carter *et al.*, Improved oligonucleotide site-directed mutagenesis using M13 vectors, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, Improved oligonucleotide-directed mutagenesis using M13 vectors, *Methods in Enzymol.* 154: 382-403 (1987); Eghtedarzadeh y Henikoff, Use of oligonucleotides to generate large deletions, *Nucl. Acids Res.* 14: 5115 (1986); Wells *et al.*, Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317:415-423 (1986); Nambiar *et al.*, Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223:1299-1301 (1984); Sakmar y Khorana, Total synthesis and expression of a gene for the alpha-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells *et al.*, Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985); Grundström *et al.*, Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, *Nucl. Acids Res.* 13: 3305-3316 (1985); Mandeckí, Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181 (1986); Arnold, Protein engineering for unusual environments, *Current Opinion in Biotechnology* 4:450-455 (1993); Sieber, *et al.*, *Nature Biotechnology*, 19:456-460 (2001); W. P. C. Stemmer, *Nature* 370, 389-91 (1994); y I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995). Pueden encontrarse detalles adicionales sobre muchos de los métodos anteriormente indicados en *Methods in Enzymology*, Volumen 154, que también describe controles útiles para los problemas relacionados con la resolución de problemas con varios métodos de mutagénesis.

Los oligonucleótidos, por ejemplo, para su uso en la mutagénesis de la presente invención, por ejemplo, la mutación de bibliotecas de sintetasas, o la modificación de ARNt, se sintetizan por lo general químicamente de acuerdo con el método del triéster de fosforamidita en fase sólida descrito por Beaucage y Caruthers, *Tetrahedron Letts.* 22(20):1859-1862, (1981), por ejemplo, utilizando un sintetizador automatizado, tal como se describe en Needham-VanDevanter *et al.*, *Nucleic Acids Res.* 12:6159-6168 (1984).

Además, puede encargarse esencialmente cualquier ácido nucleico personalizado o convencional a cualquiera de una diversidad de fuentes comerciales, tales como The Midland Certified Reagent Company (mcr@oligos.com), The Great American Gene Company (www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, California) y muchas otras.

La invención también se refiere a células hospedadoras eucariotas, células hospedadoras no eucariotas, y organismos para la incorporación *in vivo* de un aminoácido no natural por medio de pares ARNt/RS ortogonales. Las células hospedadoras se obtienen por ingeniería genética (que incluye, pero no se limita a, transformación, transducción o transfección) con los polinucleótidos de la presente invención o constructos que incluyen un polinucleótido de la presente invención, que incluye, pero no se limita a, un vector de la presente 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 a derivatizar se unen operativamente a elementos testigo de expresión génica que son funcionales en la célula hospedadora deseada. El vector puede estar, por ejemplo, en forma de plásmido, cósmido, fago, bacteria, virus, polinucleótido desnudo, o polinucleótido conjugado. Los vectores se introducen en las células y/o microorganismos mediante métodos convencionales que incluyen electroporación (Fromm *et al.*, *Proc. Natl. Acad. Sci. EE.UU.* 82, 5824 (1985)), infección mediante vectores virales, penetración balística de alta velocidad mediante partículas pequeñas con el ácido nucleico dentro de la matriz de perlas o partículas pequeñas, o sobre la superficie (Klein *et al.*, *Nature* 327, 70-73 (1987)), y/o similares.

Se encuentran disponibles varios métodos bien conocidos para introducir ácidos nucleicos diana en las células, cualquiera de los cuales puede utilizarse en la invención. Estos incluyen: fusión de las células receptoras con protoplastos bacterianos que contienen el ADN, electroporación, bombardeo con proyectiles, e infección con vectores virales (analizado más adelante), etc. Pueden utilizarse células bacterianas para amplificar el número de plásmidos que contienen los constructos de ADN de la presente invención. Las bacterias se cultivan hasta la fase logarítmica y los plásmidos dentro de las bacterias pueden aislarse mediante diversos métodos conocidos en la técnica (véase, por ejemplo, Sambrook). Además, se encuentran disponibles comercialmente kits para la purificación de plásmidos a partir de bacterias, (véase, por ejemplo, EasyPrep™, Flexiprep™, ambos de Pharmacia Biotech; StrataClean™ de Stratagene; y QIAprep™ de Qiagen). A continuación, se manipulan adicionalmente los plásmidos aislados y purificados para producir otros plásmidos, que se utilizan para transfectar células o para ser incorporados en vectores relacionados para infectar organismos. Los vectores típicos contienen terminadores de la transcripción y de la traducción, secuencias de iniciación de la traducción y de la transcripción, y promotores útiles para la regulación de la expresión del ácido nucleico diana concreto. Los vectores comprenden opcionalmente casetes de expresión genéricos que contienen al menos una secuencia terminadora independiente, secuencias que permiten la replicación del casete en eucariotas o procariotas, o ambos, (que incluyen, pero no se limitan a, vectores lanzadera) y marcadores de selección para sistemas procariotas y eucariotas. Los vectores resultan adecuados para la replicación y/o integración en procariotas, eucariotas, o ambos. Véase Gillam y Smith, *Gene* 8:81 (1979); Roberts, *et al.*, *Nature* 328:731 (1987); Schneider, E, *et al.*, *Protein Expr. Purif.* 6(1) 10-14 (1995); Ausubel, Sambrook, Berger (todos *supra*). Un catálogo de bacterias y bacteriófagos útiles para la clonación es proporcionado, por ejemplo, por la ATCC, por ejemplo, el *Catalogue of Bacteria and Bacteriophage* (1992) Gherna *et al.* (eds.), publicado por la ATCC. También se encuentran procedimientos básicos adicionales para la secuenciación, la clonación y otros aspectos de la biología molecular y consideraciones teóricas subyacentes en Watson *et al.* (1992) *Recombinant DNA*, segunda

edición, Scientific American Books, NY. Además, puede encargarse esencialmente cualquier ácido nucleico personalizado o convencional (y prácticamente cualquier ácido nucleico marcado, ya sea convencional o no convencional) de cualquiera de una diversidad de fuentes comerciales, tales como la Midland Certified Reagent Company (Midland, TX disponible en la web en mrcr.com), The Great American Gene Company (Ramona, CA disponible en la web en genco.com), ExpressGen Inc. (Chicago, IL disponible en la web en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchas otras.

Las células hospedadoras obtenidas por ingeniería genética pueden cultivarse en medios nutritivos convencionales modificados según resulte apropiado para actividades tales como, por ejemplo, las etapas cribado, la activación de promotores o la 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 posterior aislamiento del ácido nucleico) incluyen Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, Nueva York y las referencias citadas en el mismo; Payne *et al.* (1992) *Plant Cell and Tissue Culture in Liquid Systems*, John Wiley & Sons, Inc, Nueva York, NY; Gamborg y Phillips (eds.) (1995) *Plant Cell, Tissue and Organ Culture*; *Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg Nueva York) y Atlas y Parks (eds.) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL.

La capacidad para incorporar aminoácidos no naturales directamente a proteínas *in vivo* ofrece una gran variedad de ventajas, que incluyen, pero no se limitan a, altos rendimientos de proteínas mutantes, facilidad técnica, el potencial para estudiar las proteínas mutantes en las células o, posiblemente, en los organismos vivos y el uso de estas proteínas mutantes en tratamientos terapéuticos y usos de diagnóstico. La capacidad para incluir aminoácidos no naturales con diversos tamaños, niveles de acidez, nucleofilia, hidrofobicidad, y otras propiedades en proteínas puede ampliar en gran medida nuestra capacidad para manipular racional y sistemáticamente las estructuras de las proteínas, tanto para investigar la función de las proteínas como para crear nuevas proteínas u organismos con propiedades novedosas.

PROTEÍNAS Y POLIPÉPTIDOS DE INTERÉS

Puede realizarse la incorporación de un aminoácido no natural para diversos fines, que incluyen, pero no se limitan a, adaptar la estructura y/o función de la proteína, cambiar el tamaño, acidez, nucleofilia, enlaces de hidrógeno, hidrofobicidad, accesibilidad de la proteasa a los sitios diana, dirigirla a un resto (que incluye, pero no se limita a, para una matriz de proteínas), añadir una molécula biológicamente activa, fijar un polímero, fijar un radionúclido, modular la semivida en suero, modular la penetración en tejido (por ejemplo, tumores), modular el transporte activo, modular la especificidad o distribución en la célula, el tejido o el órgano, modular la inmunogenicidad, modular la resistencia a proteasas, etc. Las proteínas que incluyen un aminoácido no natural pueden tener propiedades catalíticas o biofísicas potenciadas o incluso completamente nuevas. Por ejemplo, las siguientes propiedades son modificadas opcionalmente por la inclusión de un aminoácido no natural en una proteína: toxicidad, biodistribución, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, capacidad catalítica, semivida (que incluye, pero no se limita a, semivida en suero), capacidad para reaccionar con otras moléculas, que incluye, pero no se limita a, de forma covalente o no covalente, y similares. Las composiciones que incluyen proteínas que incluyen al menos un aminoácido no natural resultan útiles para, incluyendo pero sin limitarse a, agentes de diagnóstico, enzimas catalíticas, enzimas industriales, proteínas de unión (que incluyen, pero no se limitan a, anticuerpos), agentes terapéuticos novedosos, e incluyendo, pero sin limitarse a, 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.

Una proteína puede tener al menos uno, que incluye, pero no se limita a, 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, incluyendo, pero sin limitarse a, poder haber 1, 2, 3, 4, 5, 6, 7, 8, 9, ó 10 o más sitios diferentes en la proteína que comprendan 1, 2, 3, 4, 5, 6, 7, 8, 9, ó 10 o más aminoácidos no naturales diferentes. Una proteína puede tener al menos un aminoácido concreto presente en la proteína, pero no todos ellos, sustituido con el aminoácido no natural. Para una determinada proteína con más de un aminoácido no natural, los aminoácidos no naturales pueden ser idénticos o diferentes (incluyendo, pero sin limitarse a, que la proteína puede incluir dos o más tipos diferentes de aminoácidos no naturales, o puede incluir dos aminoácidos no naturales iguales). Para una determinada proteína 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.

Mediante la producción de proteínas o polipéptidos de interés con al menos un aminoácido no natural en células eucariotas, las proteínas o polipéptidos incluirán por lo general modificaciones post-traduccionales eucarióticas. En determinadas formas de realización, una proteína incluye al menos un aminoácido no natural y al menos una modificación post-traduccionales que es realizada *in vivo* por una célula eucariota, en la que la modificación post-traduccionales no es realizada por una célula procariota. Por ejemplo, la modificación post-traduccionales incluye, incluyendo pero sin limitarse a, acetilación, acilación, modificación de lípidos, palmitoilación, adición de palmitato, fosforilación, modificación por unión de glicolípido, glicosilación, y similares. En otro aspecto más, la modificación

5 post-traducciona incluye el procesamiento proteolítico de precursores (que incluyen, pero no se limitan a, precursor de calcitonina, precursor del péptido relacionado con el gen de la calcitonina, hormona pre-proparatiroidea, preproinsulina, proinsulina, prepro-opiomelanocortina, pro-opiomelanocortina y similares), ensamblaje en una proteína de múltiples subunidades o ensamblaje macromolecular, traslado a otro sitio en la célula (que incluye, pero no se limita a, orgánulos, tales como el retículo endoplasmático, el aparato de Golgi, el núcleo, lisosomas, peroxisomas, mitocondrias, cloroplastos, vacuolas, etc., o a través de la ruta secretora). En determinadas formas de realización, la proteína comprende una secuencia de secreción o localización, una epitopo de identificación, un marcador FLAG, un marcador de polihistidina, una fusión de GST, o similares.

10 Son también una característica de la presente invención métodos para producir una proteína en una célula con un aminoácido seleccionado en una posición determinada. 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 seleccionado; en el que la célula comprende adicionalmente: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector, y una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente el O-ARNt con el aminoácido seleccionado. Por lo general, el O-ARNt comprende actividad supresora en presencia de una sintetasa afín en respuesta a un codón selector. Es también una característica de la presente invención una proteína producida mediante este método.

20 Las composiciones de la presente invención y las composiciones producidas mediante los métodos de la presente invención están opcionalmente en una célula. Pueden utilizarse los pares O-ARNt/O-RS o los componentes individuales de la presente invención en la maquinaria de traducción de un sistema hospedador, que da como resultado la incorporación de un aminoácido seleccionado, por ejemplo, un aminoácido no natural, a una proteína. Las solicitudes de patente USSN 10/825.867, titulada "Expanding the Eukaryotic Genetic Code" y 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", describen 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 seleccionado, tal como un aminoácido no natural, por ejemplo, un aminoácido sintético, tal como un derivado de un aminoácido leucina, que puede añadirse exógenamente al medio de crecimiento, a una proteína, en respuesta a un codón selector. Opcionalmente, las composiciones de la presente invención pueden estar en un sistema de traducción *in vitro*, o en un sistema o sistemas *in vivo*.

30 Puede producirse cualquier proteína (o parte de la misma) que incluye un aminoácido seleccionado, por ejemplo, un aminoácido no natural, (y cualquier ácido nucleico codificante correspondiente, por ejemplo, que incluye uno o más codones selectores) utilizando las composiciones y los métodos del presente documento. Cualquier polipéptido resulta adecuado para incorporar uno o más aminoácidos seleccionados. No se hace ningún intento por identificar los cientos de miles de proteínas conocidas, cualquiera de las cuales puede modificarse para que incluya uno o más aminoácidos no naturales, por ejemplo, adaptando cualquier método de mutación disponible para incluir uno o más codones selectores apropiados en un sistema de traducción pertinente. Los bancos de secuencias comunes para las proteínas conocidas incluyen GenBank EMBL, DDBJ y el NCBI. Pueden identificarse fácilmente otros bancos buscando en Internet.

40 Por lo general, las proteínas son, por ejemplo, al menos un 60%, al menos un 70%, al menos un 75%, al menos un 80%, al menos un 90%, al menos un 95%, o al menos un 99% o más idénticas a cualquier proteína disponible (por ejemplo, una proteína terapéutica, una proteína de diagnóstico, una enzima industrial o parte de las mismas, y similares), y comprenden uno o más aminoácidos seleccionados. Ejemplos de proteínas terapéuticas, de diagnóstico, y otras proteínas que pueden modificarse para que comprendan uno o más aminoácidos seleccionados, por ejemplo, un aminoácido no natural, pueden encontrarse, pero sin limitarse a, aquellas en el documento USSN 10/825.867 titulado "Expanding the Eukaryotic Genetic Code", y la solicitud de patente de EE.UU. con número de serie 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS".

50 En determinadas formas de realización, la proteína o el polipéptido de interés (o parte de los mismos) en los métodos y/o en las composiciones de la presente invención es codificado por un ácido nucleico. Por lo general, el ácido nucleico comprende al menos un codón selector, al menos dos codones selectores, al menos tres codones selectores, al menos cuatro codones selectores, al menos cinco codones selectores, al menos seis codones selectores, al menos siete codones selectores, al menos ocho codones selectores, al menos nueve codones selectores, diez o más codones selectores.

60 Los genes que codifican para las proteínas o los polipéptidos de interés pueden mutagenizarse utilizando métodos bien conocidos para un experto en la materia y descritos en el presente documento bajo el título "Mutagénesis y otras técnicas de biología molecular" para que incluyan, por ejemplo, uno o más codones selectores para la incorporación de un aminoácido seleccionado, por ejemplo, un aminoácido no natural. Por ejemplo, un ácido nucleico para una proteína de interés se mutageniza para que incluya uno o más codones selectores, proporcionando la inserción del uno o más aminoácidos seleccionados, por ejemplo, un aminoácido no natural. La invención incluye cualquiera de tales versiones variantes, por ejemplo, mutante, de cualquier proteína, por ejemplo, que incluya al menos un aminoácido seleccionado. Del mismo modo, la invención también incluye los ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifique uno o más aminoácidos seleccionados.

Para producir una proteína que incluya un aminoácido seleccionado, pueden utilizarse organismos y células hospedadoras que estén adaptados para la incorporación *in vivo* del aminoácido seleccionado por medio de pares de ARNt/RS ortogonales. Las células hospedadoras se obtienen por ingeniería genética (por ejemplo, transformación, transducción o transfección) con uno o más vectores que expresan el ARNt ortogonal, la ARNt sintetasa ortogonal, y un vector que codifica la proteína a derivatizar. Cada uno de estos componentes puede estar en el mismo vector, o cada uno puede estar en un vector distinto, 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, cósmido, fago, bacteria, virus, polinucleótido desnudo, o polinucleótido conjugado.

10 SISTEMAS ALTERNATIVOS

Se han empleado varias estrategias para introducir aminoácidos no naturales en proteínas en células hospedadoras no recombinantes, células hospedadoras mutagenizadas, o en sistemas libres de células. La derivatización de aminoácidos con cadenas laterales reactivas, tales como Lys, Cys y Tyr, dio como resultado la conversión de la lisina en N²-acetil-lisina. La síntesis química también proporciona un método sencillo para incorporar aminoácidos no naturales. Con el reciente desarrollo de la ligación enzimática y la ligación química nativa de fragmentos peptídicos, es posible producir proteínas más grandes. Véase, por ejemplo, P. E. Dawson y S. B. H. Kent, *Annu. Rev. Biochem.* 69:923 (2000). La ligación química de péptidos y la ligación química nativa se describen en la patente de EE.UU. N° 6.184.344, la publicación de patente de EE.UU. N° 2004/0138412, la publicación de patente de EE.UU. N° 2003/0208046, el documento WO 02/098902, y el documento WO 03/042235. Se ha utilizado un método de biosíntesis *in vitro* general en el que un ARNt supresor acilado químicamente con el aminoácido no natural deseado se añade a un extracto *in vitro* capaz de dar soporte a la biosíntesis de proteínas, para incorporar de manera específica de sitio más de 100 aminoácidos no naturales en diversas proteínas de prácticamente cualquier tamaño. Véase, por ejemplo, V.W. Cornish, D. Mendel y P.G. Schultz, *Angew. Chem. Int. Ed. Engl.* 1995, 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, *Science* 244:182-188 (1989), y J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Diala, Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, *J. Am. Chem. Soc.* 111:8013-8014 (1989). Se ha introducido una amplia gama de grupos funcionales en proteínas para estudios sobre estabilidad de las proteínas, plegamiento de las proteínas, mecanismo enzimático, y transducción de señales.

Se desarrolló un método *in vivo*, denominado incorporación bajo presión selectiva, para explotar la promiscuidad de las sintetasas de tipo silvestre. Véase, por ejemplo, N. Budisa, C. Minks, S. Alefelder, W. Wenger, F.M. Dong, L. Moroder y R. Huber, *FASEB J.* 13:41 (1999). Se cultiva una cepa auxotrófica, en la que se interrumpe la ruta metabólica pertinente que suministra a la célula un aminoácido natural concreto, en medio mínimo que contiene concentraciones limitadas del aminoácido natural, mientras se reprime la transcripción del gen diana. Al comienzo de una fase de crecimiento estacionario, el aminoácido natural está agotado y se sustituye con el análogo de aminoácido no natural. La inducción de la expresión de la proteína recombinante da como resultado la acumulación de una proteína que contiene el análogo no natural. Por ejemplo, utilizando esta estrategia, se han incorporado a las proteínas o, m y p-fluorofenilalaninas, y presentan dos hombros característicos en el espectro UV, que pueden identificarse fácilmente, véase, por ejemplo, C. Minks, R. Huber, L. Moroder y N. Budisa, *Anal. Biochem.* 284:29 (2000); se ha utilizado trifluorometionina para sustituir la metionina en la lisozima del bacteriófago T4 para estudiar su interacción con ligandos quitooligosacáridos mediante RMN de ¹⁹F, véase, por ejemplo, H. Duewel, E. Daub, V. Robinson y J.F. Honek, *Biochemistry* 36:3404 (1997); y se ha incorporado trifluoro-leucina en vez de leucina, lo que da como resultado un aumento de la estabilidad térmica y química de una proteína con cremallera de leucina. Véase, por ejemplo, Y. Tang, G. Ghirlanda, W.A. Petka, T. Nakajima, W.F. DeGrado y D.A. Tirrell, *Angew. Chem. Int. Ed. Engl.* 40:1494 (2001). Por otra parte, se incorporan selenometionina y tellurometionina en diversas proteínas recombinantes para facilitar la solución de las fases en la cristalografía de rayos-X. Véase, por ejemplo, W.A. Hendrickson, J.R. Horton y D.M. Lemaster, *EMBO J.*, 9:1665 (1990); J.O. Boles, K. Lewinski, M. Kunkle, J.D. Odom, B. Dunlap, L. Lebioda y M. Hatada, *Nat. Struct. Biol.* 1:283 (1994). N. Budisa, B. Steipe, P. Demange, C. Eckerskom, J. Kellermann y R. Huber, *Eur. J. Biochem.* 230: 788 (1995); y N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neufeind, L. Moroder y R. Huber, *J. Mol. Biol.*, 270:616 (1997). También se han incorporado de manera eficaz análogos de metionina con funcionalidades alqueno o alquino, que permite la modificación adicional de proteínas mediante medios químicos. Véase, por ejemplo, J.C. van Hest y D.A. Tirrell, *FEBS Lett.* 428:68 (1998); J.C. van Hest, K.L. Kiick y D.A. Tirrell, *J. Am. Chem. Soc.* 122:1282 (2000); y K.L. Kiick y D.A. Tirrell, *Tetrahedron*, 56: 9487 (2000); la patente de EE.UU. N° 6.586.207; la publicación de patente de EE.UU. 2002/0042097.

El éxito de este método depende del reconocimiento de los análogos de aminoácidos no naturales por las aminoacil-ARNt sintetasas, que, en general, requieren una alta selectividad para asegurar la fidelidad de la traducción de proteínas. Una forma de ampliar el campo de aplicación de este método es relajar la especificidad de sustrato de las aminoacil-ARNt sintetasas, que se ha conseguido en un número limitado de casos. Por ejemplo, la sustitución de Ala²⁹⁴ por Gly en la fenilalanil-ARNt sintetasa de *Escherichia coli* (PheRS) aumenta el tamaño del bolsillo de unión al sustrato, y da como resultado la acilación de ARNtPhe por p-Cl-fenilalanina (p-Cl-Phe). Véase, M. Ibba, P. Kast y H. Hennecke, *Biochemistry* 33:7107 (1994). Una cepa de *Escherichia coli* que alberga esta PheRS mutante permite incorporar p-Cl-fenilalanina o p-Br-fenilalanina en vez de fenilalanina. Véase, por ejemplo, M. Ibba y H. Hennecke, *FEBS Lett.* 364:272 (1995); y N. Sharma, R. Furter, P. Kast y D.A. Tirrell, *FEBS Lett.* 467:37 (2000).

Del mismo modo, una mutación puntual Phe130Ser cerca del sitio de unión del aminoácido de la tirosil-ARNt sintetasa de *Escherichia coli* demostró permitir la incorporación de azatirosina de manera más eficaz que la tirosina. Véase, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll y S. Nishimura, J. Biol. Chem, 275:40324 (2000).

5 Otra estrategia para incorporar aminoácidos no naturales en proteínas *in vivo* es modificar las sintetasas que tienen mecanismos de corrección de errores. Estas sintetasas no pueden discriminar y por lo tanto activar los aminoácidos que son estructuralmente similares a los aminoácidos naturales afines. Este error se corrige en un lugar distinto, que desacila el aminoácido cargado incorrectamente del ARNt para mantener la fidelidad de la traducción de proteínas. Si se inhabilita la actividad de corrección de errores de la sintetasa, los análogos estructurales que están 10 activados incorrectamente pueden escapar a la función de corrección de errores y ser incorporados. Este enfoque se ha demostrado recientemente con la valil-ARNt sintetasa (ValRS). Véase, V. Doring, H.D. Mootz, L.A. Nangle, T.L. Hendrickson, V. de Crecy-Lagard, P. Schimmel y P. Marliere, Science, 292:501 (2001). La ValRS puede aminoacilar incorrectamente el ARNtVal con Cys, Thr, o aminobutirato (Abu); estos aminoácidos no afines son hidrolizados 15 posteriormente por el dominio de corrección de errores. Después de la mutagénesis aleatoria del cromosoma de *Escherichia coli*, se seleccionó una cepa mutante de *Escherichia coli* que tenía una mutación en el sitio de corrección de errores de ValRS. Esta ValRS defectuosa para la corrección de errores carga incorrectamente el ARNtVal con Cys. Debido a que Abu se parece estéricamente a Cys (el grupo -SH de Cys está sustituido con -CH₃ en Abu), la ValRS mutante también incorpora Abu en las proteínas cuando esta cepa mutante de *Escherichia coli* se cultiva en 20 presencia de Abu. El análisis espectrométrico de masas muestra que aproximadamente un 24% de las valinas son sustituidas por Abu en cada posición de valina en la proteína nativa.

Los métodos de síntesis en fase sólida y semisintéticos también han permitido la síntesis de varias 25 proteínas que contienen aminoácidos novedosos. Por ejemplo, véanse las siguientes publicaciones y referencias citadas en las mismas, que son las siguientes: Crick, F.H.C., Barrett, L. Brenner, S. Watts-Tobin, R. General nature of the genetic code for proteins. Nature, 192:1227-1232 (1961); Hofmann, K., Bohn, H. Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragment, J. Am Chem, 88(24):5914-5919 (1966); Kaiser, E.T. Synthetic approaches to biologically active peptides and proteins including enzymes, Acc Chem Res, 22:47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E.T. Peptide segment 30 coupling catalyzed by the semisynthetic enzyme thiosubtilisin, J Am Chem Soc, 109:3808-3810 (1987); Schnolzer, M., Kent, S B H. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease, Science, 256(5054):221-225 (1992); Chaiken, I.M. Semisynthetic peptides and proteins, CRC Crit Rev Biochem, 11(3):255-301 (1981); Offord, R.E. Proteins engineering by Chemical means? Protein Eng., 1(3):151-157 (1987); y, Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J.A. A Designed Peptide Ligase for- Total 35 Synthesis of Ribonuclease A with Unnatural Catalytic Residues, Science, 266(5183):243 (1994).

Se ha utilizado la modificación química para introducir diversas cadenas laterales no naturales, incluyendo 40 cofactores, marcadores de espín y oligonucleótidos en proteínas *in vitro*. Véase, por ejemplo, Corey, D.R., Schultz, P.G. Generation of a hybrid sequence-specific single-stranded deoxyribonuclease, Science, 238(4832):1401-1403 (1987); Kaiser, E.T., Lawrence D.S., Rokita, S.E. The chemical modification of enzymatic specificity, Annu Rev Biochem, 54:565-595 (1985); Kaiser, E.T., Lawrence, D.S. Chemical mutation of enzyme active sites, Science, 226(4674):505-511 (1984); Neet, K.E., Nanci A, Koshland, D.E. Properties of thiol-subtilisin, J Biol. Chem, 243(24):6392-6401 (1968); Polgar, L. et M.L. Bender. A new enzyme containing a synthetically formed active site. Thiol-subtilisin. J. Am Chem Soc, 88:3153-3154 (1966); y, Pollack, S.J., Nakayama, G. Schultz, P.G. Introduction of 45 nucleophiles and spectroscopic probes into antibody combining sites, Science, 242(4881):1038-1040 (1988).

Definición de polipéptidos por inmunorreactividad

Debido a que los polipéptidos de la presente invención proporcionan diversas secuencias de polipéptido 50 nuevas (por ejemplo, que comprenden aminoácidos seleccionados (por ejemplo, aminoácidos no naturales) en el caso de proteínas sintetizadas en los sistemas de traducción del presente documento, o, por ejemplo, en el caso de las sintetasas novedosas, secuencias novedosas de aminoácidos convencionales), los polipéptidos también proporcionan nuevas características estructurales que pueden ser reconocidas, por ejemplo, en ensayos inmunológicos. La generación de antisueros, que se unen específicamente a los polipéptidos de la presente 55 invención, así como los polipéptidos que son unidos por dichos antisueros son una característica de la presente invención. El término "anticuerpo", tal como se utiliza en el presente documento, incluye, pero no se limita a, un polipéptido codificado sustancialmente por un gen de inmunoglobulina o genes de inmunoglobulina, o fragmentos de los mismos que específicamente se unen y reconocen un analito (antígeno). Los ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos, y monocatenarios, y similares. Los fragmentos de inmunoglobulinas, que 60 incluyen fragmentos Fab y fragmentos producidos por una biblioteca de expresión, incluyendo la presentación en fagos, también se incluyen en el término "anticuerpo" tal como se utiliza en el presente documento. Véase, por ejemplo, Paul, Fundamental Immunology 4^a ed., 1999, Raven Press, Nueva York, para la terminología y la estructura de los anticuerpos.

65 Por ejemplo, la invención incluye proteínas producidas utilizando los ARNt y/o las RS de la presente invención que se unen específicamente a o que son específicamente inmunorreactivas con un anticuerpo o antisuero

5 generado contra un inmunógeno que comprende una secuencia de aminoácidos. Para eliminar la reactividad cruzada con otros homólogos, el anticuerpo o antisuero se sustrae con la proteína disponible, tal como el polipéptido de tipo silvestre, por ejemplo, los polipéptidos "testigo". Cuando la proteína de tipo silvestre corresponde a un ácido nucleico, se genera un polipéptido codificado por el ácido nucleico y se utiliza para los fines de sustracción de anticuerpos/antisueros.

10 En un formato típico, el inmunoensayo utiliza un antisuero policlonal que se generó contra uno o más polipéptidos o una subsecuencia sustancial de los mismos (es decir, al menos aproximadamente un 30% de la secuencia de longitud completa proporcionada). El conjunto de inmunógenos polipetídicos potenciales provenientes de la proteína se denomina colectivamente en lo sucesivo "los polipéptidos inmunógenos". Los antisueros resultantes se seleccionan opcionalmente para que tengan baja reactividad cruzada contra los homólogos de sintetasa testigo y cualquier reactividad cruzada de este tipo se elimina, por ejemplo, mediante inmunoabsorción, con uno o más de los homólogos testigo, antes del uso del antisuero policlonal en el inmunoensayo.

15 Con el fin de producir antisueros para su uso en un inmunoensayo, se producen uno o más de los polipéptidos inmunógenos y se purifican como se describe en el presente documento. Por ejemplo, la proteína recombinante puede producirse en una célula recombinante. Se inmuniza una cepa endogámica de ratones (utilizadas en este ensayo porque los resultados son más reproducibles debido a la identidad genética virtual de los ratones) con la(s) proteína(s) inmunógena(s) en combinación con un adyuvante convencional, tal como el adyuvante de Freund, y un protocolo convencional de inmunización de ratones (véase, por ejemplo, Harlow y Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, Nueva York, para una descripción convencional de la generación de anticuerpos, formatos de inmunoensayos y las condiciones que pueden utilizarse para determinar la inmunorreactividad específica.

25 En el presente documento también se encuentran análisis sobre anticuerpos y referencias adicionales y pueden aplicarse en el presente documento a la definición de polipéptidos por inmunorreactividad. Como alternativa, se conjugan uno o más polipéptidos sintéticos o recombinantes provenientes de las secuencias descritas en el presente documento con una proteína transportadora y se utilizan como inmunógeno.

30 Se recogen los sueros policlonales y se titulan frente al polipéptido inmunógeno en un inmunoensayo, por ejemplo, un inmunoensayo en fase sólida con una o más de las proteínas inmunógenas inmovilizadas sobre un soporte sólido. Se seleccionan los antisueros policlonales con un título de 10^6 o mayor, se combinan y se sustraen con los polipéptidos sintetasa testigo para producir antisueros policlonales titulados combinados sustraídos.

35 Se someten a ensayo los antisueros policlonales titulados combinados sustraídos para determinar la reactividad cruzada frente a los homólogos testigo en un inmunoensayo comparativo. En este ensayo comparativo, se determinan las condiciones de unión discriminatorias para los antisueros policlonales titulados sustraídos que dan como resultado una relación señal-ruido al menos aproximadamente 5-10 veces mayor para la unión de los antisueros policlonales titulados a la proteína inmunógena en comparación a la unión a los homólogos de sintetasa testigo. Es decir, la rigurosidad de la reacción de unión se ajusta añadiendo competidores no específicos, tales como albúmina o leche desnatada en polvo, y/o ajustando las condiciones salinas, la temperatura, y/o similares. Estas condiciones de unión se utilizan en ensayos posteriores para determinar si un polipéptido de ensayo (un polipéptido que se compara con los polipéptidos inmunógenos y/o los polipéptidos testigo) se une específicamente a los antisueros policlonales sustraídos combinados.

45 En otro ejemplo, se utilizan inmunoensayos en el formato de unión competitiva para detectar un polipéptido de ensayo. Por ejemplo, como se ha señalado, los anticuerpos que presentan reactividad cruzada se eliminan de la mezcla de antisueros combinados mediante inmunoabsorción con los polipéptidos testigo. A continuación, se inmoviliza el polipéptido o polipéptidos inmunógenos en un soporte sólido que se expone a los antisueros combinados sustraídos. Se añaden al ensayo proteínas de ensayo para que compitan por la unión a los antisueros sustraídos combinados. La capacidad de la(s) proteína(s) de ensayo para competir por la unión a los antisueros sustraídos combinados en comparación con la(s) proteína(s) inmovilizada(s) se compara con la capacidad del polipéptido o polipéptidos inmunógenos añadidos al ensayo para competir por la unión (los polipéptidos inmunógenos compiten de manera eficaz con los polipéptidos inmunógenos inmovilizados por la unión a los antisueros combinados). Se calcula el porcentaje de reactividad cruzada para las proteínas de ensayo utilizando cálculos convencionales.

50 En un ensayo paralelo, se determina opcionalmente la capacidad de las proteínas testigo para competir por la unión de los antisueros sustraídos combinados en comparación con la capacidad del polipéptido o polipéptidos inmunógenos para competir por la unión a los antisueros. Una vez más, se calcula el porcentaje de reactividad cruzada para los polipéptidos testigo, utilizando cálculos convencionales. Cuando el porcentaje de reactividad cruzada es al menos 5 a 10 veces mayor para los polipéptidos de ensayo en comparación con los polipéptidos testigo y/o cuando la unión de los polipéptidos de ensayo se encuentra aproximadamente en el intervalo de la unión de los polipéptidos inmunógenos, se dice que los polipéptidos de ensayo se unen específicamente a los antisueros sustraídos combinados.

En general, pueden utilizarse los antisueros combinados e inmunoabsorbidos en un inmunoensayo de unión competitiva tal como se describe en el presente documento para comparar cualquier polipéptido de ensayo con el polipéptido o polipéptidos testigo o inmunógenos. Con el fin de realizar esta comparación, se someten a ensayo los polipéptidos inmunógenos, de ensayo y testigo en un amplio intervalo de concentraciones y se determina la cantidad de cada polipéptido requerida para inhibir el 50% de la unión de los antisueros sustraídos, por ejemplo, una proteína testigo, de ensayo o inmunógena inmovilizada, utilizando técnicas convencionales. Si la cantidad del polipéptido de ensayo requerida para la unión en el ensayo competitivo es inferior a dos veces la cantidad del polipéptido inmunógeno que se requiere, entonces se dice que el polipéptido de ensayo se une específicamente a un anticuerpo generado para la proteína inmunógena, siempre que la cantidad sea al menos aproximadamente 5 a 10 veces tan alta como para el polipéptido testigo.

Como determinación adicional de la especificidad, los antisueros combinados se inmunoabsorben opcionalmente totalmente con el polipéptido o péptidos inmunógenos (más que los polipéptidos testigo) hasta que resulte detectable poca o ninguna unión de los antisueros combinados sustraídos con polipéptido inmunógeno resultantes al polipéptido o péptidos inmunógenos utilizados en la inmunoabsorción. A continuación, se somete a ensayo este antisuero totalmente inmunoabsorbido para determinar la reactividad con el polipéptido de ensayo. Si se observa poca o ninguna reactividad (es decir, no más de 2 veces la relación señal-ruido observada para la unión de los antisueros totalmente inmunoabsorbidos al polipéptido inmunógeno), entonces el polipéptido de ensayo se une específicamente a los antisueros inducidos por la proteína inmunógena.

Pueden encontrarse detalles adicionales sobre proteínas, anticuerpos, antisueros, etc. en el documento USSN 10/825.867 titulado "Expanding the Eukaryotic Genetic Code", el documento WO 2002/085923, titulado "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; la patente de EE.UU. Nº 6.927.042 titulada "Glycoprotein synthesis", y la publicación de patente de EE.UU. Nº US 2004/0198637 titulada "Protein Arrays".

KITS

Los kits son también una característica de la presente invención. Por ejemplo, se proporciona un kit para producir una proteína que comprende al menos un aminoácido seleccionado, por ejemplo, un aminoácido no natural, en una célula, en el que el kit incluye un recipiente que contiene una secuencia de polinucleótido que codifica un O-ARNt, y/o un O-ARNt, y/o una secuencia de polinucleótido que codifica una O-RS, y/o una O-RS. En una forma de realización, el kit incluye adicionalmente al menos un aminoácido seleccionado. En otra forma de realización, el kit incluye un ARNt aminoacilado de la invención. En otra forma de realización, el kit comprende adicionalmente instrucciones para producir la proteína.

Un ejemplo adicional es un kit para producir una proteína que comprende al menos un aminoácido seleccionado, por ejemplo, un aminoácido no natural, en un sistema de traducción libre de células, en el que el kit incluye un recipiente que contiene una secuencia de polinucleótido que codifica un O-ARNt, y/o un O-ARNt, y/o una secuencia de polinucleótido que codifica una O-RS, y/o una O-RS. En una forma de realización, el kit incluye adicionalmente un aminoácido seleccionado. En otra forma de realización, el kit incluye un ARNt aminoacilado de la invención. En otra forma de realización, el kit comprende adicionalmente instrucciones para producir la proteína.

EJEMPLOS

Los siguientes ejemplos se ofrecen para ilustrar, pero no para limitar, la invención reivindicada. Un experto reconocerá diversos parámetros no críticos que pueden modificarse sin alejarse del alcance de la invención reivindicada.

Ejemplo 1:

Selección de una aminoacil-ARNt sintetasa contra la para-acetilfenilalanina

Se cribaron dos bibliotecas de ADN para detectar aminoacil-ARNt sintetasa contra para-acetilfenilalanina, un aminoácido codificado de forma no natural. Estas bibliotecas consistieron en seis mutaciones en el gen de la tirosil-ARNt sintetasa de *Methanococcus janneschii* en el plásmido pBK.

Se realizó el procedimiento de selección, que consistía en cinco rondas de selección alternas, tres positivas, dos negativas. Se combinaron las bibliotecas con una relación 1:1 y se sometieron a electroporación en la línea celular de selección positiva (GeneHog con plásmido de selección positiva, pPREP) y se sembraron en placas de medio mínimo (GMML) con los antibióticos apropiados y el aminoácido codificado de forma no natural para-acetilfenilalanina (pAF). Se incubaron las placas a 37°C durante aproximadamente 40 horas, momento en el que se recogieron las células por raspado. Se extrajo el ADN utilizando un procedimiento Mini-Prep de Qiagen, y a continuación se purificó en gel de agarosa para aislar el ADN de plásmido de la biblioteca.

A continuación, se sometió a electroporación este ADN en la línea celular de selección negativa (GeneHog con derivado de plásmido pBAD de selección negativa). Se sembraron estos transformantes en placas de LB con el

antibiótico adecuado y sin el aminoácido codificado de forma no natural (pAF). Después de aproximadamente 17 horas se recogieron estas células por raspado y se purificó el ADN del plásmido utilizando el procedimiento Mini-Prep de Qiagen y purificación en gel de agarosa.

5 Se realizaron las posteriores rondas de selección utilizando el mismo método de electroporación, sembrado, recolección y purificación de ADN. En la última (quinta) ronda de selección, se realizaron diluciones en serie de las células de selección positiva transformadas que se sembraron en placas de medio mínimo. A
10 continuación se recogieron colonias individuales y se cultivaron en un bloque de 96 pocillos durante toda la noche. A continuación se realizó la réplica en placa de este bloque en placas de medio mínimo con concentraciones variables de cloranfenicol (el antibiótico de selección positiva) con y sin aminoácido no natural pAF. Después de
15 aproximadamente 40 horas de crecimiento a 37°C, se compararon visualmente las placas para determinar qué colonias crecían a la concentración más alta de cloranfenicol pero no crecían o crecían poco en ausencia del aminoácido codificado de forma no natural pAF. Las colonias que cumplían estos criterios se cultivaron durante toda la noche. Se aisló el ADN de los cultivos mediante Mini-Prep y purificación en gel de agarosa, y se secuenciaron.

A partir de esta selección de pAF, se descubrió que 13 clones tenían secuencias de aminoácidos únicas y se sometieron a caracterización adicional para determinar la fidelidad y la procesividad de la pAF-ARNt sintetasa.

20 Para caracterizar estas sintetasas, se realizaron supresiones ámbar a pequeña escala para demostrar que el aminoácido codificado de forma no natural pAF se incorporaba a un polipéptido, y se visualizaron los resultados mediante SDS-PAGE. Se recogió una sola colonia y se cultivó durante toda la noche en caldo LB, que a continuación se utilizó para inocular 50 ml de LB. Se cultivaron las células hasta una DO de 0,3-0,4, momento en el
25 que se tomaron alícuotas de 1,5 ml como puntos de pre-inducción y se separó el cultivo en dos matraces. Se añadió pAF 1 mM a uno de los dos y se cultivaron ambos durante 30 minutos. Tras un crecimiento de 30 minutos, se indujeron ambos cultivos (+/- pAF) con L-arabinosa al 0,2% y se cultivaron 4,5 horas, y se registró la DO₆₀₀. A continuación se tomaron alícuotas de 1,5 ml de los matraces +/- pAF para el análisis mediante SDS-PAGE.

30 Se centrifugaron las alícuotas de 1,5 ml (Preinducción, + pAF, - pAF) a 10.000 xg durante 10 minutos para sedimentar las células. A continuación, se suspendieron las células en cantidades proporcionales de Bacterial Protein Extraction Reagent (BPER, Pierce) con respecto a su DO₆₀₀ en el momento de la recolección. Se añadió ADNasa I a las células lisadas y se incubaron a 4°C durante 20 minutos. A continuación se combinaron las muestras con un agente reductor y un colorante de carga y se corrieron en un gel de Bis-Tris al 4%-12% en tampón MES durante 30 minutos. Se lavó el gel dos veces en H₂O DI durante 10 minutos y se tiñó con colorante azul de
35 Coomassie. Se compararon las bandas +/- pAF para determinar la fidelidad de la pAF-ARNt RS para dar como resultado la incorporación de la pAF, y se comparó la banda +pAF con la pAF-ARNt RS anteriormente seleccionada.

40 Para comprobar la capacidad de procesamiento de las RS, se realizó el mismo procedimiento con un plásmido que contenía mioglobina C-H6 S4am (S4am-Myo). A continuación se purificó la S4am Myo mediante IMAC y se envió para la secuenciación de proteínas para determinar la cantidad de incorporación de pAF.

45 De las pAF-ARNt RS identificadas a partir de esta selección, se descubrió que una sintetasa (E9) incorporaba de manera eficaz pAF, con una eficacia de incorporación de la PAF en S4am-Myo superior a un 95%. Se determinó la incorporación mediante secuenciación de aminoácidos, mientras que la procesividad se demostró comparando las bandas de proteínas en geles de SDS-PAGE. La secuencia de nucleótidos para E9 se muestra en la SEQ ID NO:4, y la secuencia de aminoácidos de E9 se muestra en la SEQ ID NO:5.

Se identificó un mutante adicional con actividad similar a E9, y tiene la secuencia de aminoácidos mostrada en la SEQ ID NO:17.

50 **Ejemplo 2:**

Mutagénesis de ARNt

55 Se generaron tres mutantes de ARNt J17. La secuencia de ADN del J17 de tipo silvestre se muestra como SEQ ID NO:8 y en la publicación de patente de EE.UU. N° 2003/0108885 como SEQ ID NO:1 y en la EE.UU. 2003/0082575 como SEQ ID NO:1 (solicitud de patente de EE.UU. con N^{os} de serie 10/126.931 y 10/126.927, respectivamente). El ARNt de J17 tiene un par oscilante U51:G63 en el tallo de TΨC como se muestra en la Figura 1.

60 Se generaron tres mutantes de J17 (F12, F13, y F14) para producir pares de bases de Watson-Crick en las posiciones 51 y 63 del tallo de TΨC. Se realizó la mutagénesis mediante PCR solapante, y se clonaron los constructos finales en los sitios EcoRI y NdeI en un plásmido pET19 que comprendía la secuencia de polinucleótido que codifica la aminoacil-ARNt sintetasa E9 (SEQ ID NO:4) y la secuencia de polinucleótido que codifica la hormona de crecimiento humana (hGH) con una sustitución en el codón ámbar (SEQ ID NO:16). La expresión de la hGH estaba bajo el control del promotor de T7.
65

Se generaron dos fragmentos para la PCR solapante. El primer fragmento se obtuvo mediante extensión del cebador. La secuencia del cebador directo utilizado para generar cada uno de los tres mutantes era:

5
GTAACGCTGAATTCCCGCGGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAAT
CCGCATGGCGC (FTam11; SEQ ID NO: 9).

Para generar el mutante F12 (51C:63G), se utilizó el siguiente cebador inverso:

10
GATCTGCAGTGGTCCGGCGGGCCGGATTTGAACCGGCGCCATGCGGATTTAGAG
TCCGCCGTTCTGC (FTam12; SEQ ID NO: 10).

Para generar el mutante F13 (51U:63A), se utilizó el siguiente cebador inverso:

15
GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACCAGCGCCATGCGGATTTAGAGT
- CCGCCGTTCTGC (FTam13; SEQ ID NO: 11).

20
 Para generar el mutante F14 (51A:63U), se utilizó el siguiente cebador inverso:

25
GATCTGCAGTGGTCCGGCGGGCAGGATTTGAACCTGCGCCATGCGGATTTAGAGT
CCGCCGTTCTGC (FTam14; SEQ ID NO: 12).

30
 Para generar el segundo fragmento, se utilizó como molde el plásmido pET19 J17 E9 hGH que comprende la secuencia de polinucleótido para el ARNt de J17 (SEQ ID NO:8), la secuencia de polinucleótido que codifica la ARNt sintetasa E9 (SEQ ID NO:4) y la secuencia de polinucleótido que codifica la hormona de crecimiento humana con una sustitución en el codón ámbar (SEQ ID NO:16), para la amplificación con el siguiente conjunto de cebadores:

35
CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTAAGC (cebador directo; FTam15; SEQ ID NO:13) y **CAAATTCGTCCATATGGGATTCC** (FTAM 16; SEQ ID NO:14). Se utilizó el cebador directo para prolongar la secuencia desde el extremo 3' del ARNt hasta el sitio Nde I del plásmido. El producto resultante se purificó en gel.

40
 La etapa final de la PCR solapante implicaba al cebador directo **GTAACGCTGAATTC**CCGCGG (FTam17, SEQ ID NO:15), al cebador inverso FTam16 (SEQ ID NO:14), al primer fragmento y al segundo fragmento. Los productos ensamblados se digirieron con EcoR I y Nde I y se ligaron en el plásmido pET19 J17 E9 hGH digerido con EcoR I y Nde I. La secuencia de cada constructo se confirmó mediante secuenciación, y las secuencias de ADN para cada uno de los ARNt mutantes de J17 se muestran como SEQ ID NO:1 (F12), SEQ ID NO:2 (F13), y SEQ ID NO:3 (F14). Los ARNt llevan el nombre de sus cebadores inversos correspondientes.

45 **Expresión de proteínas**

50
 Se transformó cada uno de los plásmidos que codifican los ARNt (J17, F12, F13 o F14) en las células hospedadoras bacterianas *E. coli* cepa 1 y cepa 2 mediante medios químicos y se sembraron en placas de agar LB con 50 ug/ml de carbenicilina. Se incubaron las placas a 37°C durante toda la noche. Para cada ARNt, se recogió una sola colonia para iniciar un cultivo de una noche a 37°C en 1 ml de 2xYT con 50 ug/ml de carbenicilina. Este cultivo de 1 ml se utilizó para inocular dos cultivos de 10 ml de 2xYT con 50 ug/ml de carbenicilina a 37°C. Un cultivo de 10 ml se complementó con para-acetilfenilalanina 4 mM. A una DO₆₀₀ = 0,7, se indujo la expresión de hGH con IPTG 0,4 mM. Después de cultivar las células a 37°C durante 4 horas a 250 rpm, se recogieron las células por centrifugación a 5.000 xg durante 5 minutos. Se lisaron las células con reactivo B-PER (Pierce, Rockford, IL) complementado con 5 ug/ml de ADNasa I. Se analizó el lisado total de células mediante SDS PAGE al 4%-12%.

55
 La Figura 2 muestra un análisis de lisados totales de células de *E. coli* cepa 1 mediante SDS-PAGE. Se realizó la supresión de un codón selector en la hormona de crecimiento humana utilizando ARNt de J17 o de mutante de J17 (F12, F13, F14) y la aminoacil-ARNt sintetasa E9. Las células que albergaban mutantes de J17 crecieron ligeramente más despacio que las células que albergaban J17. No se observó ningún producto hGH de longitud completa mediante SDS-PAGE para los mutantes de ARNt en ausencia de para-acetilfenilalanina 4 mM. En presencia de para-acetilfenilalanina 4 mM, se producía el producto de longitud completa con cada uno de los mutantes de ARNt, lo que demuestra que estos pares mutantes de ARNt-RS E9 son ortogonales a la maquinaria de *E. coli*. En base al SDS-PAGE, el rendimiento de hGH suprimido de los mutantes de J17 fue aproximadamente 1,5~2 veces mayor que el de J17 en *E. coli* cepa 1.

60
 Se sometió a ensayo adicionalmente un mutante de J17, F13, en la línea celular bacteriana *E. coli* cepa 2

para la supresión ámbar tal como se muestra en la Figura 3. En *E. coli* cepa 2, los rendimientos de la supresión ámbar así como de la expresión se redujeron con respecto a los de *E. coli* cepa 1. En ausencia de *para*-acetilfenilalanina, no se observó ningún producto hGH de longitud completa mediante SDS-PAGE. En presencia de *para*-acetilfenilalanina 4 mM, se observó hGH de longitud completa para ambos ARNt. En base al SDS-PAGE, el rendimiento de hGH suprimido de F13 fue aproximadamente tres veces superior al de J17.

Se realizó una operación de fermentación que comparaba J17 y F13 con un volumen final de aproximadamente 1,5 l. El plásmido que codifica el ARNt de J17 y el plásmido que codifica el ARNt de F13 se transformaron en *E. coli* cepa 1. La densidad celular final de cada uno fue aproximadamente 190 g de células húmedas/l. El título de hGH fue 347 mg/l para el clon J17 y 542 mg/l para el clon F13.

5
10
15
20
25
30
35
40
45
50
55
60
65

TABLA 1

SEQ ID NO:	MARCADOR	SECUENCIA
1	ADN de F12	CCGGCGGTAGTTCAGCAGGGCAGAAACGGCGGACTCTAAATCCGCA TGGCGCCGGTTCAAATCCGGCCCGCCGACCA
2	ADN de F13	CCGGCGGTAGTTCAGCAGGGCAGAAACGGCGGACTCTAAATCCGCA TGGCGCTGGTTCAAATCCAGCCCGCCGACCA
3	ADN de F14	CCGGCGGTAGTTCAGCAGGGCAGAAACGGCGGACTCTAAATCCGCA TGGCGCAGGTTCAAATCCTGCCCCCGGACCA
4	ácido nucleico de RS de E9	ATGGACGAATTTGAAAATGATAAAGAGAAACACATCTGAAAATTATC AGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAATC TGCTGTATAGGTTTTGAAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCAAATAAAAAGATGATTTGATTTACAAAATGCTGGATTTG ATATAATTATATATTTGGCTGATTTACACGCCCTATTTAAACCAGAA AGGAGATTGGATGAGATTAGAAAATAGGAGATTATAACAAAA AAGTTTTGAAGCAATGGGTTAAAGGCAAAAATATGTTTATGGAA GTGAACATGGTCTTGATAAGGATTATACACTGAAATGCTATAGATT GGCTTTAAAAAACTACCTTAAAAAGAGCAAGAGGAGTATGGAACT TATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTGAAGTTATCTA TCCAATAATGCAGGTTAATGGGATTCATTTATGAGGGCGTTGATGTT GCAGTTGGAGGATGGAGCAGAGAAAATACACATGTTAGCAAAG GGAGCTTTTACCAAAAGGTTGTTTGTATTTCACAACCCCTGCTTAA ACGGGTTTGGATGGAGAAGGAAAGATGAGTTCTTCAAAGGGAAT TTTATAGCTGTTGATGACTCTCCAGAGAGATTAGGGCTAAGATAA AGAAAGCATACTGCCAGCTGGAGTTGTTGAAGGAAAATCCAATAA TGGAGATAGCTAAATACCTTCCCTTGAATATCCCTTAAACCATAAAAG GCCAGAAAATTTGGTGGAGATTTGACAGTTAAATAGCTATGAGGA GTTAGAGAGTTTATTTAAAAATAAGGAATTGCATCCCAATGGATTTA AAAAATGCTGTAGCTGAAGAACCTTATAAAGATTTTAGAGCCCAATT AGAAAGAGATTATAA

5
10
15
20
25
30
35
40
45
50
55
60
65

SEQ ID NO:	MARCADOR	SECUENCIA
5	aminoácido de RS de E9	MDEFEMIKRNTSEIIEELREVLKKDEKSAVIGFEPGKIHLGHYLLQIK KMIDLQNAQFDIIYLA DLHAYLNQKGEIDRIRKIGDYNNKKVFEAMGL KAKYVYGSEHGLDKDYTLNVYRLALKTTLKRARRSMELIAREDENP KVAEVIYPIMQVNGHYEGVDVAVGGMEQRKIHMLARELLPKKVVCI HNPVLTGLDGEKMSKGNFIAVDDSPPEIRAKIKKAYCPAGVVEG NPIMEIAKYFEYPLTIKRPEKFGDDLTVNSYEELSLFKNKELHPMDL KNVAEELKILEPIRKRRL
6	ADN de ARNt de HL(TGA)3	CCAGGGTAGCCAAAGCTCGGCCAACGGCGACGGACTCTAAATCCG TTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTGGGACCA
7	ADN de ARNt de HL(TGA)1	GCGGGGTTGCCGAGCCTGGCCAAAAGGGCGCGGACTTCAAATCCG GTCCCCGTAGGGGTTCCGGGGTTCAAATCCCCGCCCGCCGACCA
8	ARNtm Tyr de <i>M. jaranaschii</i> J17 CUA	CCGGCGGTAGTTTACGACGGCAGAACGGCGGACTCTAAATCCGCA TGCGCTGGTTCAAATCCGGCCCGCCGACCA
9	cebador FTam11	GTAACGCTGAATCCCGGGGTTAGTTCAGCAGGGCAGAACGGCGCG ACTCTAAATCCGCATGGCGC
10	cebador FTam12	GATCTGCAGTGGTCCGGGGGGCCGGATTGAAACCGGGCCCATGCG GATTTAGAGTCCCGCCGTTCTGC
11	cebador FTam13	GATCTGCAGTGGTCCGGGGGGCTGGATTGAAACCGGGCCCATGCG GATTTAGAGTCCCGCCGTTCTGC
12	cebador FTam14	GATCTGCAGTGGTCCGGGGGGCAGGATTGAAACCGGGCCCATGCG GATTTAGAGTCCCGCCGTTCTGC
13	cebador FTam15	CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTA AGC
14	cebador FTam16	CAAAATTCGTCATATGGGATTCC
15	cebador FTam17	GTAACGCTGAATCCCGGCG

5
10
15
20
25
30
35
40
45
50
55
60
65

SEQ ID NO:	MARCADOR	SECUENCIA
16	hGH (ADN)	ATGGGCCACCAACCAACCACTTCCCAACCAATCCCTTATCCA GGCTTTTGACAACGCTATGCTCCGCGCCCATCGTCTGCACCAGCT GGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTAGATCCCAA GGAACAGAAGTATTCATTCTGAGAACCCAGACCTCCCTCTGT TTCTCAGAGTCTATTCCGACACCCCTCCAAACAGGGAGGAACACAA CAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGTGCTCATCC AGTCGTGGCTGGAGCCCGTGAGTTCCTCAGGAGTGTCTTCGCCAA CAGCCTGTGTACGGCGCCTGTGACAGCAACGCTATGACCTCCTA AAGGACCTAGAGGAAGGCATCCAAACGCTGATGGGGAGGCTGGA AGATGGCAGCCCGGACTGGGACAGATCTTCAAGCAGACCTACAG CAAGTTCGACACAACACTCACACAACGATGACGCACACTCAAGAA CTACGGGCTGCTCTACTGCTTCAGGAAGGACATGGACAAGGTCGA GACATTCCTGGGCATCGTGCAGTGGCCGCTCTGTGGAGGGCAGCTGT GGCTTCTAA
17	mutante de E9 D286R	MDEFEMIKRN TSEIIEEEL REVLKKDEKS AVIGFEPGK IHLGHYLQIK KMIDLQNAGF DIIHLYADLH AYLNQKGELD EIRKIGDYNK KVFEAMGLKA KYVYGEHGL DKDYTLNVYR LALKTTLKRA RRSMEIARE DENPKVAEVI YPIMQVNGIH YEGVDVAVGG MEQRKIHMLA RELPKKVVC IHNPLVLTGLD GEGKMSSSKG NFLAVDDSP EIRAKIKKAY CPAGVVEGNP IMEIAKYFLE YPLTIKRPEK FGGDLTVNSY EELESFKNK ELHPMRLKNA VAEELIKILE PIRKRL

LISTADO DE SECUENCIAS

	<110> Tian, Feng	
5	<120> Composiciones de ARNt y usos de las mismas	
	<130> AMBX-0069.00PCT	
10	<150> US 60/709,364	
	<151> 2005-08-18	
	<160> 17	
15	<170> Patent In versión 3.3	
	<210> 1	
	<211> 77	
	<212> ADN	
20	<213> Secuencia Artificial	
	<220>	
	<223> Secuencia Artificial	
25	<400> 1	
	ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg cgggttcaaa	60
	tccggccccgc cggacca	77
30	<210> 2	
	<211> 77	
	<212> ADN	
	<213> Secuencia Artificial	
35	<220>	
	<223> Secuencia Artificial	
	<400> 2	
40	ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa	60
	tccagccccgc cggacca	77
45	<210> 3	
	<211> 77	
	<212> ADN	
	<213> Secuencia Artificial	
50	<220>	
	<223> Secuencia Artificial	
	<400> 3	
55	ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg cagggttcaaa	60
	tccagccccgc cggacca	77
60	<210> 4	
	<211> 921	
	<212> ADN	
	<213> Secuencia Artificial	
65	<220>	
	<223> Secuencia Artificial	
	<400> 4	

ES 2 408 581 T3

atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60
 5 agagaggttt taaaaaaga tgaaaaatct gctgttatag gttttgaacc aagtggtaaa 120
 atacatttag ggcattatct ccaaataaaa aagatgattg atttacaaaa tgctggattt 180
 gatataatta tatatttggc tgatttacac gcctatttaa accagaaagg agagttggat 240
 10 gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300
 aaatatgttt atggaagtga acatggtctt gataaggatt atacactgaa tgtctataga 360
 15 ttggctttaa aaactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420
 gatgaaaatc caaagggtgc tgaagttatc tatccaataa tgcaggtaa tgggattcat 480
 tatgagggcg ttgatgttgc agttggaggg atggagcaga gaaaaataca catgtagca 540
 20 agggagcttt taccaaaaaa ggtgtttgt attcacaacc ctgtcttaac gggtttgat 600
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctggtgatga ctctccagaa 660
 25 gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720
 ataatggaga tagctaaata ctctctttaa tctctttaa ccataaaaag gccagaaaaa 780
 tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840
 30 gaattgcac ccaatggattt aaaaaatgct gtagctgaag aacttataaa gatttttagag 900
 ccaattagaa agagattata a 921

35 <210> 5
 <211> 306
 <212> PRT
 <213> Secuencia Artificial

40 <220>
 <223> Secuencia Artificial

<400> 5

45 Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
 1 5 10 15

50 Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val

55

60

65

ES 2 408 581 T3

<211> 67
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> Secuencia Artificial

<400> 10

10 gatctgcagt ggtccggcgg gccggatttg aaccggcgcc atgcggattt agagtccgcc 60
 gttctgc 67

<210> 11
 <211> 67
 <212> ADN
 <213> Secuencia Artificial

20 <220>
 <223> Secuencia Artificial

<400> 11

25 gatctgcagt ggtccggcgg gctggatttg aaccagcgcc atgcggattt agagtccgcc 60
 gttctgc 67

<210> 12
 <211> 67
 <212> ADN
 <213> Secuencia Artificial

30 <220>
 <223> Secuencia Artificial

35 <400> 12

40 gatctgcagt ggtccggcgg gcaggatttg aacctgcgcc atgcggattt agagtccgcc 60
 gttctgc 67

<210> 13
 <211> 49
 <212> ADN
 <213> Secuencia Artificial

45 <220>
 <223> Secuencia Artificial

50 <400> 13
 cgccggacca ctgcagatcc ttagcgaaag ctaaggattt ttttaagc 49

<210> 14
 <211> 23
 <212> ADN
 <213> Secuencia Artificial

55 <220>
 <223> Secuencia Artificial

60 <400> 14
 caaatcgtc catatgggat tcc 23

65 <210> 15
 <211> 20
 <212> ADN

ES 2 408 581 T3

<213> Secuencia Artificial

<220>

<223> Secuencia Artificial

5 <400> 15
gtaacgctga attccggcg 20

10 <210> 16
<211> 600
<212> ADN
<213> *Homo sapiens*

15 <400> 16

atgggccacc	accaccacca	ccacttccca	accattccct	tatccagget	tttgacaac	60
gctatgctcc	gcgccatcg	tctgcaccag	ctggcctttg	acacctacca	ggagtttgaa	120
gaagcctaga	toccaaagga	acagaagtat	tcattcctgc	agaacccccca	gacctccctc	180
tgtttctcag	agtctattcc	gacacctcc	aacagggagg	aaacacaâca	gaaatccaac	240
ctagagctgc	tcgcatctc	cctgctgctc	atccagtcgt	ggctggagcc	cgtgcagttc	300
ctcaggagtg	tcttcgcaa	cagcctgggtg	tacggcgct	ctgacagcaa	cgtctatgac	360
ctcctaaagg	acctagagga	aggcatccaa	acgctgatgg	ggaggctgga	agatggcagc	420
ccccggactg	ggcagatctt	caagcagacc	tacagcaagt	tgcacacaaa	ctcacacaac	480
gatgacgcac	tactcaagaa	ctacgggctg	ctctactget	tcaggaagga	catggacaag	540
gtcgagacat	tcctgcgcat	cgtgcagtgc	cgctctgtgg	agggcagctg	tggcttctaa	600

<210> 17
<211> 306
<212> PRT

40 <213> Secuencia Artificial

<220>

<223> Secuencia Artificial

45 <400> 17

50

55

60

65

ES 2 408 581 T3

1 Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
 5 Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val
 10 Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
 15 Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
 20 Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
 25 Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
 30 Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys
 35 Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
 40 Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
 45 Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His
 50 Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile
 55 His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His

ES 2 408 581 T3

				180					185					190			
5	Asn	Pro	Val	Leu	Thr	Gly	Leu	Asp	Gly	Glu	Gly	Lys	Met	Ser	Ser	Ser	
			195					200					205				
10	Lys	Gly	Asn	Phe	Ile	Ala	Val	Asp	Asp	Ser	Pro	Glu	Glu	Ile	Arg	Ala	
		210					215					220					
15	Lys	Ile	Lys	Lys	Ala	Tyr	Cys	Pro	Ala	Gly	Val	Val	Glu	Gly	Asn	Pro	
	225					230					235					240	
20	Ile	Met	Glu	Ile	Ala	Lys	Tyr	Phe	Leu	Glu	Tyr	Pro	Leu	Thr	Ile	Lys	
					245					250					255		
25	Arg	Pro	Glu	Lys	Phe	Gly	Gly	Asp	Leu	Thr	Val	Asn	Ser	Tyr	Glu	Glu	
				260					265					270			
30	Leu	Glu	Ser	Leu	Phe	Lys	Asn	Lys	Glu	Leu	His	Pro	Met	Arg	Leu	Lys	
			275					280					285				
35	Asn	Ala	Val	Ala	Glu	Glu	Leu	Ile	Lys	Ile	Leu	Glu	Pro	Ile	Arg	Lys	
		290					295					300					
40	Arg	Leu															
	305																
45																	
50																	
55																	
60																	
65																	

REIVINDICACIONES

- 5 1. Composición que comprende un ARNt codificado por la secuencia de polinucleótido de la: SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; una secuencia de polinucleótido complementaria de la SEQ ID NO:1; una secuencia de polinucleótido complementaria de la SEQ ID NO:2; o una secuencia de polinucleótido complementaria de la SEQ ID NO:3.
2. Composición según la reivindicación 1, en la que dicho ARNt está aminoacilado.
- 10 3. Composición según la reivindicación 2, en la que dicho ARNt está aminoacilado con un aminoácido codificado de forma no natural.
- 15 4. Composición según la reivindicación 3, en la que dicho aminoácido codificado de forma no natural es para-acetilfenilalanina.
5. Composición según la reivindicación 2, en la que dicho ARNt está aminoacilado químicamente.
6. Composición según la reivindicación 2, en la que dicho ARNt está aminoacilado enzimáticamente.
- 20 7. Composición según la reivindicación 6, en la que dicho ARNt es aminoacilado enzimáticamente por una RS o por una ribozima.
- 25 8. Composición según la reivindicación 1, que comprende adicionalmente una RS, en la que el ARNt está aminoacilado con un aminoácido.
9. Composición según la reivindicación 8, en la que dicho aminoácido es un aminoácido codificado de forma no natural.
- 30 10. Composición según la reivindicación 9, en la que dicho aminoácido codificado de forma no natural es para-acetilfenilalanina.
11. Composición según la reivindicación 1, en la que el ARNt proviene de un ARNt de *Archael*.
- 35 12. Composición según la reivindicación 1, en la que el ARNt proviene de *M. janneschii*.
13. Composición según la reivindicación 1, que comprende adicionalmente un sistema de traducción.
- 40 14. Composición según la reivindicación 13, en la que dicho sistema de traducción es un sistema de traducción libre de células.
15. Composición según la reivindicación 13, en la que dicho sistema de traducción es un lisado de células.
- 45 16. Composición según la reivindicación 13, en la que dicho sistema de traducción es un sistema reconstituido.
17. Composición según la reivindicación 13, en la que dicho sistema de traducción es un sistema de traducción celular.
- 50 18. Célula que comprende un sistema de traducción, en la que el sistema de traducción comprende el ARNt según la reivindicación 1.
19. Célula según la reivindicación 18, en la que la célula es una célula eucariota.
- 55 20. Célula según la reivindicación 19, en la que la célula eucariota es una célula de levadura, una célula fúngica, una célula de mamífero, una célula de insecto, o una célula vegetal.
21. Célula según la reivindicación 18, en la que la célula es una célula no eucariota.
22. Célula según la reivindicación 21, en la que la célula no eucariota es una célula de *E. coli*.
- 60 23. Célula según la reivindicación 18, que comprende adicionalmente un polinucleótido que codifica un polipéptido de interés, en la que el polinucleótido comprende un codón selector que es reconocido por el ARNt.
24. Célula según la reivindicación 23, en la que dicho polipéptido de interés es la hormona de crecimiento humana.
- 65 25. Célula que comprende el ARNt según la reivindicación 1, en la que dicha célula es una célula de *E. coli*, una célula de levadura, una célula fúngica, una célula de mamífero, una célula de insecto o una célula vegetal.

- 5 26. Vector que comprende un polinucleótido que codifica un ARNt, en el que dicho polinucleótido tiene la secuencia de ácido nucleico de la: SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3; una secuencia de polinucleótido complementaria de la SEQ ID NO:1; una secuencia de polinucleótido complementaria de la SEQ ID NO:2; o una secuencia de polinucleótido complementaria de la SEQ ID NO:3.
27. Vector según la reivindicación 26, en el que el vector comprende un plásmido, un cósmido, un fago, o un virus.
- 10 28. Vector según la reivindicación 26, en el que el vector es un vector de expresión.
29. Célula que comprende el vector según la reivindicación 26.
- 15 30. Método para producir un polipéptido en una célula con un aminoácido seleccionado en una posición determinada, comprendiendo el método:
cultivar la célula en un medio apropiado, en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una polipéptido; y proporcionar el aminoácido seleccionado; en el que la célula comprende adicionalmente:
- 20 un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector, en el que dicho O-ARNt es codificado por la secuencia de polinucleótido de la: SEQ ID NO:1; SEQ ID NO:2; SEQ ID NO:3, una secuencia de polinucleótido complementaria de la SEQ ID NO:1; una secuencia de polinucleótido complementaria de la SEQ ID NO:2; o una secuencia de polinucleótido complementaria de la SEQ ID NO:3,
- 25 una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que dicha O-RS aminoacila el O-ARNt con el aminoácido seleccionado.
31. Método según la reivindicación 30 en el que dicho aminoácido seleccionado es para-acetilfenilalanina.
- 30 32. Método según la reivindicación 30 en el que dicho polipéptido es la hormona de crecimiento humana.

Figura 1

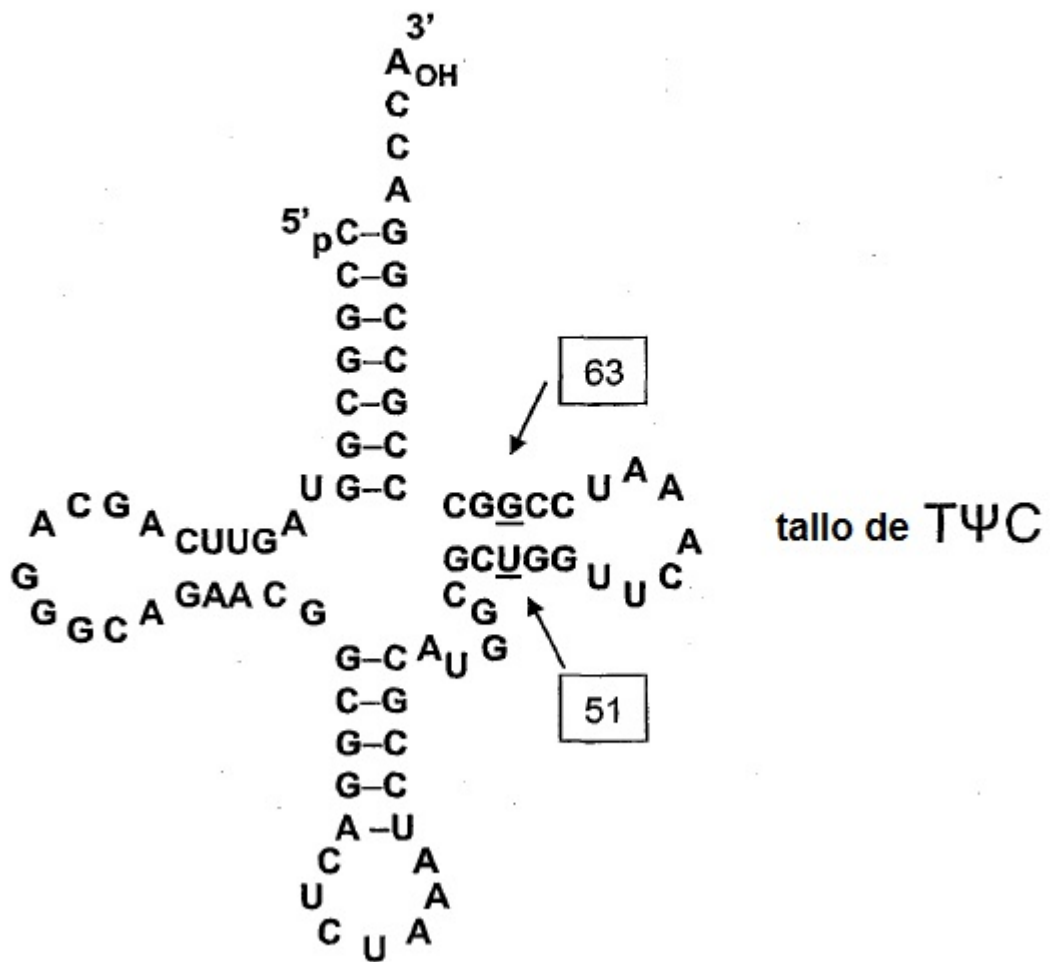


Figura 2

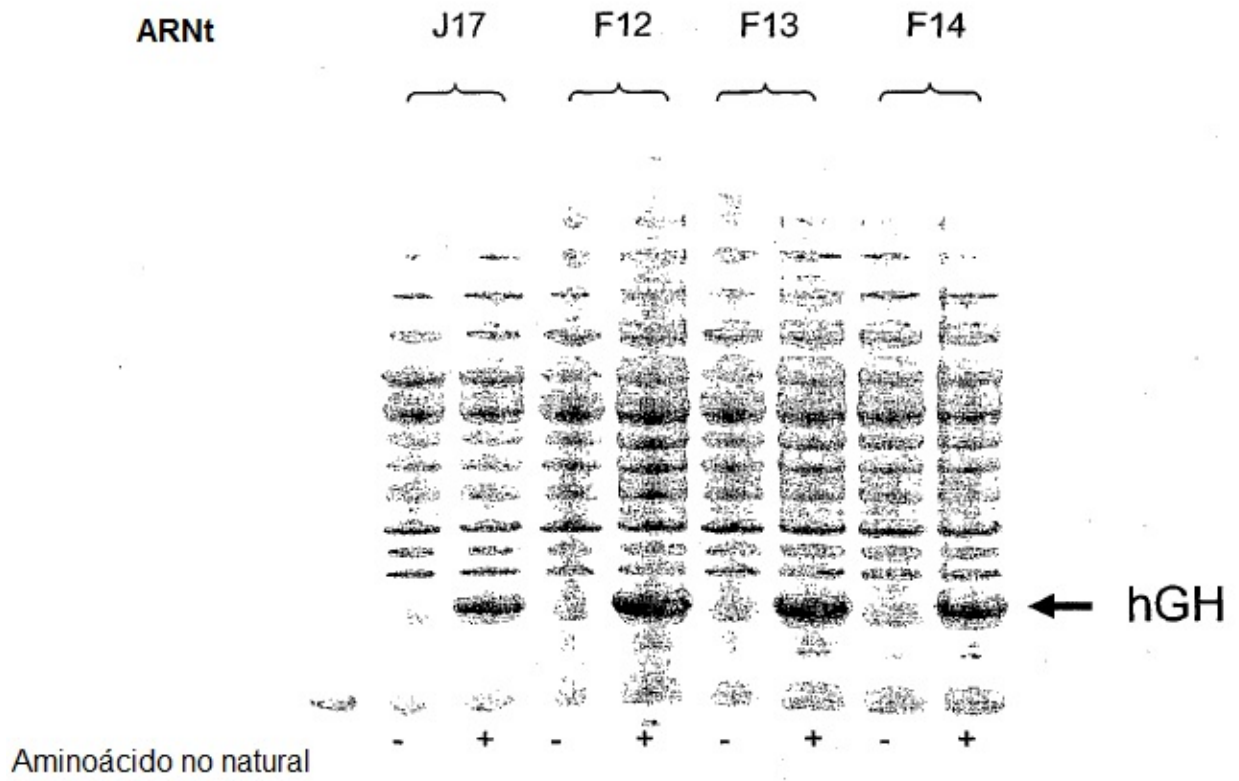


Figura 3

