

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 519 444**

51 Int. Cl.:

C12N 9/22 (2006.01)

C12N 9/10 (2006.01)

C12N 15/12 (2006.01)

C12N 15/54 (2006.01)

C12P 21/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.09.2005 E 05812008 (0)**

97 Fecha y número de publicación de la concesión europea: **06.08.2014 EP 1794323**

54 Título: **Incorporación in vivo de alquil aminoácidos a proteínas en eubacterias**

30 Prioridad:

21.09.2004 US 612220 P

24.11.2004 US 630876 P

07.12.2004 US 634151 P

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

07.11.2014

73 Titular/es:

THE SCRIPPS RESEARCH INSTITUTE (100.0%)

10550 NORTH TORREY PINES ROAD

LA JOLLA, CA 92037, US

72 Inventor/es:

DEITERS, ALEXANDER y

SCHULTZ, PETER

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 519 444 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Incorporación *in vivo* de alquil aminoácidos a proteínas en eubacterias

5 Campo de la invención

La invención pertenece al campo de la bioquímica de la traducción. La invención se refiere a composiciones y a métodos para la preparación y uso de ARNt ortogonales, aminoacil-ARNt sintetasas ortogonales, y sus pares, que incorporan alquil aminoácidos a proteínas. La invención también se refiere a métodos de producción de proteínas en células usando dichos pares y composiciones relacionadas.

Antecedentes de la invención

La capacidad para modificar químicamente proteínas en lugares específicos con moléculas no peptídicas tales como sondas espectroscópicas, adyuvantes catalíticos, o polímeros, o para entrecruzar covalentemente una proteína a otra proteína o cualquier otro resto, proporciona un poderoso medio tanto para investigar como para manipular las propiedades químicas y biológicas de las proteínas. Un enfoque habitual supone la bioconjugación de restos de superficie nucleófila sobre la proteína, por ejemplo, las cadenas laterales de lisina, histidina, o cisteína, con grupos electrófilos sobre una molécula exógena, tales como aldehídos, α -halo carboxamidas, y N-hidroxi succinimidas (Lemineux, G. A.; Bertozzi, C. R. TIBTECH 1996, 16, 506).

Desgraciadamente, un desafío al usar las dianas nucleófilas de origen natural en una proteína para modificaciones dirigidas es la poca selectividad de estas reacciones y la aparición de múltiples aminoácidos nucleófilos en proteínas, lo que da lugar a la formación de mezclas heterogéneas de proteínas marcadas. Además, las reacciones de modificación selectivas para nucleófilos con frecuencia requieren condiciones no fisiológicas, que pueden limitar las estrategias de modificación *in vivo* y/o producir la pérdida de actividad biológica de la proteína.

Existe una necesidad en la técnica de crear nuevas dianas y nuevas estrategias para las modificaciones específicas y dirigidas de proteínas. Desgraciadamente, cada organismo conocido, desde bacterias a seres humanos, codifica los mismos 20 aminoácidos habituales (con las excepciones raras de la selenocisteína (véase, por ejemplo, A. Bock y col., (1991), Molecular Microbiology 5:515-20) y la pirrolisina (véase, por ejemplo, G. Srinivasan, y col., (2002), Science 296:1459-62). Esta característica limita el uso de aminoácidos naturales en el desarrollo de nuevas químicas para la modificación dirigida de proteínas.

Una estrategia para superar esta limitación es expandir el código genético y añadir aminoácidos que tengan propiedades químicas distintas a las del repertorio biológico. Este enfoque se ha demostrado factible con el uso de ARNt "ortogonales" y las nuevas aminoacil-ARNt sintetasas "ortogonales" correspondientes para añadir aminoácidos no naturales a proteínas que usan la maquinaria biosintética de proteínas *in vivo* de la eubacteria *Escherichia coli* (*E. coli*) y otros organismos (por ejemplo, Wang y col., (2001), Science 292:498-500; Chin y col., (2002) Journal of the American Chemical Society 124:9026-9027; Chin y Schultz, (2002), ChemBioChem 11:1135-1137; Chin, y col., (2002), PNAS United States of America 99:11020-11024; y Wang y Schultz, (2002), Chem. Comm., 1-10). Véanse también, las Publicaciones internacionales WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS"; WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE"; WO 2005/019415, presentada el 7 de julio de 2004; WO 2005/007870, presentada el 7 de julio de 2004; y WO 2005/007624, presentada el 7 de julio de 2004.

Existe una necesidad en la técnica de nuevos métodos para conseguir modificaciones de proteínas altamente específicas y dirigidas. Existe una necesidad en la técnica del desarrollo de componentes de traducción ortogonales que incorporen aminoácidos no naturales *in vivo* a proteínas en *E. coli*, en donde los aminoácidos no naturales se pueden incorporar en una posición definida, y en donde los aminoácidos no naturales tienen propiedades químicas distintas que les permiten servir como dianas para la modificación específica por la exclusión de reacciones cruzadas o reacciones secundarias con otras partes de las proteínas. Esta necesidad en la técnica es especialmente aplicable a *E. coli*, puesto que los sistemas de expresión de proteínas eubacterianas pueden producir grandes cantidades de material proteico recombinante para su estudio científico o aplicaciones terapéuticas. Esta invención satisface estas y otras necesidades, como será evidente tras la revisión de la siguiente divulgación.

El documento de Estados Unidos 2003/108885 A1 describe composiciones y métodos para generar componentes de la maquinaria biosintética de proteínas que incluyen ARNt ortogonales, aminoacil-ARNt sintetasas ortogonales, y pares ortogonales de ARNt/sintetasas.

Wang y col., 2001, Science, Vol. 292, pp. 498-500, describe un método de expansión del código genético de *E. coli*.

Kobayashi y col., 2003, Nature Structural Biology, Vol. 10, No. 6, pp. 425-432, describe una base estructural para las especificidades del ARNt ortogonal de tirosil-ARNt sintetasas para la expansión del código genético.

El documento WO 02/085923 describe métodos y composiciones para la incorporación *in vivo* de aminoácidos no naturales.

Sumario de la invención

En un aspecto, la presente invención proporciona una célula eubacteriana que comprende una primera aminoacil-ARNt sintetasa ortogonal (O-RS) que actúa en la célula, en la que la O-RS preferentemente aminoacila un primer ARNt ortogonal (O-ARNt) con un primer aminoácido no natural que es un alquil aminoácido; en el que el alquil aminoácido es una tirosina para-sustituida o una fenilalanina para-sustituida, en la que la tirosina o la fenilalanina están sustituidas en la posición para con un grupo etinilo o propinilo; y en el que la O-RS aminoacila el O-ARNt con el primer aminoácido no natural con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25.

En un aspecto adicional, la presente invención proporciona un método de producción de una proteína que comprende un alquil aminoácido no natural en una célula eubacteriana, en el que el alquil aminoácido está en una posición específica, el método que comprende:

(a) el suministro de una célula eubacteriana que comprende:

- (i). una aminoacil-ARNt sintetasa ortogonal (O-RS);
- (ii). un ARNt ortogonal (O-ARNt),
- (iii). en el que dicha O-RS preferentemente aminoacila dicho O-ARNt con dicho alquil aminoácido con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25;
- (iv). un ácido nucleico que codifica dicha proteína, en el que el ácido nucleico comprende al menos un codón selector que es reconocido por el O-ARNt; y
- (v). dicho alquil aminoácido, alquil aminoácido que es una tirosina para-sustituida o una fenilalanina para-sustituida, en la que la tirosina o la fenilalanina están sustituidas en la posición para con un grupo etinilo o propinilo; y

(b) el crecimiento de dicha célula;

(c) la incorporación de dicho alquil aminoácido a dicha posición específica en la proteína codificada por el ácido nucleico durante la traducción de la proteína, en el que la posición específica en la proteína corresponde a la posición del codón selector en dicho ácido nucleico, produciendo así dicha proteína que comprende dicho alquil aminoácido en dicha posición específica.

De acuerdo con la presente invención, la célula eubacteriana puede ser una célula de *E. coli*.

De acuerdo con la presente invención:

- (i). la O-RS puede ser idéntica en al menos un 95 % a una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o
- (ii). la O-RS puede ser idéntica en al menos un 95 % a una tirosil-ARNt sintetasa de *Methanococcus jannaschii*; o
- (iii). la O-RS puede ser idéntica en al menos un 95 % a la tirosil-ARNt sintetasa de tipo silvestre de *Methanococcus jannaschii* que tiene la secuencia de aminoácidos de la SEQ ID NO: 2; o
- (iv). la O-RS puede ser idéntica en al menos un 95 % a la tirosil-ARNt sintetasa de tipo silvestre de *Methanococcus jannaschii* que tiene la secuencia de aminoácidos de la SEQ ID NO: 2, en la que la O-RS tiene una secuencia de aminoácidos que comprende:

- (a) alanina, histidina, serina o treonina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;
- (b) prolina, glutamina, lisina, arginina, serina o alanina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;
- (c) alanina, leucina, prolina o serina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina, histidina, treonina o prolina en una posición que corresponde a la posición del aminoácido 162 de la SEQ ID NO: 21; o

(v). la O-RS puede ser idéntica en al menos un 95 % a la tirosil-ARNt sintetasa de tipo silvestre de *Methanococcus jannaschii* que tiene la secuencia de aminoácidos de la SEQ ID NO: 2, en la que la O-RS tiene una secuencia de aminoácidos que comprende:

- (a) alanina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;
- (b) prolina o glutamina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;
- (c) alanina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina o prolina en una posición que corresponde a la posición del aminoácido 162 de la SEQ ID NO: 21; o

(vi). la O-RS puede comprender una secuencia de aminoácidos seleccionada entre las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, y sus variantes conservativas.

De acuerdo con la presente invención, la célula puede comprender un polinucleótido que codifica la O-RS, en la que la O-RS comprende una secuencia de aminoácidos seleccionada entre las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, y sus variantes conservativas.

De acuerdo con la presente invención, el polinucleótido se puede seleccionar entre las secuencias de nucleótidos de las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 o 19.

De acuerdo con la presente invención:

- (a) el O-ARNt puede ser un ARNt supresor ámbar; o
- (b) el O-ARNt puede comprender o puede estar codificado por una secuencia de polinucleótidos expuesta en la SEQ ID NO: 1.

De acuerdo con la presente invención, dicho alquil aminoácido puede ser para-propargiloxifenilalanina.

De acuerdo con la presente invención, el ácido nucleico puede comprender al menos un codón selector, en el que dicho codón selector es reconocido por dicho primer O-ARNt.

De acuerdo con la presente invención, la célula o el método pueden comprender una segunda O-RS y un segundo O-ARNt, en el que la segunda O-RS preferentemente aminoacila el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y en el que el segundo O-ARNt reconoce un codón selector que es diferente del codón selector reconocido por el primer O-ARNt.

De acuerdo con la presente invención, la célula puede comprender dicho alquil aminoácido. Dicho alquil aminoácido puede ser para-propargiloxifenilalanina.

De acuerdo con la presente invención, la célula puede comprender un sistema de traducción. Dicho sistema de traducción puede comprender:

- (a) dicha O-RS;
- (b) dicho O-ARNt;
- (c) un ácido nucleico que codifica un polipéptido de interés, el ácido nucleico que comprende al menos un codón selector, en el que el codón selector es reconocido por dicho O-ARNt; y
- (d) un alquil aminoácido, en el que dicha O-RS es capaz de cargar dicho O-ARNt con dicho alquil aminoácido.

De acuerdo con el método de la presente invención, dicha proteína puede comprender una proteína terapéutica de tipo silvestre, una proteína diagnóstica, una enzima industrial, o un fragmento de las mismas.

De acuerdo con el método de la presente invención, habiendo obtenido dicha proteína, la proteína se puede formular en una composición con un vehículo farmacéuticamente aceptable.

De acuerdo con el método de la presente invención, dicha proteína se puede modificar en dicha posición específica. Dicha proteína puede comprender una unión triazol en dicha posición específica.

En un aspecto adicional, la presente invención proporciona un polipéptido que es al menos idéntico en un 90 % a la tirosil aminoacil-ARNt sintetasa de *Methanococcus jannaschii* de la SEQ ID NO: 2, en el que el polipéptido tiene una secuencia de aminoácidos que comprende:

- (a) alanina, histidina, serina o treonina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;
- (b) prolina, glutamina, lisina, arginina, serina o alanina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;
- (c) alanina, leucina, prolina o serina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina, histidina, treonina o prolina en una posición que corresponde a la posición del aminoácido 162 de la

SEQ ID NO: 21;

y en el que el polipéptido es una aminoacil-ARNt sintetasa capaz de aminoacilar preferentemente un ARNt ortogonal (O-ARNt) con un alquilil aminoácido, alquilil aminoácido que es una tirosina para-sustituida o una fenilalanina para-sustituida, en la que la tirosina o la fenilalanina están sustituidas en la posición para con un grupo etinilo o propinilo;

y en el que la O-RS aminoacila el O-ARNt con el primer aminoácido no natural con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25.

El polipéptido de la presente invención puede tener una secuencia de aminoácidos que comprende:

- (a) alanina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;
- (b) prolina o glutamina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;
- (c) alanina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina o prolina en una posición que corresponde a la posición del aminoácido 162 de la SEQ ID NO: 21.

En un aspecto adicional, la presente invención proporciona un polipéptido que comprende una secuencia de aminoácidos expuesta en las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, o una de sus variantes conservativas, en el que dicho polipéptido es una aminoacil-ARNt sintetasa capaz de aminoacilar preferentemente un ARNt ortogonal (O-ARNt) en una célula eubacteriana con un alquilil aminoácido, alquilil aminoácido que es una tirosina que está sustituida en posición para con un grupo etinilo o propinilo, o el alquilil aminoácido es una fenilalanina que está sustituida en posición para con un grupo etinilo o propinilo; y

en el que la O-RS aminoacila el O-ARNt con el primer aminoácido no natural con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25.

En un aspecto adicional, la presente invención proporciona un polinucleótido que codifica un polipéptido de la invención. El polinucleótido se puede seleccionar entre las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 y 19.

En un aspecto adicional, la presente invención proporciona un vector que comprende un polinucleótido de la invención. El vector puede ser un vector de expresión.

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

Definiciones

Antes de describir la invención en detalle, se debe entender que esta invención no está limitada a sistemas biológicos particulares, que naturalmente pueden variar. También se debe entender que la terminología usada en este documento tiene como único fin describir realizaciones particulares, y no se pretende que sea limitante. Como se usa en esta memoria descriptiva y en las reivindicaciones adjuntas, las formas singulares “un”, “una”, “el” y “la” incluyen los referentes en plural a menos que el contenido dicte claramente lo contrario. Así, por ejemplo, la referencia a “una célula” incluye combinaciones de dos o más células; la referencia a “un polinucleótido” incluye, como cuestión práctica, muchas copias de ese polinucleótido.

A menos que se defina lo contrario en este documento y a continuación en el resto de la memoria descriptiva, todos los términos técnicos y científicos usados en este documento tienen el mismo significado que se entiende habitualmente para el experto en la materia a la que pertenece esta invención.

Ortogonal: Como se usa en este 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 actúa con componentes endógenos de una célula con una eficiencia reducida en comparación con una molécula correspondiente que sea endógena a la célula o al sistema de traducción, o que no puede funcionar con los componentes endógenos de la célula. En este contexto de ARNt y aminoacil-ARNt sintetasa, ortogonal se refiere a una incapacidad o eficiencia reducida, por ejemplo, eficiencia inferior al 20 %, eficiencia inferior al 10 %, eficiencia inferior al 5 %, o eficiencia inferior al 1 %, de un ARNt ortogonal para funcionar con una ARNt-sintetasa endógena en comparación con un ARNt endógeno para funcionar con la ARNt-sintetasa endógena, o de una aminoacil-ARNt sintetasa ortogonal para funcionar con un ARNt endógeno en comparación con una ARNt-sintetasa endógena para funcionar con el ARNt endógeno. La molécula ortogonal carece de una molécula complementaria endógena funcionalmente normal en la célula. Por ejemplo, un ARNt ortogonal en una célula es aminoacilado por cualquier RS endógena de la célula con una eficiencia reducida o incluso de cero, cuando se compara con la aminoacilación de un ARNt endógeno por la RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno de una célula de interés con una eficiencia reducida o incluso de cero, en comparación con la aminoacilación del ARNt endógeno por una RS endógena. Se puede introducir una segunda molécula ortogonal en la célula que funcione con la primera molécula ortogonal. Por ejemplo,

un par ARNt/RS ortogonal incluye componentes complementarios introducidos que funcionan juntos en la célula con una eficiencia (por ejemplo, eficiencia del 45 %, eficiencia del 50 %, eficiencia del 60 %, eficiencia del 70 %, eficiencia del 75 %, eficiencia del 80 %, eficiencia del 90 %, eficiencia del 95 %, o eficiencia del 99 % o superior) en comparación con la de un control, por ejemplo, un par ARNt/RS endógeno correspondiente, o un par ortogonal activo (por ejemplo, un par tirosil-ARNt ortogonal/RS).

Tirosil-ARNt ortogonal: Como se usa en este documento, un tirosil-ARNt ortogonal (tirosil-O-ARNt) es un ARNt que es ortogonal a un sistema de traducción de interés, en el que el ARNt es: (1) idéntico o sustancialmente similar a un tirosil-ARNt de origen natural, (2) derivado de un tirosil-ARNt de origen natural mediante mutagénesis natural o artificial, (3) derivado de cualquier proceso que tenga en cuenta una secuencia de tipo silvestre o una secuencia mutante de tirosil-ARNt de (1) o (2), (4) homólogo a un tirosil-ARNt silvestre o mutante; (5) homólogo a cualquier ARNt ejemplar que esté diseñado como sustrato para una tirosil-ARNt sintetasa en la Tabla 4, o (6) una variante conservativa de cualquier ARNt ejemplar que esté diseñado como sustrato para una tirosil-ARNt sintetasa en la Tabla 4. El tirosil-ARNt puede existir cargado con un aminoácido, o en estado no cargado. También se debe entender que un “tirosil-O-ARNt” opcionalmente está cargado (aminoacilado) por una sintetasa afín con un aminoácido distinto de la tirosina, por ejemplo, con el aminoácido no natural para-propargiloxifenilalanina. De hecho, se debe apreciar que de manera ventajosa se usa un tirosil-O-ARNt de la invención para insertar esencialmente cualquier aminoácido, ya sea natural o artificial, en un polipéptido en crecimiento, durante la traducción, en respuesta a un codón selector.

Tirosil aminoácido sintetasa ortogonal: Como se usa en este documento, una tirosil aminoácido sintetasa ortogonal (tirosil-O-RS) es una enzima que preferentemente aminoacila el tirosil-O-ARNt con un aminoácido en un sistema de traducción de interés. El aminoácido que carga la tirosil-O-RS sobre el tirosil-O-ARNt puede ser cualquier aminoácido, ya sea natural, no natural o artificial, y en este documento no está limitado. La sintetasa es opcionalmente la misma u homóloga a una tirosil aminoácido sintetasa de origen natural, o la misma u homóloga a una sintetasa designada como O-RS en la Tabla 4. Por ejemplo, la O-RS puede ser una variante conservativa de una tirosil-O-RS de la Tabla 4, y/o puede tener una identidad de secuencia de al menos el 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, 98 %, 99 % o superior a una O-RS la Tabla 4.

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

Preferentemente aminoacila: Como se usa en este documento en referencia a sistemas de traducción ortogonales, una O-RS “preferentemente aminoacila” un ARNt afín cuando la O-RS carga el O-ARNt con un aminoácido más eficazmente de lo que carga cualquier ARNt endógeno en un sistema de expresión. Es decir, cuando el O-ARNt y cualquier ARNt endógeno dado están presentes en un sistema de traducción en relaciones molares aproximadamente iguales, la O-RS cargará el O-ARNt con mayor frecuencia de lo que cargará el ARNt endógeno. Preferentemente, la relación relativa de O-ARNt cargado por la O-RS a ARNt endógeno cargado por la O-RS es elevada, produciendo preferentemente que la O-RS cargue exclusivamente el O-ARNt, o casi exclusivamente, cuando el O-ARNt y el ARNt endógeno están presentes en concentraciones molares iguales en el sistema de traducción. La relación relativa entre el O-ARNt y el ARNt endógeno que es cargado por la O-RS, cuando el O-ARNt y la O-RS están presentes a concentraciones molares iguales, es superior a 1:1, preferentemente al menos de 2:1 aproximadamente, más preferentemente de 5:1, aún más preferentemente de 10:1, incluso más preferentemente de 20:1, aún más preferentemente de 50:1, incluso más preferentemente de 75:1, aún más preferentemente de 95:1, 98:1, 99:1, 100:1, 500:1, 1000:1, 5000:1 o superior.

La O-RS “preferentemente aminoacila un O-ARNt con un aminoácido no natural” cuando (a) la O-RS preferentemente aminoacila el O-ARNt en comparación con un ARNt endógeno, y (b) cuando esa aminoacilación es específica para el aminoácido no natural, en comparación con la aminoacilación del O-ARNt por la O-RS con cualquier aminoácido natural. Es decir, cuando los aminoácidos naturales y no naturales están presentes en cantidades molares iguales en un sistema de traducción que comprende la O-RS y el O-ARNt, la O-RS cargará el O-ARNt con el aminoácido no natural con mayor frecuencia que con el aminoácido natural. Preferentemente, la relación relativa de O-ARNt cargado con el aminoácido no natural a O-ARNt cargado con el aminoácido natural es elevada. Más preferentemente, la O-RS carga el O-ARNt exclusivamente, o casi exclusivamente, con el aminoácido no natural. La relación relativa entre la carga del O-ARNt con el aminoácido no natural y la carga del O-ARNt con el aminoácido natural, cuando los aminoácidos naturales y no naturales están presentes en el sistema de traducción en concentraciones molares iguales, es superior a 1:1, preferentemente al menos de 2:1 aproximadamente, más preferentemente de 5:1, aún más preferentemente de 10:1, incluso más preferentemente de 20:1, aún más preferentemente de 50:1, incluso más preferentemente de 75:1, aún más preferentemente de 95:1, 98:1, 99:1, 100:1, 500:1, 1000:1, 5000:1 o superior.

Codón selector: El término “codón selector” se refiere a codones reconocidos por el O-ARNt en el proceso de traducción y no reconocidos por un ARNt endógeno. El bucle anti-codónico del O-ARNt reconoce el codón selector sobre el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido no natural, tal como un alquilil aminoácido, en este sitio en el polipéptido. Los codones selectores pueden incluir, por ejemplo, codones antisentido, tales como, codones de parada, por ejemplo, codones ámbar, ocre, y ópalo; codones de cuatro o más bases; codones raros;

codones derivados de pares de bases naturales o no naturales y/o similares.

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

Actividad de supresión: Como se usa en este documento, el término “actividad de supresión” se refiere, en general, a la capacidad de un ARNt (por ejemplo, un ARNt supresor) de permitir la lectura de traducción a través de un codón (por ejemplo, un codón selector que es un codón ámbar o un codón de 4 o más bases) que de lo contrario produciría la terminación de la traducción o un error en la traducción (por ejemplo, desplazamiento del marco de lectura). La actividad de supresión de un ARNt supresor se puede expresar como porcentaje de actividad de lectura de traducción a través de los codones de parada observada en comparación con un segundo ARNt supresor, o comparado con un sistema de control, por ejemplo, un sistema de control que carece de una O-RS.

La presente invención proporciona varios medios mediante los cuales se puede cuantificar la actividad de supresión. El porcentaje de supresión de un O-ARNt y O-RS particulares frente a un codón selector (por ejemplo, un codón ámbar) de interés se refiere al porcentaje de actividad de un marcador de ensayo expresado específico (por ejemplo, LacZ), que incluye un codón selector, en un ácido nucleico que codifica el marcador de ensayo expresado, en un sistema de traducción de interés, en el que el sistema de traducción de interés incluye una O-RS y un O-ARNt, en comparación con una construcción de control positivo, en el que el control positivo carece del O-ARNt, la O-RS y el codón selector. Así, por ejemplo, si una construcción marcadora de control positivo activo que carece de codón selector tiene una actividad observada de X en un sistema de traducción determinado, en unidades relevantes para el ensayo marcador en cuestión, entonces el porcentaje de supresión de una construcción de ensayo que comprende el codón selector es el porcentaje de X que presenta la construcción marcadora de ensayo esencialmente bajo las mismas condiciones ambientales bajo las que se expresó el marcador de control positivo, excepto porque la construcción marcadora de ensayo se expresa en un sistema de traducción que también incluye el O-ARNt y la O-RS. Normalmente, el sistema de traducción que expresa el marcador de ensayo también incluye un aminoácido que es reconocido por la O-RS y el O-ARNt. Opcionalmente, la medida del porcentaje de supresión se puede afinar comparando el marcador de ensayo con una construcción marcadora de control “de fondo” o “negativa”, que incluye el mismo codón selector que el marcador de ensayo, pero en un sistema que no incluya el O-ARNt, la O-RS y/o el aminoácido relevante reconocido por el O-ARNt y/o la O-RS. Este control negativo es útil para normalizar las medidas del porcentaje de supresión y tener en cuenta los efectos de la señal de fondo procedente del marcador en el sistema de traducción de interés.

La eficiencia de supresión se puede determinar mediante una serie de ensayos conocidos en la técnica. Por ejemplo, se puede usar un ensayo informador de β -galactosidasa, por ejemplo, un plásmido lacZ derivado (en el que la construcción tiene un codón selector en la secuencia de ácidos nucleicos de lacZ) se introduce en células de un organismo apropiado (por ejemplo, un organismo en el que se pueden usar los componentes ortogonales) junto con un plásmido que comprende un O-ARNt de la invención. También se podría introducir una sintetasa afín (ya sea como polipéptido o 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 0,5 aproximadamente, y los ensayos de β -galactosidasa se llevan a cabo, por ejemplo, utilizando el kit de β -galactosidasa β -Galactosidase Assay Kit BetaFluor™ (Novagen). El porcentaje de supresión se puede calcular como porcentaje de actividad para una muestra con respecto a un control comparable, por ejemplo, el valor observado de la construcción lacZ derivada, en donde la construcción tiene un codón sentido correspondiente en la posición deseada en vez de un codón selector.

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

Aminoácido no natural: Como se usa en este documento, el término “aminoácido no natural” se refiere a cualquier aminoácido, aminoácido modificado, y/o análogo de aminoácido, tal como un alquil aminoácido, que no es uno de los 20 aminoácidos habituales de origen natural o selenocisteína o pirrolisina.

Derivado de: Como se usa en este documento, el término “derivado de” se refiere a un componente que se aísla o se prepara usando una molécula u organismo específico, o información procedente de la molécula u organismo específico. Por ejemplo, un polipéptido derivado de un segundo polipéptido comprende una secuencia de aminoácidos que es idéntica o sustancialmente similar a la secuencia de aminoácidos del segundo polipéptido. En el caso de polipéptidos, las especies derivadas se pueden obtener, por ejemplo, mediante mutagénesis natural, mutagénesis artificial dirigida o mutagénesis artificial aleatoria. La mutagénesis usada para derivar polipéptidos puede ser intencionadamente dirigida o intencionadamente aleatoria. La mutagénesis de un polipéptido para crear

un polipéptido diferente derivado del primero puede ser un acontecimiento aleatorio (por ejemplo, provocado por la infidelidad de la polimerasa) y la identificación del polipéptido derivado puede ser fortuita. La mutagénesis de un polipéptido normalmente entraña la manipulación del polinucleótido que codifica el polipéptido.

5 Selección positiva o marcador de detección: Como se usa en este documento, el término “selección positiva o marcador de detección” se refiere a un marcador que cuando está presente, por ejemplo, expresado, activado o similar, produce la identificación de una célula que comprende el rasgo, por ejemplo, células con el marcador de selección positivo, con respecto a aquellas sin el rasgo.

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

15 Informador: Como se usa en este documento el término “informador” se refiere a un componente que se puede usar para identificar y/o seleccionar componentes diana de un sistema de interés. Por ejemplo, un informador puede incluir una proteína, por ejemplo, una enzima, que confiere resistencia o sensibilidad a antibióticos (por ejemplo, β -lactamasa, cloranfenicol acetiltransferasa (CAT), y similares), un marcador de detección fluorescente (por ejemplo, proteína verde fluorescente (por ejemplo, (GFP), YFP, EGFP, RFP, etc.), un marcador luminiscente (por ejemplo, una proteína luciferasa de luciérnaga), un marcador de detección basado en afinidad, o genes marcadores seleccionables positivos o negativos, tales como lacZ, β -gal/lacZ (β -galactosidasa), ADH (alcohol deshidrogenasa), his3, ura3, leu2, lys2, o similares.

25 Eucariota: Como se usa en este documento, el término “eucariota” se refiere a organismos que pertenecen al reino Eucarya. Los eucariotas generalmente se pueden distinguir de los procariotas por su organización normalmente multicelular (pero no exclusivamente multicelular, por ejemplo, levaduras), la presencia de un núcleo rodeado de membrana y otros orgánulos rodeados de membrana, material genético lineal (por ejemplo, cromosomas lineales), la ausencia de operones, la presencia de intrones, la modificación con caperuzas de los mensajeros y ARNm poli-A, y otras características bioquímicas, tales como una estructura ribosómica distintiva. Los organismos eucariotas incluyen, por ejemplo, animales (por ejemplo, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (por ejemplo, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidios, protistas, etc.

35 Procariota: Como se usa en este documento, el término “procariota” se refiere a organismos que pertenecen al reino Monera (también denominado Procarya). Los organismos procariotas en general se pueden distinguir de los eucariotas por su organización unicelular, reproducción asexual por gemación o fisión, la falta de un núcleo rodeado de membrana u otros orgánulos rodeados de membrana, un cromosoma circular, la presencia de operones, la ausencia de intrones, la modificación con caperuzas de los mensajeros y ARNm poli-A, y otras características bioquímicas, tales como una estructura ribosómica distintiva. Los Procarya incluyen los subreinos eubacterias y arqueas (a veces denominados “Archaebacteria”). Las cianobacterias (algas verde azuladas) y los micoplasmas a veces se encuentran en clasificaciones separadas bajo el Reino Monera.

45 Bacterias: Como se usa en este documento, los términos “bacteria” y “eubacterias” se refieren a organismos procariotas que se pueden distinguir de las arqueas. De forma similar, arqueas se refiere a procariotas que se pueden distinguir de las eubacterias. Las eubacterias y las arqueas se pueden distinguir por una serie de criterios morfológicos y bioquímicos. Por ejemplo, para asignar un organismo a eubacterias o arqueas se usan las diferencias en las secuencias de ARN ribosómico, estructura de la ARN polimerasa, la presencia o ausencia de intrones, sensibilidad a antibióticos, la presencia o ausencia de peptidoglicanos en la pared celular y otros componentes de la pared celular, las estructuras ramificadas frente a estructuras no ramificadas de los lípidos de membrana, y la presencia/ausencia de histonas y proteínas de tipo histona.

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

60 Variante conservativa: Como se usa en este documento, el término “variante conservativa”, en el contexto de un componente de traducción, se refiere a un componente de traducción, por ejemplo, una variante conservativa del O-ARNt o una variante conservativa de O-RS, que funcionalmente se comporta de forma similar a un componente base que es similar a la variante conservativa, por ejemplo, un ARNt u O-RS, que tiene variaciones en la secuencia comparado con un O-ARNt o O-RS de referencia. Por ejemplo, una O-RS aminoacilará un O-ARNt complementario o una variante conservativa del O-ARNt con un aminoácido no natural, por ejemplo, un alquil aminoácido tal como para-propargiloxifenilalanina, a pesar de que el O-ARNt y la variante conservativa del O-ARNt no tengan la misma secuencia. La variante conservativa puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones, o cinco o más variaciones en la secuencia, siempre que la variante conservativa sea

complementaria al O-ARNt u O-RS correspondientes.

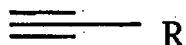
Agente de selección o de detección: Como se usa en este documento, el término “agente de selección o de detección” se refiere a un agente que, cuando está presente, permite la selección/detección de ciertos componentes dentro de una población. Por ejemplo, un agente de selección o de detección puede ser, pero no está limitado a, por ejemplo, un nutriente, un antibiótico, una longitud de onda luminosa, un anticuerpo, un polinucleótido expresado, o similares. El agente de selección se puede modificar, por ejemplo, en concentración, intensidad, etc.

En respuesta a: Como se usa en este documento, el término “en respuesta a” se refiere al proceso en el que un O-ARNt de la invención reconoce un codón selector y media en la incorporación del alquil aminoácido, que se acopla al ARNt, en la cadena polipeptídica en crecimiento.

Codificar: Como se usa en este documento, el término “codificar” se refiere a cualquier proceso mediante el cual la información de una macromolécula polimérica o cadena de secuencia se usa para dirigir la producción de una segunda molécula o cadena de secuencia que es diferente de la primera molécula o cadena de secuencia. Como se usa en este documento, el término se utiliza de forma amplia, y puede tener diferentes aplicaciones. En un aspecto, el término “codificar” describe el proceso de replicación semi-conservativa del ADN, en donde una cadena de una molécula de ADN de doble cadena se usa como molde para codificar una cadena hermana complementaria sintetizada *de novo* por una ADN polimerasa que depende del ADN.

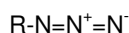
En otro aspecto, el término “codificar” se refiere a cualquier proceso mediante el cual se usa la información en una molécula para dirigir la producción de una segunda molécula que tiene una naturaleza química diferente de la primera molécula. Por ejemplo, una molécula de ADN puede codificar una molécula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

Alquino: Como se usa en este documento, el término “alquino” (a veces también denominado “acetileno”) se refiere a estructuras químicas que contienen un triple enlace entre dos átomos de carbono (como se muestra en la Figura 1B), que tiene la estructura general:



en la que R es cualquier átomo o estructura. Cuando se usa como sustituyente, el resto alquino se denomina grupo “alquinilo”. Los átomos de carbono del alquinilo presentan hibridación sp^2 y sólo forman enlaces a otros dos átomos; uno de estos enlaces será un enlace sencillo mientras que el segundo enlace es un triple enlace. Por ejemplo, un alquil aminoácido es un aminoácido que contiene un triple enlace entre dos centros carbonados. Debido a que los sustituyentes alquinilo no aparecen sobre aminoácidos en la naturaleza, todo alquil aminoácido es un aminoácido no natural.

Azido: Como se usa en este documento, el término “azido” se refiere al grupo químico $-N_3$, que tiene la estructura general:



El grupo azido normalmente está unido a un átomo de carbono.

Por ejemplo, un colorante azido es una molécula colorante con un grupo sustituyente azido (véanse, por ejemplo, los colorantes azido **2** y **3**, en las Figuras 6A y 6B). El término “azido” se refiere a un compuesto químico que contiene el grupo azido (por ejemplo, bencil azida, azida sódica, etc.).

Breve descripción de las Figuras

La Figura 1A proporciona la estructura química (1) del alquil aminoácido no natural para-propargiloxifenilalanina (también conocido como ácido 2-amino-3-[4-(prop-2-inoxi)fenil]propiónico según la nomenclatura de la IUPAC). La Figura 1B proporciona la química de la reacción generalizada de la formación irreversible de triazoles por reacción de cicloadición [3+2] de un azido y un alquino en presencia de cobre a temperatura ambiente.

La Figura 2 proporciona las secuencias de nucleótidos y de aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre (MjTyrRS). Las posiciones de los aminoácidos (y los tripletes

codónicos correspondientes) diana en la mutagénesis dirigida o mutados de otra forma en la para-propargiloxifenilalanina (pPro-Phe) ARNt-sintetasa están recuadradas.

La Figura 3 proporciona una tabla que describe las ocho (8) especies de para-propargiloxifenilalanina (pPro-Phe) ARNt-sintetasa identificadas y aisladas después de la mutagénesis de un polinucleótido que codifica la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre. Los aminoácidos codificados por los codones indicados en la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre y las para-propargiloxifenilalanina (pPro-Phe) ARNt-sintetasas están indicados. También están indicados los codones en las posiciones mutantes. La numeración de la posición de los aminoácidos de los mutantes es según la numeración de los aminoácidos de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* de tipo silvestre, como se muestra en la **Figura 2**.

La Figura 4 proporciona un gel de SDS-PAGE teñido con GelCode® Blue (Pierce Biotechnology, Inc.) de mioglobina mutante Ser4→pPro-Phe4 purificada. El Carril 1 contiene proteína expresada en *E. coli* cultivada en medio mínimo en presencia de para-propargiloxifenilalanina (pPro-Phe); el Carril 2 contiene una muestra de proteína generada en ausencia de pPro-Phe. El panel inferior muestra una transferencia de Western de los mismos materiales de muestra usando un anticuerpo dirigido contra His6 para detectar la etiqueta de hexahistidina en el extremo C-terminal de la mioglobina.

La Figura 5 proporciona un espectro de masas en tándem del péptido tríptico HGVTVLALGY*ILK que contiene el alquil aminoácido no natural (denotado Y*) mostrado con sus masas de fragmentos iónicos esperadas. Las flechas indican la serie de iones b e y observados para el péptido.

Las Figuras 6A y 6B proporcionan las estructuras químicas (**2** y **3**, respectivamente) de colorantes funcionalizados con azido. El colorante **2** en la **Figura 6A** contiene un fluoróforo de dansilo, y el colorante **3** en la **Figura 6B** contiene un fluoróforo de fluoresceína.

La Figura 7A proporciona la química de la reacción generalizada de la formación irreversible de un triazol a partir de la reacción de cicloadición [3+2] entre la mioglobina mutante que contiene un alquil aminoácido en el sitio del codón ámbar manipulado genéticamente (4TAG) y un colorante funcionalizado con azido (como se proporciona en las **Figuras 6A y 6B**). La **Figura 7B** proporciona unas imágenes de un gel de fluorescencia bajo radiación UV de la mioglobina marcada resuelta, donde la reacción de cicloadición [3+2] une covalentemente el colorante **2** o el colorante **3**.

Las Figuras 8A y 8B proporcionan las estructuras y los nombres de ejemplos de alquil aminoácidos no naturales. **La Figura 8A** proporciona alquil aminoácidos no naturales que se pueden sintetizar químicamente a partir de precursores no naturales. **La Figura 8B** proporciona alquil aminoácidos no naturales que se pueden sintetizar potencialmente a partir de sustratos de aminoácidos de origen natural ya existentes.

Descripción detallada de la invención

Existe una gran necesidad de reacciones químicas que modifiquen proteínas en condiciones fisiológicas de una forma altamente selectiva (Lemieux y Bertozzi (1996) TIBTECH, 16:506). La mayoría de las reacciones usadas actualmente en la técnica para la modificación selectiva de proteínas suponen la formación de un enlace covalente entre los homólogos de reacción nucleófilos y electrófilos que son selectivos para residuos nucleófilos de origen natural en las cadenas laterales aminoácidas de la proteína, por ejemplo, la reacción de α -halo cetonas con las cadenas laterales de histidina o cisteína. La selectividad en estos casos está determinada por el número y la accesibilidad de los restos nucleófilos en la proteína. Desafortunadamente, las proteínas de origen natural con frecuencia contienen sitios de reacción mal situados (por ejemplo, inaccesibles) o dianas de reacción múltiples (por ejemplo, restos de lisina, histidina y cisteínas), lo que produce una mala selectividad en las reacciones de modificación, dificultando la modificación de proteínas altamente dirigida mediante reactivos nucleófilos/electrófilos. Además, los sitios de modificación normalmente están limitados a las cadenas laterales nucleófilas de origen natural de lisina, histidina o cisteína. La modificación en otros sitios es difícil o imposible.

Una solución a este problema es la incorporación biosintética programada específica de sitio de aminoácidos no naturales con una nueva reactividad a proteínas usando componentes de traducción ortogonal (Wang y Schultz (2002) Chem. Commun., 1:1; y van Maarseveen y Back (2003) Angew. Chem., 115:6106). En este documento presentamos un nuevo método altamente eficiente para la modificación selectiva de proteínas que supone la incorporación genética de aminoácidos no naturales que contienen alquilino a proteínas producidas en bacterias (por ejemplo, *E. coli*) en respuesta al codón sin sentido ámbar, TAG. Estas cadenas laterales de alquil aminoácidos a continuación se pueden modificar específica y regiosselectivamente. Debido a la química de reacción única del grupo alquilino, las proteínas se pueden modificar con una selectividad extremadamente elevada.

Para introducir selectivamente el grupo funcional alquilino en sitios únicos (por ejemplo, en el sitio deseado) en proteínas producidas en un sistema de expresión bacteriano, hemos desarrollado pares ARNt/aminoacil-ARNt

sintetasa ortogonales que funcionan en eubacterias que codifican genéticamente el alquil aminoácido para-propargiloxifenilalanina (pPro-Phe; véase **Figura 1A**). En resumen, hemos identificado nuevos mutantes de la tirosil-ARNt sintetasa de *Methanococcus jannaschii* que carga selectivamente un ARNt supresor ámbar con para-propargiloxifenilalanina (pPro-Phe) en células de *Escherichia coli*. Estos pares ARNt-sintetasa evolucionados se

Modificación dirigida de una proteína

En este documento presentamos un método altamente eficiente para la modificación selectiva de proteínas que supone la incorporación genética de aminoácidos no naturales que contienen alquínulo a proteínas producidas en eubacterias (por ejemplo, *E. coli*) en respuesta al codón sin sentido ámbar, TAG. Las nuevas composiciones y métodos descritos en este documento emplean un sistema ARNt/aminoacil-ARNt sintetasa ortogonal, en el que el sistema ortogonal usa componentes derivados de *Methanococcus jannaschii*, y en el que estos componentes se usan en un sistema hospedador eubacteriano para producir la proteína de interés. La incorporación del alquil aminoácido a la proteína se puede programar para que se produzca en cualquier posición deseada por manipulación genética del polinucleótido que codifica la proteína de interés para que contenga un codón selector que señala la incorporación del alquil aminoácido.

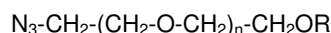
Estas cadenas laterales del alquil aminoácido sobre la proteína de interés a continuación se pueden modificar específica y regiosselectivamente mediante una reacción de cicloadición de Huisgen [3+2] con derivados azido (véase **Figura 1B**) (Padwa, en Comprehensive Organic Synthesis; [Trost, B. M., Ed.] Pergamon: Oxford, 1991, Vol. 4, p 1069-1109; Huisgen, en 1,3-Dipolar Cycloaddition Chemistry, [Padwa, A., Ed.] Wiley: Nueva York, 1984; p 1-176). Debido a que este método supone una cicloadición en lugar de una sustitución nucleófila, las proteínas se pueden modificar con una selectividad extremadamente elevada. Esta reacción tiene la ventaja de que se puede llevar a cabo a temperatura ambiente en condiciones acuosas con una regiosselectividad excelente (1,4 > 1,5) mediante la adición de cantidades catalíticas de sales de Cu (I) a la mezcla de reacción (Tornøe y col., (2002) J. Org. Chem., 67:3057-3064; Rostovtsev y col., (2002) Angew. Chem., Int. Ed., 41:2596-2599).

Una diana reactiva de alquínulo tiene la ventaja de ser completamente exógena a sistemas *in vivo*, es altamente selectiva en sus químicas de reacción (por ejemplo, altamente reactiva con restos que contienen azido), y se puede conjugar usando condiciones de reacción relativamente suaves que permitan reacciones de conjugación que implican a las proteínas tanto *in vitro* como *in vivo*, y conservando la actividad biológica de la proteína. Para demostrar (pero no limitar) la presente invención, el resto alquénulo se incorpora a una proteína modelo de mioglobina, y a continuación la proteína se bioconjugua a colorantes fluorescentes azido (véase, **Figuras 6A y 6B**) mediante una reacción de cicloadición [3+2] con la formación de una unión triazol estable (véase, **Figura 1B**).

Aunque los ejemplos usan dos colorantes fluorescentes azido para ilustrar la cicloadición [3+2] entre los restos alquil aminoácido y azido (véase, Ejemplo 4), no se pretende que la invención esté limitada al uso de estos dos colorantes azido, o a cualquier colorante o marcador, o de hecho a cualquier tipo único de material conjugable. Un resto que contiene azido de la invención virtualmente puede ser cualquier molécula que sea un derivado azido. Dichas moléculas incluyen, pero no están limitadas a, colorantes, fluoróforos, agentes de reticulación, derivados sacarídicos, polímeros (por ejemplo, derivados de polietilenglicol), foto-reticulantes, compuestos citotóxicos, marcadores de afinidad, derivados de biotina, resinas, perlas, una segunda proteína o polipéptido (o más), polinucleótido(s) (por ejemplo, ADN, ARN, etc.), quelantes metálicos, cofactores, ácidos grasos, carbohidratos, y similares. Estas moléculas azido se pueden conjugar a un aminoácido no natural con un grupo alquínulo, por ejemplo, para-propargiloxifenilalanina (véase, **Figura 1A**).

En este documento se proporciona una descripción detallada de la síntesis de los colorantes azido mostrados en las **Figuras 6A y 6B**. Véanse, los Ejemplos 6 y 7, respectivamente. No obstante, dentro de los medios del experto en la materia está el sintetizar un derivado azido de cualquier molécula particular de interés. Por ejemplo, hay disponibles numerosos textos y protocolos que describen cómo sintetizar compuestos azido. Para una referencia general véase: Patai, Saul, "The chemistry of the azido group" en The Chemistry of Functional Groups, Londres, Nueva York, Interscience Publishers, 1971.

En otro aspecto, la invención proporciona composiciones y métodos para la generación de polipéptidos PEGilados usando derivados azido de polietilenglicol (azido-PEG) para su uso en reacciones de conjugación con polipéptidos que contienen alquínulo. La estructura generalizada de un azido polietilenglicol es:



en la que R es H o CH₃, y en la que n es un número entero, por ejemplo, entre 50 y 10.000, 75 y 5000, 100 y 2000, 100 y 1000, etc. En diversas realizaciones de la invención, el azido polietilenglicol tiene un peso molecular, por ejemplo, de entre 5000 aproximadamente y 100.000 Daltons aproximadamente (es decir, entre 5 kDa aproximadamente y 100 kDa aproximadamente), entre 20.000 aproximadamente y 50.000 Daltons aproximadamente, entre 20.000 aproximadamente y 10.000 Daltons aproximadamente (por ejemplo, 20.000 Daltons), etc. Las técnicas para la síntesis de un azido polietilenglicol son muy conocidas por el experto en la

materia. Por ejemplo, una molécula de polietilenglicol que contiene un grupo electrófilo (por ejemplo, un bromuro o un éster de N-hidroxisuccinimida) se puede hacer reaccionar con una molécula nucleófila que contenga un grupo azido (por ejemplo, azida sódica o 3-azido propilamina) para generar un azido polietilenglicol.

- 5 El azido-PEG tiene utilidad con la invención cuando se bioconjug a una proteína que contienen alquínulo mediante una unión triazol. La derivación de los compuestos terapéuticos a base de proteína con polietilenglicol (PEGilación) con frecuencia puede mejorar las propiedades farmacocinéticas y farmacodinámicas de las proteínas y de esta forma, mejorar la eficacia y minimizar la frecuencia de dosificación. Las diversas ventajas de la PEGilación de compuestos terapéuticos proteicos se describen y se ilustran, por ejemplo, en Deiters y col., "Site-specific
10 PEGylation of proteins containing unnatural amino acids", *Bioorganic & Medicinal Chemistry Letters* 14:5743-5745 (2004).

- Además, se contemplan otras ventajas asociadas a la generación de polipéptidos que comprenden alquínulo aminoácidos no naturales que también contienen una unión éster. Por ejemplo, un polipéptido PEGilado creado mediante el uso de un alquínulo aminoácido con una unión éster puede permitir la liberación lenta del polipéptido por saponificación de las uniones éster *in vivo* o *in vitro*. Además, el uso de un soporte polimérico (una resina azido) en lugar de una molécula azido-PEG permite la purificación de la proteína por afinidad. La unión covalente a triazol permite etapas de lavado muy fuertes, y el uso de un éster de alquínulo aminoácido permite la liberación de la proteína mediante el tratamiento con una base. De forma significativa, dicho esquema de purificación por afinidad ya no requiere la presencia de una etiqueta artificial (por ejemplo, hexahistidina) o un epítipo sobre la proteína de interés para su purificación. Dependiendo del aminoácido no natural usado, se puede liberar un polipéptido esencialmente silvestre (nativo) de la resina de afinidad después de la etapa de escisión.

- Los alquínulo aminoácidos no naturales con uniones éster se pueden sintetizar e incorporar a proteínas, por ejemplo, ácido 3-[(prop-2-inoxi)carbonil]-2-aminopropanoico y ácido 4-[(prop-2-inoxi)carbonil]-2-aminobutanoico (véase, **Figura 8B**). Después de la bioconjugación mediante cicloadición [3+2], las uniones éster se pueden escindir por saponificación *in vivo* o *in vitro*; una aplicación sería, por ejemplo, la liberación lenta de la parte peptídica de una proteína PEGilada.

- Los polipéptidos pueden incluir polipéptidos que contienen alquínulo, y además, incluyen las formas conjugadas de estos polipéptidos. Por ejemplo, un polipéptido que comprende una unión triazol y un colorante azido fluorescente acoplado covalentemente (véase, por ejemplo, **Figuras 6A, 6B y 7A**), en donde el polipéptido anteriormente comprendía un grupo alquínulo y el colorante anteriormente comprendía un grupo azido, y los dos se han conjugado mediante una cicloadición [3+2] para formar la unión triazol. De forma alternativa, la proteína que contiene alquínulo puede comprender un azido polietilenglicol (véase estructura química 6).

Tecnología del ARNt/aminoacil-ARNt sintetasa ortogonal

- La comprensión de las nuevas composiciones y métodos de la presente invención se facilita al entender las actividades asociadas a los pares ARNt ortogonal y aminoacil-ARNt sintetasa ortogonal. Descripciones de las tecnologías de ARNt y aminoacil-ARNt sintetasa ortogonales se pueden encontrar, por ejemplo, en las Publicaciones internacionales WO 2002/085923, WO 2002/086075, WO 2004/09459, WO 2005/019415, WO 2005/007870 y WO 2005/007624.

- Con el fin de añadir aminoácidos no naturales reactivos adicionales, tales como alquínulo aminoácidos, al código genético, son necesarios nuevos pares ortogonales que comprendan una aminoacil-ARNt sintetasa y un ARNt adecuado que puedan funcionar eficientemente en la maquinaria de traducción del hospedador, pero que sean "ortogonales" al sistema de traducción en cuestión, lo que significa que funciona independientemente de las sintetisas y de los ARNt endógenos al sistema de traducción. Las características deseadas del par ortólogo incluyen ARNt que decodifica o reconoce únicamente un codón específico, por ejemplo, un codón selector, que no es decodificado por ningún ARNt endógeno, y aminoacil-ARNt sintetisas que preferentemente aminoacilan (o "cargan") su ARNt afín únicamente con un aminoácido no natural específico. El O-ARNt normalmente tampoco es aminoacilado por sintetisas endógenas. Por ejemplo, en *E. coli*, un par ortogonal incluirá una aminoacil-ARNt sintetasa que no produzca una reacción cruzada con ninguno de los ARNt endógenos que, por ejemplo, en *E. coli* son 40, y un ARNt endógeno que no es aminoacilado por ninguna de las sintetisas endógenas de las que, por ejemplo, hay 21 en *E. coli*.

- En este documento se describen pares ortogonales para la codificación genética y la incorporación de alquínulo aminoácidos a proteínas en eubacterias, por ejemplo, en *E. coli*, en donde los componentes ortogonales no producen reacciones cruzadas con componentes endógenos de la maquinaria de traducción de la célula hospedadora de *E. coli*, pero reconocen el aminoácido no natural deseado y lo incorporan a proteínas en respuesta a un codón sinsentido ámbar, TAG. Los componentes ortogonales incluyen aminoacil-ARNt sintetisas ortogonales derivadas de la tirosil-ARNt sintetasa de *Methanococcus jannaschii*, y el supresor ámbar mutante tirosil-ARNt_{CUA}. En este sistema, las aminoacil-ARNt sintetisas mutantes aminoacilan el ARNt supresor con pPro-Phe y no con ninguno de los 20 aminoácidos habituales.

En este documento se describen métodos para identificar y producir pares ortogonales de ARNt/aminoacil-ARNt sintetasa adicionales, por ejemplo, pares O-ARNt/O-RS que se pueden usar para incorporar un alquínil aminoácido a una proteína. Un O-ARNt de la invención puede mediar en la incorporación de alquínil aminoácido a una proteína que está codificada por un polinucleótido, que comprende un codón selector que es reconocido por el O-ARNt, por ejemplo, *in vivo*. El bucle anti-codónico del O-ARNt reconoce el codón selector sobre un ARNm e incorpora su aminoácido, por ejemplo, un alquínil aminoácido en este sitio en el polipéptido. Una aminoacil-ARNt sintetasa ortogonal de la invención preferentemente aminoacila (o carga) su O-ARNt únicamente con un alquínil aminoácido específico.

Por ejemplo, como se demuestra en este documento, el alquínil aminoácido para-propargiloxifenilalanina (pPro-Phe; véase Figura 1A, estructura 1), que se puede dirigir para su modificación de una forma altamente selectiva, se incorporó selectiva y eficazmente a una proteína en una célula eubacteriana (*Escherichia coli*; *E. coli*) en respuesta a un codón selector, por ejemplo, el codón TAG. Una vez incorporado a una proteína, pPro-Phe se puede dirigir químicamente dentro de la célula, por ejemplo, se puede dirigir para la modificación con un colorante que lleva un grupo azido. El grupo azido sobre una molécula de colorante puede reaccionar con el alquínil aminoácido y dirigir la proteína para el marcaje con el colorante de una forma altamente selectiva.

La capacidad para incorporar a proteínas un alquínil aminoácido con especificidad de sitio puede facilitar el estudio de proteínas, así como permitir la manipulación genética de proteínas con nuevas propiedades. Por ejemplo, la expresión de proteínas que contienen alquínilo puede facilitar el estudio de proteínas mediante marcaje específico, alteración de la función catalítica de enzimas, entrecruzamiento de proteínas con otras proteínas, moléculas pequeñas y biomoléculas, etc.

ARNt ortogonal/aminoacil-ARNt sintetasa ortogonales y sus pares

Los sistemas de traducción adecuados para la preparación de proteínas que incluyen uno o más aminoácidos no naturales se describen, por ejemplo, en las Publicaciones internacionales número WO 2002/086075, titulada "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL-tRNA SYNTHETASE PAIRS"; WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS"; y WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE"; WO 2005/019415, presentada el 7 de julio de 2004; WO 2005/007870, presentada el 7 de julio de 2004 y WO 2005/007624, presentada el 7 de julio de 2004. Dichos sistemas de traducción generalmente comprenden células (que pueden ser células no eucariotas tales como *E. coli*, o células eucariotas tales como levaduras) que incluyen un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), y un aminoácido no natural (por ejemplo, en la presente invención, un alquínil aminoácido), en el que la O-RS aminoacila el O-ARNt con el alquínil aminoácido. Un par ortogonal de la invención incluye un O-ARNt, por ejemplo, un ARNt supresor, un ARNt con desplazamiento del marco de lectura, y similares, y una O-RS. En la invención también se proporcionan componentes individuales.

En general, cuando un par ortogonal reconoce un codón selector y carga un aminoácido en respuesta al codón selector, se dice que el par ortogonal "suprime" el codón selector. Es decir, un codón selector que no es reconocido por la maquinaria endógena del sistema de traducción (por ejemplo, de la célula) normalmente no se traduce, lo que puede producir el bloqueo de la producción de un polipéptido que de lo contrario se traduciría a partir del ácido nucleico. Un O-ARNt de la invención reconoce un codón selector que incluye una eficiencia de supresión de al menos, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 %, o un 90 % aproximadamente o superior en presencia de una sintetasa afín en respuesta a un codón selector en comparación con la eficiencia de supresión de un O-ARNt que comprende o que está codificado por una secuencia de polinucleótidos como se expone en el listado de secuencias de este documento. La O-RS aminoacila el O-ARNt con un aminoácido no natural de interés, tal como un alquínil aminoácido. La célula usa el par O-ARNt/O-RS para incorporar el aminoácido no natural a una cadena polipeptídica en crecimiento, por ejemplo, mediante un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende un codón selector que es reconocido por el O-ARNt. En ciertos aspectos deseables, la célula puede incluir un par O-ARNt/O-RS adicional, en el que el O-ARNt adicional es cargado mediante la O-RS adicional con un aminoácido no natural diferente. Por ejemplo, uno de los O-ARNt puede reconocer un codón de cuatro bases y el otro puede reconocer un codón de parada. De forma alternativa, varios codones de parada diferentes o varios codones de cuatro bases diferentes pueden reconocer específicamente diferentes codones selectores.

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

En una realización, la eficiencia de supresión de la O-RS y el O-ARNt juntos es, por ejemplo, de aproximadamente 5 veces, 10 veces, 15 veces, 20 veces, o 25 veces o superior más que la eficiencia de supresión del O-ARNt que carece de la O-RS. En un aspecto, la eficiencia de supresión de la O-RS y el O-ARNt juntos es de al menos, por

ejemplo, el 35 %, 40 %, 45 %, 50 %, 60 %, 75 %, 80 %, o 90 % aproximadamente o superior de la eficiencia de supresión de un par sintetasa ortogonal como se expone en el listado de secuencias de este documento.

- 5 Como se ha indicado, la invención opcionalmente incluye varios pares O-ARNt/O-RS en una célula u otro sistema de traducción, lo que permite la incorporación de más de un aminoácido no natural, por ejemplo, un alquil aminoácido u otro aminoácido no natural. Por ejemplo, la célula además puede incluir un par O-ARNt/O-RS adicional diferente y un segundo aminoácido no natural, en el que este O-ARNt adicional reconoce un segundo codón selector y este O-RS adicional preferentemente aminoacila el O-ARNt con el segundo aminoácido no natural. Por ejemplo, una célula que incluye un par O-ARNt/O-RS (en el que el O-ARNt reconoce, por ejemplo, un codón selector ámbar), además
- 10 puede comprender un segundo par ortogonal, por ejemplo, leucilo, lisilo, glutamilo, etc., (en el que el segundo O-ARNt reconoce un codón selector diferente, por ejemplo, un codón ópalo, un codón de cuatro bases, o similar). De forma deseable, los diferentes pares ortogonales se derivan de diferentes fuentes, lo que puede facilitar el reconocimiento de diferentes codones selectores.
- 15 El O-ARNt y/o la O-RS pueden ser de origen natural o se pueden derivar, por ejemplo, por mutación de un O-ARNt y/o RS de origen natural, por ejemplo, por generación de librerías de ARNt y/o librerías de RS, a partir de diferentes organismos y/o mediante el uso de cualquiera de una serie de estrategias de mutación disponibles. Por ejemplo, una estrategia para la producción de un par ortogonal ARNt/aminoacil-ARNt sintetasa supone la importación de un par ARNt/sintetasa heterólogo (a la célula hospedadora) a partir de, por ejemplo, una fuente distinta de la célula
- 20 hospedadora, o de varias fuentes, a la célula hospedadora. Las propiedades de la sintetasa candidata heteróloga incluyen, por ejemplo, que no carga ningún ARNt de la célula hospedadora, y las propiedades del ARNt candidato heterólogo incluyen, por ejemplo, que no resulta aminoacilado por ninguna sintetasa de la célula hospedadora. Además, el ARNt heterólogo es ortogonal a todas las sintetasas de la célula hospedadora.
- 25 Una segunda estrategia para generar un par ortogonal supone la generación de librerías mutantes a partir de las cuales detectar y/o seleccionar un O-ARNt u O-RS. Estas estrategias también se pueden combinar.

ARNt ortogonal (O-ARNt)

- 30 Un ARNt ortogonal (O-ARNt) de la invención de forma deseable media en la incorporación de un aminoácido no natural, tal como un alquil aminoácido, a una proteína que está codificada por un polinucleótido que comprende un codón selector que es reconocido por el O-ARNt, por ejemplo, *in vivo* o *in vitro*. En ciertas realizaciones, un O-ARNt de la invención incluye una eficiencia de supresión de al menos, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 %, o un 90 % aproximadamente o superior en presencia de una sintetasa afín en respuesta a un codón selector en comparación con un O-ARNt que comprende o que está codificado por una secuencia de polinucleótidos como se expone en las secuencias de O-ARNt en el listado de secuencias de este documento.
- 35

- La eficiencia de supresión se puede determinar mediante una serie de ensayos conocidos en la técnica. Por ejemplo, se puede usar un ensayo informador de β -galactosidasa, por ejemplo, un plásmido lacZ derivado (en el que la construcción tiene un codón selector en la secuencia de ácidos nucleicos de lacZ) se introduce en células de un organismo apropiado (por ejemplo, un organismo en el que se pueden usar los componentes ortogonales) junto con un plásmido que comprende un O-ARNt de la invención. También se podría introducir una sintetasa afín (ya sea como polipéptido o 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 0,5 aproximadamente, y los ensayos de β -galactosidasa se
- 40 llevan a cabo, por ejemplo, utilizando el kit de β -galactosidasa β -Galactosidase Assay Kit BetaFluor™ (Novagen). El porcentaje de supresión se puede calcular como porcentaje de actividad para una muestra con respecto a un control comparable, por ejemplo, el valor observado de la construcción lacZ derivada, en donde la construcción tiene un codón sentido correspondiente en la posición deseada en vez de un codón selector.
- 45

- 50 Ejemplos de O-ARNt se exponen en el listado de secuencias de este documento. Véase también, las tablas, ejemplos y figuras de este documento para secuencias de moléculas de O-ARNt y O-RS ilustrativas. Véase también, la sección titulada "Secuencia de ácidos nucleicos y polipéptidos y sus variantes" de este documento. En una molécula de ARN, tal como un ARNm de O-RS, o una molécula de O-ARNt, la timina (T) es sustituida por uracilo (U) con respecto a una secuencia determinada (o viceversa para un ADN codificante), o su complemento. También
- 55 puede haber presentes modificaciones adicionales a las bases.

- En este documento también se describen variaciones conservativas de los O-ARNt. Por ejemplo, variaciones conservativas de O-ARNt incluyen aquellas moléculas que funcionan como los O-ARNt particulares, por ejemplo, como en el listado de secuencias de este documento y que mantienen la estructura del ARNt con forma de L gracias a una auto-complementariedad adecuada, pero que no tienen una secuencia idéntica a aquellas, por ejemplo, en el
- 60 listado de secuencias, figuras o ejemplos de este documento (y, de forma deseable, son distintas de las moléculas de ARNt de tipo silvestre). Véase, la sección de este documento titulada "Secuencia de ácidos nucleicos y polipéptidos y sus variantes".

- 65 La composición que comprende un O-ARNt además puede incluir una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que la O-RS preferentemente aminoacila el O-ARNt con un aminoácido no natural tal como un alquil

aminoácido. En ciertas realizaciones, una composición que incluye un O-ARNt además puede incluir un sistema de traducción (por ejemplo, *in vitro* o *in vivo*). Un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende un codón selector que es reconocido por el O-ARNt, o puede haber presente en la célula una combinación de uno o más de estos. Véase también, la sección de este documento titulada “Aminoacil-ARNt sintetasas ortogonales”.

Los métodos de producción de un ARNt ortogonal (O-ARNt) también son una característica de la invención. También es una característica de la invención un O-ARNt producido por el método. En ciertas realizaciones de la invención, los O-ARNt se pueden producir generando una librería de mutantes. La librería de ARNt mutantes se puede generar usando diversas técnicas de mutagénesis conocidas en la materia. Por ejemplo, los ARNt mutantes se pueden generar mediante mutaciones específicas de sitio, mutaciones en puntos aleatorios, recombinación homóloga, barajado de ADN y otros métodos de mutagénesis recursiva, construcción de quimeras y cualquiera de sus combinaciones.

Se pueden introducir mutaciones adicionales en (una) posición(es) específica(s), por ejemplo, en (una) posición(es) no conservativa(s), o en una posición conservativa, en (una) posición(es) aleatoria(s), o una combinación de ambas en un bucle o una región deseada de un ARNt, por ejemplo, un bucle anti-codónico, el tallo aceptor, el brazo o bucle D, el bucle variable, el brazo o bucle TPC, otras regiones de la molécula de ARNt, o una de sus combinaciones. Normalmente, las mutaciones en un ARNt incluyen la mutación del bucle anti-codónico de cada miembro de la librería de ARNt mutantes para permitir el reconocimiento de un codón selector. El método además puede incluir la adición de una secuencia adicional (CCA) a un extremo del O-ARNt. Normalmente, un O-ARNt posee una mejora de la ortogonalidad para un organismo deseado en comparación con el material de partida, por ejemplo, la pluralidad de secuencias de ARNt, mientras preserva su afinidad hacia una RS deseada.

Los métodos opcionalmente incluyen analizar la similitud (y/o o la homología inferida) de secuencias de ARNt y/o aminoacil-ARNt sintetasas para determinar candidatos potenciales para un O-ARNt, O-RS y/o sus pares, que parecen ser ortogonales para un organismo específico. Para el análisis se pueden usar programas de ordenador conocidos en la técnica y descritos en este documento, por ejemplo, se pueden usar los programas BLAST y Pileup. En un ejemplo, para seleccionar los posibles componentes de traducción ortogonales para su uso en *E. coli*, se selecciona una sintetasa y/o un ARNt que no presenten una similitud de secuencia próxima a organismos eubacterianos.

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

En ciertas realizaciones, en la selección negativa, se introduce un codón(es) selector(es) en un polinucleótido que codifica un marcador de selección negativo, 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 una barnasa, en una posición no esencial (por ejemplo, aun así produciendo una barnasa funcional), etc. La detección/selección se realiza opcionalmente mediante el crecimiento de la población de células en presencia de un agente selectivo (por ejemplo, un antibiótico, tal como ampicilina). En una realización, se modifica la concentración del agente de selección.

Por ejemplo, para medir la actividad de ARNt supresores, se usa un sistema de selección basado en la supresión *in vivo* del codón selector, por ejemplo, mutaciones sinsentido o de desplazamiento del marco de lectura introducidas en un polinucleótido que codifica un marcador de selección negativo, por ejemplo, un gen para la β -lactamasa (*bla*). Por ejemplo, se construyen variantes de polinucleótidos, por ejemplo, variantes de *bla*, con un codón selector en una determinada posición (por ejemplo, A184). Las células, por ejemplo, bacterias, se transforman con estos polinucleótidos. En el caso de un ARNt ortogonal, que no puede ser cargado eficientemente por las sintetasas endógenas de *E. coli*, la resistencia a antibióticos, por ejemplo, resistencia a ampicilina, debe ser parecida o inferior a la de una bacteria transformada sin plásmido. Si el ARNt no es ortogonal, o si en el sistema se co-expresa una sintetasa heteróloga capaz de cargar el ARNt, se puede observar un mayor nivel de resistencia a antibióticos, por ejemplo, ampicilina. Las células, por ejemplo, bacterias, se seleccionan para que no puedan crecer sobre placas de agar LB con concentraciones de antibiótico aproximadamente iguales a las células transformadas sin plásmidos.

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

En una realización, el reservorio de ARNt que son ortogonales a un organismo deseado a continuación se somete a una selección positiva en la que se coloca un codón selector en un marcador de selección positivo, por ejemplo,

codificado por un gen de resistencia a fármacos, tal como un gen de β -lactamasa. La selección positiva se lleva a cabo sobre una célula que comprende un polinucleótido que codifica o que comprende un miembro de un reservorio de ARNt que son ortogonales a la célula, un polinucleótido que codifica un marcador de selección positivo, y un polinucleótido que codifica una RS afín. En ciertas realizaciones, la segunda población de células comprende células que no fueron eliminadas por la selección negativa. Los polinucleótidos se expresan en la célula y la célula se crece en presencia de un agente de selección, por ejemplo, ampicilina. A continuación los ARNt se seleccionan por su capacidad para ser aminoacilados por la sintetasa afín co-expresada y para insertar un aminoácido en respuesta a este codón selector. Normalmente, estas células muestran una mejora en la eficiencia de supresión en comparación con las células que albergan ARNt no funcional(es), o ARNt que no pueden ser reconocidos eficazmente por la sintetasa de interés. La célula que alberga los ARNt no funcionales o ARNt que no son reconocidos eficazmente por la sintetasa de interés, son sensibles a los antibióticos. Por tanto, los ARNt que: (i) no son sustratos para las sintetasas del hospedador endógeno, por ejemplo, *Escherichia coli*; (ii) pueden ser aminoacilados por la sintetasa de interés; y (iii) son funcionales en la traducción, sobreviven a ambas selecciones.

Por consiguiente, el mismo marcador puede ser un marcador positivo o negativo, dependiendo del contexto en el que se detecte. Es decir, el marcador es un marcador positivo si hay una detección favorable, pero un marcador negativo si hay una detección en contra.

La rigurosidad de la selección, por ejemplo, la selección positiva, la selección negativa o tanto la selección positiva como la selección negativa, en los métodos descritos anteriormente, opcionalmente incluye variar la rigurosidad de la selección. Por ejemplo, debido a que la barnasa es una proteína extremadamente tóxica, la rigurosidad de la selección negativa se puede controlar introduciendo una serie de diferentes codones selectores en el gen de la barnasa y/o mediante el uso de un promotor inducible. En otro ejemplo, la concentración del agente de selección o de detección se modifica (por ejemplo, concentración de ampicilina). En un aspecto de la invención, la rigurosidad se modifica debido a que la actividad deseada puede ser baja durante las primeras rondas. Así, se aplican criterios de selección menos rigurosos en las primeras rondas y se aplican criterios más rigurosos en las últimas rondas de selección. En ciertas realizaciones, la selección negativa, la selección positiva o tanto la selección negativa como la selección positiva se pueden repetir varias veces. Se pueden usar múltiples marcadores de selección negativa, marcadores de selección positiva o marcadores de selección tanto negativa como positiva diferentes. En ciertas realizaciones, el marcador de selección positiva y negativa puede ser el mismo.

Se pueden usar otros tipos de selecciones/detecciones en la invención para producir componentes de traducción ortogonales, por ejemplo, un O-ARNt, una O-RS, y un par O-ARNt/O-RS que carga un aminoácido no natural tal como un alquil aminoácido en respuesta a un codón selector. Por ejemplo, el marcador de selección negativa, el marcador de selección positiva o los marcadores de selección tanto positiva como negativa pueden incluir un marcador que emite fluorescencia o cataliza una reacción luminiscente en presencia de un reactivo adecuado. En otra 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 detección basado en afinidad. Véase también, Francisco, J. A., y col., (1993) Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A. 90:10444-8.

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

Aminoacil-ARNt sintetasa ortogonal (O-RS)

Una O-RS de la invención preferentemente aminoacila un O-ARNt con un aminoácido no natural tal como un alquil aminoácido, por ejemplo, para-propargiloxifenilalanina, *in vitro* o *in vivo*. Se puede proporcionar una O-RS de la invención al sistema de traducción, por ejemplo, una célula, mediante un polipéptido que incluya una O-RS y/o mediante un polinucleótido que codifica una O-RS o un fragmento de la misma. Por ejemplo, una O-RS ilustrativa comprende una secuencia de aminoácidos como la que se expone en el listado de secuencias y los ejemplos de este documento, o una de sus variaciones conservativas. En otro ejemplo, una O-RS, o un fragmento de la misma, está codificada por una secuencia de polinucleótidos que codifica un aminoácido que comprende una secuencia en el listado de secuencias o ejemplos de este documento, o una de sus secuencias de polinucleótidos complementarias. Véase, por ejemplo, las tablas y ejemplos de este documento para las secuencias de moléculas de O-RS ilustrativas. Véase también, la sección titulada "Secuencia de ácidos nucleicos y polinucleótidos y sus variantes" del presente documento.

También son una característica de la invención los 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 método incluye someter a selección,

por ejemplo, selección positiva, a una población de células de una primera especie, en la que las células comprenden individualmente: 1) un miembro de una pluralidad de aminoacil-ARNt sintetasas (RS), (por ejemplo, la pluralidad de RS puede incluir RS mutantes, RS derivadas de una especie distinta de la primera especie o tanto RS mutantes como RS derivadas de una especie distinta de la primera especie); 2) el ARNt ortogonal (O-ARNt) (por ejemplo, de una o más especies); y 3) un polinucleótido que codifica un marcador de selección (por ejemplo, positiva) y comprende al menos un codón selector. Las células son seleccionadas o detectadas por aquellas que muestran una mejora en la eficiencia de supresión en comparación con células que carecen o con una cantidad reducida del miembro de la pluralidad de RS. La eficiencia de supresión se puede medir mediante técnicas conocidas en la materia y como se describe en este documento. Las células que tienen una mejora en la eficiencia de supresión comprenden una RS activa que aminoacila el O-ARNt. El nivel de aminoacilación (*in vitro* o *in vivo*) mediante la RS activa de un primer grupo de ARNt procedente de la primera especie se compara con el nivel de aminoacilación (*in vitro* o *in vivo*) mediante la RS activa de un segundo grupo de ARNt procedente de la segunda especie. El nivel de aminoacilación se puede determinar mediante una sustancia detectable (por ejemplo, un aminoácido marcado o un aminoácido no natural, por ejemplo, una para-propargiloxifenilalanina marcada). Normalmente se selecciona la RS activa que aminoacila más eficientemente el segundo grupo de ARNt comparado con el primer grupo de ARNt, proporcionando así una aminoacil-ARNt sintetasa ortogonal eficiente (optimizada) para su uso con el O-ARNt. También es una característica de la invención una O-RS, identificada por el método.

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

La O-RS identificada se puede manipular adicionalmente para alterar la especificidad de sustrato de la sintetasa, de manera que sólo se cargue al O-ARNt un aminoácido no natural deseado, por ejemplo, un alquil aminoácido, pero no cualquiera de los 20 aminoácidos habituales. Los métodos para generar una aminoacil-ARNt sintetasa ortogonal con una especificidad de sustrato para un aminoácido no natural incluyen la mutación de la sintetasa, por ejemplo, en el sitio activo de la sintetasa, en el sitio del mecanismo de edición en la sintetasa, en diferentes sitios mediante la combinación de diferentes dominios de sintetasas, o similares, y la aplicación de un proceso de selección. Se usa una estrategia, que se basa en la combinación de una selección positiva seguida de una selección negativa. En la selección positiva, la supresión del codón selector introducido en una posición(es) no esencial de un marcador positivo permite que las células sobrevivan bajo la presión de la selección positiva. En presencia de aminoácidos tanto naturales como no naturales, los supervivientes por tanto codifican 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(es) no esencial de un marcador negativo retira sintetasas con especificidades para los aminoácidos naturales. Los supervivientes de la selección negativa y positiva codifican sintetasas que aminoacilan (cargan) el ARNt supresor ortogonal únicamente con aminoácidos no naturales. A continuación estas sintetasas se pueden someter a mutagénesis adicional, por ejemplo, barajado de ADN u otros métodos de mutagénesis recursivos.

Se puede generar una librería de O-RS mutantes usando diversas técnicas de mutagénesis conocidas en la materia. Por ejemplo, las RS mutantes se pueden generar mediante mutaciones específicas de sitio, mutaciones de punto aleatorio, recombinación homóloga, barajado de ADN u otros métodos de mutagénesis recursiva, construcción de quimeras o cualquiera de sus combinaciones. Por ejemplo, se puede producir una librería de RS mutantes a partir de otras dos o más "sublibrerías" menos diversas, por ejemplo, más pequeñas. Las librerías quiméricas de RS también están incluidas en la invención. Cabe señalar que para pares ortogonales opcionalmente se construyen y se detectan las librerías de ARNt sintetasas procedentes de varios organismos (por ejemplo, microorganismos tales como eubacterias o arqueobacterias) tales como librerías que comprenden diversidad natural (véase, por ejemplo, Patente de Estados Unidos nº 6.238.884 de Short y col.; Patente de Estados Unidos nº 5.756.316 de Schallenger y col.; Patente de Estados Unidos nº 5.783.431 de Petersen y col.; Patente de Estados Unidos nº 5.824.485 de Thompson y col.; Patente de Estados Unidos nº 5.958.672 de Short y col.).

Una vez que las sintetasas se someten a la estrategia de selección/detección positiva y negativa, a continuación estas sintetasas se pueden someter a mutagénesis adicional. Por ejemplo, se puede aislar un ácido nucleico que codifica la O-RS; a partir del ácido nucleico se puede generar un grupo de polinucleótidos que codifican O-RS mutadas (por ejemplo, mediante mutagénesis aleatoria, mutagénesis específica de sitio, recombinación o cualquiera de sus combinaciones); y se pueden repetir estas etapas individuales o una combinación de estas etapas hasta que se obtenga una O-RS mutada que preferentemente aminoacila el O-ARNt con el aminoácido no natural, por ejemplo, un alquil aminoácido. En un aspecto de la invención; estas etapas se llevan a cabo varias veces, por ejemplo, al menos dos veces.

En los métodos de la invención también se pueden usar niveles adicionales de rigurosidad en la selección/detección, para producir O-ARNt, O-RS, o sus pares. La rigurosidad en la selección o detección se puede modificar sobre una o

las dos etapas del método para producir una O-RS. Esto puede incluir, por ejemplo, variar la cantidad de agente de selección/detección usado, etc. También se pueden llevar a cabo rondas adicionales de selección positiva y/o negativa. La selección o detección también puede comprender uno o más de un cambio en la permeabilidad del aminoácido, un cambio en la eficiencia de traducción, un cambio en la fidelidad de traducción, etc. Normalmente, el

5 uno o más cambios se basan en una mutación de uno o más genes en un organismo en el que se usa un par ARNt/ARNt sintetasa ortogonal para producir una proteína.

Detalles generales adicionales para producir O-RS, y alterar la especificidad de sustrato de la sintetasa se pueden encontrar en la Publicación internacional número WO 2002/086075, titulada "METHODS AND COMPOSITIONS FOR THE PRODUCTION OF ORTHOGONAL tRNA AMINOACYL-tRNA SYNTHETASE PAIRS"; y WO

10 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE".

Fuente y organismo hospedador

Los componentes de traducción ortogonales (O-ARNt y O-RS) de la invención pueden derivar de cualquier organismo (o combinación de organismos) para su uso en un sistema de traducción de un hospedador de cualquier otra especie, con la salvedad de que los componentes O-ARNt/O-RS y el sistema hospedador funcionen de manera ortogonal. No es un requisito que el O-ARNt y la O-RS deriven del mismo organismo. En un aspecto, los

15 componentes ortogonales derivan de genes de arqueas (es decir, arqueobacterias) para su uso en un sistema hospedador eubacteriano.

Por ejemplo, el O-ARNt ortogonal puede derivar de un organismo arquea, 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*, *Aeuryopyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei* (Mm), *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus* (Ss), *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o similares, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares, mientras que la O-RS ortogonal puede derivar de un organismo o combinación de organismos, por ejemplo, una arqueobacteria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, Halobacterium tal como *Haloferax volcanii* y especies de Halobacterium NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuryopyrum pernix*, *Methanococcus maripaludis*, *Methanopyrus kandleri*, *Methanosarcina mazei*, *Pyrobaculum aerophilum*, *Pyrococcus abyssi*, *Sulfolobus solfataricus*, *Sulfolobus tokodaii*, *Thermoplasma acidophilum*, *Thermoplasma volcanium*, o similares, o una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares. En una

25 realización, también se pueden usar fuentes eucariotas, por ejemplo, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos, artrópodos, etc.) o similares como fuentes de O-ARNt y O-RS.

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

40

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

45

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

50

Codones selectores

Los codones selectores de la invención expanden el marco de lectura del codón genético de la maquinaria biosintética de la proteína. Por ejemplo, un codón selector incluye, por ejemplo, un codón único de tres bases, un codón sinsentido, tal como un codón de parada, por ejemplo, un codón ámbar (UAG), o un codón ópalo (UGA), un codón no natural, un codón de al menos cuatro bases, un codón raro, o similares. En un gen deseado se puede introducir una serie de codones selectores, por ejemplo, uno o más, dos o más, más de tres, etc. Al usar codones selectores diferentes se pueden utilizar múltiples pares ARNt/sintetasa ortogonales para permitir la incorporación simultánea específica de sitio de varios aminoácidos no naturales, por ejemplo, incluido al menos un alquil aminoácido, usando estos codones selectores diferentes.

55

60

En una realización, los métodos suponen el uso de un codón selector que es un codón de parada para la incorporación de un alquil aminoácido *in vivo* en una célula. Por ejemplo, se produce un O-ARNt que reconoce el codón de parada y se aminoacila mediante una O-RS con un alquil aminoácido. Este O-ARNt no es reconocido por las aminoacil-ARNt sintetasa naturales del hospedador. Se puede usar mutagénesis dirigida de sitio convencional

65

para introducir el codón de parada en el sitio de interés en un polinucleótido que codifica un polipéptido de interés. Véase, por ejemplo, Sayers, J.R., y col. (1988), 5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*, 791-802. Cuando la O-RS, el O-ARNt y el ácido nucleico que codifica un polipéptido de interés se combinan, por ejemplo, *in vivo*, el alquil aminoácido se incorpora en respuesta al codón de parada para proporcionar un polipéptido que contiene el alquil aminoácido activo en la posición especificada. En una realización de la invención, el codón de parada usado como codón selector es un codón ámbar, UAG, y/o un codón ópalo UGA. En un ejemplo, un código genético en el que se usan tanto UAG como UGA como codones selectores puede codificar 22 aminoácidos al tiempo que preserva el codón sinsentido ocre, UAA, que es la señal de terminación más abundante.

La incorporación de alquil aminoácidos activos *in vivo* se puede realizar sin una perturbación significativa de la célula hospedadora. Por ejemplo, en células no eucariotas, tales como *Escherichia coli*, debido a que la eficiencia de supresión para el codón UAG depende de la competencia entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y el factor de liberación 1 (RF1) (que se une al codón UAG e inicia la liberación del péptido en crecimiento a partir del ribosoma), la eficiencia de supresión se puede modular, por ejemplo, incrementando el nivel de expresión de O-ARNt, por ejemplo, el ARNt supresor, o usando una cepa deficiente en RF1. En células eucariotas, debido a que la eficiencia de supresión para el codón UAG depende de la competencia entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y un factor de liberación de eucariotas (por ejemplo, eRF) (que se une a un codón de parada e inicia la liberación del péptido en crecimiento a partir del ribosoma), la eficiencia de supresión se puede modular, por ejemplo, incrementando el nivel de expresión de O-ARNt, por ejemplo, el ARNt supresor. Además, también puede haber presentes compuestos adicionales, por ejemplo, agentes reductores tales como ditiotritol (DTT).

Los alquil aminoácidos también pueden estar codificados con codones raros. Por ejemplo, cuando se reduce la concentración de arginina en una reacción de síntesis de proteínas *in vitro*, el codón raro de arginina, AGG, ha demostrado ser eficiente para la inserción de Ala mediante un ARNt sintético acilado con alanina. Véase, por ejemplo, Ma y col., *Biochemistry*, 32:7939 (1993). En este caso, el ARNt sintético compete con el ARNtArg de origen natural, que existe como una especie minoritaria en *Escherichia coli*. Además, algunos organismos no usan todos los codones tripletes. En *Micrococcus luteus* se ha utilizado un codón AGA sin asignar para la inserción de aminoácidos en un extracto de transcripción/traducción *in vitro*. Véase, por ejemplo, Kowal y Oliver, *Nucl. Acid. Res.*, 25:4685 (1997). Los componentes de la invención se pueden generar para el uso de estos codones raros *in vivo*.

Los codones selectores también pueden comprender codones extendidos, por ejemplo, codones de cuatro o más bases, tales como codones de cuatro, cinco, seis o más bases. Ejemplos de codones de cuatro bases incluyen, por ejemplo, AGGA, CUAG, UAGA, CCCU, y similares. Ejemplos de codones de cinco bases incluyen, por ejemplo, AGGAC, CCCCC, CCCUC, CUAGA, CUACU, UAGGC y similares. Los métodos de la invención incluyen el uso de codones extendidos basados en la supresión del desplazamiento del marco de lectura. Codones de cuatro o más bases pueden insertar, por ejemplo, uno o múltiples aminoácidos no naturales tales como un alquil aminoácido, en la misma proteína. En otras realizaciones, los bucles anti-codónicos pueden decodificar, por ejemplo, un codón de al menos cuatro bases, al menos un codón de cinco bases, o al menos un codón de seis bases o superior. Puesto que existen 256 posibles codones de cuatro bases, usando un codón de cuatro o más bases en la misma célula se pueden codificar múltiples aminoácidos no naturales. Véase también, Anderson y col., (2002) Exploring the Limits of Codon and Anticodon Size, *Chemistry and Biology*, 9:237-244; y, Magliery, (2001) Expanding the Genetics Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in *Escherichia coli*, *J. Mol. Biol.* 307: 755-769.

Por ejemplo, se han usado codones de cuatro bases para incorporar aminoácidos no naturales a proteínas usando métodos biosintéticos *in vitro*. Véase, por ejemplo, Ma y col., (1993) *Biochemistry*, 32:7939; y Hohsaka y col., (1999) *J. Am. Chem. Soc.*, 121:34. Se han usado CGGG y AGGU para incorporar simultáneamente 2-naftilalanina y un derivado NBD de lisina a estreptavidina *in vitro* con dos ARNt supresores químicamente acilados con desplazamiento del marco de lectura. Véase, por ejemplo, Hohsaka y col., (1999) *J. Am. Chem. Soc.*, 121:12194. En un estudio *in vivo*, Moore y col. examinaron la capacidad de derivados de ARNt^{Leu} con anticodones NCUA para suprimir codones UAGN (N puede ser U, A, G, o C), y comprobaron que el cuadruplete UAGA puede ser decodificado por un ARNt^{Leu} con un anticodón UCUA con una eficiencia del 13 al 26 % con poca decodificación en el marco 0 o -1. Véase Moore y col., (2000) *J. Mol. Biol.*, 298:195. En una realización, en la invención se pueden usar codones extendidos basados en codones raros o codones sinsentido, que pueden reducir la lectura sinsentido a través del codón de parada y la supresión del desplazamiento del marco de lectura en otros sitios no deseados.

Para un sistema determinado, un codón selector también puede incluir uno de los codones naturales de tres bases, en el que el sistema endógeno no usa (o raramente usa) el codón natural de 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 de tres bases es un codón raro.

Los codones selectores opcionalmente incluyen pares de bases no naturales. Estos pares de bases no naturales expanden adicionalmente el alfabeto genético existente. Un par extra de bases incrementa el número de codones tripletes de 64 a 125. Las propiedades de los pares de tres bases incluyen un apareamiento de bases estables selectivo, una incorporación enzimática eficiente al ADN con una alta fidelidad mediante una polimerasa, y la

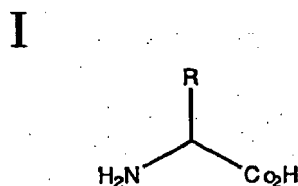
extensión del cebador continua y eficiente después de la síntesis de un par no natural de bases naciente. Descripciones de pares de bases no naturales que se pueden adaptar para los métodos y composiciones incluyen, por ejemplo, Hirao, y col., (2002) An unnatural base pair for incorporating amino acid analogues into protein, *Nature Biotechnology*, 20:177-182. Véase también Wu, Y., y col., (2002) *J. Am. Chem. Soc.* 124:14626-14630. A continuación se enumeran otras publicaciones relevantes.

Para su uso *in vivo*, el nucleósido no natural es permeable a la membrana y se fosforila para formar el trifosfato correspondiente. Además, la información genética añadida es estable y no es destruida por las enzimas celulares. Esfuerzos previos por parte de Benner y otros se aprovechan de los patrones de los puentes de hidrógeno que son diferentes de los de los pares canónicos de Watson-Crick, cuyo ejemplo más notable es el par iso-C:iso-G. Véase, por ejemplo, Switzer y col., (1989) *J. Am. Chem. Soc.*, 111:8322; y Piccirilli y col., (1990) *Nature*, 343:33; Kool, (2000) *Curr. Opin. Chem. Biol.*, 4:602. En general estas bases experimentan un cierto grado de apareamientos erróneos con las bases naturales y no se pueden replicar enzimáticamente. Kool y sus colaboradores han demostrado que las interacciones de empaquetamiento hidrófobo entre bases pueden sustituir a los puentes de hidrógeno para dar lugar a la formación del par de bases. Véase Kool, (2000) *Curr. Opin. Chem. Biol.*, 4:602; y Guckian y Kool, (1998) *Angew. Chem. Int. Ed. Engl.*, 36, 2825. En un esfuerzo por desarrollar un par de bases no naturales que satisfaga todos los requerimientos anteriores, Schultz, Romesberg y sus colaboradores han sintetizado y estudiado sistemáticamente una serie de bases hidrófobas no naturales. Se comprobó que un auto-par PICS:PICS era más estable que los pares de bases naturales, y se puede incorporar eficazmente al ADN mediante el fragmento Klenow de la ADN polimerasa I de *Escherichia coli* (KF). Véase, por ejemplo, McMinn y col., (1999) *J. Am. Chem. Soc.*, 121:11586; y Ogawa y col., (2000) *J. Am. Chem. Soc.*, 122:3274. Se puede sintetizar un auto-par 3MN:3MN mediante el KF con una eficiencia y una selectividad suficientes para su función biológica. Véase, por ejemplo, Ogawa y col., (2000) *J. Am. Chem. Soc.*, 122:8803. No obstante, las dos bases actúan como terminadores de la cadena para la replicación posterior. Recientemente se ha desarrollado una ADN polimerasa mutante que se puede usar para replicar el auto-par PICS. Además, se puede replicar un auto-par 7AI. Véase, por ejemplo, Tae y col., (2001) *J. Am. Chem. Soc.*, 123:7439. También se ha desarrollado un nuevo par de metalobase, Dipic:Py, que forma un par estable tras la unión de Cu (II). Véase, Meggers y col., (2000) *J. Am. Chem. Soc.*, 122:10714. Debido a que los codones extendidos y los codones no naturales son intrínsecamente ortogonales a los codones naturales, los métodos de la invención pueden aprovechar esta propiedad para generar ARNt ortogonales para ellos.

También se puede usar un sistema de salto de la traducción para incorporar un alquínil aminoácido a un polipéptido deseado. En un sistema de salto de la traducción, se inserta una secuencia grande en un gen pero no se traduce en una proteína. La secuencia contiene una estructura que sirve como indicación para inducir al ribosoma a saltarse la secuencia y reanudar la traducción aguas abajo de la inserción.

Aminoácidos no naturales

Como se usa en este documento, un aminoácido no natural se refiere a cualquier aminoácido, aminoácido modificado, o análogo de aminoácido distinto de la selenocisteína y/o pirrolisina y los siguientes 20 α -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 α -aminoácido se ilustra en la Fórmula I:



Un aminoácido no natural normalmente es cualquier estructura que tenga la Fórmula I en la que el grupo R es cualquier sustituyente distinto del usado en los 20 aminoácidos naturales. Véase, por ejemplo, *Biochemistry* de L. Stryer, 3ª ed. 1988, Freeman and Company, Nueva York, para las estructuras de los 20 aminoácidos naturales. Cabe señalar que los aminoácidos no naturales de la invención pueden ser compuestos de origen natural distintos de los 20 α -aminoácidos anteriores.

Debido a que los aminoácidos no naturales de la invención normalmente difieren de los aminoácidos naturales en la cadena lateral, los aminoácidos no naturales forman enlaces amida con otros aminoácidos, por ejemplo, naturales o no naturales, de la misma forma en que se forman en las proteínas de origen natural. No obstante, los aminoácidos no naturales tienen grupos en las cadenas laterales que los diferencian de los aminoácidos naturales.

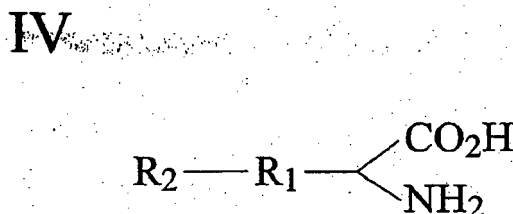
En este documento son de interés particular los aminoácidos no naturales que comprenden un grupo alquínilo reactivo, por ejemplo, un aminoácido no natural que comprende un resto alquino que reacciona específica y regioselectivamente con un resto azido. Por ejemplo, en un alquínil aminoácido, R en la Fórmula I incluye cualquier

estructura que contenga alquino. Por ejemplo, la para-propargiloxifenilalanina (abreviada por pPro-Phe; véase **Figura 1A**) es un alquinil aminoácido no natural deseado que tiene su utilidad en la invención. No se pretende que la invención esté limitada al uso de pPro-Phe con componentes de traducción ortogonales. Por ejemplo, están contemplados otros varios alquinil aminoácidos (véase **Figuras 8A y 8B**), incluyendo, pero no limitado a, por

ácido 2-amino-4-pentinoico
 ácido 2-amino-3-(4-etinilfenil)propanoico
 ácido 2-amino-3-[4-(prop-2-inil)fenil]propanoico
 ácido 2-amino-3-(prop-2-inilo)propanoico
 ácido 2-amino-3-(prop-2-inil)propanoico
 ácido 3-[(prop-2-inilo)carbonil]-2-aminopropanoico
 ácido 4-[(prop-2-inilo)carbonil]-2-aminobutanoico

En otros aminoácidos no naturales, por ejemplo, R en la Fórmula I opcionalmente comprende un alquilo, arilo, acilo, hidracina, ciano, halo, hidracida, alqueno, éter, borato, boronato, fosfo, fosfono, fosfina, enona, imina, éster, hidroxilamina, amina, y similares, o cualquiera de sus combinaciones. Otros aminoácidos no naturales de interés incluyen, pero no están limitados a, aminoácidos que comprenden un reticulante fotoactivable, aminoácidos con espín marcado, aminoácidos fluorescentes, aminoácidos unidos a metales, aminoácidos que contienen metales, aminoácidos radiactivos, aminoácidos con nuevos grupos funcionales, aminoácidos que interaccionan covalente o no covalentemente con otras moléculas, aminoácidos fotoseparables y/o fotoisomerizables, aminoácidos que contienen biotina o análogos de biotina, aminoácidos que contienen grupos ceto, aminoácidos glicosilados, un resto sacarídico unido a la cadena lateral del aminoácido, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles o fotoescindibles, aminoácidos con una cadena lateral elongada en comparación con aminoácidos naturales (por ejemplo, poliéteres o hidrocarburos de cadena larga, por ejemplo, superior a 5 aproximadamente, superior a 10 carbonos aproximadamente, etc.), aminoácidos que contienen azúcares unidos a carbono, aminoácidos que contienen amino

En otro aspecto, la invención proporciona alquinil aminoácidos que tienen la estructura general ilustrada por la Fórmula IV a continuación:

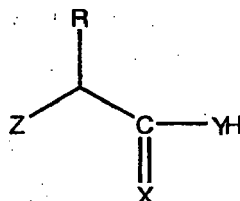


Un alquino aminoácido que tiene esta estructura normalmente es cualquier estructura en la que R₁ es un sustituyente usado en uno de los 20 aminoácidos naturales y R₂ es un sustituyente alquinilo. Así, este tipo de alquinil aminoácido se puede considerar un derivado de un aminoácido natural.

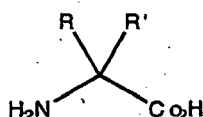
Como se ha indicado anteriormente, no se pretende que la invención esté limitada al uso del alquinil aminoácido no natural para-propargiloxifenilalanina (pPro-Phe). De hecho, cualquier alquinil aminoácido que se pueda usar en un sistema de traducción ortogonal de la invención en una eubacteria está dentro del alcance de la invención. Se conocen diversos otros alquinil aminoácidos, por ejemplo, los alquinil aminoácidos suministrados en la Figura 8. Puesto que algunas de estas estructuras de alquinil aminoácidos son muy similares a la pPro-Phe, se contempla que algunos de estos aminoácidos se puedan incorporar a proteínas en eubacterias usando los componentes ortogonales ARNt y aminoacil-ARNt sintetasa suministrados en este documento, por ejemplo, el O-ARNt de la SEQ ID NO: 1 y la O-RS de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, o sus variantes conservativas. Así, la invención también proporciona métodos para la incorporación de otros alquinil aminoácidos además de la pPro-Phe. Independientemente de si los componentes ortogonales proporcionados en la Tabla 4 (véase Ejemplo 9) son capaces de incorporar alquinil aminoácidos además de la pPro-Phe, la divulgación proporciona enseñanzas suficientes para construir componentes de ARNt ortogonales que incorporen estos otros alquinil aminoácidos, y además esos componentes ortogonales están dentro del alcance de la presente invención.

Además de aminoácidos no naturales que contienen nuevas cadenas laterales tales como el grupo alquinilo, los alquinil aminoácidos no naturales opcionalmente también pueden comprender estructuras con la cadena principal modificada, por ejemplo, como se ilustra mediante las estructuras de las Fórmulas II y III:

II



III



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

Por ejemplo, muchos aminoácidos no naturales (incluyendo algunos alquil aminoácidos) están basados en aminoácidos naturales, tales como la tirosina, serina, cisteína, aspartato, glutamato, y similares. Por ejemplo, los alquil aminoácidos:

ácido 2-amino-3-(prop-2-inilo)propanoico;
 ácido 2-amino-3-(prop-2-inil)propanoico;
 ácido 3-[(prop-2-inilo)carbonil]-2-aminopropanoico; y
 ácido 4-[(prop-2-inilo)carbonil]-2-aminobutanoico,

pueden ser todos derivados de aminoácidos naturales.

Los análogos de tirosina incluyen tirosinas para-sustituídas, tirosinas orto-sustituídas y tirosinas meta-sustituídas, en los que la tirosina sustituida comprende un grupo alquil, 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 políeter, un grupo nitro, o similares. Además, también están contemplados anillos arilo multi-sustituídos. Los análogos de glutamina de la invención incluyen, pero no están limitados a, los α-hidroxi derivados, derivados γ-sustituídos, derivados cíclicos y derivados de glutamina sustituidos con amida. Los análogos ilustrativos de fenilalanina incluyen, pero no están limitados a, fenilalaninas para-sustituídas, fenilalaninas orto-sustituídas, y fenilalaninas meta-sustituídas, en los que el sustituyente comprende un grupo alquil, un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído, un grupo nitro, un grupo tiol, o un grupo ceto, o similares. Ejemplos específicos de aminoácidos no naturales incluyen, pero no están limitados a, una p-propargiloxifenilalanina, una 3,4-dihidroxi-L-fenilalanina (DHP), una 3,4,6-trihidroxi-L-fenilalanina, una 3,4,5-trihidroxi-L-fenilalanina, 4-nitro-fenilalanina, una p-acetil-L-fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metilfenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una nitro-3-tirosina, una tiol-3-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 fosfonoserina, una fosfonotirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina, y una isopropil-, L-fenilalanina y similares. En este documento se proporcionan las estructuras de diversos aminoácidos no naturales, véase, por ejemplo, las **Figuras 1A, 8A y 8B**. Véase también la Solicitud internacional publicada WO 2004/094593, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE".

Síntesis Química de Aminoácidos no naturales

Muchos de los aminoácidos no naturales proporcionados más arriba están disponibles en el mercado, por ejemplo, en Sigma (EE.UU.) o Aldrich (Milwaukee, WI, EE.UU.). Aquellos que no estén disponibles en el mercado
 5 opcionalmente se sintetizan como se proporciona en diversas publicaciones usando métodos convencionales conocidos por los expertos en la materia. Para técnicas de síntesis orgánica, véase, por ejemplo, Organic Chemistry de Fessenden and Fessenden, (1982, 2ª Edición, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry de March (3ª Edición, 1985, Wiley and Sons, Nueva York); y Advanced Organic Chemistry de Carey and Sundberg (3ª Edición, Partes A y B, 1990, Plenum Press, Nueva York). Publicaciones adicionales que describen la síntesis de
 10 aminoácidos no naturales incluyen, por ejemplo, el documento WO 2002/085923 titulado "*In vivo* incorporation of Unnatural Amino Acids"; Matsoukas y col., (1995) J. Med. Chem., 38, 4660-4669; King, F.E. & Kidd, D.A.A. (1949) A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319; Friedman, O.M. & Chatterji, R. (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750-3752; Craig, J.C. y col. (1988) Absolute Configuration of the
 15 Enantiomers of 7-Chloro-4-[[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmon, M. & Frappier, F. (1991) Glutamine analogues as Potential Antimalarials, Eur. J. Med. Chem. 26, 201-5; Koskinen, A.M.P. & Rapoport, H. (1989) Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859-1866; Christie, B.D. & Rapoport, H. (1985) Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino
 20 Acid Decarbonylation and Iminium Ion Cyclization. J. Org. Chem. 1989:1859-1866; Barton y col., (1987) Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron Lett. 43:4297-4308; y, Subasinghe y col., (1992) Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35:4602-7. Véase también la Publicación internacional WO
 25 2004/058946, titulada "PROTEIN ARRAYS", presentada el 22 de diciembre de 2003.

Captación celular de aminoácidos no naturales

La captación de aminoácidos no naturales por parte de una célula es una de las cuestiones que normalmente se
 30 consideran cuando se diseñan y se seleccionan aminoácidos no naturales, por ejemplo, para su incorporación a una proteína. Por ejemplo, la alta densidad de carga de α -aminoácidos sugiere que es improbable que estos compuestos sean permeables a la célula. Los aminoácidos naturales son captados por la célula a través de una serie de sistemas transportadores basados en proteínas que con frecuencia presentan grados de especificidad variables por los aminoácidos. Se puede realizar una detección rápida que evalúe qué aminoácidos no naturales, si es el caso,
 35 son captados por las células. Véase, por ejemplo, los ensayos de toxicidad en, por ejemplo, la Publicación internacional WO 2004/058946, titulada "PROTEIN ARRAYS", presentada el 22 de diciembre de 2003; y Liu y Schultz (1999) Progress toward the evolution of an organism with an expanded genetic code. PNAS 96:4780-4785. A pesar de que la captación se analiza fácilmente con diferentes ensayos, una alternativa al diseño de aminoácidos no naturales que sean susceptibles a las vías de captación celular es el suministro de vías biosintéticas para crear
 40 aminoácidos *in vivo*.

Biosíntesis de aminoácidos no naturales

Muchas vías biosintéticas ya existen en las células para la producción de aminoácidos y otros compuestos. Aunque
 45 en la naturaleza, por ejemplo, en una célula, pueda no existir un método biosintético para un aminoácido no natural particular, la invención proporciona dichos métodos. Por ejemplo, las vías biosintéticas para aminoácidos no naturales se generan opcionalmente en células hospedadoras mediante la adición de nuevas enzimas o modificando las vías celulares existentes en el hospedador. Nuevas enzimas adicionales opcionalmente son enzimas de origen natural o enzimas desarrolladas artificialmente. Por ejemplo, la biosíntesis de p-aminofenilalanina (como se presenta
 50 en un ejemplo del documento WO 2002/085923, más arriba) se basa en la adición de una combinación de enzimas conocidas procedentes de otros organismos. Los genes para estas enzimas se pueden introducir en una célula transformando la célula con un plásmido que comprenda los genes. Los genes, cuando se expresan en la célula, proporcionan una vía enzimática para sintetizar el compuesto deseado. En los ejemplos siguientes se proporcionan ejemplos de los tipos de enzimas que se añaden opcionalmente. Secuencias de enzimas adicionales se encuentran,
 55 por ejemplo, en el GenBank. Las enzimas desarrolladas artificialmente también se añaden opcionalmente a una célula de la misma forma. De esta manera, la maquinaria y los recursos celulares de las células se manipulan para producir aminoácidos no naturales.

De hecho, se puede usar cualquiera de varios métodos para la producción de nuevas enzimas para su uso en vías
 60 biosintéticas, o para la evolución de vías existentes, para la producción de aminoácidos no naturales, *in vitro* o *in vivo*. En la presente invención se pueden aplicar muchos métodos de desarrollo de enzimas disponibles y otros componentes de las vías biosintéticas para producir aminoácidos no naturales (o, de hecho, desarrollar sintetetasas que tengan nuevas especificidades por sustratos u otras actividades de interés). Por ejemplo, opcionalmente se usa el barajado de ADN para desarrollar nuevas enzimas y/o vías de dichas enzimas para la producción de aminoácidos
 65 no naturales (o producción de nuevas sintetetasas), *in vitro* o *in vivo*. Véase, por ejemplo, Stemmer (1994), Rapid evolution of a protein *in vitro* by DNA shuffling, Nature 370(4):389-391; y, Stemmer, (1994), DNA shuffling by random

fragmentation and reassembly: In vitro recombination for molecular evolution, Proc. Natl. Acad. Sci. EE.UU., 91:10747-10751. Un enfoque relacionado baraja familias de genes relacionados (por ejemplo, homólogos) para desarrollar rápidamente enzimas con las características deseadas. Un ejemplo de dichos métodos de “barajado de familias de genes” se encuentra en Cramer y col. (1998) “DNA shuffling of a family of genes from diverse species accelerates directed evolution Nature, 391(6664): 288-291. También se pueden generar nuevas enzimas (ya sean componentes de una vía biosintética o sintetasas) mediante el uso de un procedimiento de recombinación de ADN conocido como “truncamiento gradual para la creación de enzimas híbridas” (“ITCHY”, por sus siglas en inglés), por ejemplo, como se describe en Ostermeier y col. (1999) “A combinatorial approach to hybrid enzymes independent of DNA homology” Nature Biotech 17:1205. Este enfoque también se puede usar para generar una librería de enzimas u otras vías variantes que pueden servir como sustratos para uno o más métodos de recombinación *in vitro* o *in vivo*. Véase, también, Ostermeier y col. (1999) “Combinatorial Protein Engineering by Incremental Truncation”, Proc. Natl. Acad. Sci. USA, 96: 3562-67, y Ostermeier y col. (1999), “Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts”, Biological and Medicinal Chemistry, 7: 2139-44. Otro enfoque usa mutagénesis de conjunto exponencial para producir librerías de enzimas u otras vías variantes que, por ejemplo, se seleccionan por la capacidad para catalizar una reacción biosintética relevante para producir un aminoácido no natural (o una nueva sintetasa). En este enfoque, grupos pequeños de restos en una secuencia de interés se mezclan aleatoriamente en paralelo para identificar, en cada posición alterada, aminoácidos que dan lugar a proteínas funcionales. Ejemplos de dichos procedimientos, que se pueden adaptar a la presente invención para producir nuevas enzimas para la producción de aminoácidos no naturales (o nuevas sintetasas) se encuentran en Delegrave & Youvan (1993) Biotechnology Research 11:1548-1552. En otra aproximación más, se puede usar mutagénesis aleatoria o semi-aleatoria usando oligonucleótidos dopados o degenerados para la manipulación genética de la enzima y/o del componente de la vía, por ejemplo, usando los métodos de mutagénesis general de, por ejemplo, Arkin y Youvan (1992) “Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis” Biotechnology, 10:297-300; o Reidhaar-Olson y col. (1991) “Random mutagenesis of protein sequences using oligonucleotide cassettes” Methods Enzymol. 208:564-86. Se puede usar otro enfoque más, con frecuencia denominada mutagénesis “no estocástica”, que usa el reensamblaje de polinucleótidos y mutagénesis por saturación de sitio para producir enzimas y/o componentes de la vía, que a continuación se pueden detectar por la capacidad para realizar una o más funciones sintetasa o de la vía biosintética (por ejemplo, para la producción de aminoácidos no naturales *in vivo*). Véase, por ejemplo, Short “NON-STOCHASTIC GENERATION OF GENETIC VACCINES AND ENZYMES”, documento WO 00/46344.

Una alternativa a dichos métodos mutacionales incluye la recombinación de genomas completos de organismos y la selección de la progenie resultante para funciones de vías particulares (con frecuencia denominada “barajado de genomas completos”). Este enfoque se puede aplicar a la presente invención, por ejemplo, mediante recombinación genómica y selección de un organismo (por ejemplo, una célula de *E. coli* u otra célula) por su capacidad para producir un aminoácido no natural (o uno de sus intermedios). Por ejemplo, se pueden aplicar los métodos enseñados en las siguientes publicaciones para diseñar una vía para la evolución de nuevas vías y/o vías existentes en células para producir aminoácidos no naturales *in vivo*: Patnaik y col. (2002) “Genome shuffling of lactobacillus for improved acid tolerance” Nature Biotechnology, 20(7): 707-712; y Zhang y col. (2002) “Genome shuffling leads to rapid phenotypic improvement in bacteria” Nature, 7 de Febrero, 415(6872): 644-646.

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

Independientemente del método usado, normalmente el aminoácido no natural producido con una vía biosintética modificada genéticamente de la invención se produce a una concentración suficiente para la biosíntesis eficiente de proteína, por ejemplo, una cantidad celular natural, pero no en un grado tal como para afectar significativamente a la concentración de otros aminoácidos celulares o para agotar los recursos celulares. Las concentraciones típicas producidas de esta manera *in vivo* son de 10 mM aproximadamente a 0,05 mM aproximadamente. Una vez que una célula se ha manipulado genéticamente para producir enzimas deseadas por una vía específica y se genera un aminoácido no natural, opcionalmente se usan selecciones *in vivo* para optimizar aún más la producción del aminoácido no natural tanto para la síntesis ribosómica de proteína como para el crecimiento celular.

Componentes ortogonales para la incorporación de para-propargiloxifenilalanina (pPro-Phe)

La invención proporciona composiciones y métodos de producción de componentes ortogonales para la incorporación de un alquil aminoácido, por ejemplo, para-propargiloxifenilalanina (pPro-Phe), a una cadena polipeptídica en crecimiento en respuesta a un codón selector, por ejemplo, un codón de parada ámbar, un codón sinsentido, un codón de cuatro o más bases, etc., por ejemplo, *in vivo*. Por ejemplo, la invención proporciona ARNt

ortogonales (O-ARNt), aminoacil-ARNt sintetasas ortogonales (O-RS) y sus pares. Estos pares se pueden usar para incorporar pPro-Phe a cadenas polipeptídicas en crecimiento.

Una composición de la invención incluye una aminoacil-ARNt sintetasa ortogonal (O-RS), en la que la O-RS preferentemente aminoacila un O-ARNt con una pPro-Phe. En ciertas realizaciones, la O-RS comprende una secuencia de aminoácidos que comprende las SEQ ID NO: 4, 6, 8, 10, 12, 14 16 o 18, o una de sus variaciones conservativas. En ciertas realizaciones de la invención, la O-RS preferentemente aminoacila el O-ARNt sobre cualquier O-ARNt endógeno con un alquil aminoácido tal como pPro-Phe, en el que la O-RS tiene una preferencia por el O-ARNt, y en el que la relación de O-ARNt cargado con pPro-Phe a ARNt endógeno cargado con pPro-Phe es superior a 1:1, y más preferentemente en el que la O-RS carga el O-ARNt exclusiva o casi exclusivamente.

Una composición que incluye una O-RS opcionalmente además puede incluir un ARNt ortogonal (O-ARNt), en el que el O-ARNt reconoce un codón selector. Normalmente, un O-ARNt de la invención incluye una eficiencia de supresión de al menos, por ejemplo, un 45 %, un 50 %, un 60 %, un 75 %, un 80 %, o un 90 % aproximadamente o superior en presencia de una sintetasa afín en respuesta a un codón selector en comparación con la eficiencia de supresión de un O-ARNt que comprende o que está codificado por una secuencia de polinucleótidos como se expone en el listado de secuencias (por ejemplo, SEQ ID NO: 1) y los ejemplos de este documento. En una realización, la eficiencia de supresión de la O-RS y el O-ARNt juntos es, por ejemplo, de aproximadamente 5 veces, 10 veces, 15 veces, 20 veces, o 25 veces o superior más que la eficiencia de supresión del O-ARNt que carece de la O-RS. En un aspecto, la eficiencia de supresión de la O-RS y el O-ARNt juntos es de al menos el 45 % de la eficiencia de supresión de un par tirosil-ARNt sintetasa ortogonal derivado de *Methanococcus jannaschii*.

Una composición que incluya un O-ARNt opcionalmente puede incluir una célula (por ejemplo, una célula eubacteriana, tal como una célula de *E. coli* y similares), y/o un sistema de traducción.

La invención también proporciona una célula (por ejemplo, una célula eubacteriana) que comprende un sistema de traducción, en donde el sistema de traducción incluye un ARNt ortogonal (O-ARNt); una aminoacil-ARNt sintetasa ortogonal (O-RS); y un alquil aminoácido, por ejemplo, para-propargiloxifenilalanina (pPro-Phe). Normalmente, la O-RS preferentemente aminoacila el O-ARNt sobre cualquier ARNt endógeno con un alquil aminoácido tal como pPro-Phe, en la que la O-RS tiene una preferencia por el O-ARNt, y en el que la relación de O-ARNt cargado con pPro-Phe a ARNt endógeno cargado con pPro-Phe es superior a 1:1, y más preferentemente en el que la O-RS carga el O-ARNt exclusiva o casi exclusivamente. El O-ARNt reconoce el primer codón selector, y la O-RS preferentemente aminoacila el O-ARNt con pPro-Phe. En una realización, el O-ARNt comprende o está codificado por una secuencia de polinucleótidos como la que se expone en la SEQ ID NO: 1, o una de sus secuencias de polinucleótidos complementarias. En una realización, la O-RS comprende una secuencia de aminoácidos como la que se expone en cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16 o 18, o una de sus variaciones conservativas.

Además, una célula de la invención opcionalmente puede comprender un par O-ARNt/O-RS adicional diferente y un segundo aminoácido no natural, por ejemplo, en el que esta O-RS reconoce un segundo codón selector y esta O-RS preferentemente aminoacila el O-ARNt correspondiente con el segundo aminoácido no natural, en el que el segundo aminoácido no natural es diferente de la pPro-Phe. Opcionalmente, una célula de la invención incluye un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende un codón selector que es reconocido por el O-ARNt.

En ciertas realizaciones, una célula de la invención es una célula eubacteriana tal como *E. coli*, que incluye un ARNt ortogonal (O-ARNt), una aminoacil-ARNt sintetasa ortogonal (O-RS), un alquil aminoácido tal como pPro-Phe, 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. En ciertas realizaciones de la invención, la O-RS preferentemente aminoacila el O-ARNt con una eficiencia que es superior a la eficiencia con la cual la O-RS aminoacila cualquier ARNt endógeno.

En ciertas realizaciones de la invención, un O-ARNt de la invención comprende o es codificado por una secuencia de polinucleótidos como se expone en el listado de secuencias (por ejemplo, SEQ ID NO: 1) o en los ejemplos de este documento, o una de sus secuencias de polinucleótidos complementarias. En ciertas realizaciones de la invención; una O-RS comprende una secuencia de aminoácidos como se expone en el listado de secuencias, o una de sus variaciones conservativas. En una realización, la O-RS o un fragmento de la misma está codificada por una secuencia de polinucleótidos que codifica un aminoácido como se expone en el listado de secuencias o los ejemplos de este documento, o una de sus secuencias de polinucleótidos complementarias.

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

Los polinucleótidos también son una característica de la invención. Un polinucleótido de la invención incluye un polinucleótido artificial (por ejemplo, fabricado por el hombre, y que no es de origen natural) que comprende una secuencia de nucleótidos que codifica un polipéptido como se expone en el listado de secuencias de este documento, y/o es complementario a esa secuencia de polinucleótidos. Un polinucleótido de la invención también

puede incluir un ácido nucleico que se hibrida a un polinucleótido descrito anteriormente, en condiciones muy rigurosas, esencialmente a lo largo de toda la longitud del ácido nucleico. Un polinucleótido de la invención también incluye un polinucleótido que es idéntico, por ejemplo, en al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 %, al menos el 98 % o superior al de un ARNt de origen natural o el ácido nucleico codificante correspondiente (pero un polinucleótido de la invención es distinto del ARNt de origen natural o del ácido nucleico codificante correspondiente), en el que el ARNt reconoce un codón selector, por ejemplo, un codón de cuatro bases. En los polinucleótidos de la invención también están incluidos los polinucleótidos artificiales que son idénticos, por ejemplo, en al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 % o superior a cualquiera de los anteriores y/o a un polinucleótido que comprende una variación conservativa de cualquiera de los anteriores.

También son una característica de la invención vectores que comprenden un polinucleótido de la invención. Por ejemplo, un vector de la invención puede incluir un plásmido, un cósmido, un fago, un virus, un vector de expresión, y/o similares. También es una característica de la invención una célula que comprende un vector de la invención.

En este documento también se describen métodos de producción de componentes de un par O-ARNt/O-RS. Los componentes producidos mediante estos métodos también son una característica de la invención. Por ejemplo, métodos de producción de al menos un ARNt que es ortogonal a una célula (O-ARNt) incluyen la generación de una librería de ARNt mutantes; la mutación de un bucle anti-codónico de cada miembro de la librería de ARNt mutantes para permitir el reconocimiento de un codón selector, proporcionando así una librería de ARNt potenciales, y sometiendo una primera población de células de una primera especie a selección negativa, en el que las células comprenden un miembro de la librería de posibles O-ARNt. La selección negativa elimina células que comprenden un miembro de la librería de posibles O-ARNt que es aminoacilado por una aminoacil-ARNt sintetasa (RS) que es endógena a la célula. Esto proporciona un reservorio de ARNt que son ortogonales a la célula de la primera especie, proporcionando así al menos un O-ARNt. También se proporciona un O-ARNt producido mediante los métodos de la invención.

En ciertas realizaciones, los métodos además comprenden someter a una segunda población de células de la primera especie a selección positiva, en el que las células comprenden un miembro del reservorio de ARNt que son ortogonales a la célula de la primera especie, una aminoacil-ARNt sintetasa afín, y un marcador de selección positivo. Usando la selección positiva, las células son seleccionadas o detectadas por aquellas células que comprenden un miembro del reservorio de ARNt que es aminoacilado por la aminoacil-ARNt sintetasa afín y que muestra una respuesta deseada en presencia de un marcador de selección positivo, proporcionando así un O-ARNt. En ciertas realizaciones, la segunda población de células comprende células que no han sido eliminadas por la selección negativa.

También se proporcionan métodos para identificar una aminoacil-ARNt sintetasa ortogonal que carga un O-ARNt con un alquil aminoácido. Por ejemplo, los métodos incluyen someter una población de células de una primera especie a una selección, en donde cada una de las células comprende: 1) un miembro de una pluralidad de aminoacil-ARNt sintetasas (RS), (por ejemplo, la pluralidad de RS pueden incluir RS mutantes, RS derivadas de una especie distinta de la primera especie o tanto RS mutantes como RS derivadas de una especie distinta de la primera especie); 2) el ARNt ortogonal (O-ARNt) (por ejemplo, de una o más especies); y 3) un polinucleótido que codifica un marcador de selección positivo y comprende al menos un codón selector.

Las células (por ejemplo, una célula hospedadora) se seleccionan o se detectan para aquellas que presentan una mejora en la eficiencia de supresión en comparación con células que carecen o que tienen una cantidad reducida del miembro de la pluralidad de RS. Estas células seleccionadas/detectadas comprenden una RS activa que aminoacila el O-ARNt. También es una característica de la invención una aminoacil-ARNt sintetasa ortogonal identificada por el método.

También son una característica de la invención los métodos de producción de una proteína en una célula (por ejemplo, en una célula eubacteriana tal como una célula de *E. coli* o similar) que tiene para-propargiloxifenilalanina (pPro-Phe) en una posición específica. Por ejemplo, un método incluye el crecimiento, en un medio adecuado, de una célula, en el que la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína, proporcionando pPR, e incorporando pPR en la posición especificada en la proteína durante la traducción del ácido nucleico con el al menos un codón selector, produciendo así la proteína. La célula además comprende: un ARNt ortogonal (O-ARNt) que actúa en la célula y reconoce el codón selector; y una aminoacil-ARNt sintetasa ortogonal (O-RS) que preferentemente aminoacila el O-ARNt con pPro-Phe. También es una característica de la invención una proteína producida mediante este método.

La invención también proporciona métodos para la producción de composiciones que incluyen proteínas, en donde las proteínas comprenden, por ejemplo, pPro-Phe como se expone en las reivindicaciones. En ciertas realizaciones, la proteína comprende una secuencia de aminoácidos que es idéntica en al menos un 75 % a la de una proteína conocida, por ejemplo, una proteína terapéutica, una proteína diagnóstica, una enzima industrial, o un fragmento de las mismas. Opcionalmente, la composición comprende un vehículo farmacéuticamente aceptable.

Secuencia de ácidos nucleicos y polipéptidos y sus variantes

Como se ha descrito anteriormente y se describe a continuación, la invención se refiere a secuencias de polinucleótidos que codifican, por ejemplo, O-ARNt y O-RS, y secuencias polipeptídicas de aminoácidos, por ejemplo, O-RS, y por ejemplo, composiciones, sistemas y métodos que comprenden dichas secuencias. En este documento se desvelan ejemplos de dichas secuencias, por ejemplo, secuencias de nucleótidos y aminoácidos de O-ARNt y O-RS (véase Tabla 4, por ejemplo, SEQ ID NOS: 5, 7, 9, 11, 13, 15, 17 y 19). No obstante, el experto en la materia apreciará que la invención tal y como se expone en las reivindicaciones no está limitada a aquellas secuencias desveladas en este documento, por ejemplo, como en los Ejemplos y en el listado de secuencias. El experto en la materia apreciará que la invención también proporciona, por ejemplo, muchas secuencias relacionadas con las funciones descritas en este documento, por ejemplo, que codifican un O-ARNt o una O-RS.

La construcción y el análisis de especies de O-RS que son capaces de aminoacilar el O-ARNt con pPro-Phe se describen en el Ejemplo 1. Este ejemplo describe las ocho especies de O-RS que han sido aisladas (véase, Figura 3 y Ejemplo 9). Como se puede observar en estas secuencias de aminoácidos, se observan tendencias de consenso parcial en las sustituciones de aminoácidos en los ocho clones mutantes de O-RS. En más de un clon se encontraron al menos dos de los siguientes aminoácidos en el bolsillo de unión: Ala32, Pro107/Gln107, Ala158, Ile159, y Ala162/Pro162 (véase, SEQ ID NO: 21). Las mutaciones Tyr32→Ala32 y Asp158→Ala158 pueden producir la pérdida de puentes de hidrógeno entre Tyr32, Asp158 y el sustrato natural tirosina, desfavoreciendo así su unión. Cabría esperar que la aparición de cadenas laterales pequeñas y mayoritariamente hidrófobas facilitase la unión de pPro-Phe. Estas tendencias consenso permiten el diseño de especies de O-RS adicionales que se prevé que funcionen en un sistema ortogonal con el O-ARNt de la SEQ ID NO: 1 en un sistema hospedador eubacteriano para incorporar pPro-Phe. Estas tendencias consenso se pueden expresar de la forma siguiente:

TABLA 1

Posición del aminoácido	Aminoácidos de la tirosil-ARNt sintetasa de tipo silvestre de <i>Methanococcus jannaschii</i> (SEQ ID NO: 2)	Consenso pPro-PheRS ortogonal (SEQ ID NO: 21)
32	Tyr	Ala
107	Glu	Pro o Gln
110	Leu	Leu
158	Asp	Ala
159	Ile	Ile
162	Leu	Ala o Pro

Así, basándose en estas tendencias consenso, racionalmente se pueden diseñar al menos cuatro pPro-Phe sintetisas ortogonales adicionales (pPro-PheRS-con1 a pPro-PheRS-con4) que no están representadas en las ocho especies de pPro-PheRS identificadas experimentalmente (es decir, pPro-PheRS-1 a pPro-PheRS-8). Estas son las siguientes:

TABLA 2

SEQ ID NO:	Especies de tirosil-ARNt sintetasa de <i>Methanococcus jannaschii</i>	Posición del aminoácido					
		32	107	110	158	159	162
2	tipo silvestre	Tyr	Glu	Leu	Asp	Ile	Leu
21	pPro-PheRS-consenso	Ala	Pro/Gln	Leu	Ala	Ile	Ala/Pro
22	pPro-PheRS-con1	Ala	Pro	Leu	Ala	Ile	Ala
23	pPro-PheRS-con2	Ala	Pro	Leu	Ala	Ile	Pro
24	pPro-PheRS-con3	Ala	Gln	Leu	Ala	Ile	Ala
25	pPro-PheRS-con4	Ala	Gln	Leu	Ala	Ile	Pro

La invención proporciona polipéptidos (O-RS) y polinucleótidos que codifican O-RS. Además, los polinucleótidos de la invención incluyen, por ejemplo, un polinucleótido que comprende una secuencia de nucleótidos como se expone en las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 o 19; un polinucleótido que es complementario o que codifica una de sus secuencias de polinucleótidos. Un polinucleótido de la invención también incluye cualquier polinucleótido que codifique una secuencia de aminoácidos que comprende las SEQ ID NO: 4, 6, 8, 10, 12, 14 16 o 18. Un polinucleótido de la invención también incluye un polinucleótido que codifique un polipéptido de la invención.

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

El experto en la materia también apreciará que muchas de las variantes de las secuencias desveladas están incluidas en la invención. Por ejemplo, las variaciones conservativas de las secuencias desveladas que proporcionan una secuencia funcionalmente idéntica están incluidas en la invención. Se considera que las variantes de las

secuencias de polinucleótidos de ácidos nucleicos, en las que las variantes se hibridan a al menos una secuencia desvelada, están incluidas en la invención. También están incluidas en la invención secuencias únicas de las secuencias desveladas en este documento, como se determina, por ejemplo, mediante técnicas convencionales de comparación de secuencias.

5

Variaciones conservativas

Debido a la degeneración del código genético, las “sustituciones silenciosas” (es decir, sustituciones en una secuencia de ácidos nucleicos que no producen una alteración en el polipéptido codificado) son una característica implícita de cada secuencia de ácidos nucleicos que codifique una secuencia de aminoácidos. De forma similar, las “sustituciones conservativas de aminoácidos”, en la que uno o un número limitado de aminoácidos en una secuencia de aminoácidos son sustituidos por aminoácidos diferentes con propiedades muy similares, también se identifican fácilmente como muy similares a una construcción desvelada. Dichas variaciones conservativas de cada secuencia desvelada son una característica de la presente invención.

10

15

Las “variaciones conservativas” de una secuencia de ácidos nucleicos particular se refieren a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o, cuando el ácido nucleico no codifique una secuencia de aminoácidos, a secuencias esencialmente idénticas. El experto en la materia reconocerá que las sustituciones, delecciones o adiciones individuales que alteran, añaden o eliminan un único aminoácido o un pequeño porcentaje de aminoácidos (normalmente inferior al 5 %, más habitualmente inferior al 4 %, 2 % o 1 %) en una secuencia codificada son “variaciones conservativamente modificadas” en donde las alteraciones producen la eliminación de un aminoácido, la adición de un aminoácido, o la sustitución de un aminoácido con un aminoácido químicamente similar. Así, las “variaciones conservativas” de una secuencia polipeptídica listada de la presente invención incluyen sustituciones de un pequeño porcentaje, normalmente inferior al 5 %, más habitualmente inferior al 2 % o al 1 %, de los aminoácidos de la secuencia polipeptídica, con un aminoácido del mismo grupo de sustitución conservativo. Por último, la adición de secuencias que no alteran la actividad codificada de una molécula de ácidos nucleicos, tal como la adición de una secuencia no funcional, es una variación conservativa del ácido nucleico básico.

20

25

30

Las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares son muy conocidas en la técnica, en las que un resto aminoácido es sustituido por otro resto aminoácido que tiene propiedades químicas similares (por ejemplo, cadenas laterales aromáticas o cadenas laterales cargadas positivamente), y por tanto no modifican sustancialmente las propiedades funcionales de la molécula polipeptídica. A continuación se exponen grupos ilustrativos que contienen aminoácidos naturales de propiedades químicas similares, en los que las sustituciones dentro de un grupo son una “sustitución conservativa”.

35

TABLA 3

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

Hibridación de ácidos nucleicos

40

Se puede usar la hibridación comparativa para identificar ácidos nucleicos de la invención, incluyendo variaciones conservativas de ácidos nucleicos de la invención, y este método de hibridación comparativa es un método preferido para distinguir ácidos nucleicos de la invención. Además, los ácidos nucleicos diana que se hibridan a un ácido nucleico representado por las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 y 19, en condiciones rigurosas elevadas, ultra elevadas y ultra ultra elevadas son una característica de la invención. Ejemplos de dichos ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones de ácidos nucleicos silentes o conservativas en comparación con una secuencia de ácidos nucleicos determinada.

45

Se dice que un ácido nucleico de ensayo se hibrida específicamente a un ácido nucleico sonda cuando se hibrida al menos un 50 % y tanto a la sonda como a la diana complementaria perfectamente coincidente, es decir, con una relación de señal a ruido al menos la mitad del máximo de la hibridación de la sonda a la diana en condiciones en las que la sonda perfectamente coincidente se une a la diana complementaria perfectamente coincidente con una relación de señal a ruido que es al menos 5x-10x aproximadamente tan alta como la observada para la hibridación de cualquiera de los ácidos nucleicos diana no coincidentes.

50

55

Los ácidos nucleicos “se hibridan” cuando se asocian, normalmente en solución. Los ácidos nucleicos se hibridan debido a varias fuerzas fisicoquímicas perfectamente caracterizadas, tales como los puentes de hidrógeno, la

exclusión de disolvente, el apilamiento de bases y similares. Una guía exhaustiva para la hibridación de ácidos nucleicos se encuentra en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* parte I capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays", (Elsevier, Nueva York), así como en *Current Protocols in Molecular Biology*, Ausubel y col., eds., Current Protocols, un proyecto conjunto entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (completado hasta 2004) ("Ausubel"); Hames y Higgins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, Inglaterra; (Hames y Higgins 1) y Hames y Higgins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, Inglaterra (Hames y Higgins 2) proporcionan detalles relativos a la síntesis, marcaje, detección y cuantificación de ADN y ARN, incluidos los oligonucleótidos.

Un ejemplo de condiciones de hibridación rigurosas para la hibridación de ácidos nucleicos complementarios que tienen más de 100 restos complementarios sobre un filtro en una transferencia de Southern o Northern es formalina al 50 % con 1 mg de heparina a 42 °C, con la hibridación que se lleva a cabo durante toda la noche. Un ejemplo de condiciones de lavado rigurosas es un lavado con 0,2× SSC a 65 °C durante 15 minutos (véase, Sambrook, más arriba para una descripción del tampón de SSC). Con frecuencia, en lavado en condiciones muy rigurosas va precedido de un lavado en condiciones poco rigurosas para eliminar la señal de fondo de la sonda. Un ejemplo de lavado en condiciones poco rigurosas es 2× SSC a 40 °C durante 15 minutos. En general, una relación de señal a ruido de 5× (o superior) la observada para una sonda no relacionada en el ensayo de hibridación particular indica la detección de una hibridación específica.

"Condiciones de hibridación y lavado rigurosas" en el contexto de los experimentos de hibridación con ácidos nucleicos tales como hibridaciones de Southern y Northern dependen de la secuencia, y son diferentes con parámetros ambientales diferentes. Una guía exhaustiva para la hibridación de ácidos nucleicos se encuentra en Tijssen (1993), más arriba, y en Hames y Higgins, 1 y 2. Las condiciones de hibridación y lavado rigurosas se pueden determinar empíricamente de forma sencilla para cualquier ácido nucleico de ensayo. Por ejemplo, para determinar las condiciones de hibridación y lavado rigurosas, las condiciones de hibridación y lavado se incrementan gradualmente (por ejemplo, incrementando la temperatura, reduciendo la concentración salina, incrementando la concentración de detergente y/o incrementando la concentración de disolventes orgánicos tales como formalina en la hibridación o el lavado), hasta que se cumplen una serie de criterios seleccionados. Por ejemplo, en condiciones de hibridación y lavado muy rigurosas, las condiciones de hibridación y lavado se incrementan gradualmente hasta que una sonda se une a una diana complementaria perfectamente coincidente con una relación de señal a ruido que es al menos 5× tan alta como la observada para la hibridación de la sonda a una diana no coincidente.

Las condiciones "muy rigurosas" se seleccionan para que sean iguales al punto de fusión térmico (T_m) para una sonda particular. La T_m es la temperatura (a una concentración iónica y un pH definidos) a la cual el 50 % de la secuencia de ensayo se hibrida a una sonda perfectamente coincidente. Para los fines de la presente invención, en general, las condiciones de hibridación y lavado muy rigurosas se seleccionan para que estén 5 °C aproximadamente por debajo de la T_m para una secuencia específica a una concentración iónica y un pH definidos.

Las condiciones de hibridación y lavado "ultra rigurosas" son aquellas en las que la rigurosidad de las condiciones de hibridación y lavado se incrementan hasta que la relación de señal a ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente coincidente sea al menos 10× tan alta como la observada para la hibridación de cualquiera de los ácidos nucleicos diana no coincidentes. Un ácido nucleico diana que se hibrida a una sonda en dichas condiciones, con una relación de señal a ruido de al menos 1/2 de la del ácido nucleico diana complementario perfectamente coincidente se dice que se une a la sonda en condiciones de rigurosidad ultra elevadas.

De forma similar, se pueden determinar niveles de rigurosidad incluso superiores incrementando 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 incrementan hasta que la relación de señal a ruido para la unión de la sonda al ácido nucleico complementario perfectamente coincidente sea al menos 10×, 20×, 50×, 100×, o 500× o superior tan alta como la observada para la hibridación de cualquiera de los ácidos nucleicos diana no coincidentes. Un ácido nucleico diana que se hibrida a una sonda bajo dichas condiciones, con una relación de señal a ruido de al menos 1/2 de la del ácido nucleico diana complementario perfectamente coincidente se dice que se une a la sonda en condiciones de rigurosidad ultra ultra elevadas.

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

Secuencias únicas

Un ácido nucleico puede comprender una secuencia única en un ácido nucleico seleccionado entre las secuencias de O-ARNt, y O-RS desveladas en este documento. La secuencia única es única en comparación con un ácido nucleico correspondiente a cualquier secuencia conocida de ácidos nucleicos de O-ARNt u O-RS. El alineamiento se puede realizar usando, por ejemplo, BLAST configurado en los parámetros por defecto. Cualquier secuencia única

es útil, por ejemplo, como sonda para identificar los ácidos nucleicos de la invención.

De forma similar, un polipéptido puede comprender una secuencia única en un polipéptido seleccionado entre las secuencias de O-RS desveladas en este documento. Aquí, la secuencia única es única en comparación con un polipéptido correspondiente a cualquiera de las secuencias polipeptídicas conocidas.

Un ácido nucleico diana se puede hibridar en condiciones rigurosas a un oligonucleótido codificante único que codifique una secuencia única en un polipéptido seleccionado entre las secuencias de O-RS en el que la secuencia única es única en comparación con un polipéptido correspondiente a cualquiera de los polipéptidos control (por ejemplo, secuencias parentales a partir de las que se derivaron las sintetasas de la invención, por ejemplo, mediante mutación). Las secuencias únicas se determinan como se ha indicado anteriormente.

Comparación, identidad, y homología de secuencia

Los términos “idénticos” o “porcentaje de identidad”, en el contexto de dos o más secuencias de ácidos nucleicos o de polipéptidos, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje especificado de restos aminoácidos o nucleótidos que son iguales, cuando se comparan y se alinean para su máxima correspondencia, como se mide usando uno de los algoritmos de comparación de secuencias descritos a continuación (u otros algoritmos disponibles para el experto en la materia) o mediante inspección visual.

La frase “sustancialmente idéntico” en el contexto de dos o más ácidos nucleicos o polipéptidos (por ejemplo, ADN que codifica un O-ARNt u O-RS, o la secuencia de aminoácidos de una O-RS) se refiere a dos o más secuencias o subsecuencias que tienen una identidad de nucleótidos o restos de aminoácidos de al menos el 60 % aproximadamente, el 80 % aproximadamente, el 90-95 % aproximadamente, el 98 % aproximadamente, el 99 % aproximadamente o superior, cuando se compara y se alinea para su máxima correspondencia, como se mide usando un algoritmo de comparación de secuencias o mediante inspección visual. Dichas secuencias “sustancialmente idénticas” normalmente se considera que son “homólogas”, sin hacer referencia a la ascendencia real. Preferentemente, la “identidad sustancial” existe a lo largo de una región de las secuencias que tiene una longitud de al menos 50 restos aproximadamente, más preferentemente a lo largo de una región de al menos 100 restos aproximadamente, y lo más preferentemente, las secuencias son sustancialmente idénticas a lo largo de al menos 150 restos aproximadamente, o a lo largo de toda la longitud de las dos secuencias a comparar.

Las proteínas y/o secuencias de proteínas son “homólogas” cuando derivan, natural o artificialmente, de una proteína o secuencia de proteína ancestral común. De forma similar, los ácidos nucleicos y/o secuencias de ácidos nucleicos son homólogos cuando derivan, natural o artificialmente, de un ácido nucleico o secuencia de ácido nucleico ancestral común. Por ejemplo, cualquier ácido nucleico de origen natural se puede modificar mediante cualquier método de mutagénesis disponible para que incluya uno o más codones selectores. Cuando se expresa, este ácido nucleico sometido a mutagénesis codifica un polipéptido que comprende uno o más aminoácidos no naturales, por ejemplo, un alquil aminoácido. Naturalmente, el proceso de mutación adicionalmente puede alterar uno o más codones convencionales, modificando también de esta forma uno o más aminoácidos convencionales en la proteína mutante resultante. La homología generalmente se infiere a partir de la similitud de secuencia entre dos o más ácidos nucleicos o proteínas (o sus secuencias). El porcentaje de similitud preciso entre secuencias que es útil a la hora de establecer la homología varía con el ácido nucleico y la proteína en cuestión, pero de forma rutinaria, para establecer la homología, se usa una similitud de secuencia de tan solo el 25 %. Para establecer la homología también se pueden usar niveles de similitud de secuencias superiores, por ejemplo, del 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, o 99 % o superior. Los métodos para determinar los porcentajes de similitud de secuencia (por ejemplo, BLASTP y BLASTN usando los parámetros por defecto) se describen en este documento y están disponibles de forma general.

Para la comparación de secuencias y la determinación de la homología, normalmente una secuencia actúa como secuencia de referencia con la cual se comparan las secuencias de ensayo. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de ensayo y de referencia se introducen en un ordenador, se designan las coordenadas de la subsecuencia, si es necesario, y se designan los parámetros del algoritmo de secuencias en el programa. 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, basándose en los parámetros designados en el programa.

El alineamiento óptimo de secuencias para su comparación se puede llevar a cabo, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482 (1981), mediante el algoritmo de homología por alineamiento de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), mediante la búsqueda por el método de similitud de Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), mediante implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA, y TFASTA del paquete de *software* Wisconsin Genetics Software Package, Genetics Computer group, 575 Science Dr., Madison, WI), o mediante inspección visual (véase, en general, Current Protocols in Molecular Biology, Ausubel y col., eds., Current Protocols, un proyecto conjunto entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., completado hasta 2004).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y la similitud de secuencia es el algoritmo BLAST, que se describe en Altschul y col., J. Mol. Biol. 215:403-410 (1990). El *software* para el realizar los análisis BLAST está disponible para el público en el National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). Este algoritmo supone en primer lugar la identificación de pares de secuencia con una puntuación elevada (HSPs) mediante la identificación de palabras cortas de longitud W en la secuencia de búsqueda, que pueden cumplir o satisfacer ciertas puntuaciones umbral de valor positivo T cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. T se refiere al umbral de puntuación de una palabra vecina (neighborhood word score Threshold, por sus términos en inglés (Altschul y col., más arriba)). Estas coincidencias de la palabra vecina inicial sirven de base para iniciar búsquedas con el fin de encontrar HSP más largos que las contengan. A continuación las coincidencias de palabras se extienden en ambas direcciones a lo largo de cada secuencia tanto como permita incrementar la puntuación del alineamiento acumulado. Las puntuaciones acumuladas se calculan usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de restos coincidentes; siempre > 0) y N (puntuación de penalización para restos no coincidentes; siempre < 0). Para secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulada. La extensión de las coincidencias de palabras en cada dirección se detienen cuando: la puntuación del alineamiento acumulado disminuye en una cantidad X respecto a su valor máximo conseguido; la puntuación acumulada llega a cero o por debajo de cero, debido a la acumulación de uno o más alineamientos de restos de puntuación negativa; o se alcanza al final de cualquiera de las secuencias. Los parámetros, W, T, y X, del algoritmo BLAST determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabras (W) de 11, una expectativa (E) de 10, un límite de 100, M=5, N=-4, y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873-5877 (1993)). Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más baja (P(N)), que proporciona una indicación de la probabilidad mediante la cual se produciría un emparejamiento al azar entre dos secuencias de nucleótidos o de aminoácidos. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más baja en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es inferior a 0,1 aproximadamente, más preferentemente inferior a 0,01 aproximadamente, y lo más preferentemente inferior a 0,001 aproximadamente.

Mutagénesis y otras técnicas de biología molecular

Los polinucleótidos y polipéptidos de la invención y que se usan en la invención se pueden manipular utilizando técnicas de biología molecular. Los textos generales que describen técnicas de biología molecular incluyen Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook y col., Molecular Cloning - A Laboratory Manual (3ª Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 2001 ("Sambrook") y Current Protocols in Molecular Biology, F.M. Ausubel y col., eds., Current Protocols, un proyecto conjunto entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (completado hasta 2004) ("Ausubel"). Estos textos describen la mutagénesis, el uso de vectores, promotores y muchos otros temas relevantes relacionados, por ejemplo, con la generación de genes que incluyen codones selectores para la producción de proteínas que incluyen alquil aminoácidos (por ejemplo, pPro-Phe), ARNt ortogonales, sintetasas ortogonales, y sus pares.

En la invención se usan varios tipos de mutagénesis, por ejemplo, para mutar moléculas de ARNt, para producir librerías de ARNt, para producir librerías de sintetasas, para insertar codones selectores que codifiquen un alquil aminoácido en una proteína o polipéptido de interés. Incluyen, pero no están limitados a, mutagénesis dirigida de sitio de punto aleatorio, recombinación homóloga, barajado de ADN u otros métodos de mutagénesis recursiva, construcción de quimeras, mutagénesis usando moldes que contienen uracilo, mutagénesis dirigida a oligonucleótidos, mutagénesis de ADN modificada con fosforotioato, mutagénesis usando dúplex de ADN incompletos o similares, o cualquiera de sus combinaciones. Los métodos adicionales adecuados incluyen la reparación de desemparejamientos puntuales, mutagénesis usando cepas hospedadoras con reparación deficiente, restricción-selección y restricción-purificación, mutagénesis por delección, mutagénesis por síntesis génica total, reparación de roturas en la doble cadena, y similares. La mutagénesis, por ejemplo, que implica construcciones quiméricas, también está incluida en la presente invención. En una realización, la mutagénesis puede estar guiada por información conocida de la molécula de origen natural o de la molécula de origen natural alterada o mutada, por ejemplo, la secuencia, las comparaciones de secuencia, las propiedades físicas, la estructura cristalina, o similares.

Las células hospedadoras se manipulan genéticamente (por ejemplo, se transforman, transducen o transfectan) con los polinucleótidos de la invención o construcciones que incluyen un polinucleótido de la invención, por ejemplo, un vector de la invención, que puede ser, por ejemplo, un vector de clonación o un vector de expresión. Por ejemplo, las regiones codificantes para el ARNt ortogonal, la ARNt-sintetasa ortogonal, y la proteína a derivar están unidas funcionalmente a elementos para el control de la expresión de genes que son funcionales en la célula hospedadora deseada. Los vectores típicos contienen terminadores de la transcripción y la traducción, secuencias de iniciación de

la transcripción y la traducción, y promotores útiles para regular la expresión del ácido nucleico diana particular. Los vectores opcionalmente comprenden 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 (por ejemplo, vectores lanzadera) y marcadores de selección tanto para sistemas procariotas como eucariotas. Los vectores son adecuados para la replicación y/o integración en procariotas, eucariotas, o preferentemente en ambos. Véase, Gilman & Smith, *Gene* 8:81 (1979); Roberts, y col., *Nature*, 328:731 (1987); Schneider, B., y col., *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (todos ellos más arriba). El vector puede estar, por ejemplo, en forma de plásmido, bacteria, virus, polinucleótido desnudo, o polinucleótido conjugado. Los vectores se introducen en células y/o microorganismos mediante métodos convencionales que incluyen electroporación (From y col., *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infección mediante vectores víricos, penetración balística a alta velocidad mediante partículas pequeñas con el ácido nucleico dentro de la matriz de cuentas o partículas pequeñas, o sobre la superficie (Klein y col., *Nature* 327, 70-73 (1987)), y/o similares.

Un catálogo de bacterias y bacteriófagos útiles para la clonación es suministrado, por ejemplo, por el ATCC, por ejemplo, *The ATCC Catalogue of Bacteria and Bacteriophage* (1996) Ghema y col. (eds) publicado por el ATCC. Procedimientos básicos adicionales para la secuenciación, clonación y otros aspectos de la biología molecular y las consideraciones teóricas subyacentes también se pueden encontrar en Sambrook (más arriba), Ausubel (más arriba), y en Watson y col. (1992) *Recombinant DNA*, 2ª Edición Scientific American Books, NY. Además, se puede solicitar de forma personalizada o convencional esencialmente cualquier ácido nucleico (y virtualmente cualquier ácido nucleico marcado, ya sea convencional o no convencional) en cualquiera de varios fuentes comerciales, tales como la Midland Certified Reagent Company (Midland, TX mcr.com), The Great American Gene Company (Ramona, CA disponible en Internet en genco.com), ExpressGen Inc. (Chicago, IL disponible en Internet en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchas otras.

Las células hospedadoras manipuladas genéticamente se pueden cultivar en medios nutrientes convencionales modificados según sea apropiado para actividades tales como, por ejemplo, etapas de detección, activación de los promotores o selección de transformantes. Estas células opcionalmente se pueden cultivar en organismos transgénicos. Otras referencias útiles, por ejemplo, para el aislamiento y cultivo celular (por ejemplo, para el posterior aislamiento del ácido nucleico) incluyen Freshney (1994) *Culture of Animal Cells*, a *Manual of Basic Technique*, 3ª Edición, Wiley- Liss, Nueva York y las referencias allí citadas; Payne y col. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. Nueva York, NY; Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture*; *Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlín Heidelberg Nueva York) y Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, FL.

Proteínas y polipéptidos de interés

Una ventaja significativa de los alquil aminoácidos (pero no limitado a éstos) es que las proteínas que comprenden el alquil aminoácido se pueden usar para el entrecruzamiento o conjugación de las proteínas con cualquiera de varias moléculas pequeñas, biomoléculas u otras proteínas, etc. Las proteínas o polipéptidos de interés con al menos un alquil aminoácido son una característica de la invención. La invención también incluye polipéptidos a proteínas con al menos un alquil aminoácido producido usando las composiciones y métodos de la invención. Con la proteína también puede haber presente un excipiente (por ejemplo, un excipiente farmacéuticamente aceptable).

Opcionalmente, una proteína de la invención puede incluir una modificación post-traducciona (además de posibles modificaciones posteriores del resto alquil aminoácido) en un aminoácido en una única posición o en múltiples posiciones, o la proteína puede tener una pluralidad de tipos de modificaciones diferentes.

Los métodos de producción de una proteína en una célula con un alquil aminoácido en una posición específica, como se expone en las reivindicaciones, también son una característica de la invención. Por ejemplo, un método incluye el crecimiento de la célula en un medio adecuado, en donde la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína; y, el suministro del alquil aminoácido; donde la célula además comprende: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector; y una aminoacil-ARNt sintetasa ortogonal (O-RS) que preferentemente aminoacila el O-ARNt con el alquil aminoácido.

En ciertas realizaciones, la O-RS comprende una tendencia a la aminoacilación del O-ARNt afín sobre cualquier ARNt endógeno en un sistema de expresión. La relación relativa entre el O-ARNt y el ARNt endógeno que es cargado por la O-RS, cuando el O-ARNt y la O-RS están presentes en concentraciones molares iguales, es superior a 1:1, preferentemente de al menos 2:1 aproximadamente, más preferentemente de 5:1, aún más preferentemente de 10:1, incluso más preferentemente de 20:1, aún más preferentemente de 50:1, incluso más preferentemente de 75:1, aún más preferentemente de 95:1, 98:1, 99:1, 100:1, 500:1, 1000:1, 5000:1 o superior.

La invención también proporciona métodos para la producción de composiciones que incluyen proteínas, en donde las proteínas comprenden un alquil aminoácido, como se expone en las reivindicaciones. En ciertas realizaciones, la proteína comprende una secuencia de aminoácidos que es idéntica en al menos un 75 % a la de una proteína terapéutica, una proteína diagnóstica, una enzima industrial, o un fragmento de las mismas.

Las composiciones preparadas mediante los métodos de la invención opcionalmente se encuentran en una célula. Los pares O-ARNt/O-RS o componentes individuales a continuación se pueden usar en la maquinaria de traducción de un sistema hospedador, que da lugar a la incorporación del alquil aminoácido a una proteína. Las Publicaciones internacionales número WO 2004/094593, presentada el 16 de abril de 2004, titulada "EXPANDING THE EUKARYOTIC GENETIC CODE"; y WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", describen este proceso. Por ejemplo, cuando se introduce un par O-ARNt/O-RS en un hospedador, por ejemplo, una célula de *Escherichia coli*, el par da lugar a la incorporación *in vivo* de un alquil aminoácido tal como para-propargiloxifenilalanina a una proteína en respuesta a un codón selector. La para-propargiloxifenilalanina que se añade al sistema es un aminoácido sintético, tal como un derivado de una fenilalanina o tirosina, que se puede añadir exógenamente al medio de crecimiento. Opcionalmente, las composiciones de la presente invención pueden ser un sistema de traducción *in vitro*, o (un) sistema(s) *in vivo*.

Una célula de la invención proporciona la capacidad de sintetizar proteínas que comprenden aminoácidos no naturales en cantidades grandes y útiles. En un aspecto, la composición opcionalmente incluye, por ejemplo, al menos 10 µg, al menos 50 µg, al menos 75 µg, al menos 100 µg, al menos 200 µg, al menos 250 µg, al menos 500 µg, al menos 1 mg, al menos 10 mg o superior de la proteína que comprende un alquil aminoácido, una cantidad que se puede conseguir con métodos *in vivo* de producción de proteínas (en este documento se proporcionan detalles sobre la producción y purificación de proteína recombinante). En otro aspecto, la proteína opcionalmente está presente en la composición a una concentración de, por ejemplo, al menos 10 µg de proteína por litro, al menos 50 µg de proteína por litro, al menos 75 µg de proteína por litro, al menos 100 µg de proteína por litro, al menos 200 µg de proteína por litro, al menos 250 µg de proteína por litro, al menos 500 µg de proteína por litro, al menos 1 mg de proteína por litro, o al menos 10 mg de proteína por litro o superior, por ejemplo en un lisado celular, un tampón, un tampón farmacéutico, u otra suspensión líquida (por ejemplo, en un volumen de, por ejemplo, cualquiera entre 1 nl aproximadamente y 100 l aproximadamente). La producción de grandes cantidades (por ejemplo, superior a lo que normalmente es posible con otros métodos, por ejemplo, traducción *in vitro*) de una proteína en una célula que incluye al menos un alquil aminoácido es una característica de la invención.

La incorporación de un alquil aminoácido se puede realizar, por ejemplo, para introducir cambios a medida en una estructura y/o función de una proteína, por ejemplo, para modificar el tamaño, la acidez, la nucleofilia, puentes de hidrógeno, hidrofobicidad, accesibilidad de proteasas a sitios diana, direccionamiento a un resto (por ejemplo, para una matriz de proteínas), etc. Las proteínas que incluyen un alquil aminoácido pueden tener unas propiedades catalíticas o físicas mejoradas o incluso completamente nuevas. Por ejemplo, mediante la inclusión de un alquil aminoácido en una proteína opcionalmente se modifican las siguientes propiedades: toxicidad, biodistribución, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, capacidad catalítica, semi-vida (por ejemplo, semi-vida en suero), capacidad para reaccionar con otras moléculas, por ejemplo, covalentemente o no covalentemente, y similares. Las composiciones que incluyen proteínas que incluyen al menos un alquil aminoácido son útiles, por ejemplo, para nuevos compuestos terapéuticos, diagnósticos, enzimas catalíticas, enzimas industriales, proteínas de unión (por ejemplo, anticuerpos), y para, por ejemplo, el estudio de la estructura y la función de proteínas. Véase, por ejemplo, Dougherty, (2000) Unnatural Amino Acids as Probes of Protein Structure and Function, Current Opinion in Chemical Biology, 4:645-652.

Una composición puede incluir al menos una proteína con al menos uno, por ejemplo, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, o al menos diez o más aminoácidos no naturales, por ejemplo, alquil aminoácidos y/u otros aminoácidos no naturales. Los aminoácidos no naturales pueden ser iguales o diferentes, por ejemplo, puede haber 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más sitios diferentes en la proteína que comprendan 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más aminoácidos no naturales diferentes. De manera alternativa, una composición incluye una proteína con al menos uno, pero no todos, de un aminoácido particular presente en la proteína que está sustituido con el alquil aminoácido. Para una proteína determinada con más de un aminoácido no natural, los aminoácidos no naturales pueden ser idénticos o diferentes (por ejemplo, la proteína puede incluir dos o más tipos de aminoácidos no naturales, o puede incluir dos aminoácidos no naturales idénticos). Para una proteína determinada con más de dos aminoácidos no naturales, los aminoácidos no naturales pueden ser idénticos, diferentes o una combinación de múltiples aminoácidos no naturales del mismo tipo con al menos un aminoácido no natural diferente.

Esencialmente se puede producir cualquier proteína (o un fragmento de la misma) que incluya un alquil aminoácido (o cualquier ácido nucleico codificante correspondiente, por ejemplo, que incluya uno o más codones selectores) usando las composiciones y métodos de este documento. No se ha intentado identificar los cientos de miles de proteínas conocidas, cualquiera de ellas que se puede modificar para que incluya uno o más aminoácidos no naturales, por ejemplo, mediante la adaptación de cualquiera de los métodos de mutación disponibles para que incluya uno o más codones selectores adecuados en un sistema de traducción pertinente. Los repositorios de secuencias comunes para proteínas conocidas incluyen el GenBank, EMBL, DDBJ y el NCBI. Se pueden identificar fácilmente otros repositorios buscando en internet.

Normalmente, las proteínas son idénticas en, 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 superior a cualquier proteína disponible (por ejemplo, una proteína terapéutica, una proteína diagnóstica, una enzima industrial, o uno de sus fragmentos, y similares), y comprenden uno o más aminoácidos no naturales. Ejemplos de proteínas terapéuticas, diagnósticas, y otras proteínas que se pueden modificar para que comprendan uno o más alquínil aminoácidos se pueden encontrar en, pero no está limitado a, aquellas de las Publicaciones internacionales WO 2004/094593, presentada el 16 de abril de 2004, titulada "Expanding the Eukaryotic Genetic Code"; y WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Ejemplos de proteínas terapéuticas, diagnósticas, y otras proteínas que se pueden modificar para que comprendan uno o más alquínil aminoácidos incluyen, pero no están limitados a, por ejemplo, α -1-antitripsina, angiostatina, factor antihemolítico, anticuerpos (se pueden encontrar más detalles acerca de los anticuerpos a continuación), apolipoproteína, apoproteína, factor natriurético auricular, polipéptido natriurético auricular, péptidos auriculares, quimiocinas C-X-C (por ejemplo, T39765, NAP-2, ENA-78, Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), calcitonina, quimioquinas CC (por ejemplo, proteína quimiotáctica de monocitos-1, proteína quimiotáctica de monocitos-2, proteína quimiotáctica de monocitos-3, proteína inflamatoria de monocitos 1-alfa, proteína inflamatoria de monocitos 1-beta, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), ligando CD40, ligando C-kit, colágeno, factor estimulante de colonias (CSF), factor del complemento 5a, inhibidor del complemento, receptor de complemento 1, citocinas, (por ejemplo, péptido epitelial activador de neutrófilos 78, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1), factor de crecimiento epidérmico (EGF), eritropoyetina ("EPO"), toxinas exfoliantes A y B, Factor IX, Factor VII, Factor VIII, Factor X, factor de crecimiento de fibroblastos (FGF), fibrinógeno, fibronectina, G-CSF, GM-CSF, glucocerebrosidasa, gonadotropina, factores de crecimiento, proteínas Hedgehog (por ejemplo, Sonic, Indian, Desert), hemoglobina, factor de crecimiento de hepatocitos (HGF), hirudina, seroalbúmina humana, insulina, factor de crecimiento similar a la insulina (IGF), interferones (por ejemplo, IFN- α , IFN- β , IFN- γ), interleucinas (por ejemplo, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, etc.), factor de crecimiento de queratinocitos (KGF), lactoferrina, factor inhibidor de la leucemia, luciferasa, neurturina, factor inhibidor de neutrófilos (NIF), oncostatina M, proteína osteogénica, hormona paratiroidea, PD-ECSF, PDGF, hormonas peptídicas (por ejemplo, hormona del crecimiento humana), pleyotropina, Proteína A, Proteína G, exotoxinas pirógenas A, B, y C, relaxina, renina, SCF, receptor soluble del complemento I, Soluble I-CAM 1, receptores solubles de interleucina (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), receptor soluble del TNF, somatomedina, somatostatina, somatotropina, estreptoquinasa, superantígenos, es decir, enterotoxinas estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutasa (SOD), toxina del síndrome de choque tóxico (TSST-1), timosina alfa 1, activador del plasminógeno tisular, factor de necrosis tumoral beta (TNF beta), receptor del factor de necrosis tumoral (TNFR), factor de necrosis tumoral-alfa (TNF alfa), factor de crecimiento endotelial vascular (VEGEF), uroquinasa y muchos otros.

Una clase de proteínas que se puede preparar usando las composiciones y métodos para la incorporación *in vivo* de alquínil aminoácidos descritos en este documento incluye moduladores transcripcionales y fragmentos de los mismos. Moduladores transcripcionales ilustrativos incluyen genes y proteínas moduladoras transcripcionales que modulan el crecimiento, la diferenciación, y la regulación celular, o similares. Los moduladores transcripcionales se encuentran en procariotas, virus, y eucariotas, incluyendo hongos, plantas, levaduras, insectos y animales, incluyendo mamíferos, lo que proporciona un amplio espectro de dianas terapéuticas. Se apreciará que los activadores transcripcionales y de expresión regulan la transcripción mediante numerosos mecanismos, por ejemplo, mediante la unión a receptores, estimulando una cascada de transducción de señales, regulando la expresión de los factores de transcripción, uniéndose a promotores y potenciadores, uniéndose a proteínas que se unen a promotores y potenciadores, desenrollando el ADN, procesando el pre-ARNm, poliadenilando el ARN, y degradando el ARN.

Una clase de proteínas (por ejemplo, proteínas con uno o más alquínil aminoácidos) incluye proteínas biológicamente activas tales como citocinas, moléculas inflamatorias, factores de crecimiento, sus receptores, y productos oncogénicos, por ejemplo, interleucinas (por ejemplo, IL-1, IL-2, IL-8, etc.), interferones, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- α , TGF- β , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1, y hialurina/CD44; moléculas para la transducción de señales y sus correspondientes productos oncogénicos por ejemplo, Mos, Ras, Raf, y Met; y activadores y supresores transcripcionales, por ejemplo, p53, Tat, Fos, Myc, Jun, Myb, Rel, y receptores de hormonas esteroideas tales como aquellos para estrógenos, progesterona, testosterona, y el ligando receptor de LDL y corticosterona.

Las enzimas (por ejemplo, enzimas industriales) o sus fragmentos pueden incluir al menos un alquínil aminoácido. Ejemplos de enzimas incluyen, pero no están limitados a, por ejemplo, amidasas, racemasas de aminoácidos, acilasas, deshalogenasas, dioxigenasas, diarilpropano peroxidasas, epimerasas, epóxido hidrolasas, esterases, isomerasas, quinasas, glucosa isomerasas, glicosidasas, glicosil transferasas, haloperoxidasas, monooxigenasas (por ejemplo, p450), lipasas, lignina peroxidasas, nitrilo hidratasas, nitrilasas, proteasas, fosfatasas, subtilisinas, transaminasa y nucleasas.

Muchas de estas proteínas están disponibles en el mercado (véase, por ejemplo, el catálogo de Sigma BioSciences 2002 y su lista de precios), y las secuencias correspondientes de proteínas y genes y, normalmente, muchas de sus variantes son muy conocidas (véase, por ejemplo, el GenBank). Cualquiera de ellas se puede modificar mediante la

inserción de uno o más alquiniil aminoácidos de acuerdo con la invención, por ejemplo, para alterar la proteína con respecto a una o más propiedades terapéuticas, diagnósticas o enzimáticas de interés. Ejemplos de propiedades terapéuticamente relevantes incluyen semi-vida en suero, semi-vida útil, estabilidad, inmunogenicidad, actividad terapéutica, detectabilidad (por ejemplo, mediante la inclusión de grupos informadores (por ejemplo, marcadores o sitios de unión a marcadores) en aminoácidos no naturales, por ejemplo, alquiniil aminoácidos), reducción de la DL₅₀ u otros efectos secundarios, capacidad para introducirse en el cuerpo a través del tracto gástrico (por ejemplo, disponibilidad oral) o similares. Ejemplos de propiedades diagnósticas incluyen la semi-vida útil, estabilidad, actividad diagnóstica, detectabilidad, o similares. Ejemplos de propiedades enzimáticas relevantes incluyen la semi-vida útil, estabilidad, actividad enzimática, capacidad de producción, o similares.

También se pueden modificar otras varias proteínas para que incluyan uno o más alquiniil aminoácidos usando células eubacterianas y métodos de la invención como se expone en las reivindicaciones. Por ejemplo, la sustitución de uno o más aminoácidos naturales en una o más proteínas de vacuna con un alquiniil aminoácido, por ejemplo, en proteínas procedentes de hongos infecciosos, por ejemplo, *Aspergillus*, especies de *Candida*; bacterias, particularmente *E. coli*, que sirve de modelo para bacterias patógenas, así como bacterias médicamente importantes tales como *Staphylococci* (por ejemplo, *aureus*), o *Streptococci* (por ejemplo, *pneumoniae*); protozoos tales como esporozoa (por ejemplo, *Plasmodia*), rizópodos (por ejemplo, *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); virus tales como virus ARN (+) (sus ejemplos incluyen Poxvirus, por ejemplo, *vaccinia*; picornavirus, por ejemplo, *polio*; Togavirus, por ejemplo, *rubella*; Flavivirus, por ejemplo, HCV, y Coronavirus), virus ARN (-) (por ejemplo, Rhabdovirus, por ejemplo, VSV; Paramyxoviruses, por ejemplo, RSV; Orthomyxoviruses, por ejemplo, *influenza*; Bunyavirus y Arenavirus), virus de ADNds (Reovirus, por ejemplo), virus de ARN a ADN, es decir, Retrovirus, por ejemplo, VIH y HTLV, y determinados virus de ADN a ARN tales como el de la hepatitis B.

También son dianas adecuadas para su modificación con alquiniil aminoácidos las proteínas relacionadas con la agricultura tales como proteínas de resistencia a insectos (por ejemplo, las proteínas Cry), enzimas de producción de almidón y lípidos, toxinas de plantas e insectos, proteínas de resistencia a toxinas, proteínas de detoxificación de micotoxinas, enzimas para el crecimiento de las plantas (por ejemplo, ribulosa-1,5-bifosfato carboxilasa/oxigenasa, "RUBISCO"), lipoxigenasa (LOX) y fosfoenolpiruvato (PEP) carboxilasa.

La proteína o polipéptido de interés (o un fragmento del mismo) en los métodos y/o composiciones de la invención están codificados por un ácido nucleico. Normalmente, el ácido nucleico comprende al menos un codón selector, al menos dos codones selectores, al menos tres codones selectores, al menos cuatro codones selectores, al menos cinco codones selectores, al menos seis codones selectores, al menos siete codones selectores, al menos ocho codones selectores, al menos nueve codones selectores, diez o más codones selectores.

Los genes que codifican para proteínas o polipéptidos de interés se pueden someter a mutagénesis usando métodos muy conocidos por el experto en la materia y descritos en este documento bajo el epígrafe "Mutagénesis y otras técnicas de biología molecular" para que incluyan, por ejemplo, uno o más codones selectores para la incorporación de un alquiniil aminoácido. Por ejemplo, un ácido nucleico para una proteína de interés se somete a mutagénesis para que incluya uno o más codones selectores, proporcionando la inserción de uno o más alquiniil aminoácidos. La invención incluye cualquiera de dichas versiones variantes, por ejemplo, mutantes, de cualquier proteína, por ejemplo, que incluye al menos un alquiniil aminoácido. De forma similar, la invención también incluye ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifique uno o más alquiniil aminoácidos.

Para preparar una proteína que incluye un alquiniil aminoácido, se pueden usar células u organismos hospedadores que estén adaptados para la incorporación *in vivo* del alquiniil aminoácido mediante pares ARNt/RS ortogonales. Las células hospedadoras se manipulan genéticamente (por ejemplo, se transforman, transducen o transfectan) con uno o más vectores que expresan el ARNt ortogonal, la ARNt-sintetasa ortogonal, y un vector que codifica la proteína a derivar. Cada uno de estos componentes puede estar sobre el mismo vector, o cada uno de ellos puede estar sobre un vector separado, o dos componentes pueden estar sobre un vector y el tercer componente sobre un segundo reactor. El vector puede estar, por ejemplo, en forma de plásmido, bacteria, virus, polinucleótido desnudo, o polinucleótido conjugado.

Definición de polipéptidos mediante inmunorreactividad

Los polipéptidos descritos en este documento proporcionan varias nuevas secuencias polipeptídicas (por ejemplo, polipéptidos que comprenden aminoácidos en el caso de proteínas sintetizadas en los sistemas de traducción de este documento, o por ejemplo, en el caso de las nuevas sintetisas, nuevas secuencias de aminoácidos convencionales), los polipéptidos también proporcionan nuevas características estructurales que se pueden reconocer, por ejemplo, en ensayos inmunológicos. Se puede generar antisuero que se una específicamente a los polipéptidos descritos en este documento. El término "anticuerpo", como se usa en este documento, incluye, pero no está limitado a un polipéptido codificado sustancialmente por un gen de una inmunoglobulina o genes de inmunoglobulinas, o sus fragmentos que se unen y reconocen específicamente un analito (antígeno). Sus ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos, y de cadena sencilla, y similares. En el término "anticuerpo" como se usa en este documento también están incluidos fragmentos de inmunoglobulinas, incluyendo

fragmentos F_{ab} y fragmentos producidos mediante una librería de expresión, incluyendo la presentación de fagos. Véase, por ejemplo, Paul, Fundamental Immunology, 4ª Ed., 1999, Raven Press, Nueva York, para la estructura y terminología de los anticuerpos.

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

Uso de O-ARNt y O-RS y pares O-ARNt/O-RS

- 20 Las composiciones preparadas mediante los métodos de la invención opcionalmente están en una célula. Los pares O-ARNt/O-RS o componentes individuales descritos en este documento se pueden usar en la maquinaria de traducción de un sistema hospedador, que da lugar a la incorporación de un alquil aminoácido a una proteína. La Publicación internacional número WO 2002/085923 de Schultz, y col., titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", describe este proceso. Por ejemplo, cuando se introduce un par O-ARNt/O-RS en un hospedador, por ejemplo, *Escherichia coli*, el par da lugar a la incorporación *in vivo* de un alquil aminoácido, que se puede añadir exógenamente al medio de crecimiento, en una proteína, por ejemplo, una proteína de ensayo mioglobina o una proteína terapéutica, en respuesta a un codón selector, por ejemplo, un codón sin sentido ámbar. Opcionalmente, las composiciones de la invención pueden estar en un sistema de traducción *in vitro*, o en (un) sistema(s) celular(es) *in vivo*. Se pueden usar proteínas con el alquil aminoácido en cualquiera de un amplio espectro de aplicaciones. De forma muy notable, el resto alqueno incorporado a la proteína puede servir como diana para cualquiera de una amplia variedad de modificaciones, por ejemplo, entrecruzamiento con otras proteínas, con moléculas pequeñas tales como marcadores o colorantes y/o biomoléculas. Con estas modificaciones, la incorporación del alquil aminoácido puede producir proteínas terapéuticas mejoradas y se pueden usar para alterar o mejorar la función catalítica de enzimas. En algunos aspectos, la incorporación y modificación posterior del alquil aminoácido a una proteína puede facilitar el estudio de la estructura de la proteína, de las interacciones con otras proteínas, y similares.

KITS

- 40 En este documento se desvelan kits. Por ejemplo, un kit para la producción de una proteína que comprende al menos un alquil aminoácido en una célula, en el que el kit incluye un contenedor que contiene una secuencia de polinucleótidos que codifica un O-ARNt y/o un O-ARNt, y/o una secuencia de polinucleótidos que codifica una O-RS, y/o una O-RS. En una realización, el kit además incluye un alquil aminoácido tal como para-propargiloxifenilalanina. En otra realización, el kit además comprende instrucciones para la producción de la proteína.

Ejemplos

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

EJEMPLO 1

Evolución de un par ARNt/sintetasa ortogonal para la incorporación de un alquil aminoácido a proteínas en *E. coli*.

- 55 Al desarrollar la especificidad de pares ARNt-sintetasa ortogonales, hemos incorporado de forma selectiva y eficiente una serie de aminoácidos no naturales a proteínas en respuesta a codones sin sentido y de desplazamiento del marco de lectura tanto en procariotas como en eucariotas (Anderson y col. (2004) Proc. Natl. Acad. Sci. U.S.A., 101:7566; Alfonta y col. (2003) J. Am. Chem. Soc., 125:14662; Wang y col. (2003) Proc. Natl. Acad. Sci. U.S.A., 100:56; Chin y col. (2003) Science 301:964; Chin y col. (2002) Proc. Natl. Acad. Sci., 99:11020; y Wang y col. (2001) Science 292:498). La presente invención proporciona composiciones y métodos para la incorporación biosintética a proteínas de aminoácidos que contienen restos alquilo reactivos usando la maquinaria de traducción de *E. coli*. La biosíntesis usando los componentes de traducción de *E. coli* se puede producir *in vivo* (por ejemplo, en la célula de *E. coli*) o *in vitro* usando extractos celulares en bruto o componentes de traducción purificados. El grupo alquilo que se incorpora a proteínas se conjuga fácil y específicamente con restos que contienen azido, proporcionando así una diana útil para la modificación/manipulación de proteínas.

La química de los grupos alquínico y azido (mostrada en la Figura 1B) es completamente ortogonal a las químicas de todos los grupos funcionales endógenos presentes en proteínas. Un ejemplo de su reactividad única es la formación irreversible de triazoles mediante una cicloadición [3+2] (véase, Figura 1B; y Padwa, en *Comprehensive Organic Synthesis*; [Trost, B. M., Ed.] Pergamon: Oxford, 1991, Vol. 4, p 1069; Huisgen, en *1,3-Dipolar Cycloaddition Chemistry*, [Padwa, A., Ed.] Wiley: Nueva York, 1984; p 1). Cuando esta reacción se lleva a cabo en presencia de cobre (I) a temperatura ambiente en un medio acuoso (condiciones suficientemente suaves para modificar muestras biológicas), evoluciona de una forma completamente regioselectiva (Rostovtsev y col. (2002) *Angew. Chem. Int. Ed.*, 41:2596) y se puede usar para modificar selectivamente proteínas en las que se han introducido grupos funcionales alquínico y azido (Deiters y col. (2003) *J. Am. Chem. Soc.*, 125:11782; Wang y col. (2003) *J. Am. Chem. Soc.*, 125:3192; Link y Tirrell (2003) *J. Am. Chem. Soc.*, 125:11164).

La invención descrita en este documento proporciona pares ARNt/ARNt-sintetasa ortogonales derivados de componentes de *Methanococcus jannaschii* que incorporan selectivamente el alquínico aminoácido para-propargiloxifenilalanina (abreviado pPro-Phe; también conocido como ácido 2-amino-3-[4-(prop-2-inoxi)fenil]-propiónico según la nomenclatura de la IUPAC; estructura mostrada en la Figura 1A; y con la designación de la estructura química 1) en un sistema hospedador de *E. coli*. El presente estudio demuestra que la pPro-Phe se incorpora selectivamente a proteínas expresadas en *E. coli* usando los nuevos reactivos ARNt y ARNt-sintetasa ortogonales proporcionados en este documento.

En este documento presentamos la evolución de un par ARNt-sintetasa ortogonal derivado de un par tirosil-ARNt/ARNt-sintetasa de *M. jannaschii* (MjTyrRS/ARNt_{CUA}), en donde el par ortogonal no tiene afinidad o tiene muy poca afinidad por cualquiera de los aminoácidos convencionales (es decir, de origen natural). La ARNt-sintetasa ortogonal derivada carga selectivamente el ARNt_{CUA} supresor ámbar con pPro-Phe, y además, el ARNt supresor aminoacilado (es decir, el ARNt "cargado") es usado como sustrato por el aparato de traducción endógeno de *E. coli* para incorporar la pPro-Phe en respuesta a un codón de parada ámbar TAG (un codón selector) que se encuentra en un transcrito. La ortogonalidad (Steer y Schimmel (1999) *Biol. Chem.*, 274:35601) de este par ARNt/sintetasa garantiza que ni el ARNt ni la sintetasa experimenten reacciones cruzadas con los ARNt o las sintetastas endógenos de *E. coli* y que el aminoácido no natural únicamente se libere en respuesta a un codón sinsentido ámbar, TAG.

Se generó una librería de $\sim 10^7$ tirosil-ARNt sintetastas diferentes de *M. jannaschii* mediante mutagénesis de la tirosil-ARNt sintetasa de *M. jannaschii* de tipo silvestre. Para crear la librería de MjTyrRS, en primer lugar las cinco posiciones seleccionadas para la mutación se convirtieron a codones de alanina. El gen MjTyrRS se expresó bajo el control del promotor y terminador GlnRS de *E. coli* en el plásmido pBK-JYRS, un plásmido derivado de pBR322 con resistencia a kanamicina. Los restos Tyr32, Glu107, Asp158, Ile159, y Leu162 se sustituyeron con Ala mediante mutagénesis dirigida de sitio para dar el plásmido pBK-JYA5. Se usaron ocho nucleótidos con NNK (N = A + T + G + C y K = G + T) en los sitios de mutación para la amplificación por PCR del mutante Ala5 MjTyrRS (pBK-JYA5) y se volvió al ligar en el pBK-JYA5 digerido con Nde I-Pst I para generar la librería de MjTyrRS. Los vectores ligados se transformaron en células competentes DH10B de *E. coli* para dar una librería de $1,6 \times 10^9$ unidades formadoras de colonias.

Las secuencias de polinucleótidos y de aminoácidos de la molécula tirosil-ARNt sintetasa de *M. jannaschii* de tipo silvestre se muestran en la **Figura 2**, y también se proporcionan en la **SEQ ID NO: 3 y 2**, respectivamente. La mutagénesis consistió en la distribución aleatoria de cinco restos con sitios activos (Tyr32, Glu107, Asp158; Ile159, y Leu162), basándose en la estructura cristalina de la tirosil-ARNt sintetasa homóloga de *Bacillus stearothermophilus*.

Después de la mutagénesis, el reservorio de sintetastas se pasó por rondas de selección positiva y negativa. La selección positiva se basa en la supresión de un codón de parada ámbar en un sitio permisivo (Asp112) en el gen de la cloranfenicol acetiltransferasa (CAT). Cuando las células de *E. coli* que albergan la librería mutante de MjTyrRS, el gen CAT mutado y un ARNt_{CUA} supresor ámbar Mj co-expresado se crecen en medio mínimo en presencia de pPro-Phe (1 mM) y cloranfenicol (80 µg/ml), las únicas células supervivientes son aquellas células que contienen una sintetasa mutante que aminoacila el ARNt_{CUA} con un aminoácido endógeno o pPro-Phe. A continuación los genes de la librería de sintetastas se transformaron en células que contienen un gen mutado que codifica la proteína tóxica barnasa, que tiene tres mutaciones ámbar en sitios permisivos (Gln2, Asp44, Gly65). El vector que porta el informador barnasa también contenía el ARNt supresor. El crecimiento de estas células en ausencia de pPro-Phe seleccionado contra sintetastas fue capaz de aceptar aminoácidos endógenos como sustratos. Después de tres rondas de selección, se detectaron 96 clones por dependencia de la velocidad de crecimiento en presencia o ausencia de pPro-Phe, y se identificaron y secuenciaron ocho clones candidatos. Las sustituciones de aminoácidos observadas en estos clones se muestran en la **Figura 3**. Las secuencias de polinucleótidos y aminoácidos de los ocho clones también se proporcionan en la **SEQ ID NO: 4 a 19**.

Se observan tendencias consenso en las sustituciones de aminoácido en los ocho clones de O-RS mutantes. Se encontró una preponderancia de los siguientes aminoácidos en el bolsillo de unión de la mayoría de los clones: Ala32, Pro107/Gln107, Ala158, Ile159, y Ala162/Pro162. Las mutaciones Tyr32→Ala32 y Asp158→7Ala158 pueden producir la pérdida de puentes de hidrógeno entre Tyr32, Asp158 y el sustrato natural tirosina, desfavoreciendo así su unión. Cabría esperar la aparición de pequeñas cadenas laterales y mayoritariamente hidrófobas para facilitar la unión de pPro-Phe. También se observó una mutación Leu110→Phe110 adicional en uno de los clones (pPro-

PheRS-1).

La sintetasa pPro-PheRS-1 se seleccionó para su posterior caracterización. Esta sintetasa confiere a *E. coli* resistencia a cloranfenicol con valores de Cl_{50} de 110 y 5 $\mu\text{g/ml}$ en presencia y ausencia de pPro-Phe, respectivamente. Esta gran diferencia entre la resistencia a cloranfenicol con y sin pPro-Phe sugiere una especificidad *in vivo* sustancial de la pPro-PheRS-1 para el aminoácido no natural pRPO-Phe.

EJEMPLO 2

10 Incorporación específica de sitio de un alquil aminoácido a una proteína en *E. coli*

El ARNt_{CUA} supresor mutante ámbar y el par ortogonal pPro-PheRS-1 se usaron en *E. coli* para incorporar selectivamente pPro-Phe a mioglobina de cachalote, una hemoproteína monomérica de 153 restos que ha sido objeto de una serie de estudios estructurales, mecanísticos, y de plegamiento de proteínas (Reedy y Gibney (2004) Chem. Rev., 104:617, y las referencias allí citadas; Uzawa y col. (2004) Proc. Natl. Acad. Sci. U.S.A., 101:1171, y las referencias allí citadas; Wright, y Baldwin en Frontiers in Molecular Biology: Mechanisms of Protein Folding, [Pain, R., ed.] Oxford University Press, Londres, 2000, pág. 309).

Para producir la mioglobina modificada con alquililo, el cuarto codón del marco de lectura abierto de la mioglobina (Ser4) se mutó a TAG (codón ámbar) y se añadió una etiqueta C-terminal de 6×His (hexahistidina) al marco de lectura abierto. Para expresar la proteína mutante, el plásmido pBAD/JYAMB-4TAG (que codifica el gen de la mioglobina mutante de cachalote con un promotor de arabinosa y un terminador rrnB, el tirosil-ARNt_{CUA} sobre un promotor lpp y un terminador rrnC; y un marcador de resistencia a tetraciclina) se co-transformó con un vector pBK (que codifica la sintetasa mutante y un gen de resistencia a kanamicina) en *E. coli* DH10B. Las células se amplificaron en medio de Luria-Bertani (5 ml) suplementado con tetraciclina (25 mg/l) y kanamicina (30 mg/l), se lavaron con tampón fosfato, y se usaron para inocular 500 ml de medio de glicerol líquido mínimo (suplementado con leucina 0,3 mM) que contiene los antibióticos apropiados, pPro-Phe (1 mM), y arabinosa (0,002 %). Las células se crecieron hasta saturación y a continuación se recogieron por centrifugación. La proteína se purificó usando cromatografía de afinidad de Ni con un rendimiento de 2 mg/l después de la purificación mediante la cromatografía de Uni-afinidad y se estima homogénea en un 90 % mediante SDS-PAGE/tinción con Gelcode® Blue (Pierce Biotechnology, Inc.). Se obtuvo un rendimiento total de ~1 mg de mioglobina mutante.

La proteína producida de esta manera se visualiza en la Figura 4, carril 1, usando tanto tinción Gelcode® Blue como transferencia de Western usando un anticuerpo dirigido contra His6. En ausencia de pPro-Phe, no había mioglobina visible tras la tinción o la transferencia de Western (usando un anticuerpo dirigido contra His6), lo que indica una alta selectividad de la sintetasa desarrollada (véase Figura 4, carril 2).

EJEMPLO 3

40 Confirmación por espectrometría de masas de la incorporación de alquil aminoácido a una proteína en *E. coli*

Para confirmar aún más la identidad del aminoácido incorporado en el sitio de un codón de parada ámbar en mioglobina mutante, una digestión triptica de la mioglobina se sometió a cromatografía líquida/espectrometría de masas en tándem. La mioglobina mutante usada en este experimento contenía un codón de parada ámbar modificado genéticamente en posición 74. La incorporación de pPro-Phe en esta posición (pPro-Phe74) se sometió a ensayo con la mutación 74. Se usó el mutante de mioglobina-74TAG en lugar del Ser4→TAG (detención ámbar) descrito previamente debido a las propiedades mejoradas para el análisis de LC MS/MS.

Después de la expresión eubacteriana, la mioglobina 74TAG se purificó usando una columna de afinidad de níquel. Las bandas de proteína se visualizaron mediante tinción con Gelcode® Blue en un gel SDS-PAGE. Las bandas de gel correspondientes a la mioglobina mutante se escindieron del gel de poliacrilamida, se cortaron en cubos de 1,5 mm y se sometieron a hidrólisis con tripsina esencialmente como se describe (Shevchenko y col. (1996) Anal. Chem.; 68:850-858).

Los péptidos tripticos que contienen el aminoácido no natural se analizaron mediante HPLC de fase inversa con nanoflujo/ μ ESI/MS con un espectrómetro de masas de trampa iónica LCQ. El análisis de espectrometría de masas en tándem y cromatografía líquida (LC-MS/MS) se llevó a cabo sobre un espectrómetro de masas de trampa iónica Finnigan LCQ Deca (Thermo Finnigan) equipado con un Nanospray HPLC (Agilent 1100 series).

Los iones precursores correspondiente a los iones con una y con dos cargas del péptido HGVTVLALGY*ILK (SEQ ID NO: X) que contienen los aminoácidos no naturales (denotados por Y*) se separaron y se fragmentaron con un espectrómetro de masas de trampa iónica. Los resultados de este análisis se proporcionan en la Tabla 5. Las masas iónicas de los fragmentos se pudieron asignar de forma inequívoca, confirmando la incorporación específica de sitio de la pPro-Phe. Los análisis de LC MS/MS no indicaban la incorporación de ningún aminoácido natural en esta posición, confirmando la alta selectividad de la sintetasa desarrollada.

EJEMPLO 4**Derivación de una proteína que contiene un alquinil aminoácido mediante cicloadición [3+2]**

5 Las proteínas que contienen grupos funcionales alquinilo se pueden dirigir eficazmente para su modificación mediante el uso de una reacción de cicloadición [3+2]. El presente ejemplo describe la derivación de la alquinil mioglobina con dos moléculas colorantes que contienen azido diferentes. La mioglobina mutante usada en este ejemplo incorpora pPro-Phe en el cuarto codón (Ser4→pPro-Phe4), como se describe en el Ejemplo 1.

10 La mioglobina Ser4→pPro-Phe4 se produjo en *E. coli* como se describe en el Ejemplo 2, y a continuación se derivó con los colorantes 2 o 3 funcionalizados con azido, que contienen los fluoróforos dansilo y fluoresceína, respectivamente (como se muestra en las Figuras 6A y 6B; véase también, Deiters y col. (2003) J. Am. Chem. Soc., 125:11782; Wang y col. (2003) J. Am. Chem. Soc., 125:3192; Link y Tirrell (2003) J. Am. Chem. Soc., 125:11164). La reacción de derivación por cicloadición [3+2] se ilustra en la Figura 7A.

15 Para la reacción de cicloadición, se añadió 1 µl de CuSO₄ (disolución madre 50 mM en H₂O; 1 mM en el volumen de reacción final), 2 µl decolorante 2 o 3 (50 mM en EtOH), 2 µl de tris(1-bencil-1H-[1,2,3]triazol-4-ilmetil)amina (50 mM en DMSO), y 1 mg de alambre de Cu o 1 µl de tris(carboxietil)fosfina (100 mM en H₂O) (como agentes reductores) a 45 µl de mioglobina mutante purificada (~0,5 mg/ml) en tampón fosfato 0,1 M (pH=8). Después de 8 horas a temperatura ambiente o durante toda la noche a 4 °C, se añadieron 450 µl de H₂O y la mezcla se centrifugó a través de una membrana de diálisis (límite de 10 kDa). Después de lavar el sobrenadante con tampón fosfato 2 × 500 µl mediante centrifugación, la solución se llevó hasta un volumen de 50 µl.

25 El uso de alambre de Cu o tris(carboxietil)fosfina (2 mM) como agentes reductores en general da lugar a una eficiencia de marcaje similar. En contraste a las observaciones previas (Wang y col. (2003) J. Am. Chem. Soc., 125:3192), la presencia o ausencia de ligando tris(1-bencil-1H-[1,2,3]triazol-4-ilmetil)amina no tuvo una influencia significativa sobre el resultado de estas reacciones. A continuación se analizó una muestra de 20 µl de las proteínas marcadas con fluorescencia (Blake (2001) Curr. Opin. Pharmacol., 1:533; Wouters y col. (2001) Trends in Cell Biology 11:203; Zacharias y col. (2000) Curr. Opin. Neurobiol., 10:416) mediante SDS-PAGE y se obtuvieron las imágenes del gel. Se tomó la imagen de la mioglobina mutante modificada con el colorante dansilo 2 ($\lambda_{\text{ex}} = 337$ nm, $\lambda_{\text{em}} = 506$ nm) en el gel a 360 ± 30 nm usando un densitómetro Eagle Eye (Stratagene). La unión del colorante de fluoresceína 3 ($\lambda_{\text{ex}} = 495$ nm, $\lambda_{\text{em}} = 516$ nm) se visualizó a 450 ± 30 nm con un Storm Phosphorimager (Molecular Dynamics). Los resultados de estas imágenes fluorescentes obtenidas se muestran en la Figura 7B. La mioglobina mutante se marcó eficazmente con los dos colorantes 2 y 3. La eficiencia de marcaje fue del ~75 % como se determina por comparación de los valores A_{280}/A_{495} para la mioglobina marcada con 3 (véase, Wang y col. (2003) J.). La selectividad de esta bioconjugación se verificó por el hecho de que no se observó reacción entre la mioglobina de tipo silvestre y 2 o 3 (resultados no mostrados).

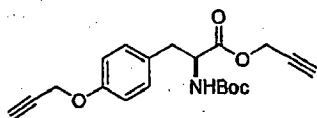
40 La descripción proporcionada en este documento demuestra que se puede incorporar eficaz y selectivamente un alquinil aminoácido, por ejemplo, para-propargiloxifenilalanina, a proteínas en un organismo, por ejemplo, *E. coli*. A continuación estos aminoácidos se pueden dirigir químicamente dentro de la proteína para su conjugación, por ejemplo, mediante cicloadición [3+2] usando restos azido, y además, en donde esta modificación dirigida es altamente específica y regioselectiva. La capacidad para incorporar alquinil aminoácidos a proteínas con especificidad de sitio proporciona una herramienta valiosa en el estudio de cualquier proteína en la que se desee la conjugación o modificación de la proteína.

EJEMPLO 5**Síntesis del alquinil aminoácido no natural para-propargiloxifenilalanina**

50 El alquinil aminoácido no natural para-propargiloxifenilalanina (abreviado pPro-Phe; véase **Figura 1A**, compuesto **1**) se sintetizó a partir de N-Boc-tirosina disponible en el mercado en tres etapas (véase, Deiters y col. (2003) J. Am. Chem. Soc., 125:11782; Wang y col. (2003) J. Am. Chem. Soc., 125:3192; Link y Tirrell (2003) J. Am. Chem. Soc., 125:11164) con un rendimiento global del 81 %.

Etapas 1

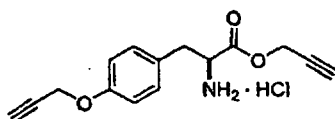
60 Se suspendió N-terc-butoxicarbonil-tirosina (2 g, 7 mmol, 1 equiv.) y K₂CO₃ (3 g, 21 mmol, 3 equiv.) en DMF anhidra (15 ml). Lentamente se añadió bromuro de propargilo (2,1 ml, 21 mmol, 3 equiv., solución al 80 % en tolueno) y la mezcla de reacción se agitó durante 18 horas a temperatura ambiente. Se añadió agua (75 ml) y Et₂O (50 ml), las capas se separaron y la fase acuosa se extrajo con Et₂O (2 × 50 ml). Las capas orgánicas combinadas se secaron (MgSO₄) y el disolvente se retiró a presión reducida. El producto (**4**), mostrado y formulado a continuación, se obtuvo en forma de aceite amarillo (2,3 g, 91 %) y se usó en la siguiente etapa sin purificación adicional.



éster propargílico del ácido N-terc-butoxicarbonilamino-3-[4-(prop-2-iniloxi)fenil]-propiónico (compuesto 4)

5 Etapa 2

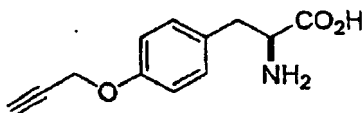
Se añadió cloruro de acetilo (7 ml) a metanol (60 ml) con cuidado a 0 °C para dar una solución 5 M de HCl anhidro en MeOH. Se añadió el producto de la etapa anterior (compuesto 4; 2 g, 5,6 mmol) y la reacción se agitó durante 4 horas al tiempo que se dejaba calentar a temperatura ambiente. Después de retirar los compuestos volátiles a presión reducida, se obtuvo un sólido amarillento (compuesto 5, mostrado y formulado a continuación; 1,6 g, 98 %) que se usó directamente en la siguiente etapa



15 éster propargílico del ácido 2-amino-3-[4-(prop-2-iniloxi)fenil]-propiónico (compuesto 5)

Etapa 3

El éster propargílico (1,6 g, 5,5 mmol) de la etapa anterior (5) se disolvió en una mezcla acuosa de NaOH 2 N (14 ml) y MeOH (10 ml). Después de agitar durante 1,5 horas a temperatura ambiente, el pH se ajustó a 7 añadiendo HCl concentrado. Se añadió agua (20 ml) y la mezcla se mantuvo a 4 °C durante toda la noche. El precipitado se filtró, se lavó con agua helada en hielo, y se secó al vacío para dar 1,23 g (90 %) de pPro-Phe (1) en forma de sólido blanco. RMN ¹H (400 MHz, D₂O; como sal de potasio en D₂O) δ 7,20 (d, J = 8,8 Hz, 2 H), 6,99 (d, J = 8,8 Hz, 2 H), 4,75 (s, 2 H), 3,50 (dd, J = 5,6, 7,2 Hz, 1 H), 2,95 (dd, J = 5,6, 13,6 Hz, 1 H), 2,82 (dd, J = 7,2, 13,6 Hz, 1 H); RMN ¹³C (100 MHz, D₂O) δ 181,3, 164,9, 155,6, 131,4, 130,7, 115,3, 57,3, 56,1, 39,3; HRMS (CI) m/z 220,0969 [C₁₂H₁₃NO₃ (M+1) requiere 220,0968].

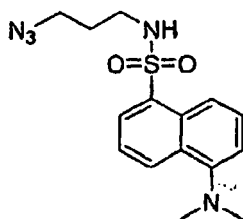


30 para-propargiloxifenilalanina (compuesto 1)

Ejemplo comparativo 6

Síntesis del colorante azido 2

El colorante azido 2 (véase **Figura 6A**; compuesto 2) se sintetizó de acuerdo con el protocolo siguiente. Se añadió 3-azidopropilamina (371 mg, 3,71 mmol, 3 equiv.) (sintetizada de acuerdo con Carboni y col. (1993) Org. Chem., 58:3736-3741) a una solución de cloruro de dansilo (500 mg, 1,85 mmol, 1 equiv.) y trietilamina (258 µl, 1,85 mmol, 1 equiv.) en CH₂Cl₂ (10 ml) a 0 °C. Después de agitar durante 1 h, la mezcla de reacción se calentó a temperatura ambiente y se agitó durante una hora más. Los compuestos volátiles se retiraron al vacío y el producto en bruto se purificó por cromatografía en gel de sílice (Et₂O/hexanos = 1:1) para dar 2 (548 mg, 89 %) en forma de aceite amarillo. RMN ¹H (400 MHz, CDCl₃) δ 8,55 (d, J = 8,4 Hz, 1 H), 8,29 (d, J = 8,8 Hz, 1 H), 8,23 (dd, J = 1,2, 7,2 Hz, 1 H), 7,56-7,49 (comp, 2 H, 7,18 (d, J = 7,6 Hz, 1 H), 5,24 (sa, 1 H), 3,21 (t, J = 6,4 Hz, 2 H), 2,95 (dt, J = 6,4 Hz, 2 H), 2,89 (s, 6 H), 1,62 (quin, J = 6,4 Hz, 2 H); RMN ¹³C (100 MHz, CDCl₃) δ 134,3, 130,4, 129,7, 129,4, 128,4, 123,3, 118,8, 115,3, 48,6, 45,4, 40,6, 28,7 (no todas las señales de los átomos de carbono cuaternario son visibles en el espectro de RMN ¹³C); HRMS (CI) m/z 334,1336 [C₁₅H₂₀N₅O₂S (M+1) requiere 334,1332].

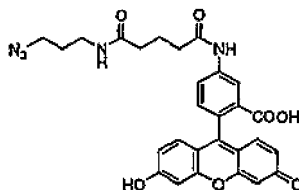


colorante azido 2 (compuesto 2)

5 EJEMPLO comparativo 7

Síntesis del colorante azido 3

El colorante azido 3 (véase **Figura 6B**; compuesto 3) se sintetizó de acuerdo con el protocolo siguiente. Se añadió EDCI (83 mg, 0,43 mmol, 1 equiv.) a una solución de fluoresceinamina (150 mg, 0,43 mmol, 1 equiv.) y ácido 4-(3-azidopropilcarbamoil)-butírico (92 mg, 0,43, 1 equiv.) en piridina (2 ml) a temperatura ambiente. El ácido 4-(3-azidopropilcarbamoil)-butírico se sintetizó haciendo reaccionar la 3-azidopropilamina con anhídrido del ácido glutárico. La suspensión se agitó durante toda la noche y la mezcla de reacción se vertió en H₂O (15 ml). La solución se acidificó (pH < 2) añadiendo HCl concentrado. Después de agitar durante 1 h, el precipitado se retiró por filtración, se lavó con HCl 1N (3 x 3 ml) y se disolvió en una pequeña cantidad de EtOAc. La adición de hexano dio lugar a la precipitación de 3 en forma de cristales naranjas, que se recogieron y se secaron al vacío (200 mg, 86 %). RMN ¹H (400 MHz, CD₃OD) δ 8,65 (s, 1 H), 8,15 (d, J = 8,4 Hz, 1 H), 7,61-7,51 (comp, 2 H), 7,40 (d, J = 8,4 Hz, 1 H), 7,35 (sa, 2 H), 7,22-7,14 (comp, 2 H), 6,85-6,56 (comp, 3 H), 3,40-3,24 (comp, 4 H), 2,54 (t, J = 7,2 Hz, 2 H), 2,39-2,30 (comp, 2 H), 2,10-1,99 (comp, 2 H), 1,82-1,72 (comp, 2 H); RMN ¹³C (100 MHz, CD₃OD) δ 175,7,174,4,172,4,167,9, 160,8,143,0,134,3,132,9,131,8,129,6,124,4,123,3,121,1,118,5 103,5, 50,2, 38,0, 37,2, 36,2, 29,8, 22,9;4 HRMS (CI) m/z 544,1835 [C₂₈H₂₅N₅O₇ (M+1) requiere 544,1827].



colorante azido 3 (compuesto 3)

EJEMPLO 8

O-RS y O-ARNt ilustrativos para la incorporación de alquil aminoácidos a *E. coli*

- Un O-ARNt ilustrativo comprende la SEQ ID NO: 1 (véase Ejemplo 9, Tabla 4). Las O-RS ilustrativas incluyen las secuencias de aminoácidos proporcionadas en las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16 y 18 (véase Figura 3 y Ejemplo 9, Tabla 4).
- Los ejemplos de polinucleótidos que codifican O-RS o fragmentos de las mismas incluyen cualquier polinucleótido que codifique una secuencia de aminoácidos que comprende las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16 y 18. Por ejemplo, los polinucleótidos proporcionados en las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 y 19 codifican O-RS ilustrativas.

EJEMPLO 9

Secuencias de nucleótidos y de aminoácidos

- Este ejemplo proporciona secuencias de nucleótidos y aminoácidos para diversos polinucleótidos y polipéptidos, respectivamente. Las secuencias proporcionadas en la Tabla 4 a continuación están destinadas únicamente a proporcionar ejemplos, y no se pretende que la invención esté limitada de ninguna forma por las secuencias proporcionadas en la Tabla 4.

TABLA 4

SEQ ID NO:	Descripción	SECUENCIA
1	mutARN _{CUA} ^{Tyr}	CCGGCGGUAGUUCAGCAGGGCAGAACGGCGG ACUCUAAAUCCGCAUGGCGCUGGUUCAAUCC CGGCCCGCCGGACCA
2	Secuencia de aminoácidos de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i> (MjTyrRS)	MDEFEMIKRNTSEIISEEELREVLKKDEKSA YIGFEPGSKIHLGHYLQIKKMIDLQNAFGDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSEFQLDKDYTLNVYRLALK TTLKRARRSMELIAREDENPKVAEVIYPIMQ VNDIHYLGVDVAVGGMEQRKIHMLARELLPK KVVCIHNPVLTGLDGEKMSSSSKGNFIAVDD SPEEIRAKIKKAYCPAGVVEGNPIMEIAKYF LEYPLTIKRPEKFGGDLTVNSYEELESLEFKN KELHPMDLKNVAEELIKILEPIRKRL
3	Secuencia de nucleótidos de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i> (MjTyrRS)	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTTACATAGGT TTTGAACCAAGTGGTAAAATACATTTAGGGCATT ATCTCCAAATAAAAAAGATGATTGATTTACAAAA TGCTGGATTTGATATAATTATATTGTTGGCTGAT TTACACGCCATTAAACCAGAAAGGAGAGTTGG ATGAGATTAGAAAAATAGGAGATTATAACAAAA AGTTTGTGAAGCAATGGGGTTAAAGGCAAAATAT GTTTATGGAAGTGAATTCAGCTTGATAAGGATT ATACACTGAATGCTTATAGATTGGCTTTAAAAAC TACCTTAAAAAGAGCAAGAAGGAGTATGGAACCT ATAGCAAGAGAGGATGAAAATCCAAAGGTTGCTG AAGTTATCTATCCAATAATGCAGGTTAATGATAT TCATTATTTAGGCGTTGATGTTGCAGTTGGAGGG ATGGAGCAGAGAAAAATACACATGTTAGCAAGGG AGCTTTTACCAAAAAAGGTTGTTTGTATTACAA CCCTGCTTTAACGGGTTTGGATGGAGAAGGAAAG ATGAGTTCTTCAAAGGGAATTTTATAGCTGTTG ATGACTCTCCAGAAGAGATTAGGGCTAAGATAAA GAAAGCATACTGCCAGCTGGAGTTGTTGAAGGA AATCCAATAATGGAGATAGCTAAATACTTCCTTG AATATCCTTTAACCATAAAAAGGCCAGAAAAATT TGGTGGAGATTTGACAGTTAATAGCTATGAGGAG TTAGAGAGTTTATTTAAAAATAAGGAATTGCATC CAATGGATTTAAAAAATGCTGTAGCTGAAGAACT TATAAAGATTTTAGAGCCAATTAGAAAGAGATTA MDEFEMIKRNTSEIISEEELREVLKKDEKSA
4	pPro-PheRS-1; Secuencia de aminoácidos del aislado 1 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa (derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>), que tiene los cambios de aminoácidos: Tyr32→Ala32 Glu107→Pro107 Leu110→Phe110 Asp158→Ala158 Leu162→Ala162	AIGFEPGSKIHLGHYLQIKKMIDLQNAFGDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSPFQFDKDYTLNVYRLAL KTTLKRARRSMELIAREDENPKVAEVIYPIM QVNAIHYAGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEKMSSSSKGNFIAV DDSPREEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGGDLTVNSYEELESLEF KNKELHPMDLKNVAEELIKILEPIRKRL

5	Secuencia de nucleótidos de pPro-PheRS-1	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTGCGATAGG TTTTGAACCAAGTGTTAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTCCGTTCCAGTTTGATAAGG ATTATACACTGAATGCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATGC AATTCATTATGCTGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATT CACAACCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTTCTTCAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACTT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAACTTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
6	pPro-PheRS-2; Secuencia de aminoácidos del aislado 2 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→Ala32 Glu107→Lys107 Asp158→Ala158 Leu162→Ala162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSEIISEEELREVLKKDEKSA AIGFEPGSKIHLGHYLQIKKMIDLQNAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSKFQLDKDYTLNVYRLAL KTTLKRARRSMELIAREDENPKVAEVIYPIM QVNAIHYAGVDVAVGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEGKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGGDLTVNSYEELESF KNKELHPMDLKNVAEELIKILEPIRKRL

7	Secuencia de nucleótidos de pPro-PheRS-2	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTGCGATAGG TTTTGAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTAAGTTCCAGCTTGATAAGG ATTATACACTGAATGTCCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATGC AATTCATTATGCCGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATT CACAACCCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTTCTTCAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAGAAAGCATACTGCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACCT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAACCTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
8	pPro-PheRS-3, Secuencia de aminoácidos del aislado 3 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→Ala32 Glu107→Arg107 Asp158→Ala158 Leu162→Pro162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSEFISEEELREVLKKDEKSA AIGFEPGSKIHLGHYLQIKKMIDLQNAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSRFQLDKDYTLNVYRLAL KTTTLKRARRSMELIAREDENPKVAEVIYPIM QVNAIHYPGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEKMSSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGDLTVNSYEELESF KNKELHPMDLKNVAEELIKILEPIRKRL

9	Secuencia de nucleótidos de pPro-PheRS-3	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTGCGATAGG TTTTGAACCAAGTGGTAAAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTCGGTTCCAGCTTGATAAGG ATTATACACTGAATGTCATATAGATTGGCTTTAA AATACCTTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATGC AATTCATTATCCGGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATT CACAACCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTTCCTTCAAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACGCCCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACCT CCTTGAATATCCTTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAACCTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
10	pPro-PheRS-4, Secuencia de aminoácidos del aislado 4 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→His32 Glu107→Ala107 Asp158→Ala158 Leu162→Pro162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSEIISEEELREVLKKDEKSA. HIGFEPSGKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSAFQLDKDYTLNVYRLAL KTTLKRARRSMELIAREDENPKVAEVIYPIM QVNAIHYPGVDDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEKMSSSSKGNFIAV DDSPEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGGDLTVNSYEELESLE KNKELHPMDLKNVAEELIKILEPIRKRL

11	Secuencia de nucleótidos de pPro-PheRS-4	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTCATATAGG TTTTGAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTGCTTTCCAGCCTGATAAGG ATTATACACTGAATGTCCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATGC AATTCATTATCCTGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTTGATT CACAACCCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTTCCTCAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACTT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAACTTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
12	pPro-PheRS-5; Secuencia de aminoácidos del aislado 5 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→Ser32 Glu107→Gln107 Asp158→Ala158 Leu162→Ala162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSEI ISEEELREVLKKDEKSA SIGFEPSGKIHLGHYLQIKKMIDLQNAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSQFQLDKDYTLNVYRLAL KTTLKRARRSMELIAREDENPKVAEVIPIM QVNAIHYAGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEGKMSSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGDLTVNSYEELESLF KNKELHPMDLKNVAEELIKILEPIRKRL

13	Secuencia de nucleótidos de pPro-PheRS-5	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTTCGATAGG TTTTGAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCCAAAATA TGTTTATGGAAGTCAGTTCCAGCTTGATAAGG ATTATACACTGAATGTCCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATGC AATTCATTATGCCGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTGTATT CACAACCCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTCTTCAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACTGCCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACTT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAACCTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
14	pPro-PheRS-6; Secuencia de aminoácidos del aislado 6 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→Thr32 Glu107→Ser107 Asp158→Leu158 Ile159→His159 Leu162→Pro162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSEIISEEELREVLKKDEKSA TIGFEPSGKIHLGHYLQIKKMIDLQNAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSSSFQLDKDYTLNVYRLAL KTTLKRARRSMELIAREDENPKVAEVIYPIM QVNLHHYPGVDVAVGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEGKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGGDLTVNSYEELESLF KNKELHPMDLKNAVAEELIKILEPIRKRL

15	Secuencia de nucleótidos de pPro-PheRS-6	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTACGATAGG TTTTGAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTTCGTTCCAGCTTGATAAGG ATTATACACTGAATGTCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATCT TCATCATTATCCGGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATT CACAAACCCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTTCTTCAAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACTT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAAC TTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
16	pPro-PheRS-7; Secuencia de aminoácidos del aislado 7 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→Ala32 Glu107→Gln107 Asp158→Pro158 Ile159→Gly159 Leu162→Thr162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSEIISEEELREVLKKDEKSA AIGFEPSGKIHLGHYLQIKKMIDLQNAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSQFQLDKDYTLNVYRLL KTTLKRARRSMELIAREDENPKVAEVIYPIM QVNPGHYTGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEGKMSSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGGDLTVNSYEEELESLF KNKELHPMDLKNAVAELIKILEPIRKRL

17	Secuencia de nucleótidos de pPro-PheRS-7	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTGCTATAGG TTTTGAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCAAATAAAAAAGATGATTGATTTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCTATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTCAGTTCCAGCTTGATAAGG ATTATACACTGAATGTCCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATCC GGGGCATTATACGGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAAAAAAGGTTGTTTGTATT CACAACCCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTCTTCAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACTT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAAATGCTGTAGCTGAA GAACTTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
18	pPro-PheKS-8; Secuencia de aminoácidos del aislado 8 de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa, que tiene los cambios de aminoácidos: Tyr32→Ala32 Glu107→Pro107 Asp158→Ser158 Ile159→Leu159 Leu162→His162 derivada de la tirosil-ARNt sintetasa de tipo silvestre de <i>M. jannaschii</i>	MDEFEMIKRNTSELIISEEELREVLKKDEKSA AIGFEPGSKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGSPFQLDKDYTLNVYRLAL KTTLKRARRSMELIAREDENPKVAEVIYPIM QVNSLHYHGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEKMSSSKGNFIAV DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAK YFLEYPLTIKRPEKFGGDLTVNSYEELESLE KNKELHPMDLKNVAEELIKILEPIRKRL

19	Secuencia de nucleótidos de pPro-PheRS-8	ATGGACGAATTTGAAATGATAAAGAGAAACACAT CTGAAATTATCAGCGAGGAAGAGTTAAGAGAGGT TTTAAAAAAGATGAAAAATCTGCTGCTATAGG TTTTGAACCAAGTGGTAAAATACATTTAGGGCAT TATCTCCAAATAAAAAAGATGATTGATTACAAA ATGCTGGATTTGATATAATTATATTGTTGGCTGA TTTACACGCCATTTAAACCAGAAAGGAGAGTTG GATGAGATTAGAAAAATAGGAGATTATAACAAAA AAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATA TGTTTATGGAAGTCCTTTCCAGCCTGATAAGG ATTATACACTGAATGTCCTATAGATTGGCTTTAAA AACTACCTTAAAAAGAGCAAGAAGGAGTATGGAA CTTATAGCAAGAGAGGATGAAATCCAAAGGTTG CTGAAGTTATCTATCCAATAATGCAGGTTAATTC TCTGCATTATCATGGCGTTGATGTTGCAGTTG GAGGGATGGAGCAGAGAAAAATACACATGTTAGC AAGGGAGCTTTTACCAGAAAAGGTTGTTTGTATT CACAACCCCTGTCTTAACGGGTTTGGATGGAGAAG GAAAGATGAGTTCTTCAAAGGGAATTTTATAGC TGTTGATGACTCTCCAGAAGAGATTAGGGCTAAG ATAAAGAAAGCATACTGCCAGCTGGAGTTGTTG AAGGAAATCCAATAATGGAGATAGCTAAATACTT CCTTGAATATCCTTTAACCATAAAAAGGCCAGAA AAATTTGGTGGAGATTGACAGTTAATAGCTATG AGGAGTTAGAGAGTTTATTTAAAAATAAGGAATT GCATCCAATGGATTTAAAAATGCTGTAGCTGAA GAACCTATAAAGATTTTAGAGCCAATTAGAAAGA GATTA
20	Péptido tríptico de mioglobina mutante (74-TAG) usado en el análisis de espectrometría de masas	HGVTVL TALGY* ILK
21	pPRO-PheRS-consenso; secuencia consenso de aminoácidos de la para-propargiloxifenilalanina aminoacil-ARNt sintetasa	MDEFEMIKRNTSEIISEEELREVLKKDEKSA AIGFEPGSKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGS [P/Q] FQLDKDYTLNVY RLALKTTTLKRARRSMELIAREDENPKVAEVI YPIMQVNAIHY [A/P] GVDVAVGGMEQRKI HMLARELLPKKVVCIHNPVLTGLDGEKMS SKGNFIAVDDSPPEIRAKIKKAYCPAGVVEG NPIMEIAKYFLEYPLTIKRPEKFGGDLTVNS YEELESFLFNKELHPMDLKNVAEELIKILE PIRKRL
22	pPro-PheRS-con1	MDEFEMIKRNTSEIISEEELREVLKKDEKSA AIGFEPGSKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGS PFQLDKDYTLNVYRLALE TTLKRARRSMELIAREDENPKVAEVIYPIMQ VNAIHYAGVDVAVGGMEQRKIHMLARELLP KKVVCIHNPVLTGLDGEKMSSSKGNFIAV DSPEEIRAKIKKAYCPAGVVEGNPIMEIAKY FLEYPLTIKRPEKFGGDLTVNSYEELESFLF NKELHPMDLKNVAEELIKILEPIRKRL

23	pPro-PheRS-con2	MDEFEMIKRNTSEIISEEELREVLKKDEKSA A IGFEPGSKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGS P FQLDKDYTLNVYRLALK TTLKRARRSMELIAREDENPKVAEVIYPIMQ VN A IHY P GVDDAVGGMEQRKIHMLARELLP KKVVCIHNPVLTGLDGEGKMSSSKGNFIAVD DSPEEIRAKIKKAYCPAGVVEGNPIMEIAKY FLEYPLTIKRPEKFGGDLTVNSYEELESFLK NKELHPMDLKNAVAEELIKILEPIRKRL
24	pPro-PheRS-con3	MDEFEMIKRNTSEIISEEELREVLKKDEKSA A IGFEPGSKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGS Q FQLDKDYTLNVYRLALK TTLKRARRSMELIAREDENPKVAEVIYPIMQ VN A IHY A GVDDAVGGMEQRKIHMLARELLP KKVVCIHNPVLTGLDGEGKMSSSKGNFIAVD DSPEEIRAKIKKAYCPAGVVEGNPIMEIAKY FLEYPLTIKRPEKFGGDLTVNSYEELESFLK NKELHPMDLKNAVAEELIKILEPIRKRL
25	pPro-PheRS-con4	MDEFEMIKRNTSEIISEEELREVLKKDEKSA A IGFEPGSKIHLGHYLQIKKMIDLQONAGFDI IILLADLHAYLNQKGELDEIRKIGDYNKKVF EAMGLKAKYVYGS Q FQLDKDYTLNVYRLALK TTLKRARRSMELIAREDENPKVAEVIYPIMQ VN A IHY P GVDDAVGGMEQRKIHMLARELLP KKVVCIHNPVLTGLDGEGKMSSSKGNFIAVD DSPEEIRAKIKKAYCPAGVVEGNPIMEIAKY FLEYPLTIKRPEKFGGDLTVNSYEELESFLK NKELHPMDLKNAVAEELIKILEPIRKRL

REIVINDICACIONES

1. Una célula eubacteriana que comprende una primera aminoacil-ARNt sintetasa ortogonal (O-RS) que actúa en la célula, en donde la O-RS preferentemente aminoacila un primer ARNt ortogonal (O-ARNt) con un primer aminoácido no natural que es un alquil aminoácido; en donde el alquil aminoácido es una tirosina para-sustituida o una fenilalanina para-sustituida, en donde la tirosina o la fenilalanina están sustituidas en la posición para con un grupo etinilo o propinilo; y en donde la O-RS aminoacila el O-ARNt con el primer aminoácido no natural con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25.

2. Un método de producción de una proteína que comprende un alquil aminoácido no natural en una célula eubacteriana, en donde el alquil aminoácido está en una posición específica, comprendiendo el método:

(a) el suministro de una célula eubacteriana que comprende:

(i). una aminoacil-ARNt sintetasa ortogonal (O-RS);

(ii). un ARNt ortogonal (O-ARNt),

en donde dicha O-RS preferentemente aminoacila dicho O-ARNt con dicho alquil aminoácido con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25;

(iii). un ácido nucleico que codifica dicha proteína, en donde el ácido nucleico comprende al menos un codón selector que es reconocido por el O-ARNt; y

(iv). dicho alquil aminoácido, alquil aminoácido que es una tirosina para-sustituida o una fenilalanina para-sustituida, en donde la tirosina o la fenilalanina están sustituidas en la posición para con un grupo etinilo o propinilo; y

(b) el cultivo de dicha célula;

(c) la incorporación de dicho alquil aminoácido a dicha posición específica en la proteína codificada por el ácido nucleico durante la traducción de la proteína, en donde la posición específica en la proteína corresponde a la posición del codón selector en dicho ácido nucleico, produciendo así dicha proteína que comprende dicho alquil aminoácido en dicha posición específica.

3. La célula de la reivindicación 1 o el método de la reivindicación 2, en donde célula eubacteriana es una célula de *E. coli*.

4. La célula de la reivindicación 1 o el método de la reivindicación 2, en donde:

(i). la O-RS es idéntica en al menos un 95 % a una aminoacil-ARNt sintetasa de *Methanococcus jannaschii*; o

(ii). la O-RS es idéntica en al menos un 95 % a una tirosil-ARNt sintetasa de *Methanococcus jannaschii*; o

(iii). la O-RS es idéntica en al menos un 95 % a la tirosil-ARNt sintetasa de tipo silvestre de *Methanococcus jannaschii* que tiene la secuencia de aminoácidos de la SEQ ID NO: 2; o

(iv). la O-RS es idéntica en al menos un 95 % a la tirosil-ARNt sintetasa de tipo silvestre de *Methanococcus jannaschii* que tiene la secuencia de aminoácidos de la SEQ ID NO: 2, en donde la O-RS tiene una secuencia de aminoácidos que comprende:

(a) alanina, histidina, serina o treonina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;

(b) prolina, glutamina, lisina, arginina, serina o alanina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;

(c) alanina, leucina, prolina o serina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y

(d) alanina, histidina, treonina o prolina en una posición que corresponde a la posición del aminoácido 162 de la SEQ ID NO: 21; o

(v). la O-RS es idéntica en al menos un 95 % a la tirosil-ARNt sintetasa de tipo silvestre de *Methanococcus jannaschii* que tiene la secuencia de aminoácidos de la SEQ ID NO: 2, en donde la O-RS tiene una secuencia de aminoácidos que comprende:

(a) alanina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;

(b) prolina o glutamina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;

- (c) alanina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina o prolina en una posición que corresponde a la posición del aminoácido 162 de la SEQ ID NO: 21; o

5 (vi). la O-RS comprende una secuencia de aminoácidos seleccionada entre las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, y sus variantes conservativas.

5. La célula de la reivindicación 1 o el método de la reivindicación 2, en donde la célula comprende un polinucleótido que codifica la O-RS, en donde la O-RS comprende una secuencia de aminoácidos seleccionada entre las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, y sus variantes conservativas.

6. La célula o el método de la reivindicación 5, en donde el polinucleótido se selecciona entre las secuencias de nucleótidos de las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 o 19.

15 7. La célula de la reivindicación 1 o el método de la reivindicación 2, en donde:

- (a) el O-ARNt es un ARNt supresor ámbar; o
- (b) el O-ARNt comprende o está codificado por una secuencia de polinucleótidos expuesta en la SEQ ID NO: 1.

20 8. La célula de la reivindicación 1 o el método de la reivindicación 2, en donde dicho alquil aminoácido es para-propargiloxifenilalanina.

9. La célula de la reivindicación 1 o el método de la reivindicación 2, que comprende al menos un codón selector, en donde dicho codón selector es reconocido por dicho primer O-ARNt.

25 10. La célula o el método de la reivindicación 9, que comprende una segunda O-RS y un segundo O-ARNt, en donde la segunda O-RS preferentemente aminoacila el segundo O-ARNt con un segundo aminoácido no natural que es diferente del primer aminoácido no natural, y en donde el segundo O-ARNt reconoce un codón selector que es diferente del codón selector reconocido por el primer O-ARNt.

30 11. La célula de la reivindicación 1 que comprende dicho alquil aminoácido.

12. La célula de la reivindicación 11 en la que dicho alquil aminoácido es para-propargiloxifenilalanina.

35 13. La célula de la reivindicación 1 que comprende un sistema de traducción.

14. La célula de la reivindicación 13, comprendiendo dicho sistema de traducción:

- (a) dicha O-RS;
- (b) dicho O-ARNt;
- (c) un ácido nucleico que codifica un polipéptido de interés, comprendiendo el ácido nucleico al menos un codón selector, en donde el codón selector es reconocido por dicho O-ARNt; y
- (d) un alquil aminoácido, en donde dicha O-RS es capaz de cargar dicho O-ARNt con dicho alquil aminoácido.

45 15. El método de la reivindicación 2, en el que dicha proteína comprende una proteína terapéutica de tipo silvestre, una proteína diagnóstica, una enzima industrial, o un fragmento de las mismas.

50 16. El método de la reivindicación 15, en el que, habiendo obtenido dicha proteína, la proteína se formula en una composición con un vehículo farmacéuticamente aceptable.

17. El método de la reivindicación 2, en el que dicha proteína se modifica en dicha posición específica.

55 18. El método de la reivindicación 17, en el que dicha proteína comprende una unión triazol en dicha posición específica.

19. Un polipéptido que es al menos idéntico en un 90 % a la tirosil aminoacil-ARNt sintetasa de *Methanococcus jannaschii* de la SEQ ID NO: 2, en donde el polipéptido tiene una secuencia de aminoácidos que comprende:

- (a) alanina, histidina, serina o treonina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;
- (b) prolina, glutamina, lisina, arginina, serina o alanina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;
- (c) alanina, leucina, prolina o serina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina, histidina, treonina o prolina en una posición que corresponde a la posición del aminoácido 162 de la

SEQ ID NO: 21;

y en donde el polipéptido es una aminoacil-ARNt sintetasa capaz de aminoacilar preferentemente un ARNt ortogonal (O-ARNt) con un alquil aminoácido, alquil aminoácido que es una tirosina para-sustituida o una fenilalanina para-sustituida, en donde la tirosina o la fenilalanina están sustituidas en la posición para con un grupo etinilo o propinilo; y

en donde la O-RS aminoacila el O-ARNt con el primer aminoácido no natural con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25.

20. El polipéptido de la reivindicación 19, que tiene una secuencia de aminoácidos que comprende:

- (a) alanina en una posición que corresponde a la posición del aminoácido 32 de la SEQ ID NO: 21;
- (b) prolina o glutamina en una posición que corresponde a la posición del aminoácido 107 de la SEQ ID NO: 21;
- (c) alanina en una posición que corresponde a la posición del aminoácido 158 de la SEQ ID NO: 21; y
- (d) alanina en una posición que corresponde a la posición del aminoácido 162 de la SEQ ID NO: 21.

21. Un polipéptido que comprende una secuencia de aminoácidos expuesta en las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, o una de sus variantes conservativas, en donde dicho polipéptido es una aminoacil-ARNt sintetasa capaz de aminoacilar preferentemente un ARNt ortogonal (O-ARNt) en una célula eubacteriana con un alquil aminoácido, alquil aminoácido que es una tirosina que está sustituida en posición para con un grupo etinilo o propinilo, o el alquil aminoácido es una fenilalanina que está sustituida en posición para con un grupo etinilo o propinilo; y en donde la O-RS aminoacila el O-ARNt con el primer aminoácido no natural con una eficiencia de supresión de al menos el 50 % de la eficiencia de supresión de un par sintetasa ortogonal, par que es el O-ARNt de la SEQ ID NO: 1 y una polipéptido sintetasa que comprende una secuencia de aminoácidos seleccionada entre una cualquiera de las SEQ ID NO: 4, 6, 8, 10, 12, 14, 16, 18, 21, 22, 23, 24 y 25.

22. Un polinucleótido que codifica un polipéptido de una cualquiera de las reivindicaciones 19 a 21.

23. El polinucleótido de la reivindicación 22, en donde el polinucleótido se selecciona entre las SEQ ID NO: 5, 7, 9, 11, 13, 15, 17 y 19.

24. Un vector que comprende un polinucleótido de la reivindicación 22.

25. El vector de la reivindicación 24, en donde el vector es un vector de expresión.

26. Una célula que comprende un vector de la reivindicación 24.

FIG. 1A

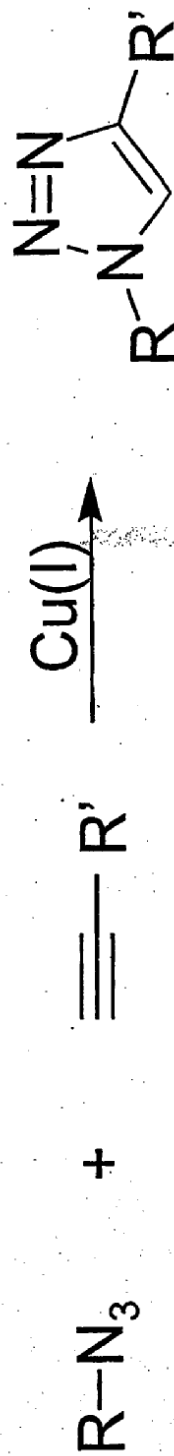
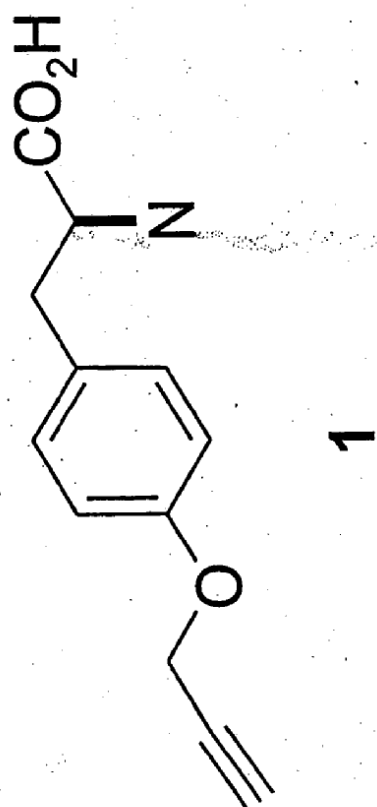


FIG. 1B

FIG. 2

1/1 1/11 31/11 61/21 91/31
 ATG GAC GAA TTT GAA ATG ATA AAG AGA AAC ACA TCT GAA ATT ATC AGC GAG GAA GAG TTA AGA GAG GTT TTA AAA AAA GAT GAA AAA TCT GCT TAC ATA GGT TTT GAA CCA AGT GGT AAA
 Met asp glu phe glu met ile lys arg asn thr ser glu ile ile ser glu glu glu leu arg glu val leu lys lys asp glu lys ser ala tyr ile gly phe glu pro ser gly lys
 121/41 151/51 271/91 301/101 331/111 451/151 571/191
 ATA CMT TTA GGG CMT TAT CTC CAA ATA AAG ATC ATT GAT TTA CAA AAT GCT GGA TTT GAT ATA ATT ATA TTG TTG GCT GAT TTA CAC GCC TAT TTA AAC CAG AAA GGA GAG TTG GAT
 ile his leu gly his tyr leu glu ile lys met ile asp leu glu asn ala gly phe asp ile ile ile leu leu ala asp leu his ala tyr leu asn gln lys gly glu leu asp
 241/81 271/91 301/101 421/141 541/181 661/221 781/261
 GAG ATT AGA AAA ATA GCA GAT TAT AAC AAA AAG GTT TTT GAA CCA ATG GGG TTA AAG GCA AAA TAT GAT TTA GCA AGT GAA TTC CAG CTT GAT AAG GAT TAT ACA CTG AAT GTC TAT AGA
 glu ile arg lys ile gly asp tyr asn lys lys val phe glu ala met gly leu lys ala lys tyr val tyr gly ser glu phe gln leu asp lys asp lys asp tyr thr leu asn val tyr arg
 361/121 391/131 511/171 631/211 751/251 871/291
 TTG GCT TTA AAA ACT ACC TTA AAA AAG GCA AGA AGG AGT ATG GAA CTT ATA GCA AGA GAG GAT GAA AAT CCA AAG GTT GCT GAA GTT ATC TAT CCA ATA ATG CAG GTT AAT GAT ATT CAT
 leu ala leu lys thr thr leu lys arg ala arg arg ser met glu leu ile ala arg glu asp glu asn pro lys val ala glu val ile tyr pro ile met gln val asn asp ile his
 481/161 601/201 721/241 841/281
 TAT TTA GGC GTT GAT GTT GCA GTT GGA GGG ATG GAG CAG AGA AAG ATA CAC ATG TTA GCA AGG GAG CTT TTA CCA AAA AAG GTT GTT TGT ATT CAC AAC CCT GTC TTA ACG GGT TTG GAT
 tyr leu gly val asp val ala val gly gly met glu gln arg lys ile his met leu ala arg glu leu pro lys lys val val cys ile his asn pro val leu thr thr gly leu asp
 601/201 721/241 841/281
 GGA GAA GGA AAG ATG ACT TCT TCA AAA GGG AAT TTT ATA OCT GTT GAT GAC TCT CCA GAA GAG AAT AGG CCT AAG ATA AAG AAA GCA TAC TGC CCA GCT GGA GTT GTT GAA GGA AAT CCA
 gly glu gly lys met ser ser lys gly asn phe ile ala val asp asp ser pro glu glu ile arg ala lys ile lys lys ala tyr cys pro ala gly val val glu gly asn pro
 721/241 841/281
 ATA ATG GAG ATA GCT AAA TAC TTC CTT GAA TAT CCT TTA ACC ATA AAA AGG CCA GAA AAA TTT GGT GGA GAT TTG ACA GTT AAT AGC TAT GAG GAG TTA GAG AGT TTA TTT AAA AAT AAG
 ile met glu ile ala lys tyr phe leu glu tyr pro leu thr ile lys arg pro glu lys phe gly gly asp leu thr val asn ser tyr glu glu leu glu ser leu phe lys asn lys
 841/281 901/301
 GAA TTG CMT CCA ATG GAT TTA AAA AAT CCT CTA GCT GAA GAA CTT ATA AAG ATT TTA GAG CCA ATT AGA AAG AGA TTA (SEQ ID NO:3)
 glu leu his pro met asp leu leu lys asn ala val ala glu glu leu ile lys leu glu pro ile arg lys arg leu (SEQ ID NO:2)

Especies de tirosil aminoacil- ARNt sintetasa de <i>Methanococcus jannaschii</i>	Posición del aminoácido						SEC ID Nº:
	32	107	110	158	159	162	
Tipo silvestre	Tyr (TAC)	Glu (GAA)	Leu (CTT)	Asp (GAT)	Ile (ATT)	Leu (TTA)	2
pPRO-PheRS-1	Ala (GCG)	Pro (CCG)	Phe (TTT)	Ala (GCA)	Ile (ATT)	Ala (GCT)	4
pPRO-PheRS-2	Ala (GCG)	Lys (AAG)	Leu (CTT)	Ala (GCA)	Ile (ATT)	Ala (GCC)	6
pPRO-PheRS-3	Ala (GCG)	Arg (CGG)	Leu (CTT)	Ala (GCA)	Ile (ATT)	Pro (CCG)	8
pPRO-PheRS-4	His (CAT)	Ala (GCT)	Leu (CTT)	Ala (GCA)	Ile (ATT)	Pro (CCT)	10
pPRO-PheRS-5	Ser (TCG)	Gln (CAG)	Leu (CTT)	Ala (GCA)	Ile (ATT)	Ala (GCC)	12
pPRO-PheRS-6	Thr (ACG)	Ser (TCG)	Leu (CTT)	Leu (CTT)	His (CAT)	Pro (CCG)	14
pPRO-PheRS-7	Ala (GCT)	Gln (CAG)	Leu (CTT)	Pro (CCG)	Gly (GGG)	Thr (ACG)	16
pPRO-PheRS-8	Ala (GCT)	Pro (CCT)	Leu (CTT)	Ser (TCT)	Leu (CTG)	His (CAT)	18
pPRO-PheRS-consenso	Ala	Pro/Gln	Leu	Ala	Ile	Ala/Pro	21

FIG. 3

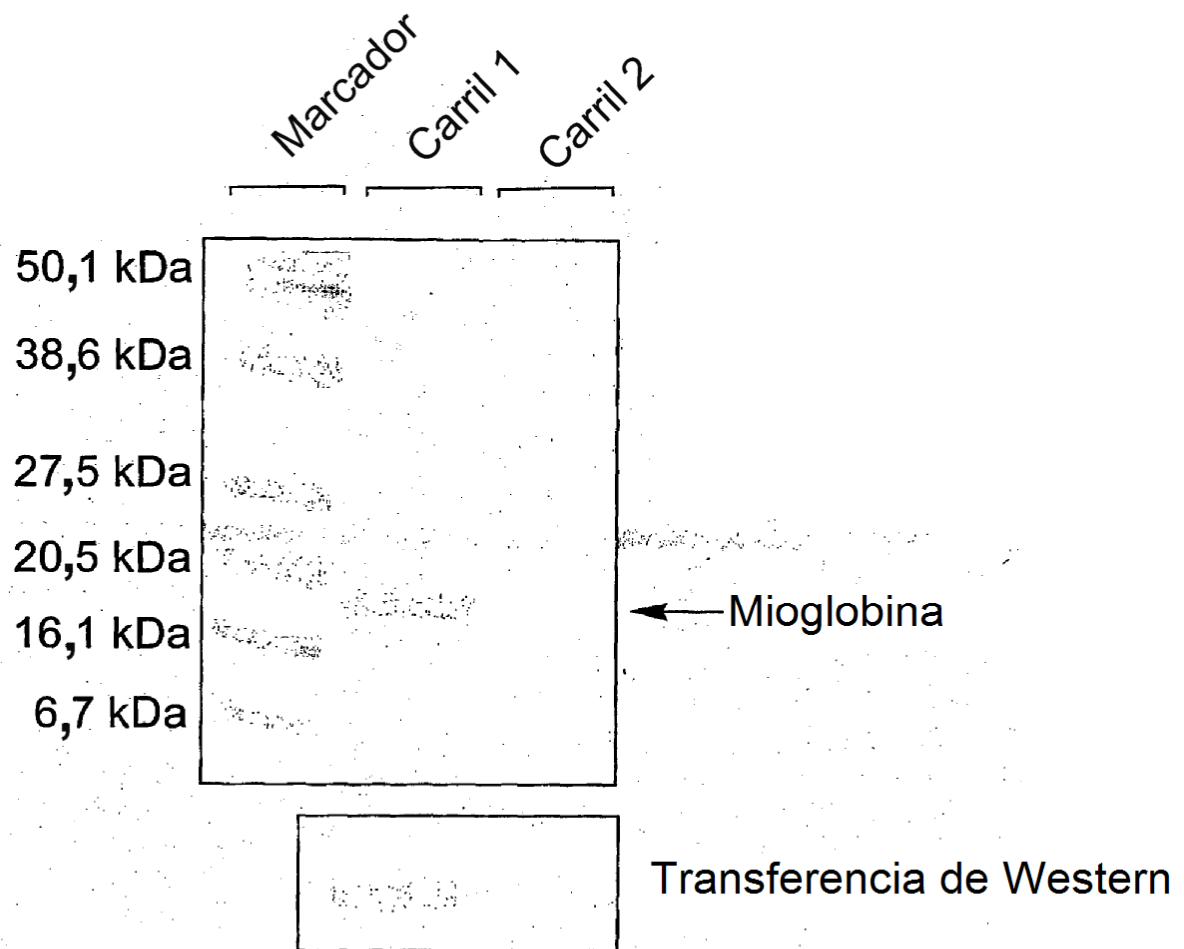
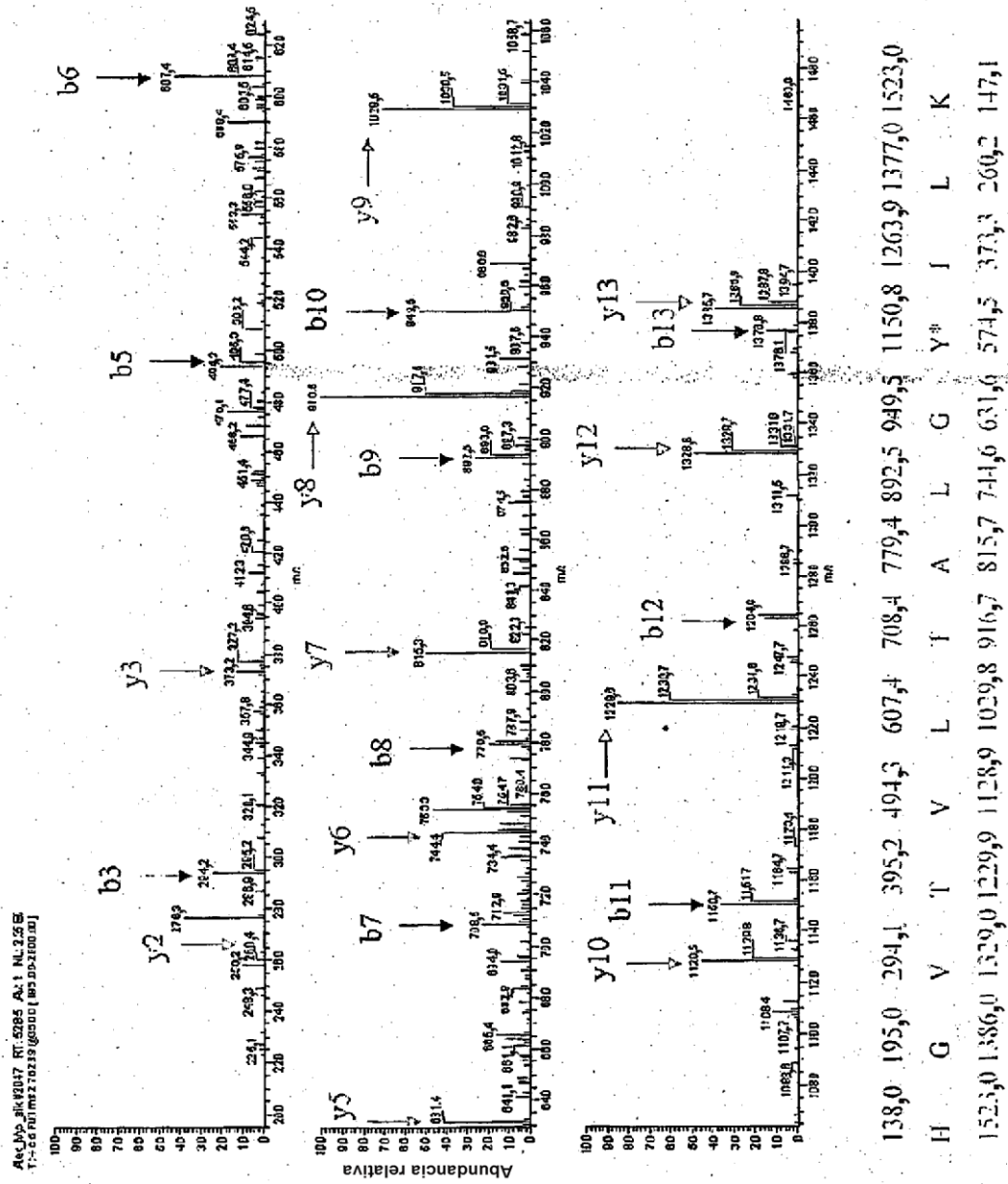
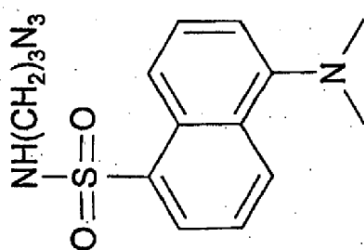


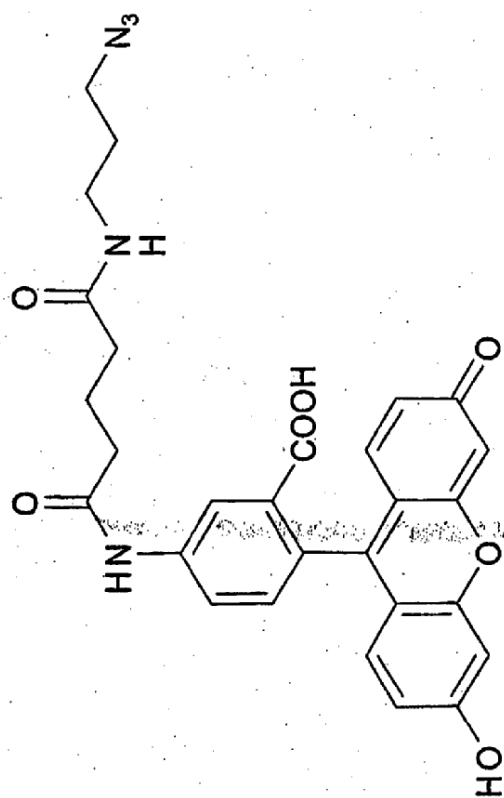
FIG. 4





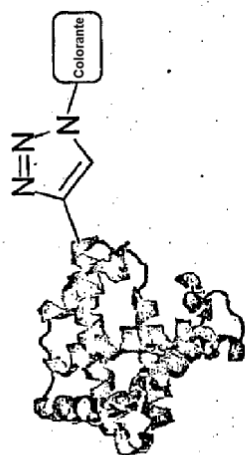
2

FIG. 6A



3

FIG. 6B



$CuSO_4$, alambre de Cu
 Tampón PB, pH 8,
 4 °C o temperatura ambiente

+ 2 or 3



Myoglobina (4TAG)

Colorante 2 (360 nm):
 Colorante 3 (450 nm):



FIG. 7A

FIG. 7B

Alquino aminoácidos potenciales
derivados de aminoácidos naturales:

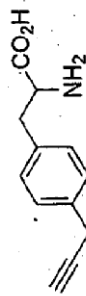
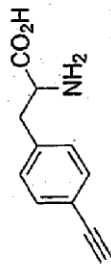
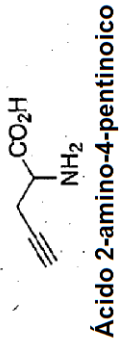


Fig. 8A

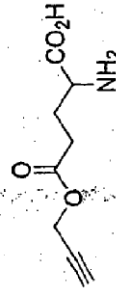
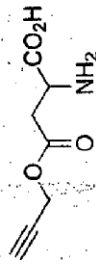
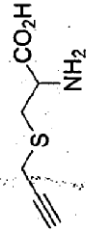
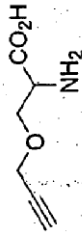


Fig. 8B