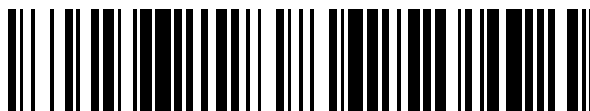


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 773 664**

51 Int. Cl.:

G01N 33/53 (2006.01)
A61K 38/17 (2006.01)
C12N 5/00 (2006.01)
C12N 9/00 (2006.01)
C12N 15/85 (2006.01)
C12N 15/67 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **16.04.2004 E 14171211 (7)**

97 Fecha y número de publicación de la concesión europea: **13.11.2019 EP 2793027**

54 Título: **Adiciones al código genético de aminoácidos reactivos artificiales**

30 Prioridad:

18.06.2003 US 479931 P
05.08.2003 US 493014 P
19.08.2003 US 496548 P

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
14.07.2020

73 Titular/es:

THE SCRIPPS RESEARCH INSTITUTE (100.0%)
10550 North Torrey Pines Road
La Jolla, CA 92037, US

72 Inventor/es:

DEITERS, ALEXANDER;
CROPP, ASHTON. T.;
CHIN, JASON W;
ANDERSON, CHRISTOPHER J. y
SCHULTZ, PETER, G.

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 773 664 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

Adiciones al código genético de aminoácidos reactivos artificiales

5 **Campo de la invención**

La invención pertenece al campo de la bioquímica de traducción en células eucariotas. La presente divulgación se refiere a métodos de producción y composiciones de ARNt ortogonales, sintetasas ortogonales y pares de los mismos, en células eucariotas. La invención se refiere a proteínas y métodos para producir proteínas en células eucariotas que incluyen aminoácidos artificiales.

Antecedentes de la invención

El código genético de todo organismo conocido, desde bacterias a seres humanos, codifica los mismos veinte aminoácidos comunes. Diferentes combinaciones de los mismos veinte aminoácidos naturales forman proteínas que contienen prácticamente todos los procesos complejos de la vida, desde la fotosíntesis hasta la transducción de señales, y la respuesta inmunológica. A fin de estudiar y modificar la estructura y función de las proteínas, los científicos han intentado manipular tanto el código genético como la secuencia de aminoácidos de las proteínas. Sin embargo, ha sido difícil eliminar las restricciones impuestas por el código genético que limitan las proteínas a veinte componentes básicos estándar genéticamente codificados (con la rara excepción de la selenocisteína (véase, por ejemplo, A. Bock et al., (1991), *Molecular Microbiology* 5:515-20) y la pirrolisina (véase, por ejemplo, G. Srinivasan, et al., (2002), *Science* 296:1459-62).

Se ha hecho algún progreso para eliminar estas restricciones, aunque este progreso ha sido limitado y la capacidad para controlar racionalmente la estructura y función de las proteínas se encuentra aún en su fase temprana. Por ejemplo, los químicos han desarrollado métodos y estrategias para sintetizar y manipular la estructura de moléculas pequeñas (véase, por ejemplo, E. J. Corey, y X.-M. Cheng, *The Logic of Chemical Synthesis* (Wiley-Interscience, Nueva York, 1995)). La síntesis total (véase, por ejemplo, B. Merrifield, (1986), *Science* 232:341-7 (1986)), y las metodologías semisintéticas (véanse, por ejemplo, D. Y. Jackson et al., (1994) *Science* 266:243-7; y, P. E. Dawson, y S. B. Kent, (2000), *Annual Review of Biochemistry* 69:923-60), han hecho posible sintetizar péptidos y proteínas pequeñas, pero estas metodologías tienen una utilidad limitada con proteínas por encima de 10 kilo Daltons (kDa). Los métodos de mutagénesis, aunque potentes, se encuentran restringidos a un número limitado de cambios estructurales. En varios casos, ha sido posible incorporar competitivamente análogos estructurales cercanos de aminoácidos comunes a lo largo de las proteínas. Véanse, por ejemplo, R. Furter, (1998), *Protein Science* 7:419-26; K. Kirshenbaum, et al., (2002), *ChemBioChem* 3:235-7; y, V. Doring et al., (2001), *Science* 292:501-4.

En un intento por extender la capacidad para manipular la estructura y función de las proteínas, se desarrollaron métodos *in vitro* utilizando ARNt ortogonales químicamente acilados que permitieron incorporar selectivamente aminoácidos artificiales en respuesta a un codón sin sentido *in vitro* (véase, por ejemplo, J. A. Ellman, et al., (1992), *Science* 255:197-200). Los aminoácidos con nuevas estructuras y propiedades físicas se incorporaron selectivamente en proteínas para estudiar el plegamiento y estabilidad de las proteínas y el reconocimiento y la catálisis biomolecular. Véanse, por ejemplo, D. Mendel, et al., (1995), *Annual Review of Biophysics and Biomolecular Structure* 24:435-462; y, V. W. Cornish, et al. (31 de marzo de 1995), *Angewandte Chemie-International Edition in English* 34:621-633. Sin embargo, la naturaleza estequiométrica de este proceso limitó gravemente la cantidad de proteína que podía generarse.

Se han microinyectado en células aminoácidos artificiales. Por ejemplo, se introdujeron aminoácidos artificiales en el receptor nicotínico de acetilcolina en ovocitos *Xenopus* (por ejemplo, M.W. Nowak, et al. (1998), *In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system*, *Method Enzymol.* 293:504-529) mediante microinyección de un ARNt de *Tetrahymena thermophila* químicamente mal acilado (por ejemplo, M.E. Saks, et al. (1996), *An engineered Tetrahymena tRNAGln for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression*, *J. Biol. Chem.* 271:23169-23175), y en el ARNm relevante. Esto ha permitido estudios biofísicos detallados del receptor en ovocitos mediante la introducción de aminoácidos que contienen cadenas laterales con propiedades físicas o químicas únicas. Véase, por ejemplo, D.A. Dougherty (2000), *Unnatural amino acids as probes of protein structure and function*, *Curr. Opin. Chem. Biol.* 4:645-652. Desafortunadamente, esta metodología se limita a proteínas en células que pueden microinyectarse, y debido a que el ARNt relevante se encuentra químicamente acilado *in vitro*, y no puede reacilarse, el rendimiento de la proteína es muy bajo.

Para superar estas limitaciones, se añadieron nuevos componentes a la maquinaria biosintética de las proteínas del procarionte *Escherichia coli* (*E. coli*) (por ejemplo, L. Wang, et al., (2001), *Science* 292:498-500), que permitió la codificación genética de los aminoácidos artificiales *in vivo*. Varios aminoácidos nuevos con nuevas propiedades químicas, físicas o biológicas, incluyendo etiquetas de fotoafinidad y aminoácidos fotoisomerizables, aminoácidos cetos, y aminoácidos glucosilados se han incorporado eficientemente y con alta fidelidad en proteínas en *E. coli* en respuesta al codón ámbar, TAG, utilizando esta metodología. Véase, por ejemplo, J. W. Chin et al., (2002), *Journal of the American Chemical Society* 124:9026-9027; J. W. Chin, y P. G. Schultz, (2002), *ChemBioChem* 11:1135-1137; J. W. Chin, et al., (2002), *PNAS United States of America* 99:11020-11024; y, L. Wang, y P. G. Schultz, (2002),

Chem. Comm., 1-10. Sin embargo, la maquinaria de traducción de procariotas y eucariotas no se encuentra altamente conservada; por lo tanto, los componentes de la maquinaria biosintética añadidos a *E. Coli* no pueden utilizarse frecuentemente para incorporar específicamente en el sitio aminoácidos artificiales en proteínas en células eucariotas. Por ejemplo, el par tirosil-ARNt sintetasa/ARNt de *Methanococcus jannaschii* que se usó en *E.coli* no es ortogonal en células eucariotas. Además, la transcripción de ARNt en eucariotas, pero no en procariotas, se realiza mediante ARN polimerasa III y esto establece restricciones en la secuencia primaria de los genes estructurales del ARNt que pueden transcribirse en células eucariotas. Además, en contraste con las células procarióticas, los ARNt en células eucariotas necesitan exportarse desde el núcleo, donde se encuentran transcritos, al citoplasma, para funcionar en la traducción. Por último, el ribosoma eucariota 80S es distinto al ribosoma procarita 70S. Speers A E et al: J. Am. Chem. Soc., Washington, EE.UU., 125 págs. 4686-4687 (2003) divulga el perfil de proteínas basado en la actividad *in vivo* usando cicloadición [3 + 2] catalizada por cobre (I). Deiters et al., Bioorganic & Medicinal Chemistry Letters, 2005, Vol. 15, n.º 5, págs. 1521-1524, publicado después de la fecha de presentación de la presente solicitud de patente, describen la incorporación *in vivo* de un alquino en proteínas en *Escherichia coli*. Deiters et al., Journal of the American Chemical Society, 2003, Vol. 125, n.º 39, págs. 11782-11783, describe la adición de aminoácidos con reactividad novedosa al código genético de *Escherichia coli*. Wang Qian et al., Journal of the American Chemical Society, 2003, Vol. 125, n.º 11, págs. 3192-3193, describen la bioconjugación mediante cicloadición [3 + 2] azida-alquino catalizada con cobre (I).

Por lo tanto, existe la necesidad de desarrollar componentes mejorados de la maquinaria biosintética para extender el código genético eucariota. Esta invención cumple estas y otras necesidades, como será aparente tras revisar la siguiente divulgación.

Compendio

En un primer aspecto, la invención proporciona una composición que comprende una célula eucariota, cuya célula comprende un anticuerpo o fragmento de anticuerpo, en la que:

- (i) el anticuerpo o fragmento de anticuerpo comprende al menos un aminoácido artificial que comprende un grupo reactivo para la unión de una molécula mediante una reacción de cicloadición [3 + 2] y en la que el al menos un aminoácido artificial se incorpora al anticuerpo o anticuerpo fragmento por una aminoacil ARNt sintetasa ortogonal (O-RS) y un par de ARNt ortogonal *in vivo* en la célula eucariota; o
- (ii) el anticuerpo o fragmento de anticuerpo comprende al menos un aminoácido artificial que comprende un primer grupo reactivo incorporado por una O-RS y un par de ARNt ortogonal y al menos una modificación postraducciona, en la que la al menos una modificación postraducciona comprende la unión de una molécula que comprende un segundo grupo reactivo mediante una reacción de cicloadición [3 + 2] al al menos un aminoácido artificial que comprende un primer grupo reactivo, y en la que:

(a) la O-RS, o una porción de la misma, está codificada por una secuencia polinucleotídica como se expone en una cualquiera de las SEQ ID NO: 20-35;

(b) la O-RS comprende una secuencia de aminoácidos como se expone en una cualquiera de las SEQ ID NO: 48-63; o

(c) la O-RS comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la de la tirosilaminoacil-ARNt sintetasa (TyrRS) de SEQ ID NO: 2 y comprende las siguientes sustituciones aminoacídicas:

valina, isoleucina, leucina, glicina, serina, alanina, o treonina en una posición correspondiente a Tyr37 de SEQ ID NO: 2; y

treonina, serina, arginina, asparagina, o glicina en una posición correspondiente a Asp182 de SEQ ID NO: 2, y opcionalmente comprende además una o más de las siguientes sustituciones aminoacídicas:

aspartato en una posición correspondiente a Asn126 de SEQ ID NO: 2; metionina, alanina, valina, o tirosina en una posición correspondiente a Phe183 de SEQ ID NO: 2; y

serina, metionina, valina, cisteína, treonina, o alanina en una posición correspondiente a Leu186 de SEQ ID NO: 2.

En algunas realizaciones, la molécula se selecciona del grupo que consiste en una molécula orgánica pequeña, un compuesto citotóxico y una proteína o polipéptido.

En algunas realizaciones, el al menos un aminoácido artificial es p-azido-L-fenilalanina, p-benzoil-L-fenilalanina o p-propargiloxifenilalanina.

En algunas realizaciones, el primer grupo reactivo es un resto alquínico o azido y el segundo grupo reactivo es un resto azido o alquínico.

En algunas realizaciones, la al menos una modificación postraducciona se realiza *in vivo* en una célula eucariota.

En un segundo aspecto, la presente invención proporciona un método para producir en una célula eucariota al menos anticuerpo o fragmento de anticuerpo que comprende al menos un aminoácido artificial, comprendiendo el método:

cultivar, en un medio apropiado, una célula eucariota que comprende un ácido nucleico que comprende al menos un codón selector y codifica el anticuerpo o fragmento de anticuerpo; en el que el medio comprende un aminoácido artificial y la célula eucariota 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 aminoácido artificial, en el que la O-RS comprende una secuencia de aminoácidos que corresponde a la SEQ ID NO.: 48-53.

En un tercer aspecto, la presente invención proporciona un método para producir un conjugado que comprende al menos un anticuerpo o fragmento de anticuerpo que comprende al menos un aminoácido artificial, en el que dicho anticuerpo o fragmento de anticuerpo se produce en una célula eucariota, comprendiendo el método:

cultivar, en un medio apropiado, una célula eucariota que comprende un ácido nucleico que comprende al menos un codón selector y codifica el anticuerpo o fragmento de anticuerpo; en el que el medio comprende un aminoácido artificial y la célula eucariota 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 aminoácido artificial;
incorporar en el anticuerpo o fragmento de anticuerpo el aminoácido artificial en la célula eucariota, en el que el aminoácido artificial comprende un primer grupo reactivo; y
poner en contacto el anticuerpo o fragmento de anticuerpo con una molécula que comprende un segundo grupo reactivo; en el que el primer grupo reactivo reacciona con el segundo grupo reactivo para unir la molécula al aminoácido artificial a través de una reacción de cicloadición, y en el que:

(a) la O-RS, o una porción de la misma, está codificada por una secuencia polinucleotídica como se expone en una cualquiera de las SEQ ID NO: 20-35;

(b) la O-RS comprende una secuencia de aminoácidos como se expone en una cualquiera de las SEQ ID NO: 48-63; o

(c) la O-RS comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la de la tirosilaminoacil-ARNt sintetasa (TyrRS) de SEQ ID NO: 2 y comprende las siguientes sustituciones aminoacídicas:

valina, isoleucina, leucina, glicina, serina, alanina, o treonina en una posición correspondiente a Tyr37 de SEQ ID NO: 2; y

treonina, serina, arginina, asparagina, o glicina en una posición correspondiente a Asp182 de SEQ ID NO: 2, y opcionalmente comprende además una o más de las siguientes sustituciones aminoacídicas:

aspartato en una posición correspondiente a Asn126 de SEQ ID NO: 2;
metionina, alanina, valina, o tirosina en una posición correspondiente a Phe183 de SEQ ID NO: 2; y
serina, metionina, valina, cisteína, treonina, o alanina en una posición correspondiente a Leu186 de SEQ ID NO: 2.

En algunas realizaciones la reacción de cicloadición es una reacción de cicloadición [3 + 2].

En algunas realizaciones, la molécula se selecciona del grupo que consiste en una molécula orgánica pequeña, un compuesto citotóxico y una proteína o polipéptido.

En algunas realizaciones, el primer grupo reactivo es un resto alquilino o azido y el segundo grupo reactivo es un resto azido o alquilino.

En algunas realizaciones, el al menos un aminoácido artificial es p-azido-L-fenilalanina, p-benzoil-L-fenilalanina o p-propargiloxifenilalanina.

En algunas realizaciones, la reacción de cicloadición comprende al menos una modificación postraducciona realizada *in vivo* en la célula eucariota.

DEFINICIONES

Antes de describir la presente invención en detalle, debe entenderse que esta invención no se limita a los dispositivos o sistemas biológicos particulares, que, por supuesto, pueden variar. También debe entenderse que la terminología usada en el presente documento es únicamente con el fin de describir realizaciones particulares, y no pretende ser limitante. Como se utilizan en esta memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un/a", "uno" y "el/la" incluyen las referencias plurales salvo que el contenido indique claramente otra cosa. Por lo tanto, por ejemplo, la referencia a "una célula" incluye una combinación de dos o más células; la

referencia a "bacterias" incluye mezclas de bacterias, y similares.

A menos que se defina de otra manera en el presente documento, o más adelante en el resto de la memoria descriptiva, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado entendido normalmente por los expertos habituales en la materia a la que pertenece la presente invención.

Homólogo: Las proteínas y/o secuencias proteínicas son "homólogas" cuando se obtienen, de forma natural o artificial, de una proteína o secuencia proteínica ancestral común. De manera similar, los ácidos nucleicos y/o secuencias de ácido nucleico son homólogos cuando se obtienen, de forma natural o artificial, de un ácido nucleico o secuencia de ácido nucleico ancestral común. Por ejemplo, cualquier ácido nucleico de origen natural puede modificarse mediante cualquier método de mutagénesis disponible para incluir uno o más codones selectores. Al expresarse, este ácido nucleico mutagenizado codifica un polipéptido que comprende uno o más aminoácidos artificiales. El proceso de mutación, por supuesto, puede alterar adicionalmente uno o más codones estándar, cambiando así también uno o más aminoácidos estándar en la proteína mutante resultante. La homología generalmente se deduce de la similitud de secuencia entre dos o más ácidos nucleicos o proteínas (o secuencias de las mismas). El porcentaje preciso de similitud entre secuencias que es útil a la hora de establecer la homología varía con el ácido nucleico y la proteína en cuestión, pero se utiliza rutinariamente una similitud de secuencia de tan solo el 25 % para establecer la homología. Niveles mayores de similitud de secuencia, por ejemplo, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 %, o 99 % o más, también pueden usarse para establecer la homología. Los métodos para establecer los porcentajes de similitud de secuencia (por ejemplo, BLASTP y BLASTN usando parámetros por defecto) se describen en este documento y están disponibles en general.

Ortogonal: Como se usa en el presente documento, el término "ortogonal" se refiere a una molécula (por ejemplo, un ARNt ortogonal (O-ARNt) y/o una aminoacil ARNt sintetasa ortogonal (O-RS)) que funciona con componentes endógenos de una célula con eficiencia reducida en comparación con una molécula correspondiente endógena a la célula o al sistema de traducción, o que no funciona con componentes endógenos de la célula. En el contexto de los ARNt y aminoacil-ARNt sintetetasas, ortogonal se refiere a una incapacidad o eficiencia reducida, por ejemplo, menos del 20 % de eficacia, menos del 10 % de eficacia, menos del 5 % de eficiencia, o menos del 1 % de eficiencia, 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 funcional en la célula. Por ejemplo, un ARNt ortogonal en una célula se aminoacila mediante cualquier RS endógena de la célula con eficiencia reducida o incluso nula, cuando se compara con la aminoacilación de un ARNt endógeno mediante la RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno en una célula de interés con eficiencia reducida o incluso nula, en comparación con la aminoacilación del ARNt endógeno mediante una RS endógena. Puede introducirse una segunda molécula ortogonal en la célula que funciona con la primera molécula ortogonal. Por ejemplo, un par de ARNt/RS ortogonal incluye componentes complementarios introducidos que funcionan conjuntamente en la célula con una eficiencia (por ejemplo, el 50 % de eficiencia, el 60 % de eficiencia, el 70 % de eficiencia, el 75 % de eficiencia, el 80 % de eficiencia, el 90 % de eficiencia, el 95 % de eficiencia, o el 99 % o más de eficiencia) con respecto a la correspondiente al par de ARNt/RS endógeno.

Complementariedad: El término "complementariedad" se refiere a componentes de un par ortogonal, O-ARNt y O-RS que pueden funcionar conjuntamente, por ejemplo, cuando la O-RS aminoacila el O-ARNt.

Preferentemente aminoacila: El término "preferentemente aminoacila" se refiere a una eficiencia, por ejemplo, el 70 % de eficiencia, el 75 % de eficiencia, el 85 % de eficiencia, el 90 % de eficiencia, el 95 % de eficiencia, o el 99 % o más de eficiencia, a la que una O-RS aminoacila un O-ARNt con un aminoácido artificial en comparación con la O-RS que aminoacila un ARNt de origen natural o un material de inicio utilizado para generar el O-ARNt. El aminoácido artificial se incorpora en una cadena polipeptídica creciente con alta fidelidad, por ejemplo, a más del 75 % de eficacia para un codón selector dado, a más de aproximadamente el 80 % de eficacia para un codón selector dado, a más de aproximadamente el 90 % de eficacia para un codón selector dado, a más de aproximadamente el 95 % de eficacia para un codón selector dado, o a más de aproximadamente el 99 % o más de eficacia para un codón selector dado.

Codón selector: La expresión "codón selector" se refiere a codones reconocidos por el ARNt-O en el proceso de traducción y no reconocidos por un ARNt endógeno. El bucle anticodónico de ARNt-O reconoce el codón selector en el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido artificial, en este sitio en el polipéptido. Los codones selectores pueden incluir, por ejemplo, codones sin sentido, tales como, codones de terminación, por ejemplo, codones ámbar, ocre y ópalo; codones de cuatro o más bases; codones raros; codones derivados de pares de bases naturales o artificiales 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 dado, por ejemplo, proporcionando un mecanismo para incorporar un aminoácido en una cadena

polipeptídica en respuesta a un codón selector. Por ejemplo, un ARNt supresor puede leer a través de, por ejemplo, un codón de terminación, un codón de cuatro bases, un codón raro y/o similares.

ARNt reciclable: El término "ARNt reciclable" se refiere a un ARNt que se aminoacila y puede reaminoacilarse repetidamente con un aminoácido (por ejemplo, un aminoácido artificial) para la incorporación del aminoácido (por ejemplo, el aminoácido artificial) en una o más cadenas polipeptídicas durante la traducción.

Sistema de traducción: El término "sistema de traducción" se refiere al conjunto colectivo de componentes que incorporan un aminoácido de origen natural en una cadena polipeptídica creciente (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetasa, ARNm, aminoácidos, y similares. Los componentes de la presente divulgación (por ejemplo, ORS, O-ARNt, aminoácidos artificiales, etc.) pueden añadirse a un sistema de traducción *in vitro* o *in vivo*, por ejemplo, una célula eucariota, por ejemplo, una célula de levadura, una célula de mamífero, una célula vegetal, una célula de alga, una célula de hongo, una célula de insecto, y/o similares.

Aminoácido artificial: Como se usa en el presente documento, el término "aminoácido artificial" se refiere a cualquier aminoácido, aminoácido modificado, y/o análogo de aminoácido que no es de los 20 aminoácidos comunes de origen natural, seleno cisteína o pirrolisina.

Derivado de: Como se usa en el presente documento, el término "derivado de" se refiere a un componente que se aísla de o se prepara utilizando información de una molécula u organismo específico.

RS inactiva: Como se usa en el presente documento, el término "RS inactiva" se refiere a una sintetasa que ha mutado de modo que ya no puede aminoacilar su ARNt análogo natural con un aminoácido.

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

Marcador de cribado o selección negativa: Como se usa en el presente documento, el término "marcador de cribado o selección negativa" se refiere a un marcador que, cuando está presente, por ejemplo, expresado, activado o similares, permite la identificación de una célula que no posee la propiedad deseada (por ejemplo, en comparación con una célula que posee la propiedad deseada).

Indicador: Como se usa en el presente documento, el término "indicador" se refiere a un componente que puede usarse para seleccionar componentes objetivo de un sistema de interés. Por ejemplo, un indicador puede incluir un marcador de cribado fluorescente (por ejemplo, proteína verde fluorescente), un marcador luminiscente (por ejemplo, una proteína de luciferasa de luciérnaga), un marcador de cribado basado en afinidad, o genes marcadores elegibles tales como his3, ura3, leu2, lys2, lacZ, β -gal/lacZ (β -galactosidasa), Adh (alcohol dehidrogenasa) o similares.

Eucariota: Como se usa en el presente documento, el término "eucariota" se refiere a organismos que pertenecen al dominio filogenético Eucarya tal como animales (por ejemplo, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas, (por ejemplo, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidia, protistas, etc.

No eucariota: Como se usa en el presente documento, el término "no eucariota" se refiere a organismos no eucariotas. Por ejemplo, un organismo no eucariota puede pertenecer al dominio filogenético Eubacteria (por ejemplo, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, etc.), o al dominio filogenético Archaea (por ejemplo, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y las especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pirococcus furiosus*, *Pirococcus horikoshii*, *Aeuropyrum pernix*, etc.).

Anticuerpo: El término "anticuerpo", como se usa en el presente documento, incluye, pero sin limitación, un polipéptido sustancialmente codificado por un gen de inmunoglobulina o genes de inmunoglobulina, o fragmentos de los mismos, que se unen específicamente y reconocen un analito (antígeno). Ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos y monocatenarios, y similares. Fragmentos de inmunoglobulinas, incluyendo fragmentos Fab y fragmentos producidos por una biblioteca de expresión, incluyendo presentación en fagos, también se incluyen en el término "anticuerpo" como se usa en el presente documento. Véase, por ejemplo, Paul, Fundamental Immunology, 4ª edición, 1999, Raven Press, Nueva York, para la estructura y terminología del anticuerpo.

Variante conservativa: El término "variante conservativa" se refiere a un componente de traducción, por ejemplo, un O-ARNt de variante conservativa o una O-RS de variante conservativa, que funcionalmente funciona como el componente a partir del cual se basa la variante conservativa, por ejemplo, un O-ARNt o una O-RS, pero tiene

variaciones en la secuencia. Por ejemplo, una O-RS aminoacilará un O-ARNt complementario o un O-ARNt de variante conservativa con un aminoácido artificial, aunque el O-ARNt y el O-ARNt de variante conservativa no tienen la misma secuencia. La variante conservativa puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones, o cinco o más variaciones en secuencia, siempre que la variante conservativa sea complementaria al O-ARNt o a la O-RS correspondiente.

Agente de selección o de cribado: Como se usa en el presente documento, el término "agente de selección o cribado" se refiere a un agente que, cuando está presente, permite una selección/cribado de ciertos componentes de una población. Por ejemplo, un agente de selección o cribado incluye, pero sin limitación, por ejemplo, un nutriente, un antibiótico, una longitud de onda de luz, un anticuerpo, un polinucleótido expresado (por ejemplo, una proteína moduladora de la transcripción), o similares. El agente de selección se puede variar, por ejemplo, por concentración, intensidad, etc.

Sustancia detectable: El término "sustancia detectable", como se usa en el presente documento, se refiere a un agente que, cuando se activa, se altera, se expresa o similares, permite la selección/cribado de ciertos componentes de una población. Por ejemplo, la sustancia detectable puede ser un agente químico, por ejemplo, ácido 5-fluorouracilo (5-FOA), que en ciertas condiciones, por ejemplo, la expresión de un indicador URA3, se vuelve detectable, por ejemplo, un producto tóxico que destruye células que expresan el indicador URA3.

Breve descripción de los dibujos

La **figura 1, Paneles A, B y C** ilustra esquemáticamente un esquema general de selección positiva y negativa para expandir el código genético de una célula eucariota, por ejemplo, *S. cerevisiae*. El **Panel A** ilustra esquemáticamente la transcripción activada de genes indicadores, que se impulsan mediante supresión ámbar de los codones TAG en GAL4. El dominio de enlace de ADN se indica por el recuadro de rayas y los dominios de activación principal y críptica se indican en el recuadro sombreado. El **Panel B** ilustra ejemplos de genes indicadores, por ejemplo, HIS3, LacZ, URA3 en MaV203. El **Panel C** ilustra esquemáticamente los plásmidos que pueden utilizarse en el esquema de selección, por ejemplo, pEcTyrRS/tRNA_{CUA} y pGADGAL4xxTAG.

La **figura 2** ilustra los fenotipos dependientes de EcTyrRS y ARNt_{CUA} de la 1.^a generación de indicadores GAL4 en el medio selectivo. DB-AD es una fusión entre el dominio de unión a ADN de GAL4 y el dominio de activación. DB-TAG tiene un codón TAG que reemplaza un codón tirosina en el enlazador artificial entre DB y AD. A5 es una versión inactiva de EcTyrRS en la que 5 residuos en el sitio activo han mutado en alanina.

La **figura 3, El Panel A y B** ilustran los fenotipos dependientes de EcTyrRS y ARNt_{CUA} de la 2.^a generación de indicadores GAL4 en el medio selectivo. El dominio de enlace de ADN se indica por el recuadro de rayas y los dominios de activación principal y críptica se indican en el recuadro sombreado. El **Panel A** ilustra construcciones, cada una con mutación de aminoácidos individuales en GAL4. El **panel B** ilustra construcciones, cada una con dos mutaciones de aminoácidos en GAL4.

Figura 4 Los Paneles A, B y C ilustran pGADGAL4 (T44TAG, R110TAG) con y sin EcTyrRS y diversos indicadores en MaV203. El **Panel A** muestra los resultados en presencia de X-gal, -Ura, o -Leu, -Trp. El **Panel B** muestra los resultados en presencia de concentraciones variables de 3-AT. El **Panel C** muestra los resultados en el presente de porcentajes variables de 5-FOA.

Figure 5 Los Paneles A y B ilustran la hidrólisis de ONPG con diversos mutantes GAL4, por ejemplo, donde los residuos T44 (A) y R110 (B) son sitios permisivos. El **Panel A** ilustra la medición de hidrólisis de ONPG con diversos tipos de mutaciones en el sitio T44. El **Panel B** ilustra la medición de hidrólisis de ONPG con diversos tipos de mutaciones en el sitio R110. "GAL4" son MaV203 transformadas con pCL1 y estaba fuera de escala ~600 unidades de hidrólisis de ONPG. "Ninguno" son MaV203 transformadas con plásmidos que codifican GAL4 DB y GAL4 AD por separado.

La **figura 6** muestra la selección de clones de EcTyrRS activos. MaV203 que contienen una mezcla 1:10 de pEcTyrRS-ARNt_{CUA}: pA5-ARNt_{CUA} se pusieron en placas a una dilución 10³ en placas (-Leu, -Trp) (izquierda) o placas (-Leu, -Trp, -His + 3-AT 50 mM) (derecha) y se procesaron usando recubrimiento de XGAL.

Figura 7, Paneles A y B. El **Panel A** ilustra una vista estereoscópica del sitio activo de tirosil-ARNt sintetasa de *B. Stearothermophilus* con tirosina unida. Los residuos mutados se muestran y corresponden a los residuos de tirosil-ARNt sintetasa de *E. coli* Tyr³⁷ (residuo TyrRS de *B. Stearothermophilus* Tyr³⁴), Asn¹²⁶ (Asn¹²³), Asp¹⁸² (Asp¹⁷⁶), Phe¹⁸³ (Phe¹⁷⁷), y Leu¹⁸⁶ (Leu¹⁸⁰). El **Panel B** ilustra las fórmulas estructurales de ejemplos de aminoácidos artificiales (de izquierda a derecha) *p*-acetil-L-fenilalanina (1), *p*-benzoil-L-fenilalanina (2), *p*-azido-L-fenilalanina (3), O-metil-L-tirosina (4), y *p*-yodo-L-tirosina (5).

Figura 8, Paneles A, B, C y D. El **Panel A** ilustra vectores y construcciones indicadoras que se pueden usar en la selección/cribado de ARNt ortogonales, aminoacil sintetisas ortogonales o pares de ARNt/RS ortogonales en células eucariotas. El **Panel B** ilustra los fenotipos de levaduras que albergan indicadores sensibles a GAL4

HIS3, URA3 y *lacZ* en respuesta a aminoacil-ARNt sintetasas activas (TyrRS) o inactivas (A5RS) en medios selectivos. El **Panel C** ilustra un ejemplo de un esquema de selección utilizado para seleccionar sintetasas mutantes que codifican aminoácidos adicionales en una célula eucariota, por ejemplo, *S. cerevisiae*, donde UAA es un aminoácido artificial. El Panel D ilustra los fenotipos de levadura aislados de una selección con *p*-acetil-L-fenilalanina.

La **figura 9** ilustra la expresión de proteínas de la superóxido dismutasa humana (hSOD) (33TAG) HIS en *S. cerevisiae* que codifica genéticamente aminoácidos artificiales como se indica en la figura 7, **Panel B**.

Figura 10, Los **Paneles A-H** ilustran el espectro de masas en tándem del péptido tríptico VY*GSIK (SEQ ID NO:87) que contiene los aminoácidos artificiales (representados Y*) como se indica en la **figura 7, Panel B**. El **Panel A** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido artificial *p*-acetil-L-fenilalanina (1). El **Panel B** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido artificial *p*-benzoyl-L-fenilalanina (2). El **Panel C** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido artificial *p*-azido-L-fenilalanina (3). El **Panel D** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido artificial O-metil-L-tirosina (4). El Panel E ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido artificial *p*-yodo-L-tirosina (5). El **Panel F** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido triptófano (W) en la posición Y*. El **Panel G** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido tirosina (Y) en la posición Y*. El **Panel H** ilustra el espectro de masas en tándem del péptido tríptico con el aminoácido leucina (L) en la posición Y*.

La **figura 11** ilustra ejemplos de dos aminoácidos artificiales (1) para-propargiloxifenilalanina y (2) para-azidofenilalanina.

La **figura 12, Paneles A, B y C** ilustran la expresión de SOD en presencia o ausencia de los aminoácidos artificiales 1 y 2 indicados en la **figura 11**. El **Panel A** ilustra el experimento de la tinción Gelcode Blue. El **Panel B** ilustra una transferencia de Western con un anticuerpo anti-SOD. El **Panel C** ilustra una transferencia de Western con anticuerpo anti-6xHis.

La **figura 13, Paneles A, B y C** ilustran el etiquetado de proteínas por cicloadición [3+2]. El **Panel A** ilustra las etiquetas de colorante sintetizadas 3-6. El **Panel B** ilustra la reacción entre la SOD y el colorante. El **Panel C** ilustra el barrido de fluorescencia en gel y la tinción de Gelcode Blue.

La **figura 14** ilustra el crecimiento de células eucariotas, por ejemplo, células de *S. cerevisiae*, transformadas con mutantes de sintetasa en ausencia o presencia de 1 o 2 como se indica en la **figura 11** en medios SD que carecen de uracilo.

La **figura 15, Paneles A y B**, ilustran el espectro de masas en tándem del péptido tríptico VY*GSIK (SEQ ID NO:87) que contiene los aminoácidos artificiales azida (Az) (**Panel A**) o alquino (Al) (**Panel B**) en la posición Y* que se muestran con sus masas iónicas de fragmento esperadas. La flecha indica series observadas de iones b (azul) e y (rojo) para cada péptido.

La **figura 16** ilustra esquemáticamente la incorporación *in vivo* de un aminoácido artificial, por ejemplo, para-propargiloxifenilalanina, en una cadena polipeptídica creciente y la bioconjugación con moléculas orgánicas pequeñas mediante una reacción de cicloadición [3+2] a través de este aminoácido artificial.

La **figura 17, Paneles A, B y C** ilustra la PEGilación de una proteína que comprende un aminoácido artificial utilizando una cicloadición [3+2]. El **Panel A** ilustra la reacción de una propargil amida PEG con una proteína que comprende un aminoácido azido (por ejemplo, N3-SOD) en presencia de Cu(I) y tampón de fosfato (PB). El **Panel B** ilustra la PEGilación de la proteína mediante análisis de gel. El **Panel C** ilustra la síntesis de la propargil amida PEG.

Descripción detallada

La capacidad para modificar genéticamente las estructuras de proteínas directamente en células eucariotas, más allá de las restricciones químicas impuestas por el código genético, proporciona una poderosa herramienta molecular tanto para sondear como para manipular los procesos celulares. La invención proporciona componentes de traducción que expanden el número de aminoácidos codificados genéticamente en células eucariotas. Estos incluyen aminoacil-ARNt sintetasas (por ejemplo, sintetasa ortogonal (O-RS)), de acuerdo con las reivindicaciones.

Normalmente, los O-ARNt de la invención se expresan y se procesan eficientemente, y funcionan en la traducción en una célula eucariota, pero no se aminoacilan significativamente por las aminoacil-ARNt sintetasas del huésped. En respuesta a un codón selector, un O-ARNt de la presente divulgación suministra un aminoácido artificial, que no codifica ninguno de los veinte aminoácidos comunes, a una cadena polipeptídica creciente durante la traducción de ARNm.

Una O-RS de la invención aminoacila preferentemente un O-ARNt de la presente divulgación con un aminoácido artificial en una célula eucariota, pero no aminoacila ninguno de los ARNt del huésped citoplásmico. Además, la especificidad de una aminoacil-ARNt sintetasa de la invención proporciona la aceptación de un aminoácido artificial mientras que excluye cualquier aminoácido endógeno. Los polipéptidos que incluyen secuencias de aminoácidos de las O-RS a modo de ejemplo, o porciones de las mismas, también son una característica de la invención. Además, los polinucleótidos que codifican los componentes de traducción, O-ARNt, O-RS y porciones de los mismos, se divulgan.

La presente divulgación también proporciona métodos para producir los componentes de traducción deseados, por ejemplo, O-RS, y/o un par ortogonal (ARNt ortogonal y aminoacil-ARNt sintetasa ortogonal), que utiliza un aminoácido artificial para su uso en una célula eucariota (y los componentes producidos por dichos métodos). Por ejemplo, un par tirosil-ARNt sintetasa/ARNt_{CUA} de *E. coli* es un par O-ARNt/O-RS de la presente divulgación. Además, la presente divulgación también presenta métodos para seleccionar/cribar componentes de traducción en una célula eucariota, y una vez seleccionados/cribados, utilizar esos componentes en una célula eucariota diferente (una célula eucariota que no se utilizó para la selección/cribado). Por ejemplo, los métodos de selección/cribado para producir los componentes de traducción para células eucariotas que pueden prepararse en levadura, por ejemplo, *Saccharomyces cerevisiae*, y después los componentes seleccionados pueden utilizarse en otra célula eucariota, por ejemplo, otra célula de levadura, una célula de mamífero, una célula de insecto, una célula vegetal, una célula de hongo, etc.

La invención proporciona además métodos para producir una proteína en una célula eucariota, como se define en las reivindicaciones, donde la proteína comprende un aminoácido artificial. La proteína se produce utilizando los componentes de traducción de la invención. La invención proporciona también proteínas (y proteínas producidas mediante los métodos de la invención), que incluyen aminoácidos artificiales. La proteína o polipéptido de interés también puede incluir una modificación de post-traducción, por ejemplo, que se añade mediante una cicloadición [3+2], o una reacción nucleófila-electrófila, que no se produce por una célula eucariota, etc. En determinadas realizaciones, también se divulgan métodos para producir una proteína moduladora de la transcripción con un aminoácido artificial (y proteínas producidas mediante dichos métodos). Las composiciones, que incluyen proteínas, incluyen un aminoácido artificial, tal como se define en las reivindicaciones.

AMINOACIL-ARNt SINTETASAS ORTOGONALES (O-RS)

A fin de incorporar específicamente un aminoácido artificial en una proteína o polipéptido de interés en una célula eucariota, la especificidad del sustrato de la sintetasa se altera de manera que solo se carga en el ARNt el aminoácido artificial deseado, pero ninguno de los veinte aminoácidos comunes. Si la sintetasa ortogonal es promiscua, dará como resultado proteínas mutantes con una mezcla de aminoácidos naturales y sintéticos en la posición objetivo. La invención proporciona composiciones y métodos para producir aminoacil-ARNt sintetasas ortogonales que tienen especificidad de sustrato modificada para un aminoácido artificial específico.

Una célula eucariota que incluye una aminoacil-ARNt sintetasa ortogonal (O-RS) es una característica de la invención. La O-RS aminoacila preferentemente un ARNt ortogonal (O-ARNt) con un aminoácido artificial en la célula eucariota. En determinadas realizaciones, la O-RS utiliza más de un aminoácido artificial, por ejemplo, dos o más, tres o más, etc. Por lo tanto, una O-RS de la invención puede tener la capacidad de aminoacilar preferentemente un O-ARNt con aminoácidos artificiales diferentes. Esto permite un nivel adicional de control seleccionando el aminoácido artificial o combinación de aminoácidos artificiales que se colocan con la célula y/o seleccionando las diferentes cantidades de aminoácidos artificiales que se colocan con la célula para su incorporación.

Una O-RS de la invención tiene opcionalmente una o más propiedades enzimáticas mejoradas o aumentadas para el aminoácido artificial en comparación con un aminoácido natural. Estas propiedades incluyen, por ejemplo, un Km más alto, un Km más bajo, un kcat más alto, un kcat más bajo, un kcat/km más bajo, un kcat/km más alto, etc., para el aminoácido artificial, en comparación con un aminoácido de origen natural, por ejemplo, uno de los 20 aminoácidos comunes conocidos.

Opcionalmente, la O-RS puede proporcionarse a la célula eucariota mediante un polipéptido que incluye una O-RS o mediante un polipéptido que codifica una O-RS o una porción de la misma. Por ejemplo, una O-RS o una porción de la misma, se codifica por una secuencia polinucleotídica como se expone en cualquiera de las SEQ ID NO.: 20-35, o por una secuencia polinucleotídica complementaria de las mismas. En otro ejemplo, una O-RS comprende una secuencia de aminoácidos como se expone en una cualquiera de las SEQ ID NO.: 48-63. Véanse, por ejemplo, las Tablas 5, 6 y 8, y el Ejemplo 6 en el presente documento para las secuencias de moléculas O-RS a modo de ejemplo.

Una O-RS también puede comprender una secuencia de aminoácidos que es, por ejemplo, al menos el 90 %, al menos el 95 %, al menos el 98 %, al menos el 99 %, o incluso al menos el 99,5 %, idéntica a la de una tirosil aminoacil-ARNt sintetasa (TyrRS) de origen natural (por ejemplo, como se expone en la SEQ ID NO.: 2) y comprende dos o más aminoácidos del grupo A-E. El grupo A incluye valina, isoleucina, leucina, glicina, serina, alanina, o treonina en una posición correspondiente a Tyr37 de TyrRS de *E. coli*; el grupo B incluye aspartato en una posición

correspondiente a Asn126 de TyrRS de *E. coli*; el grupo C incluye treonina, serina, arginina, asparagina o glicina en una posición correspondiente a Asp182 de TyrRS de *E. coli*; el grupo D incluye metionina, alanina, valina, o tirosina en una posición correspondiente a Phe183 de TyrRS de *E. coli*; y, el grupo E incluye serina, metionina, valina, cisteína, treonina, o alanina en una posición correspondiente a Leu186 de TyrRS de *E. coli*. Las combinaciones de estos grupos son una característica de la invención como se define en las reivindicaciones. Por ejemplo, en una realización, la O-RS tiene dos o más aminoácidos seleccionados de valina, isoleucina, leucina, o treonina aparece en una posición correspondiente a Tyr37 de TyrRS de *E. coli*; treonina, serina, arginina, o glicina en una posición correspondiente a Asp182 de TyrRS de *E. coli*; metionina, o tirosina en una posición correspondiente a Phe183 de TyrRS de *E. coli*; y, serina, o alanina en una posición correspondiente a Leu186 de TyrRS de *E. coli*. En otra realización, la O-RS incluye dos o más aminoácidos seleccionados de glicina, serina, o alanina en una posición correspondiente a Tyr37 de TyrRS de *E. coli*, aspartato en una posición correspondiente a Asn126 de TyrRS de *E. coli*, asparagina en una posición correspondiente a Asp182 de TyrRS de *E. coli*, alanina, o valina, en una posición correspondiente a Phe183 de TyrRS de *E. coli*, y/o metionina, valina, cisteína, o treonina, en una posición correspondiente a Leu186 de TyrRS de *E. coli*. Véase también, por ejemplo, la Tabla 4, la Table 6 y la Tabla 8, en el presente documento.

Además de la O-RS, una célula eucariota de la invención puede incluir componentes adicionales, por ejemplo, uno o más aminoácidos artificiales. La célula eucariota incluye también un ARNt ortogonal (O-ARNt) (por ejemplo, derivado de un organismo no eucariota, tal como *Escherichia coli*, *Bacillus stearothermophilus*, y/o similares), donde el O-ARNt reconoce un codón selector y preferentemente se aminoacila con el aminoácido artificial por la O-RS. Un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende un codón selector que se reconoce por el O-ARNt, o una combinación de uno o más de éstos, también puede estar presente en la célula.

En un aspecto, el O-ARNt media la incorporación del aminoácido artificial en una proteína con, por ejemplo, al menos el 45 %, al menos el 50 %, al menos el 60 %, al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 %, o el 99 %, o la eficiencia de un ARNt que comprende o se procesa a partir de una secuencia polinucleotídica como se expone en la SEQ ID NO.: 65. En otra realización, el O-ARNt comprende la SEQ ID NO.:65, y la O-RS comprende una secuencia polipeptídica como se expone en cualquiera de las SEQ ID NO.: 48-63. Véanse también, por ejemplo, la Tabla 5 y el Ejemplo 6, en el presente documento, para secuencias de moléculas a modo de ejemplo de O-RS y O-ARNt.

En un ejemplo, una célula eucariota comprende una aminoacil-ARNt sintetasa ortogonal (O-RS), un ARNt ortogonal (O-ARNt), un aminoácido artificial, y un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, cuyo polinucleótido comprende un codón selector reconocido por el O-ARNt. La O-RS aminoacila preferentemente el ARNt ortogonal (O-ARNt) con el aminoácido artificial en la célula eucariota, y la célula produce el polipéptido de interés en ausencia del aminoácido artificial, con un rendimiento que es de, por ejemplo, menos del 30 %, menos del 20 %, menos del 15 %, menos del 10 %, menos del 5 %, menos del 2,5 %, etc., del rendimiento del polipéptido en presencia del aminoácido artificial.

Los métodos para producir una O-RS, que se divulgan incluyen opcionalmente generar un grupo de sintetasas mutantes procedentes de la estructura de una sintetasa de tipo silvestre, y después seleccionar RS mutadas basándose en su especificidad por un aminoácido artificial relativo a los veinte aminoácidos comunes. Para aislar tal sintetasa, los métodos de selección son: (i) sensitivos, dado que la actividad de las sintetasas deseadas procedentes de las rondas iniciales puede ser baja y la población pequeña; (ii) "modulables", dado que es deseable variar la rigurosidad de selección a diferentes rondas de selección; y, (iii) generales, de manera que los métodos pueden utilizarse para diferentes aminoácidos artificiales.

Los métodos para producir una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente un ARNt ortogonal con un aminoácido artificial en una célula eucariota incluyen normalmente aplicar una combinación de una selección positiva seguida por una selección negativa. En la selección positiva, la supresión del codón selector introducido en la posición o posiciones no esenciales de un marcador positivo permite que las células eucariotas sobrevivan bajo la presión de la selección positiva. En presencia de aminoácidos artificiales, las supervivientes codifican, por lo tanto, sintetasas activas que cargan el ARNt ortogonal supresor con un aminoácido artificial. En la selección negativa, la supresión de un codón selector introducido en la posición o posiciones no esenciales de un marcador negativo retira las sintetasas con especificidad de aminoácido natural. Las supervivientes a la selección negativa y positiva codifican las sintetasas que aminoacilan (cargan) el ARNt ortogonal de supresión solamente (o al menos preferentemente) con aminoácidos artificiales.

Por ejemplo, el método incluye: (a) someter a selección positiva, en presencia de un aminoácido artificial, una población de células eucariotas de una primera especie, donde las células eucariotas comprenden cada una: i) un miembro de una biblioteca de aminoacil-ARNt sintetasas (RS), ii) un ARNt ortogonal (O-ARNt), iii) un polinucleótido que codifica un marcador de selección positiva, y iv) un polinucleótido que codifica un marcador de selección negativa; en el que las células que sobreviven a la selección positiva comprenden una RS activa que aminoacila el ARNt ortogonal (O-ARNt) en presencia de un aminoácido artificial; y, (b) someter las células que sobreviven a la selección positiva a selección negativa en ausencia del aminoácido artificial para eliminar las RS activas que

aminoacilan el O-ARNt con un aminoácido natural, proporcionando así la O-RS que aminoacila preferentemente el O-ARNt con el aminoácido artificial.

- El marcador de selección positiva puede ser cualquiera de diversas moléculas. En una realización, el marcador de selección positiva es un producto que proporciona un suplemento nutritivo para el crecimiento y la selección se realiza en un medio carente del suplemento nutritivo. Ejemplos de polinucleótidos que codifican marcadores de selección positiva incluyen, pero sin limitación, por ejemplo, un gen indicador basado en complementar la auxotrofia del aminoácido de una célula, un gen *his3* (por ejemplo, cuando el gen *his3* codifica una imidazolo glicerol fosfato deshidratasa, detectada proporcionando 3-aminotriazolo (3-AT)), un gen *ura3*, un gen *leu2*, un gen *lys2*, un gen *lacZ*, un gen *adh*, etc. Véase, por ejemplo, G.M. Kishore, y D.M. Shah, (1988), Amino acid biosynthesis inhibitors as herbicides, *Annual Review of Biochemistry* 57:627-663. En una realización, la producción de *lacZ* se detecta por la hidrólisis de orto-nitrofenil- β -D-galactopiranosido (ONPG). Véanse, por ejemplo, I.G. Serebriiskii, y E.A. Golemis, (2000), Uses of *lacZ* to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, *Analytical Biochemistry* 285:1-15. Los marcadores de selección positiva adicionales incluyen, por ejemplo, luciferasa, proteína verde fluorescente (GFP), YFP, EGFP, RFP, el producto de un gen resistente a antibióticos (por ejemplo, cloranfenicol acetiltransferasa (CAT)), una proteína moduladora de la transcripción (por ejemplo, GAL4), etc. Opcionalmente, un polinucleótido que codifica el marcador de selección positiva comprende un codón selector.
- Un polinucleótido que codifica el marcador de selección positiva puede encontrarse unido operativamente a un elemento de respuesta. Un polinucleótido adicional que codifica una proteína moduladora de la transcripción que modula la transcripción del elemento de respuesta, y comprende al menos un codón selector, también puede estar presente. La incorporación del aminoácido artificial en la proteína moduladora de la transcripción mediante el O-ARNt aminoacilado con el aminoácido artificial da como resultado la transcripción del polinucleótido (por ejemplo, el gen indicador) que codifica el marcador de selección positiva. Por ejemplo, véase la **figura 1A**. Opcionalmente, el codón selector se localiza en o sustancialmente cercano a una porción del polinucleótido que codifica un dominio de unión a ADN de la proteína moduladora de la transcripción.
- Un polinucleótido que codifica el marcador de selección negativa también puede encontrarse unido operativamente a un elemento de respuesta a partir del cual se media la transcripción mediante la proteína moduladora de la transcripción. Véase, por ejemplo, A. J. DeMaggio, et al., (2000), The yeast split-hybrid system, *Method Enzymol.* 328:128-137; H.M. Shih, et al., (1996), A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP, *Proc. Natl. Acad. Sci. U.S. A.* 93:13896-13901; M. Vidal, et al., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. [comentario], *Proc. Natl. Acad. Sci. U.S. A.* 93:10321-10326; y, M. Vidal, et al., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. [comentario], *Proc. Natl. Acad. Sci. U.S. A.* 93:10315-10320. La incorporación de un aminoácido natural en la proteína moduladora de la transcripción mediante el O-ARNt aminoacilado con un aminoácido natural da como resultado la transcripción del marcador de selección negativa. Opcionalmente, el marcador de selección negativa comprende un codón selector. En una realización, el marcador de selección positiva y/o el marcador de selección negativa de la presente divulgación puede comprender al menos dos codones selectores, que pueden comprender cada uno o ambos al menos dos codones selectores diferentes o al menos dos de los mismos codones selectores.
- La proteína moduladora de la transcripción es una molécula que se une (directa o indirectamente) a una secuencia de ácido nucleico (por ejemplo, un elemento de respuesta) y modula la transcripción de una secuencia que se encuentra unida operativamente a un elemento de respuesta. Una proteína moduladora de la transcripción puede ser una proteína activadora de la transcripción (por ejemplo, GAL4, receptores de hormona nuclear, AP1, CREB, miembros de la familia LEF/tcf, SMAD, VP16, SP1, etc.), una proteína represora de la transcripción (por ejemplo, receptores de hormona nuclear, familia Groucho/tle, familia Engrailed, etc.), o una proteína que puede tener ambas actividades dependiendo del entorno (por ejemplo, LEF/tcf, proteínas homobox, etc.). Un elemento de respuesta es normalmente una secuencia de ácido nucleico reconocida por la proteína moduladora de la transcripción o un agente adicional que actúa en concierto con la proteína moduladora de la transcripción.
- Otro ejemplo de una proteína moduladora de la transcripción es la proteína activadora de la transcripción, GAL4 (véase, por ejemplo, la **figura 1A**). Véanse, por ejemplo, A. Laughon, et al., (1984), Identification of two proteins encoded by the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:268-275; A. Laughon, y R.F. Gesteland, (1984), Primary structure of the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:260-267; L. Keegan, et al., (1986), Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein, *Science* 231:699-704; y, M. Ptashne, (1988), How eukaryotic transcriptional activators work, *Nature* 335:683-689. Los 147 aminoácidos N-terminales de esta proteína de 881 aminoácidos forman un dominio de unión a ADN (DBD) que se une a una secuencia de ADN específicamente. Véanse, por ejemplo, M. Carey, et al., (1989), An amino-terminal fragment of GAL4 binds DNA as a dimer, *J. Mol. Biol.* 209:423-432; y, E. Giniger, et al., (1985), Specific DNA binding of GAL4, a positive regulatory protein of yeast, *Cell* 40:767-774. El DBD está unido, por medio de una secuencia de proteína intermedia, a un dominio de activación de 113 aminoácidos C-terminales (AD) que puede activar la transcripción cuando se une al ADN. Véanse, por ejemplo, J. Ma, y M. Ptashne,

(1987), Deletion analysis of GAL4 defines two transcriptional activating segments, Cell 48:847-853; y, J. Ma, y M. Ptashne, (1987), The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, Cell 50:137-142. Al colocar los codones ámbar hacia, por ejemplo, el DBD N-terminal de un único polipéptido que contiene tanto el DBD N-terminal de GAL4 como su AD C-terminal, la supresión ámbar por medio del par O-ARNt/O-RS puede unirse a la activación de transcripción por medio de GAL4 (**figura 1, Panel A**). Los genes indicadores GAL4 activados pueden utilizarse para realizar las selecciones tanto positivas como negativas con el gen (**figura 1, Panel B**).

El medio utilizado para la selección negativa puede comprender un agente de selección o cribado que se convierte en una sustancia detectable mediante el marcador de selección negativa. En un aspecto de la presente divulgación, la sustancia detectable es una sustancia tóxica. Un polinucleótido que codifica un marcador de selección negativa puede ser, por ejemplo, un gen ura3. Por ejemplo, el indicador URA3 puede colocarse bajo el control de un promotor que contiene sitios de unión a ADN de GAL4. Cuando se produce el marcador de selección negativa, por ejemplo, mediante la traducción de un polinucleótido que codifica GAL4 con codones selectores, GAL4 activa la transcripción de URA3. La selección negativa se logra en un medio que comprende ácido 5-fluoroorótico (5-FOA), que se convierte en una sustancia detectable (por ejemplo, una sustancia tóxica que destruye la célula) por medio del producto génico del gen ura3. Véanse, por ejemplo, J.D. Boeke, et al., (1984), A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoroorotic acid resistance, Molecular & General Genetics 197:345-346; M. Vidal, et al., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system. [comentario], Proc. Natl. Acad. Sci. U.S. A. 93:10321-10326; y, M. Vidal, et al., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions, [comentario], Proc. Natl. Acad. Sci. U.S. A. 93:10315-10320. Véase también, la **figura 8C**.

Como con el marcador de selección positiva, el marcador de selección negativa también puede ser cualquiera de diversas moléculas. En una realización, el marcador de selección positiva y/o el marcador de selección negativa es un polipéptido que hace fluorescente o cataliza una reacción luminiscente en presencia de un reactivo adecuado. Por ejemplo, los marcadores de selección negativa incluyen, pero sin limitación, por ejemplo, luciferasa, proteína verde fluorescente (GFP), YFP, EGFP, RFP, el producto de un gen resistente a antibióticos (por ejemplo, cloranfenicol acetiltransferasa (CAT)), el producto de un gen lacZ, una proteína moduladora de la transcripción, etc. En un aspecto de la presente divulgación, el marcador de selección positiva y/o el marcador de selección negativa se detecta mediante clasificación celular activada por fluorescencia (FACS), o mediante luminiscencia. En otro ejemplo, el marcador de selección positiva y/o el marcador de selección negativa comprende un marcador de cribado basado en la afinidad. El mismo polinucleótido puede codificar tanto el marcador de selección positiva como el marcador de selección negativa.

También pueden utilizarse niveles adicionales de rigurosidad de selección/cribado en los métodos de la presente divulgación. La rigurosidad de selección o cribado se puede variar en una o ambas etapas del método para producir una O-RS. Esto puede incluir, por ejemplo, variar la cantidad de elementos de respuesta en un polinucleótido que codifica el marcador de selección positiva y/o negativa, añadir una cantidad variable de una sintetasa inactiva a una o ambas etapas, variar la cantidad del agente de selección/cribado que se utiliza, etc. También pueden realizarse rondas adicionales de selecciones positiva y/o negativa.

La selección o cribado también puede comprender uno o más cribados o selecciones positivas o negativas que incluyen, por ejemplo, un cambio en la permeabilidad aminoacídica, un cambio en la eficiencia de traducción, un cambio en la fidelidad de traducción, etc. Normalmente, el uno o más cambios se basan en una mutación en uno o más polinucleótidos que comprenden o codifican componentes de un par de ARNt ortogonal/ARNt sintetasa que se utiliza para producir la proteína.

También pueden utilizarse estudios de enriquecimiento del modelo para seleccionar rápidamente una sintetasa activa a partir de un exceso de síntetasas inactivas. Pueden hacerse estudios de selección positiva y/o negativa del modelo. Por ejemplo, las células eucariotas que comprenden aminoacil-ARNt síntetasas potenciales activas se mezclan con un exceso en grados variables de aminoacil-ARNt síntetasas inactivas. Se hace una comparación promedio entre las células cultivadas en un medio no selectivo y se analizan, por ejemplo, por recubrimiento X-GAL, y aquellas cultivadas y capaces de sobrevivir en un medio selectivo, (por ejemplo, en ausencia de histidina y/o uracilo) y se analizan, por ejemplo, por un ensayo X-GAL. Para una selección de un modelo negativo, se mezclan aminoacil-ARNt síntetasas potenciales activas con un exceso en grados variables de aminoacil-ARNt síntetasas inactivas y se realiza la selección con una sustancia de selección negativa, por ejemplo, 5-FOA.

Normalmente, la biblioteca de RS (por ejemplo, una biblioteca de RS mutantes) comprende RS derivadas de al menos una aminoacil-ARNt sintetasa (RS), por ejemplo, de un organismo no eucariota. En una realización, la biblioteca de RS se deriva de una RS inactiva, por ejemplo, donde la RS inactiva se genera mutando una RS activa, por ejemplo, en el sitio activo en la sintetasa, en el sitio de mecanismo de edición en la sintetasa, en diferentes sitios combinando diferentes dominios de síntetasas, o similares. Por ejemplo, los residuos en el sitio activo de la RS mutan en, por ejemplo, residuos de alanina. El polinucleótido que codifica la RS mutada en alanina se utiliza como plantilla para mutagenizar los residuos de alanina en los 20 aminoácidos. La biblioteca de las RS mutantes se selecciona/criba para producir la O-RS. En otra realización, la RS inactiva comprende un bolsillo de unión a aminoácidos, y uno o más aminoácidos que comprenden el bolsillo de unión se sustituyen con uno o más

aminoácidos diferentes. En un ejemplo, los aminoácidos sustituidos se sustituyen con alaninas. Opcionalmente, el polinucleótido que codifica la RS mutada en alanina se utiliza como plantilla para mutagenizar los residuos alanina en los 20 aminoácidos y se criba/selecciona.

- 5 El método para producir una O-RS puede incluir además producir la biblioteca de RS utilizando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, las RS mutantes pueden generarse mediante mutaciones específicas del sitio, mutaciones puntuales aleatorias, recombinación homóloga, transposición de ADN u otros métodos recursivos de mutagénesis, construcción quimérica o cualquier combinación de las mismas. Por ejemplo, una biblioteca de RS mutantes puede producirse a partir de dos o más "sub-bibliotecas" diferentes, por ejemplo, más
10 pequeñas, menos diversas. Una vez que las sintetasas se someten a la estrategia de selección positiva y negativa/cribado, estas sintetasas pueden someterse a continuación a mutagénesis adicional. Por ejemplo, puede aislarse un ácido nucleico que codifica la O-RS; puede generarse un conjunto de polinucleótidos que codifican las O-RS mutadas (por ejemplo, mediante mutagénesis aleatoria, mutagénesis específica del sitio, recombinación o cualquier combinación de las mismas) a partir del ácido nucleico; y, estas etapas individuales, o una combinación de estas etapas, pueden repetirse hasta obtener una O-RS mutada que aminoacila preferentemente el O-ARNt con el aminoácido artificial. En un aspecto de la presente divulgación, las etapas se realizan al menos dos veces.

- Pueden encontrarse detalles adicionales para producir una O-RS en el documento WO 2002/086075 titulado "Methods and compositions for the production of orthogonal tRNA-aminoacyl-tRNA synthetase pairs". Véanse
20 también, Hamano-Takaku et al., (2000) A mutant Escherichia coli Tyrosyl-tRNA Synthetase Utilizes the Unnatural Amino Acid Azatyrine More Efficiently than Tyrosine, Journal of Biological Chemistry, 275(51):40324-40328; Kiga et al. (2002), An engineered Escherichia coli tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cell-free system, PNAS 99(15): 9715-9723; y, Francklyn et al., (2002), Aminoacyl-tRNA synthetases: Versatile players in the changing theater of
25 translation; RNA, 8:1363-1372.

ARNt ORTOGONALES

- Se divulgan células eucariotas que incluyen un ARNt ortogonal (O-ARNt). El ARNt ortogonal media la incorporación
30 de un aminoácido artificial en una proteína que es codificada por un polinucleótido que comprende un codón selector reconocido por el O-ARNt *in vivo*. En determinadas realizaciones, un O-ARNt de la presente divulgación media la incorporación de un aminoácido artificial en una proteína con, por ejemplo, al menos el 40 %, al menos el 45 %, al menos el 50 %, al menos el 60 %, al menos el 75 %, al menos el 80 %, o incluso el 90 % o más, tan eficientemente como un ARNt que comprende o se procesa en una célula de una secuencia polinucleotídica como se expone en la
35 SEQ ID NO.: 65. Véase, la Tabla 5, en el presente documento.

- Un ejemplo de un O-ARNt de la presente divulgación es la SEQ ID NO.: 65. (Véase el Ejemplo 6 y la Tabla 5, en el presente documento). La SEQ ID NO.: 65 es una transcripción de precorte y empalme/procesamiento que se procesa opcionalmente en la célula, por ejemplo, utilizando el corte y empalme endógeno de la célula y la
40 maquinaria de procesamiento, y se modifica para formar un O-ARNt activo. Normalmente, una población de dichas transcripciones de precorte y empalme forma una población de ARNt activos en la célula (los ARNt activos pueden estar en una o más formas activas). La presente divulgación también proporciona variaciones conservativas del O-ARNt y sus productos celulares procesados. Por ejemplo, las variaciones conservativas de O-ARNt incluyen las moléculas que funcionan como el O-ARNt de la SEQ ID NO.: 65 y mantienen la estructura con forma de L del ARNt,
45 por ejemplo, en forma procesada, pero no tienen la misma secuencia (y son diferentes a las moléculas de ARNt de tipo silvestre). Normalmente, un O-ARNt de la presente divulgación es un O-ARNt reciclable, debido a que el O-ARNt puede reaminoacilarse *in vivo* para mediar de nuevo la incorporación del aminoácido artificial en una proteína que está codificada por un polinucleótido en respuesta a un codón selector.

- La transcripción del ARNt en eucariotas, pero no en procariotas, se realiza mediante ARN polimerasa II, lo que establece restricciones en la secuencia primaria de los genes estructurales del ARNt que pueden transcribirse en células eucariotas. Además, en las células eucariotas, los ARNt han de exportarse del núcleo, donde se encuentran transcritos, al citoplasma, para funcionar en la traducción. Los ácidos nucleicos que codifican un O-ARNt de la presente divulgación, o un polinucleótido complementario del mismo, también se divulgan. En un aspecto de la
55 presente divulgación, un ácido nucleico que codifica un O-ARNt de la presente divulgación incluye una secuencia promotora interna, por ejemplo, una caja A (por ejemplo, TRGCNNAGY) y una caja B (por ejemplo, GGTTCGANTCC, SEQ ID NO:88). El O-ARNt de la presente divulgación también puede modificarse de manera post-transcripcional. Por ejemplo, la modificación de post-transcripción de los genes de ARNt en eucariotas incluye la eliminación de las secuencias flanqueantes en 5' y 3' mediante Rnasa P y de una 3'-endonucleasa, respectivamente. La adición de una secuencia 3'-CCA también es una modificación de post-transcripción de un gen de ARNt en eucariotas.
60

- En una realización, se obtiene un O-ARNt sometiendo a selección negativa a una población de células eucariotas de una primera especie, donde las células eucariotas comprenden un miembro de una biblioteca de ARNt. La selección
65 negativa elimina las células que comprenden un miembro de la biblioteca de ARNt que se aminoacila mediante una aminoacil-ARNt sintetasa (RS) endógena a las células eucariotas. Esto proporciona un grupo de ARNt que son

ortogonales a la célula eucariota de la primera especie.

Como alternativa, o en combinación con otros métodos descritos anteriormente para incorporar un aminoácido artificial en un polipéptido, puede utilizarse un sistema de trans-traducción. Este sistema implica una molécula llamada ARNtm presente en *Escherichia coli*. Esta molécula de ARN se relaciona estructuralmente con un alanil ARN^t y se aminoacila por la alanil sintetasa. La diferencia entre el ARNtm y el ARN^t es que el bucle anticodónico se reemplaza por una secuencia grande especial. Esta secuencia permite que el ribosoma reinicie la traducción en las secuencias que se han paralizado utilizando un marco de lectura abierto dentro del ARNtm como plantilla. En la presente divulgación, puede generarse un ARNtm ortogonal que se aminoacila preferentemente con una sintetasa ortogonal y se carga con un aminoácido artificial. Al transcribir un gen mediante el sistema, el ribosoma se paraliza en un sitio específico; el aminoácido artificial se introduce en ese sitio, y la traducción se reinicia utilizando la secuencia codificada dentro del ARNtm ortogonal.

Pueden encontrarse métodos adicionales para producir ARN^t ortogonales recombinantes, por ejemplo, en la solicitud de patente internacional WO 2002/086075, titulada "Methods and compositions for the production of orthogonal tRNA-aminoacyl-tRNA synthetase pairs". Véanse también, Forster et al., (2003) Programming peptidomimetic synthetases by translating genetic codes designed de novo PNAS 100(11):6353-6357; y, Feng et al., (2003), Expanding tRNA recognition of a tRNA synthetase by a single amino acid change, PNAS 100(10): 5676-5681.

PARES DE ARN^t ORTOGONAL Y AMINOACIL-ARN^t SINTETASA ORTOGONAL

Un par ortogonal está compuesto por un O-ARN^t, por ejemplo, un ARN^t supresor, un ARN^t de desplazamiento del marco de lectura, o similares, y una O-RS. El O-ARN^t no se acila mediante sintetasa endógenas y es capaz de mediar la incorporación de un aminoácido artificial en una proteína que es codificada por un polinucleótido que comprende un codón selector que es reconocido por el O-ARN^t *in vivo*. La O-RS reconoce el O-ARN^t y aminoacila preferentemente el O-ARN^t con un aminoácido artificial en una célula eucariota. Se incluyen en la invención métodos para producir pares ortogonales, así como los pares ortogonales producidos por dichos métodos y las composiciones de pares ortogonales para uso en células eucariotas. El desarrollo de múltiples pares de ARN^t ortogonal/sintetasa puede permitir la incorporación simultánea de múltiples aminoácidos artificiales utilizando diferentes codones en una célula eucariota.

Un par de O-ARN^t ortogonal/O-RS en una célula eucariota puede producirse importando un par, por ejemplo, un par supresor sin sentido, de un organismo diferente con ineficiente aminoacilación de la especie cruzada. El O-ARN^t y la O-RS se expresan eficientemente y se procesan en la célula eucariota y el O-ARN^t se exporta eficientemente del núcleo al citoplasma. Por ejemplo, tal par tal es el par de tirosil-ARN^t sintetasa/ARN^t_{CUA} de *E. coli* (véanse, por ejemplo, H. M. Goodman, et al., (1968), Nature 217:1019-24; y, D. G. Barker, et al., (1982), FEBS Letters 150:419-23). La tirosil-ARN^t sintetasa de *E. coli* aminoacila eficientemente su ARN^t_{CUA} análogo de *E. coli*, cuando ambos se expresan en el citoplasma de *S. cerevisiae*, pero no aminoacila los ARN^t de *S. cerevisiae*. Véanse, por ejemplo, H. Edwards, y P. Schimmel, (1990), Molecular & Cellular Biology 10:1633-41; y, H. Edwards, et al., (1991), PNAS United States of America 88:1153-6. Además, el tirosil ARN^t_{CUA} de *E. coli* es un sustrato deficiente para las aminoacil-ARN^t sintetasa de *S. cerevisiae* (véase, por ejemplo, V. Trezeguet, et al., (1991), Molecular & Cellular Biology 11:2744-51), pero funciona eficientemente en la traducción de proteínas en *S. cerevisiae*. Véanse, por ejemplo, H. Edwards, y P. Schimmel, (1990) Molecular & Cellular Biology 10:1633-41; H. Edwards, et al., (1991), PNAS United States of America 88:1153-6; y, V. Trezeguet, et al., (1991), Molecular & Cellular Biology 11:2744-51. Además, la TyrRS de *E. coli* no tiene un mecanismo de edición para la corrección de un aminoácido artificial ligado al ARN^t.

El O-ARN^t y la O-RS pueden ser de origen natural o pueden derivarse mediante mutación de un ARN^t y/o RS de origen natural, que genera bibliotecas de ARN^t y/o bibliotecas de RS de diversos organismos. Véase la sección titulada "Fuentes y huéspedes" en el presente documento. En diversas realizaciones, el O-ARN^t y la O-RS se obtienen de al menos un organismo. En otra realización, el O-ARN^t se obtiene de un ARN^t de origen natural o de origen natural mutado de un primer organismo y la O-RS se obtiene de RS de origen natural o de origen natural mutada de un segundo organismo. En una realización, el primer y segundo organismos no eucariotas son el mismo. Como alternativa, el primer y segundo organismos no eucariotas pueden ser diferentes.

Véanse las secciones tituladas en el presente documento "aminoacil-ARN^t sintetasa ortogonales" y "O-ARN^t" para los métodos para producir O-RS y O-ARN^t. Véase también, la solicitud de patente internacional WO 2002/086075, titulada "Methods and compositions for the production of orthogonal tRNA-aminoacyl-tRNA synthetase pairs".

FIDELIDAD, EFICIENCIA, Y RENDIMIENTO

Fidelidad se refiere a la precisión con la que una molécula deseada, por ejemplo, un aminoácido artificial o un aminoácido, se incorpora en un polipéptido creciente en una posición deseada. Los componentes de traducción de la invención incorporan aminoácidos artificiales, con alta fidelidad, en proteínas en respuesta a un codón selector. Por ejemplo, el uso de los componentes de la invención, la eficiencia de incorporación de un aminoácido artificial deseado en una cadena polipeptídica creciente en una posición deseada (por ejemplo, en respuesta a un codón

selector) es, por ejemplo, mayor del 75 %, mayor del 85 %, mayor del 95 %, o incluso mayor del 99 % o más eficiente en comparación con la incorporación no deseada de un aminoácido natural específico que se incorpora en la cadena polipeptídica creciente en la posición deseada.

- 5 La eficiencia también puede referirse al grado con el cual la O-RS aminoacila el O-ARNt con el aminoácido artificial en comparación con un control relevante. Las O-RS de la invención pueden definirse por su eficiencia. En determinadas realizaciones de la invención, una O-RS se compara con otra O-RS. Por ejemplo, una O-RS de la invención aminoacila un O-ARNt con un aminoácido artificial, por ejemplo, al menos el 40 %, al menos el 50 %, al menos el 60 %, al menos el 75 %, al menos el 80 %, al menos el 90 %, al menos el 95 %, o incluso el 99 % o más,
- 10 tan eficientemente como una O-RS que tiene una secuencia de aminoácidos, por ejemplo, como se expone en la SEQ ID NO.: 86 o 45 (u otra RS específica en la Tabla 5) aminoacila un O-ARNt. En otra realización, una O-RS de la invención aminoacila el O-ARNt con el aminoácido artificial al menos 10 veces, al menos 20 veces, al menos 30 veces, etc., más eficientemente que la O-RS aminoacila el O-ARNt con un aminoácido natural.
- 15 El uso de los componentes de traducción de la invención, el rendimiento del polipéptido de interés que comprende el aminoácido artificial es, por ejemplo, al menos el 5 %, al menos el 10 %, al menos el 20 %, al menos el 30 %, al menos el 40 %, el 50 % o más, del obtenido para el polipéptido de origen natural de interés de una célula en la que el polinucleótido carece del codón selector. En otro aspecto, la célula produce el polipéptido de interés en ausencia del aminoácido artificial, con un rendimiento que es de, por ejemplo, menos del 30 %, menos del 20 %, menos del
- 20 15 %, menos del 10 %, menos del 5 %, menos del 2,5 %, etc., del rendimiento del polipéptido en presencia del aminoácido artificial.

ORGANISMOS FUENTE Y HUÉSPED

- 25 Los componentes de traducción ortogonales de la invención se derivan normalmente de organismos no eucariotas para su uso en células eucariotas o sistemas de traducción. Por ejemplo, el O-ARNt ortogonal puede derivarse de un organismo no eucariota, por ejemplo, una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares, o una arqueobacteria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-1,
- 30 *Archaeoglobus fulgidus*, *Pirococcus furiosus*, *Pirococcus horikoshii*, *Aeuryopyrum pernix*, o similares, mientras que una O-RS ortogonal puede derivarse de un organismo no eucariota, por ejemplo, una eubacteria, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares, o una arqueobacteria, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pirococcus furiosus*, *Pirococcus horikoshii*, *Aeuryopyrum*
- 35 *pernix*, o similares. Como alternativa, también pueden utilizarse fuentes eucariotas, por ejemplo, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos, artrópodos, etc.), o similares, por ejemplo, donde los componentes son ortogonales a una célula o sistema de traducción de interés, o donde se modifican (por ejemplo, mutan) para ser ortogonales a la célula o el sistema de traducción.
- 40 Los componentes individuales de un par de O-ARNt/O-RS pueden derivarse del mismo organismo o de diferentes organismos. En una realización, el par de O-ARNt/O-RS es del mismo organismo. Por ejemplo, el par de O-ARNt/O-RS puede derivarse de un par de tirosil-ARNt sintetasa/ARNt_{CUA} de *E. coli*. Como alternativa, el O-ARNt y la O-RS del par de O-ARNt/O-RS son opcionalmente de diferentes organismos.
- 45 El O-ARNt ortogonal, la O-RS o el par de O-ARNt/O-RS puede seleccionarse o cribarse y/o utilizarse en una célula eucariota para producir un polipéptido con un aminoácido artificial. Una célula eucariota puede provenir de diversas fuentes, por ejemplo, una planta (por ejemplo, una planta compleja tal como monocotiledóneas o dicotiledóneas), un alga, un protista, un hongo, una levadura (por ejemplo, *Saccharomyces cerevisiae*), un animal (por ejemplo, un mamífero, un insecto, un artrópodo, etc.) o similares. Las composiciones de células eucariotas con componentes de
- 50 traducción de la invención también son una característica de la invención.

La presente divulgación también proporciona el cribado eficiente en una especie para su uso opcional en esas especies y/o en una segunda especie (opcionalmente, sin selección/cribado adicional). Por ejemplo, los componentes de O-ARNt/O-RS se seleccionan o se criban en una especie, por ejemplo, una especie fácilmente manipulada (tal como una célula de levadura, etc.) y se introducen en una segunda especie eucariota, por ejemplo, una planta (por ejemplo, una planta compleja tal como monocotiledóneas o dicotiledóneas), un alga, un protista, un hongo, una levadura, un animal (por ejemplo, un mamífero, un insecto, un artrópodo, etc.), o similares, para su uso en la incorporación *in vivo* de un aminoácido artificial en la segunda especie.

- 60 Por ejemplo, puede seleccionarse *Saccharomyces cerevisiae* (*S. cerevisiae*) como la primera especie eucariota, dado que es unicelular, tiene un tiempo rápido de generación, y genética relativamente bien caracterizada. Véase, por ejemplo, D. Burke, et al., (2000) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Además, dado que la maquinaria de traducción de los eucariotas está altamente conservada (véanse, por ejemplo, (1996) *Translational Control*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Y. Kwok, y J.T.
- 65 Wong, (1980), *Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes*, *Canadian Journal of Biochemistry* 58:213-218; y, (2001) *The*

Ribosome. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), los genes aaRS para la incorporación de aminoácidos artificiales descubiertos en *S cerevisiae* pueden introducirse en organismos eucariotas mayores y utilizarse, en sociedad con los ARNt análogos (véanse, por ejemplo, K. Sakamoto, et al., (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* 30:4692-4699; y, C. Kohrer, et al., (2001), Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins, *Proc. Natl. Acad. Sci. U.S. A.* 98:14310-14315) para incorporar aminoácidos artificiales.

En un ejemplo, el método para producir O-ARNt/O-RS en una primera especie como se describe en el presente documento, incluye además introducir un ácido nucleico que codifica el O-ARNt y un ácido nucleico que codifica la O-RS en una célula eucariota de una segunda especie (por ejemplo, un mamífero, un insecto, un hongo, un alga, una planta y similares). En otro ejemplo, un método para producir una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente un ARNt ortogonal con un aminoácido artificial en una célula eucariota incluye: (a) someter a selección positiva, en presencia de un aminoácido artificial, una población de células eucariotas de una primera especie (por ejemplo, levadura y similares). Cada una de las células eucariotas comprende: i) un miembro de una biblioteca de aminoacil-ARNt sintetetasas (RS), ii) un ARNt ortogonal (O-ARNt), iii) un polinucleótido que codifica un marcador de selección positiva, y iv) un polinucleótido que codifica un marcador de selección negativa. Las células que sobreviven a la selección positiva comprenden una RS activa que aminoacila el ARNt ortogonal (O-ARNt) en presencia de un aminoácido artificial. Las células que sobreviven a la selección positiva se someten a selección negativa en ausencia del aminoácido artificial para eliminar las RS activas que aminoacilan el O-ARNt con un aminoácido natural. Esto proporciona una O-RS que aminoacila preferentemente el O-ARNt con el aminoácido artificial. Un ácido nucleico que codifica el O-ARNt y un ácido nucleico que codifica la O-RS (o los componentes de O-ARNt y/o O-RS) se introducen en una célula eucariota de una segunda especie, por ejemplo, un mamífero, un insecto, un hongo, un alga, una planta y/o similares. Normalmente, se obtiene el O-ARNt sometiendo a selección negativa a una población de células eucariotas de una primera especie, donde las células eucariotas comprenden un miembro de una biblioteca de ARNt. La selección negativa elimina las células que comprenden un miembro de la biblioteca de ARNt que se aminoacila mediante una aminoacil-ARNt sintetasa (RS) endógena a las células eucariotas, lo que proporciona un grupo de ARNt que son ortogonales a la célula eucariota de la primera especie y la segunda especie.

CODONES SELECTORES

Los codones selectores de la invención expanden el marco de codones genéticos de la maquinaria biosintética de proteínas. Por ejemplo, un codón selector incluye, por ejemplo, un codón de tres bases único, un codón sin sentido, tal como un codón de terminación, por ejemplo, un codón ámbar (UAG), un codón ópalo (UGA), un codón artificial, al menos un codón de cuatro bases, un codón raro, o similares. Puede introducirse un número de codones selectores en un gen deseado, por ejemplo, uno o más, dos o más, más de tres, etc., una vez que el gen puede incluir múltiples copias de un codón selector dado, o puede incluir múltiples codones selectores diferentes, o cualquier combinación de los mismos.

En una realización, los métodos de la presente invención, como se definen en las reivindicaciones, implican el uso de un codón selector que es un codón de terminación para la incorporación de aminoácidos artificiales *in vivo* en una célula eucariota. Por ejemplo, se produce un O-ARNt que reconoce el codón de terminación, por ejemplo, UAG, y se aminoacila mediante una O-RS con un aminoácido artificial deseado. Este ARNt-O no es reconocido por las aminoacil-ARNt sintetetasas del huésped de origen natural. Puede usarse mutagénesis dirigida convencional para introducir el codón de terminación, por ejemplo, TAG, en el sitio de interés en un polipéptido de interés. Véanse, por ejemplo, Sayers, J.R., et al. (1988), 5',3' Exonuclease in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*, 791-802. Cuando la RS-O, el O-ARNt y el ácido nucleico que codifica el polipéptido de interés se combinan *in vivo*, el aminoácido artificial se incorpora en respuesta al codón UAG para dar un polipéptido que contiene el aminoácido artificial en la posición especificada.

La incorporación de aminoácidos artificiales *in vivo* puede realizarse sin perturbación significativa de la célula huésped eucariota. Por ejemplo, debido a que la eficacia de supresión para el codón UAG depende de la competición entre el ARNt-O, por ejemplo, el ARNt supresor ámbar, y de un factor de liberación eucariota (por ejemplo, eRF) (que se une a un codón de terminación e inicia la liberación del polipéptido creciente del ribosoma), la eficacia de supresión puede modularse, por ejemplo, aumentando el nivel de expresión de ARNt-O, por ejemplo, el ARNt supresor.

Los codones selectores también comprenden codones extendidos, por ejemplo, codones de cuatro bases o más, 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, CCCC, CCCUC, CUAGA, CUACU, UAGGC, y similares. Una característica de la invención incluye el uso de codones extendidos basados en la supresión del desplazamiento del marco de lectura. Los codones de cuatro o más bases pueden insertar, por ejemplo, uno o múltiples aminoácidos artificiales en la misma proteína. Por ejemplo, en presencia de O-ARNt mutados, por ejemplo, un ARNt supresor de desplazamiento del marco de lectura, con bucles anticodónicos, por ejemplo, con al menos 8-10 nt bucles anticodónicos, el codón de cuatro o más bases se

lee como un aminoácido único. En otras realizaciones, los bucles anticodónicos pueden decodificar, por ejemplo, al menos un codón de cuatro bases, al menos un codón de cinco bases, o al menos un codón de seis bases o más. Dado que hay 256 codones de cuatro bases posibles, pueden codificarse múltiples aminoácidos artificiales en la misma célula utilizando un codón de cuatro o más bases. Véanse, Anderson et al., (2002) Exploring the Limits of Codon and Anticodon Size, Chemistry and Biology, 9:237-244; Magliery, (2001) Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli, J. Mol. Biol. 307: 755-769.

Por ejemplo, se han utilizado codones de cuatro bases para incorporar aminoácidos artificiales en proteínas utilizando métodos bioartificiales *in vitro*. Véanse, por ejemplo, Ma et al., (1993) Biochemistry, 32:7939; y Hohsaka et al., (1999) J. Am. Chem. Soc., 121:34. Se utilizaron CGGG y AGGU para incorporar simultáneamente 2-naftilalanina y un derivado NBD de lisina en estreptavidina *in vitro* con dos ARNt supresores de desplazamiento del marco de lectura químicamente acilados. Véase, por ejemplo, Hohsaka et al., (1999) J. Am. Chem. Soc., 121:12194. En un estudio *in vivo*, Moore et al., examinaron la capacidad de los derivados ARNtLeu con anticodones NCUA para suprimir los codones UAGN (N puede ser U, A, G o C), y se encontró que el cuadruplete UAGA puede decodificarse mediante un ARNtLeu con un anticodón UCUA con una eficiencia del 13 al 26 % con poca decodificación en el marco 0 o -1. Véase, Moore et al., (2000) J. Mol. Biol., 298:195. En una realización, pueden utilizarse en la invención los codones extendidos basándose en codones raros o codones sin sentido, que pueden reducir la lectura sin sentido y la supresión del desplazamiento del marco de lectura en otros sitios no deseados.

Para un sistema dado, un codón selector también puede incluir uno de los codones naturales de tres bases, donde el sistema endógeno no utiliza (o rara vez utiliza) el codón de base natural. Por ejemplo, este incluye un sistema que carece de un ARNt que reconozca el codón de tres bases natural y/o un sistema en el que el codón de tres bases sea un codón raro.

Los codones selectores incluyen opcionalmente pares de bases sintéticas. Estos pares de bases artificiales expanden adicionalmente el alfabeto genético existente. Un par base extra aumenta el número de codones triples de 64 a 125. Las propiedades de los pares de tercera base incluyen un emparejamiento de bases estable y selectivo, incorporación enzimática eficiente en el ADN con alta fidelidad mediante una polimerasa, y la eficiente continua extensión inicial después de la síntesis del par de bases artificial naciente. Las descripciones de pares de base artificiales que pueden adaptarse para los métodos y composiciones incluyen, por ejemplo, Hirao, et al., (2002) An unnatural base pair for incorporating amino acid analogues into protein, Nature Biotechnology, 20:177-182. Otras publicaciones relevantes se enumeran a continuación.

Para su uso *in vivo*, el nucleósido artificial es de membrana permeable y se encuentra fosforilado para formar el trifosfato correspondiente. Además, la información genética aumentada es estable y no se destruye por las enzimas celulares. Los esfuerzos previos de Benner y otros tomaron ventaja de los patrones de unión de hidrógeno que son diferentes a los de pares Watson-Crick canónicos, cuyo ejemplo más notable es el par iso-C:iso-G. Véanse, por ejemplo, Switzer et al., (1989) J. Am. Chem. Soc., 111:8322; y Piccirilli et al., (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. Estas bases en general no se emparejan en ningún grado con las bases naturales y no pueden replicarse enzimáticamente. Koll y colaboradores demostraron que las interacciones de empaquetamiento hidrófobo entre las bases pueden reemplazar el enlace de hidrógeno para impulsar la formación de un 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 artificial que satisfaga todos los requisitos anteriores, Schultz, Romesberg y colaboradores han sintetizado y estudiado sistemáticamente una serie de bases hidrófobas no naturales. Se encontró que un auto-par PICS:PICS es más estable que los pares de bases naturales, y puede incorporarse eficientemente en el ADN mediante un fragmento Klenow de ADN polimerasa I de Escherichia coli (KF). Véanse, por ejemplo, McMinn et al., (1999) J. Am. Chem. Soc., 121:11586; y Ogawa et al., (2000) J. Am. Chem. Soc., 122:3274. Un auto-par 3MN:3MN puede sintetizarse mediante KF con eficiencia y selectividad suficientes para la función biológica. Véase, por ejemplo, Ogawa et al., (2000) J. Am. Chem. Soc., 122:8803. Sin embargo, ambas bases actúan como un exterminador de cadena para la repetición adicional. Se ha desarrollado recientemente una ADN polimerasa mutante que puede utilizarse para repetir el auto-par PICS. Además, puede replicar un auto-par 7AI. Véase, por ejemplo, Tae et al., (2001) J. Am. Chem. Soc., 123:7439. También se ha desarrollado un nuevo par de metalobase, Dipic:Py, que forma un par estable tras unirse a Cu(II). Véase, Meggers et al., (2000) J. Am. Chem. Soc., 122:10714. Debido a que los codones extendidos y los codones artificiales son intrínsecamente ortogonales a los codones naturales, los métodos de la presente divulgación pueden tomar ventaja de esta propiedad para generar ARNt ortogonales para ellos.

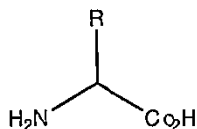
También puede utilizarse un sistema de derivación de traducción para incorporar un aminoácido artificial en un polipéptido deseado. En un sistema de derivación de traducción, se inserta una secuencia grande en un gen pero no se traslada en la proteína. La secuencia contiene una estructura que sirve como clave para inducir que el ribosoma salte sobre la secuencia y reinicie la traducción en la corriente descendente de la inserción.

AMINOÁCIDOS ARTIFICIALES

Como se usa en el presente documento, un aminoácido artificial se refiere a cualquier aminoácido, aminoácido

modificado, o análogo de aminoácido diferente de selenocisteína y/o pirrolisina y los siguientes veinte aminoácidos alfa genéticamente codificados; alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. La estructura genérica de un aminoácido alfa se ilustra mediante la Fórmula I:

I

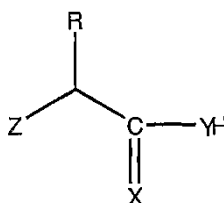


Un aminoácido artificial es normalmente cualquier estructura que tiene la Fórmula I en la que el grupo R es cualquier sustituyente diferente del utilizado en los veinte aminoácidos naturales. Véase, por ejemplo, Biochemistry de L. Stryer, 3.^a ed. 1988, Freeman and Company, Nueva York, para estructuras de los veinte aminoácidos naturales. Cabe destacar que, los aminoácidos artificiales de la invención pueden ser compuestos de origen natural diferentes a los veinte aminoácidos alfa anteriores.

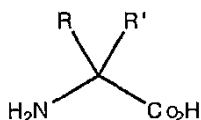
Debido a que los aminoácidos artificiales de la invención normalmente difieren de los aminoácidos naturales en la cadena lateral, los aminoácidos artificiales forman enlaces amida con otros aminoácidos, por ejemplo, naturales o artificiales, de la misma manera en la que se forman en proteínas de origen natural. Sin embargo, los aminoácidos artificiales tienen grupos de cadena lateral que los distinguen de los aminoácidos naturales. Por ejemplo, R en la Fórmula I comprende opcionalmente, alquil-, aril-, acil-, ceto-, azido-, hidroxil-, hidrazina-, ciano-, halo-, hidrazida, alquenilo, alquinilo, éter, tiol, seleno-, sulfonil-, borato, boronato, fosfo, fosfono, fosfina, heterocíclico, enona, imina, aldehído, éster, tioácido, hidroxilamina, amina, y similares, o cualquier combinación de los mismos. Otros aminoácidos artificiales de interés incluyen, pero sin limitación, aminoácidos que comprenden un reticulador fotoactivable, aminoácidos marcados por espín, aminoácidos fluorescentes, aminoácidos de unión a metal, aminoácidos que contienen metal, aminoácidos radiactivos, aminoácidos con nuevos grupos funcionales, aminoácidos que interactúan de manera covalente o no covalente con otras moléculas, aminoácidos fotoatrapado y/o fotoisomerizable, aminoácidos que contienen biotina o análogo de biotina, aminoácidos que contienen ceto, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con un átomo pesado; aminoácidos químicamente escindibles o fotoescindibles, aminoácidos con una cadena lateral alargada en comparación con los aminoácidos naturales (por ejemplo, poliéteres o hidrocarburos de cadena larga, por ejemplo, mayores de aproximadamente 5, mayores de aproximadamente 10 carbonos, etc.), aminoácidos que contienen azúcar unido a carbono, aminoácidos activos redox, aminoácidos que contienen amino tioácido, y aminoácidos que contienen uno o más restos tóxicos. En algunas realizaciones, los aminoácidos artificiales tienen un reticulador fotoactivable que se utiliza, por ejemplo, para unir una proteína a un soporte sólido. En una realización, los aminoácidos artificiales tienen un resto sacárido unido a la cadena lateral de aminoácido (por ejemplo, aminoácidos glucosilados) y/o otra modificación de carbohidratos.

Además de los aminoácidos artificiales que contienen nuevas cadenas laterales, los aminoácidos artificiales también comprenden opcionalmente estructuras de cadena principal modificadas, por ejemplo, como se ilustra mediante las estructuras de Fórmula II y III:

II



III



en la que Z comprende normalmente OH, NH₂, SH, NH-R', o S-R'; X e Y, que pueden ser iguales o diferentes,

comprenden normalmente S u O, y R y R', que son opcionalmente iguales o diferentes, se seleccionan normalmente de la misma lista de constituyentes para el grupo R descrito anteriormente para los aminoácidos artificiales que tienen la Fórmula I, así como hidrógeno. Por ejemplo, los aminoácidos artificiales de la invención comprenden opcionalmente sustituciones en el grupo amino o carboxilo como se ilustra por las Fórmulas II y III. Los aminoácidos artificiales de este tipo incluyen, pero sin limitación, α -hidroxi ácidos, α -tioácido α -aminotiocarboxilatos, por ejemplo, con cadenas laterales correspondientes a los veinte aminoácidos naturales comunes o cadenas laterales sintéticas. Además, las sustituciones en el α -carbono incluyen opcionalmente aminoácidos L, D, o α - α -disustituidos tales como D-glutamato, D-alanina, D-metil-O-tirosina, ácido aminobutírico y similares. Otras alternativas estructurales incluyen aminoácidos cíclicos, tales como análogos de prolina, así como análogos de prolina de anillo de 3, 4, 6, 7, 8, y 9 miembros, aminoácidos β y γ tales como β -alanina sustituida y ácido γ -amino butírico.

Por ejemplo, muchos aminoácidos artificiales se basan en aminoácidos naturales, tales como tirosina, glutamina, fenilalanina y similares. Los análogos de tirosina incluyen tirosinas para-sustituidas, tirosinas orto-sustituidas, y tirosinas meta sustituidas, donde la tirosina sustituida comprende, por ejemplo, un grupo ceto (por ejemplo, un grupo acetilo), un grupo benzoilo, un grupo amino, una hidrazina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo C₆-C₂₀ de cadena lineal o ramificada, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, un grupo alquinilo o similares. Además, también se contemplan anillos de arilo sustituidos de forma múltiple. Los análogos de glutamina de la invención incluyen, pero sin limitación, derivados de α -hidroxi, derivados γ -sustituidos, derivados cíclicos, y derivados de glutamina amida sustituidos. Ejemplos de análogos de fenilalanina incluyen, pero sin limitación, fenilalaninas para-sustituidas, fenilalaninas orto-sustituidas, y fenilalaninas meta sustituidas, donde el sustituyente comprende, por ejemplo, un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído, un azido, un yodo, un bromo, un grupo ceto (por ejemplo, un grupo acetilo), un benzoilo, un grupo alquinilo, o similares. Ejemplos específicos de aminoácidos artificiales incluyen, pero sin limitación, una p-acetil-L-fenilalanina, una p-propargiloxifenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metil-fenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una tri-O-acetil-GlcNAc β -serina, una L-Dopa, una fenilalanina fluorada, una isopropil-L-fenilalanina, una p-azido-L-fenilalanina, una p-acil-L-fenilalanina, una p-benzoil-L-fenilalanina, una L-fosfoserina, una fosfonoserina, una fosfonotirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina, y una isopropil-L-fenilalanina, y similares. Ejemplos de estructuras de aminoácidos artificiales se ilustran en la **figura 7, Panel B y la figura 11**. Estructuras adicionales de diversos aminoácidos artificiales se proporcionan, por ejemplo, en las figuras 16, 17, 18, 19, 26, y 29 del documento WO 2002/085923 titulado "*In vivo* incorporation of unnatural amino acids". Véanse también, la figura 1 estructuras 2-5 de Kiick et al., (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, PNAS 99:19-24, para análogos de metionina adicionales.

En una realización, se proporcionan composiciones que incluyen un aminoácido artificial (tal como p-(propargiloxi)-fenilalanina). También se proporcionan diversas composiciones que comprenden p-(propargiloxi)-fenilalanina y, por ejemplo, proteínas y/o células. En un aspecto, una composición que incluye el aminoácido artificial p-(propargiloxi)-fenilalanina incluye además un ARNt ortogonal. El aminoácido artificial puede unirse (por ejemplo, covalentemente) al ARNt ortogonal, por ejemplo, unirse covalentemente al ARNt ortogonal a través de un enlace amino-acilo, unirse covalentemente a un 3'OH o a un 2'OH de un azúcar de ribosa terminal del ARNt ortogonal, etc.

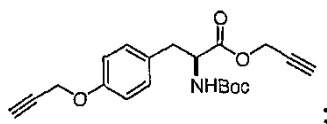
Los restos químicos a través de aminoácidos artificiales que pueden incorporarse en proteínas ofrecen diversas ventajas y manipulaciones de la proteína. Por ejemplo, la reactividad única de un grupo ceto funcional permite la modificación selectiva de proteínas con cualquier número de reactivos que contienen hidrazina o hidroxilamina *in vitro* e *in vivo*. Un aminoácido artificial de átomo pesado, por ejemplo, puede ser útil para sincronizar los datos de estructura de rayos X. La introducción de sitio específico de átomos pesados utilizando aminoácidos artificiales proporciona también selectividad y flexibilidad para seleccionar posiciones para los átomos pesados. Los aminoácidos artificiales fotorreactivos (por ejemplo, aminoácidos con cadenas laterales de benzofenona y arilazidas (por ejemplo, fenilazida), por ejemplo, permiten la fotorreticulación eficiente *in vivo* e *in vitro* de proteínas. Ejemplos de aminoácidos artificiales fotorreactivos incluyen, pero sin limitación, por ejemplo, p-azido-fenilalanina y p-benzoil-fenilalanina. La proteína con los aminoácidos artificiales fotorreactivos puede entonces reticularse a voluntad mediante la excitación del control temporal (y/o espacial) que proporciona el grupo fotorreactivo. En un ejemplo, el grupo metilo de un aminoácido artificial puede sustituirse con uno marcado isotópicamente por ejemplo, un grupo metilo, como sonda de la estructura y dinámica local, por ejemplo, con el uso de resonancia magnética nuclear y espectroscopia de vibración. Los grupos funcionales alquinilo o azido, por ejemplo, permiten la modificación selectiva de proteínas con moléculas a través de una reacción de cicloadición [3+2].

Síntesis química de aminoácidos artificiales

Muchos de los aminoácidos artificiales proporcionados anteriormente se encuentran disponibles comercialmente, por ejemplo, en Sigma (EE.UU.) o Aldrich (Milwaukee, WI, EE.UU.). Aquellos que no se encuentran disponibles comercialmente se sintetizan opcionalmente como se proporciona en el presente documento o como se proporciona en diversas publicaciones o utilizando métodos estándar conocidos por los expertos en la técnica. Para las técnicas de síntesis orgánica, véanse, por ejemplo, Organic Chemistry de Fessenden y Fessenden, (1982, segunda edición, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Tercera Edición, 1985, Wiley and Sons, Nueva York); y Advanced Organic Chemistry de Carey y Sundberg (Tercera Edición, Partes A y B, 1990, Plenum

Press, Nueva York). Publicaciones adicionales que describen la síntesis de aminoácidos artificiales incluyen, por ejemplo, el documento WO 2002/085923 titulado "*In vivo* incorporation of Unnatural Amino Acids"; Matsoukas et al., (1995) J. Med. Chem., 38, 4660-4669; King, F.E. y Kidd, D.A.A. (1949) A New Synthesis of Glutamine and of γ -Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319; Friedman, O.M. y Chatterji, R. (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81,3750-3752; Craig, J.C. et al. (1988) Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-inethylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. y Frappier, F. (1991) Glutamine analogues as Potential Antimalarials. Eur. J. Med. Chem. 26, 201-5; Koskinen, A.M.P. y Rapoport, H. (1989) Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859-1866; Christie, B.D. y Rapoport, H. (1985) Synthesis of Optically Pure Pivcolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization. J. Org. Chem. 1989:1859-1866; Barton et al., (1987) Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron Lett. 43:4297-4308; y, Subasinghe et al., (1992) Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35:4602-7. Véase también, la solicitud de patente titulada "Protein Arrays", número de expediente del mandatario P1001US00 presentado el 22 de diciembre de 2002.

En un aspecto de la presente divulgación se proporciona un método para sintetizar un compuesto de *p*-(propargiloxi)-fenilalanina. Un método comprende, por ejemplo, (a) suspender *N*-*tert*-butoxicarbonil-tirosina y K_2CO_3 en DMF anhidra; (b) añadir bromuro de propargilo a la mezcla de reacción de (a) y alquilar el grupo hidroxilo y carboxilo, dando como resultado un compuesto intermedio protegido que tiene la estructura:



y (c) mezclar el compuesto intermedio protegido con HCl anhidro en MeOH y desproteger el resto amina, sintetizando así el compuesto de *p*-(propargiloxi)-fenilalanina. En una realización, El método comprende además (d) disolver el HCl de *p*-(propargiloxi)-fenilalanina en NaOH acuoso y MeOH y agitarlo a temperatura ambiente; (e) ajustar el pH a un pH de 7; y (f) precipitar el compuesto de *p*-(propargiloxi)-fenilalanina. Véase, por ejemplo, la síntesis de propargiloxifenilalanina en el Ejemplo 4 en el presente documento.

Absorción celular de aminoácidos artificiales

La absorción de aminoácidos artificiales mediante una célula eucariota es una cuestión normalmente considerada al diseñar y seleccionar aminoácidos artificiales, por ejemplo, para su incorporación en una proteína. Por ejemplo, la alta densidad de carga de los α -aminoácidos sugiere que es improbable que estos compuestos sean de permeables en células. Los aminoácidos naturales se absorben en la célula eucariota a través de una colección de sistemas de transporte a base de proteínas. Puede hacerse una exploración rápida que evalúa los aminoácidos artificiales, si hubiera, que se absorben por las células. Véanse, por ejemplo, los ensayos de toxicidad, por ejemplo, en la solicitud titulada "Protein Arrays", número de expediente del mandatario P1001US00 presentado el 22 de diciembre de 2002; y Liu, D.R. y Schultz, P.G. (1999) Progress toward the evolution of an organism with an expanded genetic code. PNAS United States 96:4780-4785. Aunque la absorción se analiza fácilmente con diversos ensayos, una alternativa para diseñar aminoácidos artificiales que son susceptibles a las trayectorias de absorción celular es proporcionar rutas biosintéticas para crear aminoácidos *in vivo*.

Biosíntesis de aminoácidos artificiales

Existen ya muchas rutas biosintéticas en células para la producción de aminoácidos y otros compuestos. Aunque puede no existir en la naturaleza un método bioartificial para un aminoácido artificial particular, por ejemplo, en una célula eucariota, la presente divulgación proporciona dichos métodos. Por ejemplo, las rutas biosintéticas para aminoácidos artificiales se generan opcionalmente en una célula huésped añadiendo nuevas enzimas o modificando las rutas de la célula huésped existentes. Las nuevas enzimas adicionales son opcionalmente enzimas de origen natural o enzimas desarrolladas artificialmente. Por ejemplo, la biosíntesis de *p*-aminofenilalanina (según se presenta en un ejemplo en el documento WO 2002/085923 titulado "*In vivo* incorporation of unnatural amino acids") se basa en la adición de una combinación de enzimas conocidas procedentes de otros organismos. Los genes para estas enzimas pueden introducirse en una célula eucariota transformando la célula con un plásmido que comprende los genes. Los genes, cuando se expresan en la célula, proporcionan una ruta enzimática para sintetizar el compuesto deseado. Se proporcionan ejemplos de los tipos de enzimas que se añaden opcionalmente en los ejemplos a continuación. Las secuencias de enzimas adicionales se encuentran, por ejemplo, en Genbank. Las enzimas desarrolladas artificialmente también se añaden en una célula de la misma manera. De esta manera, la maquinaria celular y los recursos de una célula se manipulan para producir aminoácidos artificiales.

Se encuentran disponibles diversos métodos para producir nuevas enzimas para su uso en rutas biosintéticas o para la evolución de rutas existentes. Por ejemplo, la recombinación recursiva, por ejemplo, según se desarrolló por Maxygen, Inc. (disponible en la red informática mundial en www.maxygen.com), se utiliza opcionalmente para desarrollar nuevas enzimas y rutas. Véanse, por ejemplo, Stemmer (1994), Rapid evolution of a protein *in vitro* by DNA shuffling, *Nature* 370(4):389-391; y, Stemmer, (1994), DNA shuffling by random fragmentation and reassembly: *In vitro* recombination for molecular evolution, *Proc. Natl. Acad. Sci. USA.*, 91:10747-10751. De forma similar, DesignPath™, desarrollado por Genencor (disponible en la red informática mundial en genencor.com) se utiliza opcionalmente para el diseño de rutas metabólicas, por ejemplo, para diseñar una ruta para crear O-metil-L-tirosina en una célula. Esta tecnología reconstruye rutas existentes en organismos huésped utilizando una combinación de nuevos genes, por ejemplo, identificados a través de genómica funcional, y evolución y diseño molecular. Diversa Corporation (disponible en la red informática mundial en diversa.com) también proporciona tecnología para la selección rápida de bibliotecas de genes y rutas génicas, por ejemplo, para crear nuevas rutas.

Normalmente, el aminoácido artificial producido con una ruta biosintética diseñada de la presente divulgación se produce en una concentración suficiente para la biosíntesis de proteínas eficiente, por ejemplo, una cantidad celular natural, pero no a tal grado que afecte a la concentración de los demás aminoácidos o agote los recursos celulares. Las concentraciones típicas producidas *in vivo* de esta manera son de aproximadamente 10 mM a aproximadamente 0,05 mM. Una vez que una célula se transforma con un plásmido que comprende los genes utilizados para producir las enzimas deseadas para una ruta específica, y se genera un aminoácido artificial, se utilizan opcionalmente selecciones *in vivo* para optimizar adicionalmente la producción del aminoácido artificial tanto para la síntesis de proteínas ribosomal como para el crecimiento celular.

POLIPÉPTIDOS CON AMINOÁCIDOS ARTIFICIALES

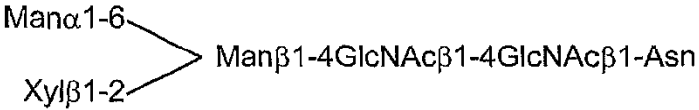
Las proteínas o polipéptidos de interés con al menos un aminoácido artificial y contenidos en una célula eucariota son una característica de la invención. La invención incluye también polipéptidos o proteínas con al menos un aminoácido artificial producido utilizando las composiciones y métodos de la invención, tal como se define en las reivindicaciones. También puede encontrarse presente con la proteína un excipiente (por ejemplo, un excipiente farmacéuticamente aceptable).

Al producir proteínas o polipéptidos de interés con al menos un aminoácido artificial en células eucariotas, las proteínas o polipéptidos incluirán normalmente modificaciones eucariotas de post-traducción. En determinadas realizaciones, una proteína incluye al menos un aminoácido artificial y se hace al menos una modificación de post-traducción *in vivo* mediante una célula eucariota, donde la modificación de post-traducción no se hace mediante una célula procariota. Por ejemplo, la modificación de post-traducción incluye, por ejemplo, acetilación, acilación, modificación de lípidos, palmitoilación, adición de palmitato, fosforilación, modificación de la unión de glucolípidos, glucosilación, y similares. En un aspecto, la modificación de post-traducción incluye la unión de un oligosacárido (por ejemplo, (GlcNAc-Man)₂-Man-GlcNAc-GlcNAc) a una asparagina por una unión de GlcNAc-asparagina. Véanse también, la Tabla 7, que enumera algunos ejemplos de oligosacáridos unidos a N de proteínas eucariotas (también pueden encontrarse presentes residuos adicionales, que no se muestran). En otro aspecto, la modificación de post-traducción incluye la unión de un oligosacárido (por ejemplo, Gal-GalNAc, Gal-GlcNAc, etc.) a una serina o treonina mediante una unión GalNAc-serina o GalNAc-treonina, o una unión GlcNAc-serina o GlcNAc-treonina.

TABLA 7: EJEMPLOS DE OLIGOSACÁRIDOS A TRAVÉS DE LA UNIÓN GlcNAc

Tipo	Estructura de base
Alto contenido en manosa	
Híbrido	
Complejo	

(continuación)

Tipo	Estructura de base
Xilosa	

En otro aspecto más, la modificación de post-traducción incluye el procesamiento proteolítico de precursores (por ejemplo, precursor de calcitonina, precursor del péptido relacionado con el gen de calcitonina, hormona preproparatiroidea, preproinsulina, proinsulina, prepro-opiomelanocortina y similares), ensamblados en una proteína de multisubunidad o ensamblaje macromolecular, traducción a otro sitio en la célula (por ejemplo, a organelos, tal como el retículo endoplásmico, el aparato de golgi, el núcleo, lisosomas, peroxisomas, mitocondrias, cloroplastos, vacuolas, etc., a través de la ruta secretora). En determinadas realizaciones, la proteína comprende una secuencia de secreción o localización, una etiqueta de epítipo, una etiqueta FLAG, una etiqueta de polihistidina, una fusión GST, o similares.

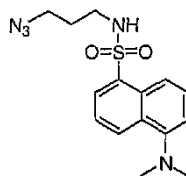
Una ventaja de un aminoácido artificial es que presenta restos químicos adicionales que pueden utilizarse para añadir moléculas adicionales. Estas modificaciones pueden hacerse *in vivo* en una célula eucariota, o *in vitro*. Por lo tanto, en determinadas realizaciones, la modificación de post-traducción es a través del aminoácido artificial. Por ejemplo, la modificación de post-traducción puede ser a través de una reacción nucleófila-electrófila. La mayoría de las reacciones utilizadas actualmente para la modificación selectiva de proteínas implican la formación de un enlace covalente entre compañeros de reacción nucleófila o electrófila, por ejemplo, la reacción de α -halocetonas con las cadenas laterales de histidina o cisteína. La selectividad en estos casos se determina por el número y accesibilidad de los residuos nucleófilos en la proteína. En las proteínas de la invención, pueden utilizarse otras reacciones más selectivas, tales como la reacción de un aminoácido ceto artificial con hidrazidas o compuestos aminooxi, *in vitro* e *in vivo*. Véanse, por ejemplo, Cornish, et al., (1996) Am. Chem. Soc., 118:8150-8151; Mahal, et al., (1997) Science, 276:1125-1128; Wang, et al., (2001) Science 292:498-500; Chin, et al., (2002) Am. Chem. Soc. 124:9026-9027; Chin, et al., (2002) Proc. Natl. Acad. Sci., 99:11020-11024; Wang, et al., (2003) Proc. Natl. Acad. Sci., 100:56-61; Zhang, et al., (2003) Biochemistry. 42:6735-6746; y, Chin, et al., (2003) Science, en prensa. Esto permite el etiquetado selectivo de prácticamente cualquier proteína con un huésped de reactivos, incluyendo fluoróforos, agentes de reticulación, derivados de sacárido, y moléculas citotóxicas. Véase también, la solicitud de patente USSN 10/686.944 titulada "Glycoprotein synthesis" presentada el 15 de octubre de 2003. Las modificaciones de post-traducción, por ejemplo, a través de un aminoácido azido, también pueden realizarse a través de ligadura de Staudinger (por ejemplo, con reactivos de triarilfosfina). Véase, por ejemplo, Kiick et al., (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, PNAS 99:19-24.

Esta invención proporciona otro método altamente eficiente para la modificación selectiva de proteínas, que implica la incorporación genética de aminoácidos artificiales, por ejemplo, que contiene un resto azida o alquinilo (véanse, por ejemplo, **2 y 1 de la figura 11**), en proteínas en respuesta a un codón selector. Estas cadenas laterales de aminoácido pueden modificarse a continuación, por ejemplo, por reacción de cicloadición [3+2] de Huisgen (véase, por ejemplo, Padwa, A. en Comprehensive Organic Synthesis, Vol. 4, (1991) Ed. Trost, B. M., Pergamon, Oxford, pág. 1069-1109; y, Huisgen, R. en 1,3-Dipolar Cycloaddition Chemistry, (1984) Ed. Padwa, A., Wiley, Nueva York, pág. 1-176) con, por ejemplo, derivados de alquinilo o azida, respectivamente. Véase, por ejemplo, la **figura 16**. Debido a que este método implica la cicloadición en lugar de una sustitución nucleófila, las proteínas pueden modificarse con una selectividad extremadamente alta. Esta reacción puede realizarse a temperatura ambiente en condiciones acuosas con excelente regioselectividad (1,4 > 1,5) mediante la adición de cantidades catalíticas de sales de Cu(I) a la mezcla de reacción. Véanse, por ejemplo, Tornøe, et al., (2002) Org. Chem. 67:3057-3064; y, Rostovtsev, et al., (2002) Angew. Chem. Int. Ed. 41:2596-2599. Otro método que puede utilizarse es el intercambio de ligando en un compuesto biarsénico con un motivo de tetracisteína, véase, por ejemplo, Griffin, et al., (1998) Science 281:269-272.

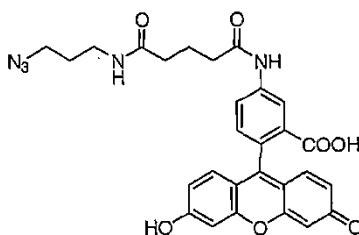
Una molécula que puede añadirse a una proteína de la invención a través de una cicloadición [3+2] incluye prácticamente cualquier molécula con un derivado azido o alquinilo. Véanse, por ejemplo, el Ejemplo 3 y 5, en el presente documento. Dichas moléculas incluyen, pero sin limitación, colorantes, fluoróforos, agentes de reticulación, derivados de sacárido, polímeros (por ejemplo, derivados de polietilenglicol), fotorreticuladores, compuestos citotóxicos, etiquetas de afinidad, derivados de biotina, resinas, perlas, una segunda proteína o polipéptido (o más), uno o más polinucleótidos (por ejemplo, ADN, ARN, etc.), quelatos metálicos, cofactores, ácidos grasos, carbohidratos, y similares. Véanse, por ejemplo, las **figuras 13A**, y el Ejemplo 3 y 5, en el presente documento. Estas moléculas pueden añadirse a un aminoácido artificial con un grupo alquinilo, por ejemplo, p-propargiloxifenilalanina, o grupo azido, por ejemplo, p-azido-fenilalanina, respectivamente. Por ejemplo, véanse la **figura 13B** y la **figura 17A**.

En otro aspecto, la invención proporciona composiciones que incluyen dichas moléculas y métodos para producir estas moléculas, por ejemplo, colorantes azido (tales como los mostrados en la estructura química 4 y la estructura

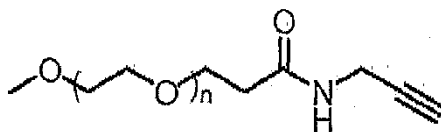
química 6), un alquilil polietilenglicol (por ejemplo, como se muestra en la estructura química 7), donde n es un número entero entre, por ejemplo, 50 y 10.000, 75 y 5.000, 100 y 2.000, 100 y 1.000, etc. En una realización de la invención, el alquilil polietilenglicol tiene un peso molecular de, por ejemplo, aproximadamente 5.000 a aproximadamente 100.000 Da, de aproximadamente 20.000 a aproximadamente 50.000 Da, de aproximadamente 20.000 a aproximadamente 10.000 Da (por ejemplo, 20.000 Da), etc.



4



6



7

También se proporcionan diversas composiciones que comprenden estos compuestos, por ejemplo, con proteínas y células. En un aspecto de la invención, una proteína que comprende un colorante azido (por ejemplo, de estructura química 4 o estructura química 6), incluye además al menos un aminoácido artificial (por ejemplo, un aminoácido de alquilil), donde el colorante azido se une al aminoácido artificial a través de una cicloadición [3+2].

En una realización, una proteína comprende el alquilil polietilenglicol de la estructura química 7. En otra realización, la composición incluye además al menos un aminoácido artificial (por ejemplo, un aminoácido azido), en el que el alquilil polietilenglicol se une a un aminoácido artificial a través de una cicloadición [3+2].

También se proporcionan métodos para sintetizar colorantes azido. Por ejemplo, tal método comprende: (a) proporcionar un compuesto colorante que comprende un resto de haluro de sulfonilo; (b) calentar el compuesto colorante a temperatura ambiente en presencia de 3-azidopropilamina y trietilamina y acoplar un resto amina de la 3-azidopropilamina en la posición haluro del compuesto colorante, sintetizando así el colorante azido. En una realización de ejemplo, el compuesto colorante comprende cloruro de dansilo, y el colorante azido comprende la composición de la estructura química 4. En un aspecto, el método comprende además purificar el colorante azido de la mezcla de reacción. Véase, *por ejemplo*, el Ejemplo 5, en el presente documento.

En otro ejemplo, un método para sintetizar un colorante azido comprende (a) proporcionar un compuesto colorante que contiene amina; (b) combinar el compuesto colorante que contiene amina con una carbodiimida y ácido 4-(3-azidopropilcarbamoil)-bútrico en un disolvente adecuado, y acoplar un grupo carbonilo del ácido al resto amina del compuesto colorante, sintetizando así el colorante azido. En una realización, la carbodiimida comprende clorhidrato de 1-etil-3-(3-dimetilaminopropil)carbodiimida (EDCI). En un aspecto, el colorante que contiene amina comprende fluoresceinamina, y el disolvente adecuado comprende piridina. Por ejemplo, el colorante que contiene amina comprende opcionalmente fluoresceinamina y el colorante azido comprende la composición de la estructura química 6. En una realización, el método comprende además (c) precipitar el colorante azido; (d) lavar el precipitado con HCl; (e) disolver el precipitado lavado en EtOAc; y (f) precipitar el colorante azido en hexanos. Véase, *por ejemplo*, el Ejemplo 5, en el presente documento.

También se proporcionan métodos para sintetizar un propargil amida polietilenglicol. Por ejemplo, el método

comprende hacer reaccionar propargilamina con polietilenglicol (PEG)-éster de hidroxisuccinimida en un disolvente orgánico (por ejemplo, CH_2Cl_2) a temperatura ambiente, dando como resultado el propargil amida polietilenglicol de la estructura química 7. En una realización, el método comprende además precipitar la propargilamida polietilenglicol utilizando acetato de etilo. En un aspecto, el método incluye además recrystalizar el propargilamida polietilenglicol en metanol; y secar el producto al vacío. Véase, *por ejemplo*, el Ejemplo 5, en el presente documento.

Una célula eucariota de la invención proporciona la capacidad de sintetizar proteínas que comprenden aminoácidos artificiales en grandes cantidades útiles. En un aspecto, la composición incluye opcionalmente, por ejemplo, al menos 10 microgramos, al menos 50 microgramos, al menos 75 microgramos, al menos 100 microgramos, al menos 200 microgramos, al menos 250 microgramos, al menos 500 microgramos, al menos 1 miligramo, al menos 10 miligramos o más de la proteína que comprende un aminoácido artificial, o una cantidad que puede lograrse con métodos de producción de proteínas *in vivo* (se proporcionan en el presente documento detalles sobre la producción y purificación de proteínas recombinantes). En otro aspecto, la proteína opcionalmente se encuentra presente en la composición en una concentración de, por ejemplo, al menos 10 microgramos de proteína por litro, al menos 50 microgramos de proteína por litro, al menos 75 microgramos de proteína por litro, al menos 100 microgramos de proteína por litro, al menos 200 microgramos de proteína por litro, al menos 250 microgramos de proteína por litro, al menos 500 microgramos de proteína por litro, al menos 1 miligramo de proteína por litro, o al menos 10 miligramos de proteína por litro o más, en, por ejemplo, un lisato celular, un tampón, un tampón farmacéutico, u otra suspensión líquida (por ejemplo, en un volumen de, por ejemplo, cualquiera de aproximadamente 1 nl a aproximadamente 100 l). La producción de grandes cantidades (por ejemplo, mayores que las normalmente posibles con otros métodos, por ejemplo, traducción *in vitro*) de una proteína en una célula eucariota que incluye al menos un aminoácido artificial es una característica de la invención.

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

En un aspecto de la invención, una composición incluye al menos una proteína con al menos uno, por ejemplo, al menos dos, al menos tres, al menos cuatro, al menos cinco, al menos seis, al menos siete, al menos ocho, al menos nueve, o al menos diez o más aminoácidos artificiales. Los aminoácidos artificiales pueden ser iguales o diferentes, por ejemplo, puede haber 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 o más sitios diferentes en la proteína que comprenden 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más aminoácidos artificiales diferentes. En otro aspecto, una composición incluye una proteína con al menos uno, pero menos que todos, de un aminoácido particular presente en la proteína sustituido con el aminoácido artificial. Para una proteína dada con más de uno aminoácido artificial, los aminoácidos artificiales pueden ser idénticos o diferentes (por ejemplo, la proteína puede incluir dos o más tipos diferentes de aminoácidos artificiales, o puede incluir dos del mismo aminoácido artificial). Para una proteína dada con más de dos aminoácidos artificiales, los aminoácidos artificiales pueden ser iguales, diferentes o una combinación de múltiples aminoácidos artificiales del mismo tipo con al menos un aminoácido artificial diferente.

Esencialmente, cualquier proteína (o una porción de la misma) que incluye un aminoácido artificial (y cualquier ácido nucleico de codificación correspondiente, por ejemplo, que incluye uno o más codones selectores) puede producirse utilizando las composiciones y métodos en el presente documento. No se hace ningún intento por identificar los cientos de miles de proteínas conocidas, cualquiera de las cuales puede modificarse para incluir uno o más aminoácidos artificiales, por ejemplo, adaptando cualquier método de mutación disponible para incluir uno o más codones selectores apropiados en un sistema de traducción relevante. Los depositarios de secuencias comunes para proteínas conocidas incluyen GenBank EMBL, DDBJ y el NCBI. Otros depositarios pueden identificarse fácilmente mediante búsqueda en Internet.

Normalmente, las proteínas son, por ejemplo, al menos un 60 %, al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, o al menos un 99 % o más idénticas a cualquier proteína disponible (por ejemplo, una proteína terapéutica, una proteína de diagnóstico, una enzima industrial, o porción de la misma, y similares), y comprenden uno o más aminoácidos artificiales. Ejemplos de proteínas terapéuticas, de diagnóstico, y otras proteínas que pueden modificarse para comprender uno o más aminoácidos artificiales incluyen, pero sin limitación, por ejemplo, alfa-1 antitripsina, angiostatina, factor antihemolítico, anticuerpos (a continuación se encuentran detalles adicionales de anticuerpos), apolipoproteína, apoproteína, factor natriurético auricular, polipéptido natriurético auricular, péptidos auriculares, quimiocinas C-X-C (por ejemplo, T39765, NAP-2, ENA-78,

Gro-a, Gro-b, Gro-c, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), calcitonina, quimiocinas CC (por ejemplo, proteína 1 quimioatrayente de monocitos, proteína 2 quimioatrayente de monocitos, proteína 3 quimioatrayente de monocitos, proteína 1 alfa inflamatoria de monocitos, proteína 1 beta inflamatoria de monocitos, RANTES, 1309, 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 1 del complemento, citocinas, (por ejemplo, péptido 78 epitelial de activación de neutrófilos, GRO α /MGSA, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1), factor del crecimiento epidérmico (EGF), eritropoyetina ("EPO", que representa un objetivo preferido para la modificación mediante la incorporación de uno o más aminoácidos artificiales), 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, albúmina sérica humana, insulina, factores de crecimiento insulínico (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 de péptido (por ejemplo, hormona del crecimiento humano), pleyotropina, proteína A, proteína G, exotoxinas pirógenas A, B, y C, relaxina, renina, SCF, receptor I de complemento soluble, I-CAM 1 soluble, receptores de interleucina solubles (IL-1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15), receptor TNF soluble, somatomedina, somatostatina, somatotropina, estreptocinasa, superantígenos, es decir, enterotoxinas de estafilococos (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutasa (SOD), toxina del síndrome del choque tóxico (TSST-1), timosina alfa 1, activador del plasminógeno tisular, factor beta de necrosis tumoral (TNF beta), receptor del factor de necrosis tumoral (TNFR), factor alfa de necrosis tumoral (TNF alfa), factor del crecimiento endotelial vascular (VEGEF), urocinasa, y muchas otras.

Una clase de proteínas que puede prepararse utilizando las composiciones y métodos para la incorporación *in vivo* de aminoácidos artificiales descritos en el presente documento incluye moduladores de la transcripción o porciones de los mismos. Moduladores de la transcripción de ejemplo incluyen genes y proteínas de moduladores de la transcripción que modulan el crecimiento celular, la diferenciación, la regulación o similares. Los moduladores de la transcripción se encuentran en procariotas, virus, y eucariotas, incluyendo hongos, plantas, levaduras, insectos, y animales, incluyendo mamíferos, que proporcionan un amplio rango de dianas terapéuticas. Se apreciará que los activadores de la expresión y la transcripción regulan la transcripción mediante muchos mecanismos, por ejemplo, mediante la unión a receptores, estimulación de una cascada de transducción de señal, regulación de la expresión de los factores de transcripción, unión a promotores y potenciadores, unión a proteínas que se unen a promotores y potenciadores, desenrollado del ADN, corte y empalme del pre-ARNm, poliadenilación de ARN, y degradación del ARN. Por ejemplo, las composiciones de proteína GAL4 o porciones de las mismas en una célula eucariota también son una característica de la invención. Normalmente, la proteína GAL4 o porciones de la misma comprende al menos un aminoácido artificial. Véase también la sección en el presente documento titulada "Aminoacil-ARNt sintetasas ortogonales".

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

También se proporcionan por la invención enzimas (por ejemplo, enzimas industriales), o porciones de las mismas con al menos un aminoácido artificial. Ejemplos de enzimas incluyen, pero sin limitación, por ejemplo, amidasas, racemasas de aminoácidos, acilasas, dehalogenasas, dioxigenasas, diarilpropano peroxidasas, epimerasas, epóxido hidrolasas, esterases, isomerasas, quinasas, glucosa isomerasas, glicosidasas, glicosil transferasas, haloperoxidasas, monooxigenasas (por ejemplo, p450s), lipasas, lignina peroxidasas, nitrilo hidratasas, nitrilasas, proteasas, fosfatasas, subtilisinas, transaminasa, y nucleasas.

Muchas de estas proteínas se encuentran disponibles comercialmente (véase, por ejemplo, el catálogo y lista de precios de Sigma BioSciences 2002), y las secuencias y genes de proteínas correspondientes y, normalmente, muchas variaciones de las mismas, son ya conocidas (véase, por ejemplo, Genbank). Cualquiera de ellas puede modificarse mediante la inserción de uno o más aminoácidos artificiales de acuerdo con la invención, por ejemplo, para alterar la proteína con respecto a una o más propiedades terapéuticas, de diagnóstico o enzimáticas de interés. Ejemplos de propiedades terapéuticamente relevantes incluyen, la semivida en suero, la semivida en almacenamiento, la estabilidad, la inmunogenicidad, la actividad terapéutica, la detectabilidad (por ejemplo, mediante la inclusión de grupos indicadores (por ejemplo, etiquetas o sitios de unión a etiqueta) en los aminoácidos artificiales), reducción de la DL₅₀ u otros efectos secundarios, capacidad para entrar al cuerpo a través del tracto gástrico (por ejemplo, disponibilidad oral), o similares. Ejemplos de propiedades de diagnóstico incluyen semivida en almacenamiento, estabilidad, actividad de diagnóstico, detectabilidad, o similares. Ejemplos de propiedades enzimáticas relevantes incluyen la semivida en almacenamiento, estabilidad, actividad enzimática, capacidad de

producción, o similares.

También puede modificarse otras diversas proteínas para incluir uno o más aminoácidos artificiales de la invención. Por ejemplo, la invención puede incluir sustituir uno o más aminoácidos naturales en una o más proteínas de vacunas con un aminoácido artificial, por ejemplo, en proteínas de hongos infecciosos, por ejemplo, *Aspergillus*, especies de *Candida*; bacterias, particularmente *E. coli*, que sirve como modelo de bacterias patógenas, así como bacterias médicamente importantes tales como *Staphylococci* (por ejemplo, *aureus*), o *Streptococci* (por ejemplo, *pneumoniae*); protozoos tales como esporozoos (por ejemplo, *Plasmodia*), *rhizopods* (por ejemplo, *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); virus tales como virus ARN (+) (los ejemplos incluyen virus de viruela, por ejemplo, *vaccinia*; Picornavirus, por ejemplo, *polio*; Togavirus, por ejemplo, *rubella*; Flavivirus, por ejemplo, VHC; y Coronavirus), virus ARN (-) (por ejemplo, Rhabdovirus, por ejemplo, VSV; Paramixovirus, por ejemplo, RSV; Ortomixovirus, por ejemplo, gripe; Buniavirus; y Arenavirus), virus de ADNds (Reovirus, por ejemplo, virus de ARNA a ADN, es decir, Retrovirus, por ejemplo, VIH y HTLV, y ciertos virus de ADN a ARN tales como Hepatitis B.

Las proteínas relacionadas con la agricultura tales como proteínas resistentes a insectos (por ejemplo, las proteínas Cry), enzimas de producción de almidones y lípidos, toxinas vegetales y de insecto, proteínas resistentes a toxinas, proteínas de destoxificación de micotoxinas, enzimas de crecimiento vegetal (por ejemplo, Ribulosa 1,5-bifosfato carboxilasa/oxigenasa, "RUBISCO"), lipoxigenasa (LOX), y fosfoenolpiruvato (PEP) carboxilasa son también objetivos adecuados para la modificación de aminoácidos artificiales.

La invención proporciona también métodos para producir en una célula eucariota al menos una proteína que comprende al menos un aminoácido artificial (y proteínas producidas mediante dichos métodos). Por ejemplo, un método incluye: cultivar, en un medio apropiado, una célula eucariota que comprende un ácido nucleico que comprende al menos un codón selector y codifica la proteína. La célula eucariota también 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 aminoácido artificial, y el medio comprende un aminoácido artificial.

En una realización, el método incluye además incorporar en la proteína el aminoácido artificial, donde el aminoácido artificial comprende un primer grupo reactivo; y poner en contacto la proteína con una molécula (por ejemplo, un colorante, un polímero, por ejemplo, un derivado de polietilenglicol, un fotorreticulador, un compuesto citotóxico, una etiqueta de afinidad, un derivado de biotina, una resina, una segunda proteína o polipéptido, un quelante metálico, un cofactor, un ácido graso, un hidrato de carbono, un polinucleótido (por ejemplo, ADN, ARN, etc.), y similares) que comprende un segundo grupo reactivo. El primer grupo reactivo reacciona con el segundo grupo reactivo para unir la molécula al aminoácido artificial a través de una cicloadición [3+2]. En una realización, el primer grupo reactivo es un resto alquilino o azido y el segundo grupo reactivo es un resto azido o alquilino. Por ejemplo, el primer grupo reactivo es el resto alquilino (por ejemplo, en el aminoácido artificial p-propargiloxifenilalanina) y el segundo grupo reactivo es el resto azido. En otro ejemplo, el primer grupo reactivo es el resto azido (por ejemplo, en el aminoácido artificial p-azido-L-fenilalanina) y el segundo grupo reactivo es el resto alquilino.

En una realización, la O-RS aminoacila el O-ARNt con el aminoácido artificial al menos un 50 % tan eficientemente como la O-RS que tiene una secuencia de aminoácidos, por ejemplo, como se expone en la SEQ ID NO.: 86 o 45. En otra realización, el O-ARNt comprende, se procesa a partir de, o se codifica por las SEQ ID NO.: 65 o 64, o una secuencia polinucleotídica complementaria de las mismas. En otra realización más, la O-RS comprende una secuencia de aminoácidos expuesta en cualquiera de las SEQ ID NO.: 36-63 (por ejemplo, 36-47, 48-63, o cualquier otro subconjunto de 36-63) y/o 86.

La proteína codificada puede comprender, por ejemplo, una proteína terapéutica, una proteína de diagnóstico, una enzima industrial, o una porción de las mismas. Opcionalmente, la proteína que se produce mediante el método se modifica adicionalmente a través del aminoácido artificial. Por ejemplo, la proteína producida mediante el método se modifica opcionalmente mediante al menos una modificación de post-traducción *in vivo*.

También se proporcionan métodos para producir una proteína moduladora de la transcripción de cribado o selección (y proteínas moduladoras de la transcripción de cribado o selección producidas por dichos métodos). Por ejemplo, un método incluye: seleccionar una primera secuencia polinucleotídica, donde la secuencia polinucleotídica codifica un dominio de unión a ácido nucleico; y mutar la primera secuencia polinucleotídica para incluir al menos un codón selector. Esto proporciona una secuencia polinucleotídica de cribado o selección. El método incluye también: seleccionar una segunda secuencia polinucleotídica, donde la segunda secuencia polinucleotídica codifica un dominio de activación de la transcripción; proporcionar una construcción que comprende la secuencia polinucleotídica de cribado o selección unida operativamente a la segunda secuencia polinucleotídica; e, introducir la construcción, un aminoácido artificial, una ARNt sintetasa ortogonal (O-RS) y un ARNt ortogonal (O-ARNt) en una célula. Con estos componentes, la O-RS aminoacila preferentemente el O-ARNt con el aminoácido artificial y el O-ARNt reconoce el codón selector e incorpora el aminoácido artificial en el dominio de unión a ácido nucleico, en respuesta al codón selector en la secuencia de polinucleótido de cribado o selección, proporcionando así la proteína moduladora de la transcripción de cribado o selección.

En determinadas realizaciones, la proteína o polipéptido de interés (o una porción del mismo) en los métodos y/o composiciones de la invención se codifica 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

Los genes que codifican las proteínas o polipéptidos de interés pueden mutagenizarse utilizando métodos ya conocidos por el experto en la técnica y descritos en el presente documento en "Mutagénesis y otras técnicas de biología molecular" para incluir, por ejemplo, uno o más codones selectores para la incorporación de un aminoácido artificial. Por ejemplo, un ácido nucleico para una proteína de interés se mutageniza para incluir uno o más codones selectores, proporcionando la inserción de uno o más aminoácidos artificiales. La invención incluye cualquier variante tal como, por ejemplo, mutante, versiones de cualquier proteína, por ejemplo, incluyendo al menos un aminoácido artificial. De manera similar, la invención incluye también ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifican uno o más aminoácidos artificiales.

En una realización de ejemplo, la invención proporciona composiciones (y composiciones producidas mediante los métodos de la invención) que incluyen un mutante Thr44, Arg110 TAG de GAL4, donde la proteína GAL4 incluye al menos un aminoácido artificial. En otra realización, la invención proporciona composiciones que incluyen un mutante Trp33 TAG de superóxido dismutasa humana (hSOD), donde la proteína hSOD incluye al menos un aminoácido artificial.

Purificación de proteínas recombinantes que comprenden aminoácidos artificiales

Las proteínas de la invención, por ejemplo, proteínas que comprenden aminoácidos artificiales, anticuerpos para proteínas que comprenden aminoácidos artificiales, etc., pueden purificarse, bien parcialmente o bien sustancialmente hasta su homogeneidad, de acuerdo con procedimientos estándar conocidos y utilizados por los expertos en la técnica. Por consiguiente, los polipéptidos de la invención pueden recuperarse y purificarse mediante cualquiera de varios métodos muy conocidos en la técnica, incluyendo, por ejemplo, precipitación con sulfato de amonio o etanol, extracción con ácido o base, cromatografía en columna, cromatografía en columna de afinidad, cromatografía de intercambio aniónico o catiónico, cromatografía de fosfocelulosa, cromatografía de interacción hidrófoba, cromatografía de hidroxilapatita, cromatografía de lectina, electroforesis en gel y similares. Pueden usarse etapas de replegamiento de proteínas, según se desee, para preparar proteínas maduras correctamente plegadas. El análisis por cromatografía líquida de alto rendimiento (HPLC), cromatografía de afinidad u otros métodos adecuados pueden emplearse en etapas de purificación final cuando se desee alta pureza. En una realización, los anticuerpos preparados contra aminoácidos artificiales (o proteínas que comprenden aminoácidos artificiales) se utilizan como reactivos de purificación, por ejemplo, para purificación basada en afinidad de proteínas que comprenden uno o más aminoácidos artificiales. Una vez purificados, parcialmente o hasta homogeneidad, según se desee, los polipéptidos se utilizan opcionalmente, por ejemplo, como componentes de ensayo, reactivos terapéuticos o como inmunógenos para la producción de anticuerpos.

Además de otras referencias apreciadas en el presente documento, son muy conocidos en la técnica diversos métodos de purificación/plegamiento de proteínas, incluyendo, por ejemplo, los expuestos en R. Scopes, Protein Purification, Springer-Verlag, N.Y. (1982); Deutscher, Methods in Enzymology Vol. 182: Guide to Protein Purification, Academic Press, Inc. N.Y. (1990); Sandana (1997) Bioseparation of Proteins, Academic Press, Inc.; Bollag et al. (1996) Protein Methods, 2.^a edición Wiley-Liss, NY; Walker (1996) The Protein Protocols Handbook Humana Press, NJ, Harris y Angal (1990) Protein Purification Applications: A Practical Approach IRL impreso en Oxford, Oxford, Inglaterra; Harris y Angal Protein Purification Methods: A Practical Approach IRL impreso en Oxford, Oxford, Inglaterra; Scopes (1993) Protein Purification: Principles and Practice 3.^a edición Springer Verlag, NY; Janson y Ryden (1998) Protein Purification: Principles, High Resolution Methods and Applications, segunda edición Wiley-VCH, NY; y Walker (1998) Protein Protocols on CD-ROM Humana Press, NJ; y las referencias citadas en los mismos.

Una ventaja de la producción de una proteína o polipéptido de interés con un aminoácido artificial en una célula eucariota es que normalmente las proteínas o polipéptidos se plegarán en sus configuraciones nativas. Sin embargo, en ciertas realizaciones de la invención, los expertos en la técnica reconocerán que, después de la síntesis, expresión y/o purificación, las proteínas pueden poseer una conformación diferente de las conformaciones deseadas de los polipéptidos relevantes. En un aspecto de la invención, la proteína expresada opcionalmente se desnaturaliza y después se renaturaliza. Esto se consigue, por ejemplo, añadiendo una chaperonina a la proteína o polipéptido de interés, y/o solubilizando las proteínas en un agente caotrópico tal como guanidina HCl, etc.

En general, en ocasiones es deseable desnaturalizar y reducir los polipéptidos expresados y después hacer que los polipéptidos se replieguen en su configuración preferida. Por ejemplo, guanidina, urea, DTT, DTE y/o una chaperonina pueden añadirse a un producto de traducción de interés. Los métodos de reducción, desnaturalización y renaturalización de proteínas son bien conocidos por los expertos en la técnica (véanse, las referencias anteriores, y Debinski, et al. (1993) J. Biol. Chem., 268: 14065-14070; Kreitman y Pastan (1993) Bioconjug. Chem., 4: 581-585; y

Buchner, et al., (1992) Anal. Biochem., 205: 263-270). Debinski, et al., por ejemplo, describen la desnaturalización y reducción de proteínas de cuerpos de inclusión en guanidina-DTE. Las proteínas pueden replegarse en un tampón redox que contiene, por ejemplo, glutatión oxidado y L-arginina. Los agentes de replegamiento pueden fluir o moverse de otro modo en contacto con el uno o más polipéptidos u otro producto de expresión, o viceversa.

5

ANTICUERPOS

En un aspecto, la invención proporciona anticuerpos a las moléculas de la invención, por ejemplo, sintetasas, ARNt, y proteínas que comprenden aminoácidos artificiales. Los anticuerpos para las moléculas de la invención son útiles como reactivos de purificación, por ejemplo, para purificar las moléculas de la invención. Además, los anticuerpos pueden utilizarse como reactivos indicadores para indicar la presencia de una sintetasa, un ARNt, o proteína que comprende un aminoácido artificial, por ejemplo, para rastrear la presencia o localización (por ejemplo, *in vivo* o *in situ*) de la molécula.

10

Un anticuerpo de la invención puede ser una proteína que comprende uno o más polipéptidos sustancial o parcialmente codificados por genes de inmunoglobulina o fragmentos de genes de inmunoglobulina. Los genes de inmunoglobulina reconocidos incluyen los genes de región constante kappa, lambda, alfa, gamma, delta, épsilon y mu, así como una miríada de genes de las regiones variables de inmunoglobulinas. Las cadenas ligeras se clasifican en kappa o lambda. Las cadenas pesadas se clasifican en gamma, mu, alfa, delta o épsilon, que a su vez definen las clases de inmunoglobulina, IgG, IgM, IgA, IgD e IgE, respectivamente. Una unidad estructural típica de inmunoglobulina (por ejemplo, un anticuerpo) comprende un tetrámero. Cada tetrámero está compuesto por dos pares idénticos de cadenas polipeptídicas, teniendo cada par una cadena "ligera" (aproximadamente 25 kD) y una cadena "pesada" (aproximadamente 50-70 kD). El extremo N de cada cadena define una región variable de aproximadamente 100 a 110 o más aminoácidos principalmente responsables del reconocimiento de antígenos. Las expresiones de cadena ligera variable (VL) y cadena pesada variable (VH) se refieren a estas cadenas ligera y pesada, respectivamente.

15

20

25

Los anticuerpos existen como inmunoglobulinas intactas, o en forma de numerosos fragmentos bien caracterizados producidos por digestión con varias peptidasas. Por lo tanto, por ejemplo, la pepsina digiere un anticuerpo por debajo de los enlaces disulfuro en la región de bisagra para producir F(ab')₂, un dímero de Fab que en sí mismo es una cadena ligera unida a V_H-C_H1 mediante un enlace disulfuro. El F(ab')₂ se puede reducir en condiciones suaves para romper la unión disulfuro en la región de bisagra, convirtiendo así el dímero F(ab')₂ en un monómero Fab'. El monómero Fab' es esencialmente un Fab con parte de la región bisagra (véase, Fundamental Immunology, 4.^a edición, W.E. Paul, ed., Raven Press, N.Y. (1999), para una descripción más detallada de otros fragmentos de anticuerpo). Aunque varios fragmentos de anticuerpos se definen en términos de la digestión de un anticuerpo intacto, un experto apreciará que dichos fragmentos Fab', etc., pueden sintetizarse *de novo* ya sea químicamente o utilizando metodología de ADN recombinante. Por lo tanto, el término "anticuerpo", como se usa en el presente documento, incluye también opcionalmente fragmentos de anticuerpo ya sea producidos mediante la modificación de anticuerpos completos o sintetizados *de novo* utilizando metodologías de ADN recombinante. Los anticuerpos incluyen anticuerpos monocatenarios, incluyendo Fv monocatenario (sFv o scFv) anticuerpos en los que una cadena pesada variable y una ligera variable se encuentran unidas entre sí (directamente o a través de un enlazador peptídico) para formar un polipéptido continuo. Los anticuerpos de la invención pueden ser, por ejemplo, policlonales, monoclonales, quiméricos, humanizados, monocatenarios, fragmentos Fab, fragmentos producidos por una biblioteca de expresión Fab, o similares.

30

35

40

45

En general, los anticuerpos de la invención son valiosos, tanto como reactivos generales como reactivos terapéuticos en diversos procesos moleculares biológicos o farmacéuticos. Los métodos para producir anticuerpos policlonales y monoclonales se encuentran disponibles, y pueden aplicarse para preparar los anticuerpos de la invención. Varios textos básicos describen procesos estándar de producción de anticuerpos, incluyendo, por ejemplo, Borrebaeck (ed) (1995) Antibody Engineering, 2.^a Edición Freeman and Company, NY (Borrebaeck); McCafferty et al. (1996) Antibody Engineering, A Practical Approach IRL at Oxford Press, Oxford, Inglaterra (McCafferty), y Paul (1995) Antibody Engineering Protocols Humana Press, Towata, NJ (Paul); Paul (ed.), (1999) Fundamental Immunology, Quinta edición Raven Press, N.Y.; Coligan (1991) Current Protocols in Immunology Wiley/Greene, NY; Harlow y Lane (1989) Antibodies: A Laboratory Manual Cold Spring Harbor Press, NY; Stites et al. (eds.) Basic and Clinical Immunology (4.^a ed.) Lange Medical Publications, Los Altos, CA, y referencias citadas en los mismos; Goding (1986) Monoclonal Antibodies: Principles and Practice (2.^a ed.) Academic Press, Nueva York, NY; y Kohler y Milstein (1975) Nature 256: 495-497.

50

55

Diversas técnicas recombinantes para la preparación de anticuerpos que no se basan, por ejemplo, en la inyección de un antígeno en un animal, se han desarrollado y pueden utilizarse en el contexto de la presente invención. Por ejemplo, es posible generar y seleccionar bibliotecas de anticuerpos recombinantes en vectores fago o similares. Véase, por ejemplo, Winter et al. (1994) Making Antibodies by Phage Display Technology Annu. Rev. Immunol. 12:433-55 y las referencias citadas en el mismo para una revisión. Véanse también, Griffiths y Duncan (1998) Strategies for selection of antibodies by phage display Curr Opin Biotechnol 9: 102-8; Hoogenboom et al. (1998) Antibody phage display technology and its applications Immunotechnology 4: 1-20; Gram et al. (1992) *in vitro* selection and affinity maturation of antibodies from a naïve combinatorial immunoglobulin library PNAS 89:3576-

60

65

3580; Huse et al. (1989) Science 246: 1275-1281; y Ward, et al. (1989) Nature 341: 544-546.

En una realización, las bibliotecas de anticuerpo pueden incluir repertorios de genes V (por ejemplo, cosechados de poblaciones de linfocitos o ensamblados *in vitro*) que se clonan para la visualización de dominios variables asociados de cadena pesada y ligera en la superficie de bacteriófagos filamentosos. Los fagos se seleccionan por unión a un antígeno. Los anticuerpos solubles se expresan de bacterias infectadas por fagos y el anticuerpo puede mejorarse, por ejemplo, a través de mutagénesis. Véanse, por ejemplo, Balint y Larrick (1993) Antibody Engineering by Parsimonious Mutagenesis Gene 137:109-118; Stemmer et al. (1993) Selection of an Active Single Chain Fv Antibody From a Protein Linker Library Prepared by Enzymatic Inverse PCR Biotechniques 14(2):256-65; Crameri et al. (1996) Construction and evolution of antibody-phage libraries by DNA shuffling Nature Medicine 2:100-103; y Crameri y Stemmer (1995) Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes BioTechniques 18:194-195.

También se conocen y se encuentran disponibles kits para la clonación y expresión de sistemas de fagos de anticuerpos recombinantes, por ejemplo, el "sistema de fagos de anticuerpos recombinantes, módulo ScFv de ratón", de Amersham-Pharmacia Biotechnology (Uppsala, Suecia). También se han producido bibliotecas de anticuerpos bacteriófagos para producir anticuerpos humanos de alta afinidad mediante transposición de cadena (véase, por ejemplo, Marks et al. (1992) By-Passing Immunization: Building High Affinity Human Antibodies by Chain Shuffling Biotechniques 10:779-782. También se reconocerá que los anticuerpos pueden prepararse por cualquier número de servicios comerciales (por ejemplo, Bethyl Laboratories (Montgomery, TX), Anawa (Suiza), Eurogentec (Bélgica y en EE.UU. en Filadelfia, PA, etc.) y muchos otros.

En determinadas realizaciones, es útil "humanizar" los anticuerpos de la invención, por ejemplo, cuando los anticuerpos van a administrarse terapéuticamente. El uso de anticuerpos humanizados tiende a reducir la incidencia de respuestas inmunológicas no deseadas contra los anticuerpos terapéuticos (por ejemplo, cuando el paciente es un ser humano). Las referencias a anticuerpos anteriores describen estrategias de humanización. Además de los anticuerpos humanizados, también son una característica de la invención los anticuerpos humanos. Los anticuerpos humanos consisten en secuencias de inmunoglobulina característicamente humanas. Los anticuerpos humanos pueden producirse para utilizar una amplia diversidad de métodos (véase, por ejemplo, Larrick et al., Pat. de EE.UU. N.º 5.001.065, para una revisión). Un enfoque general para producir anticuerpos humanos mediante tecnología de trioma se describe por Ostberg et al. (1983), Hybridoma 2: 361-367, Ostberg, Pat. de EE.UU. N.º 4.634.664, y Engelman et al., Pat. de EE.UU. N.º 4.634.666.

Se conoce diversos métodos para el uso de anticuerpos en la purificación y detección de proteínas y pueden aplicarse para detectar y purificar proteínas que comprenden aminoácidos artificiales como se aprecia en el presente documento. En general, los anticuerpos son reactivos útiles para ELISA, transferencia de Western, inmunquímica, métodos de cromatografía por afinidad, SPR, y muchos otros métodos. Las referencias indicadas anteriormente proporcionan detalles sobre cómo realizar ensayos ELISA, transferencias de Western, resonancia de plasmón superficial (SPR) y similares.

En un aspecto de la invención, los anticuerpos de la invención en sí mismos incluyen aminoácidos artificiales, que proporcionan los anticuerpos con propiedades de interés (por ejemplo, una mejor semivida, estabilidad, toxicidad, y similares). Véase también, la sección en el presente documento titulada "Polipéptidos con aminoácidos artificiales". Los anticuerpos representan casi el 50 % de todos los compuestos actualmente en pruebas clínicas (Wittrup, (1999) Phage on display Tibtech 17: 423-424, y los anticuerpos se utilizan de manera ubicua como reactivos de diagnóstico. Por consiguiente, la capacidad para modificar anticuerpos con aminoácidos artificiales proporciona una importante herramienta para modificar estos valiosos reactivos.

Por ejemplo, existen muchas aplicaciones de Mabs en el campo del diagnóstico. Los ensayos varían desde simples pruebas rápidas hasta métodos más involucrados tales como el NR-LU-10 Mab radio-marcado de DuPont Merck, Co., utilizado para el diagnóstico por imagen de tumores (Rusch et al. (1993) NR-LU-10 monoclonal antibody scanning. A helpful new adjunct to computed tomography in evaluating non-small-cell lung cancer. J Thorac Cardiovasc Surg 106: 200-4). Como se ha indicado, los Mab son reactivos centrales para ELISA, transferencia de Western, inmunquímica, métodos de cromatografía de afinidad, y similares. Cualquiera de dichos anticuerpos de diagnóstico puede modificarse para incluir uno o más aminoácidos artificiales, alterando, por ejemplo, la especificidad o avidéz del Ab por un objetivo, o alterando una o más propiedades detectables, por ejemplo, incluyendo una etiqueta detectable (por ejemplo, espectrográfica, fluorescente, luminiscente, etc.) en el aminoácido artificial.

Una clase de valiosos reactivos de anticuerpo son los Ab terapéuticos. Por ejemplo, los anticuerpos pueden ser MaAb específicos de tumor que disminuyen el crecimiento tumoral dirigiéndose a células tumorales para su destrucción por medio de citotoxicidad mediada por células dependiente de anticuerpos (ADCC) o lisis mediada por complemento (CML) (estos tipos generales de Ab se denominan a veces como "balas mágicas"). Un ejemplo es Rituxan, un Mab anti-CD20 para el tratamiento de linfoma no Hodgkin (Scott (1998) Rituximab: a new therapeutic monoclonal antibody for non-Hodgkin's lymphoma Cancer Pract 6: 195-7). Un segundo ejemplo se refiere a anticuerpos que interfieren con un componente crítico del crecimiento tumoral. La herceptina es un anticuerpo

- monoclonal anti-HER-2 para el tratamiento de cáncer de mama metastásico, y proporciona un ejemplo de un anticuerpo con este mecanismo de acción (Baselga et al. (1998) Recombinant humanized anti-HER2 antibody (Herceptin) enhances the antitumor activity of paclitaxel and doxorubicin against HER2/neu overexpressing human breast cancer xenografts [published erratum appears in Cancer Res (1999) 59(8):2020], Cancer Res 58: 2825-31).
- 5 Un tercer ejemplo se refiere a anticuerpos para el suministro de compuestos citotóxicos (toxinas, radionúclidos, etc.) directamente a un tumor u otro sitio de interés. Por ejemplo, un Mab de aplicación es CYT-356, un anticuerpo unido a 90Y que dirige la radiación directamente a células de tumor de próstata (Deb et al. (1996) Treatment of hormone-refractory prostate cancer with 90Y-CYT-356 monoclonal antibody Clin Cancer Res 2: 1289-97. Una cuarta aplicación es la terapia de profármaco de enzima dirigida al anticuerpo, donde una enzima colocada en un tumor
- 10 activa un profármaco administrado por vía sistémica en la proximidad del tumor. Por ejemplo, se ha desarrollado un anticuerpo anti-Ep-CAM1 unido a carboxipeptidasa A para el tratamiento de cáncer colorrectal (Wolfe et al. (1999) Antibody-directed enzyme prodrug therapy with the T268G mutant of human carboxypeptidase A1: *in vitro* and *in vivo* studies with prodrugs of methotrexate and the thymidylate synthase inhibitors GW1031 and GW1843 Bioconjug Chem 10: 38-48). Otros Ab (por ejemplo, antagonistas) están diseñados para inhibir específicamente las funciones celulares normales para un beneficio terapéutico. Un ejemplo es Orthoclone OKT3, un Mab anti-CD3 ofrecido por Johnson and Johnson para reducir el rechazo agudo al trasplante de órganos (Strate et al. (1990) Orthoclone OKT3 as first-line therapy in acute renal allograft rejection Transplant Proc 22: 219-20. Otra clase de productos de anticuerpo son los agonistas. Estos Mab están diseñados para aumentar específicamente las funciones celulares normales para un beneficio terapéutico. Por ejemplo, los agonistas a base de Mab de receptores acetilcolina para
- 20 neuroterapia se encuentran en desarrollo (Xie et al. (1997) Direct demonstration of MuSK involvement in acetylcholine receptor clustering through identification of agonist ScFv Nat. Biotechnol. 15: 768-71. Cualquiera de estos anticuerpos puede modificarse para incluir uno o más aminoácidos artificiales para aumentar una o más propiedades terapéuticas (especificidad, avidéz, semivida en suero, etc.).
- 25 Otra clase de productos de anticuerpo proporciona nuevas funciones. Los anticuerpos principales en este grupo son los anticuerpos catalíticos tales como las secuencias Ig que se han diseñado para imitar las capacidades catalíticas de enzimas (Wentworth and Janda (1998) Catalytic antibodies Curr Opin Chem Biol 2: 138-44. Por ejemplo, una aplicación interesante implica el uso del anticuerpo catalítico mAb-15A10 para hidrolizar cocaína *in vivo* para terapia contra la adicción (Mets et al. (1998) A catalytic antibody against cocaine prevents cocaine's reinforcing and toxic effects in rats Proc Natl Acad Sci U S A 95: 10176-81). Los anticuerpos catalíticos también pueden modificarse para incluir uno o más aminoácidos artificiales para mejorar una o más propiedades de interés.
- 30

Definición de polipéptidos mediante inmunorreactividad

- 35 Debido a que los polipéptidos de la invención proporcionan diversas nuevas secuencias de polipéptidos (por ejemplo, que comprenden aminoácidos artificiales en el caso de proteínas sintetizadas en los sistemas de traducción en el presente documento, o por ejemplo, en el caso de las nuevas sintetisas en el presente documento, nuevas secuencias de aminoácidos estándar), los polipéptidos proporcionan también nuevas características estructurales que pueden reconocerse, por ejemplo, en ensayos inmunológicos. La generación de anticuerpos o anticuerpos que
- 40 se unen específicamente a los polipéptidos de la invención, así como los polipéptidos que están enlazados por dichos anticuerpos o antisueros, son una característica de la invención.

- Por ejemplo, la invención incluye proteínas de sintetasa que se unen específicamente o que son específicamente inmunorreactivas con anticuerpos o antisueros generados contra un inmunógeno que comprende una secuencia de
- 45 aminoácidos seleccionada de una o más de (SEQ ID NO: 36-63 (por ejemplo, 36-47, 48-63, o cualquier otro subconjunto de 36-63), y/o 86). Para eliminar la reactividad cruzada con otros homólogos, el anticuerpo o antisuero se sustrae con homólogos de control de sintetasa disponibles, tal como tirosil sintetasa de *E. coli* de tipo silvestre (TyrRS) (por ejemplo, SEQ ID NO:2).

- 50 En un formato típico, el inmunoensayo utiliza antisuero policlonal que se cultivó contra uno o más polipéptidos que comprendían una o más de las secuencias correspondientes a una o más de las SEQ ID NO: 36-63 (por ejemplo, 36-47, 48-63, o cualquier otro subconjunto de 36-63), y/o 86, o una subsecuencia sustancial de las mismas (es decir, al menos aproximadamente el 30 % de la secuencia de longitud completa proporcionada). El conjunto de inmunógenos de polipéptidos potenciales derivados de SEQ ID NO: 36-63 y 86 se denominan colectivamente a
- 55 continuación "los polipéptidos inmunógenos". El antisuero resultante se selecciona opcionalmente para que tenga una reactividad cruzada contra los homólogos de sintetasa de control y cualquier reactividad cruzada de este tipo se elimina, por ejemplo, mediante inmuoabsorción, con uno o más homólogos de sintetasa de control, antes del uso del antisuero policlonal en el inmunoensayo.

- 60 A fin de producir antisuero para su uso en un inmunoensayo, se producen y se purifican uno o más de los polipéptidos inmunógenos como se describe en el presente documento. Por ejemplo, la proteína recombinante puede producirse en una célula recombinante. Una cepa criada de ratones (utilizados en este ensayo dado que los resultados son más reproducibles debido a la identidad genética virtual de los ratones) se inmuniza con la proteína o proteínas inmunógenas en combinación con un adyuvante estándar, tal como el adyuvante de Freund, y un protocolo estándar de inmunización de ratón (véase, por ejemplo, Harlow y Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, Nueva York, para una descripción estándar de la generación de
- 65

anticuerpos, formatos de inmunoensayos y condiciones que pueden utilizarse para determinar la inmunorreactividad específica. También se encuentran en el presente documento referencias adicionales y el análisis de anticuerpos y pueden aplicarse en el presente documento para preparar anticuerpos que definen/detectan polipéptidos mediante inmunorreactividad). Como alternativa, uno o más polipéptidos artificiales o recombinantes derivados de las

5 secuencias divulgadas en el presente documento se conjugan con una proteína portadora y se utilizan como inmunógeno.

Se recogen sueros policlonales y se titulan contra el polipéptido inmunógeno en un inmunoensayo, por ejemplo, un inmunoensayo en fase sólida con una o más de las proteínas inmunógenas inmovilizadas en un soporte sólido. Se

10 seleccionan sueros policlonales con una titulación de 10^6 o más, se agrupan y se sustraen con los polipéptidos de sintetasa de control para producir antisuero policlonal titulado combinado sustraído.

El antisuero policlonal titulado combinado sustraído se prueba para determinar la reactividad cruzada contra los homólogos de control en un inmunoensayo comparativo. En este ensayo comparativo, se determinan las

15 condiciones discriminatorias de unión para el antisuero policlonal titulado sustraído que dan como resultado una relación de señal-ruido de al menos aproximadamente 5-10 veces más alta para la unión del antisuero policlonal titulado a la sintetasa inmunógena en comparación con la unión a un homólogo de sintetasa de control. Es decir, la rigurosidad de la reacción o reacciones de unión/lavado se ajustan mediante la adición de competidores no

20 específicos tales como albúmina o leche seca desnatada, y/o ajustando las condiciones de sal, temperatura, y/o similares. Estas condiciones de unión/lavado se utilizan en ensayos posteriores para determinar si un polipéptido de prueba (comparándose un polipéptido con los polipéptidos inmunógenos y/o los polipéptidos de control) se encuentra específicamente unido mediante el antisuero policlonal sustraído combinado. En particular, los polipéptidos de prueba, que muestran al menos una relación de señal-ruido 2-5 veces más alta que el homólogo de sintetasa de control en condiciones discriminatorias de unión, y al menos aproximadamente una relación de señal-

25 ruido de 1/2 en comparación con el polipéptido o polipéptidos inmunógenos, comparten una similitud estructural sustancial con el polipéptido inmunógeno en comparación con sintetastas conocidas, y es, por lo tanto, un polipéptido de la invención.

En otro ejemplo, los inmunoensayos en el formato de unión competitiva se utilizan para la detección de un polipéptido de prueba. Por ejemplo, según se indica, los anticuerpos de reacción cruzada se eliminan de la mezcla

30 de antisuero combinada mediante inmunoabsorción con los polipéptidos de control. El polipéptido o polipéptidos inmunógenos se inmovilizan a continuación en un soporte sólido que se encuentra expuesto al antisuero combinado sustraído. Se añaden las proteínas de prueba al ensayo para competir por la unión al antisuero sustraído combinado. La capacidad de la proteína o proteínas de prueba para competir por la unión al antisuero sustraído

35 combinado en comparación con la proteína o proteínas inmovilizadas se compara con la capacidad del polipéptido o polipéptidos inmunógenos añadidos al ensayo, para competir por la unión (los polipéptidos inmunógenos compiten eficazmente con los polipéptidos inmunógenos inmovilizados por la unión al antisuero combinado). La reactividad cruzada porcentual para las proteínas de prueba se calcula utilizando cálculos estándar.

En un ensayo paralelo, la capacidad de las proteínas de control para competir por la unión al antisuero sustraído combinado se determina opcionalmente en comparación con la capacidad del polipéptido o polipéptidos

40 inmunógenos para competir por la unión al antisuero. De nuevo, se calcula la reactividad cruzada porcentual para los polipéptidos de control, utilizando cálculos estándar. Cuando la reactividad cruzada porcentual es al menos 5-10x tan alta para los polipéptidos de prueba en comparación con los polipéptidos de control y/o cuando la unión de los polipéptidos de control se encuentra aproximadamente en el intervalo de la unión de los polipéptidos inmunógenos,

45 se dice que los polipéptidos de prueba se unen específicamente al antisuero sustraído combinado.

En general, el antisuero inmunoabsorbido y combinado puede utilizarse en un inmunoensayo de unión competitiva como se describe en el presente documento para comparar cualquier polipéptido de prueba con el polipéptido o

50 polipéptidos inmunógenos y/o de control. A fin de hacer esta comparación, los polipéptidos inmunógenos, de prueba y de control se ensayan cada uno en un amplio intervalo de concentraciones, y se determina la cantidad de cada polipéptido requerida para inhibir el 50 % de la unión del antisuero sustraído a, por ejemplo, una proteína inmovilizada de control, de prueba o inmunógena utilizando técnicas estándar. Si la cantidad del polipéptido de prueba requerida para la unión en el ensayo competitivo es menor de dos veces la cantidad del polipéptido

55 inmunógeno requerida, entonces se dice que el polipéptido de prueba se une específicamente a un anticuerpo generado para la proteína inmunógena, siempre que la cantidad sea al menos aproximadamente 5-10x tan alta como para el polipéptido de control.

Como una determinación adicional de especificidad, el antisuero combinado opcionalmente se inmunoabsorbe completamente con el polipéptido o polipéptidos inmunógenos (en lugar de los polipéptidos de control) hasta que se

60 puede detectar poca o ninguna unión del antisuero combinado sustraído del polipéptido inmunógeno resultante al polipéptido o polipéptidos inmunógenos utilizados en la inmunoabsorción. Este antisuero completamente inmunoabsorbido se analiza a continuación para determinar la reactividad con el polipéptido de prueba. Si se observa poca o ninguna reactividad (es decir, no más de 2x la relación de señal-ruido observada para la unión del antisuero completamente inmunoabsorbido al polipéptido inmunógeno), entonces el polipéptido de prueba se une

65 específicamente mediante el antisuero emitido por la proteína inmunógena.

COMPOSICIONES FARMACÉUTICAS

Los polipéptidos o proteínas de la invención (por ejemplo, sintetasas, proteínas que comprenden uno o más aminoácidos artificiales, etc.) se emplean opcionalmente para usos terapéuticos, por ejemplo, en combinación con un vehículo farmacéutico adecuado. Dichas composiciones, por ejemplo, comprenden una cantidad terapéuticamente eficaz del compuesto, y un vehículo o excipiente farmacéuticamente aceptable. Tal vehículo o excipiente incluye, pero sin limitación, solución salina, solución salina tamponada, dextrosa, agua, glicerol, etanol, y/o combinaciones de los mismos. La formulación se prepara para adecuarse al modo de administración. En general, los métodos para administrar proteínas se conocen bien en la técnica y pueden aplicarse para la administración de los polipéptidos de la invención.

Las composiciones terapéuticas que comprenden uno o más polipéptidos de la invención se prueban opcionalmente en uno o más modelos animales apropiados de la enfermedad *in vitro* y/o *in vivo*, para confirmar la eficacia, el metabolismo tisular, y para estimar las dosis, de acuerdo con métodos ya conocidos en la técnica. En particular, las dosis pueden determinarse inicialmente mediante la actividad, estabilidad u otras medidas adecuadas de homólogos de aminoácidos artificiales en el presente documento a naturales (por ejemplo, comparación de un EPO modificado para incluir uno o más aminoácidos artificiales a una EPO de aminoácido natural), es decir, en un ensayo relevante.

La administración es mediante cualquiera de las rutas normalmente utilizadas para introducir una molécula en contacto final con células sanguíneas o de tejido. Los polipéptidos de aminoácido artificial de la invención se administran de cualquier manera adecuada, opcionalmente con uno o más vehículos farmacéuticamente aceptables. Los métodos adecuados para administrar dichos polipéptidos en el contexto de la presente invención a un paciente se encuentran disponibles y, aunque se puede usar más de una ruta para administrar una composición particular, una ruta particular puede proporcionar frecuentemente una acción o reacción más inmediata y más eficaz que otra ruta.

Los vehículos farmacéuticamente aceptables se determinan en parte por la composición particular que se administra, así como por el método particular utilizado para administrar la composición. Por consiguiente, existe una amplia variedad de formulaciones adecuadas de las composiciones farmacéuticas de la presente invención.

Las composiciones de polipéptidos pueden administrarse mediante varias rutas que incluyen, pero sin limitación: medio oral, intravenoso, intraperitoneal, intramuscular, transdérmico, subcutáneo, tópico, sublingual, o rectal. Las composiciones de polipéptidos de aminoácidos artificiales también pueden administrarse a través de liposomas. Dichas rutas de administración y formulaciones apropiadas se conocen generalmente por los expertos en la técnica.

El polipéptido de aminoácidos artificiales, en solitario o en combinación con otros componentes adecuados, puede prepararse también en formulaciones en aerosol (es decir, pueden "nebulizarse") para administrarse a través de inhalación. Las formulaciones de aerosol pueden colocarse en propulsores aceptables a presión, tales como diclorodifluorometano, propano, nitrógeno y similares.

Las formulaciones adecuadas para administración parenteral, tales como, por ejemplo, mediante ruta intraarticular (en las articulaciones), intravenosa, intramuscular, intradérmica, intraperitoneal, y subcutánea, incluyen soluciones para inyección estériles isotónicas acuosas y no acuosas, que pueden contener antioxidantes, tampones, bacteriostáticos, y solutos que hacen que la formulación sea isotónica con la sangre del receptor previsto, y suspensiones estériles acuosas y no acuosas que pueden incluir agentes de suspensión, solubilizantes, agentes espesantes, estabilizantes y conservantes. Las formulaciones de ácido nucleico empaquetado pueden presentarse en recipientes sellados de dosis unitarias o de dosis múltiples, tales como ampollas y viales.

La administración parenteral y la administración intravenosa son métodos preferidos de administración. En particular, las rutas de administración ya en uso para productos terapéuticos de homólogos de aminoácidos naturales (por ejemplo, los que se utilizan normalmente para EPO, GCSF, GMCSF, IFN, interleucinas, anticuerpos y/o cualquier otra proteína administrada farmacéuticamente), junto con las formulaciones en uso actual, proporcionan las rutas preferidas de administración y formulación para las proteínas que incluyen aminoácidos artificiales de la invención (por ejemplo, variantes pegiladas de proteínas terapéuticas actuales, etc.).

La dosis administrada a un paciente, en el contexto de la presente invención, es suficiente para realizar una respuesta terapéutica beneficiosa en el paciente en el tiempo, o, por ejemplo, para inhibir una infección por una actividad patógena, u otra apropiada, dependiendo de la aplicación. La dosis se determina mediante la eficacia de una composición/formulación particular, y la actividad, estabilidad o semivida en suero del polipéptido de aminoácido artificial empleado y la condición del paciente, así como el peso corporal o el área superficial del paciente a tratar. El tamaño de la dosis también se determina por la existencia, naturaleza, y extensión de cualquier efecto adverso secundario que acompañe a la administración de una composición/formulación particular, o similares en un paciente en particular.

Para determinar la cantidad eficaz de la composición/formulación a administrar en el tratamiento o profilaxis de la

enfermedad (por ejemplo, cánceres, enfermedades hereditarias, diabetes, SIDA, o similares), el médico evalúa los niveles en plasma circulantes, las toxicidades de la formulación, el progreso de la enfermedad, y/o cuando es relevante, la producción de anticuerpos anti-polipéptidos de aminoácidos artificiales.

- 5 La dosis administrada, por ejemplo, para un paciente de 70 kilogramos, se encuentra normalmente en el intervalo equivalente a las dosis de proteínas terapéuticas utilizadas actualmente, ajustadas para la actividad alterada de la semivida en suero de la composición relevante. Las composiciones/formulaciones de esta invención pueden complementar las condiciones de tratamiento mediante cualquier terapia convencional conocida, incluyendo la administración de anticuerpos, administración de vacunas, administración de agentes citotóxicos, polipéptidos de aminoácidos naturales, ácidos nucleicos, análogos de nucleótidos, modificadores de respuesta biológica, y similares.

10 Para su administración, las formulaciones de la presente invención se administran a una tasa determinada por la DL-50 de la formulación relevante, y/o la observación de cualquier efecto secundario de los aminoácidos artificiales en diversas concentraciones, por ejemplo, según se aplican a la masa y a la salud general del paciente. La administración puede lograrse a través de dosis únicas o divididas.

Si un paciente que experimenta la infusión de una formulación desarrolla fiebre, escalofríos o dolores musculares, recibe la dosis apropiadas de aspirina, ibuprofeno, acetaminofeno u otro fármaco para el control del dolor/fiebre. Los paciente que experimentan reacciones a la infusión tales como fiebre, dolores musculares, y escalofríos se premedican 30 minutos antes de futuras infusiones ya sea con aspirina, acetaminofeno, o, por ejemplo, difenilhidramina. Se utiliza meperidina para escalofríos y dolores musculares más graves que no responden rápidamente a antipiréticos y antihistaminas. El tratamiento se retrasa o se interrumpe de la gravedad de la reacción.

SECUENCIA Y VARIANTES DE ÁCIDOS NUCLEICOS Y POLIPÉPTIDOS

25 Como se ha descrito anteriormente y a continuación, la invención proporciona secuencias polinucleotídicas de ácido nucleico y secuencias de aminoácidos de polipéptidos, por ejemplo, O-ARNt y O-RS, y, por ejemplo, composiciones y métodos que comprenden dichas secuencias. Ejemplos de dichas secuencias, por ejemplo, O-ARNt y O-RS se divulgan en el presente documento (véanse, la Tabla 5, por ejemplo, la SEQ ID NO. 3-65, 86, y otras distintas de las SEQ ID NO.: 1 y 2). Sin embargo, el experto en la técnica apreciará que la presente divulgación no se limita a las secuencias divulgadas en el presente documento, por ejemplo, los Ejemplos y la Tabla 5. Un experto apreciará que la presente divulgación proporciona también muchas secuencias relacionadas e incluso no relacionadas con las funciones descritas en el presente documento, por ejemplo, que codifican un O-ARNt o una O-RS.

35 La presente divulgación proporciona también polipéptidos (O-RS) y polinucleótidos, por ejemplo, O-ARNt, polinucleótidos que codifican las O-RS o porciones de las mismas (por ejemplo, el sitio activo de la sintetasa), los oligonucleótidos utilizados para construir mutantes de aminoacil-ARNt sintetasa, etc. Por ejemplo, un polipéptido de la invención incluye un polipéptido que comprende una secuencia de aminoácidos como se muestra en cualquiera de las SEQ ID NO.: 48-63, un polipéptido que comprende una secuencia de aminoácidos codificada por una secuencia polinucleotídica como se describe en una cualquiera de las SEQ ID NO.: 20-35, o un polipéptido que comprende una secuencia de aminoácidos codificada por una secuencia polinucleotídica como se describe en una cualquiera de las SEQ ID NO.: 20-35.

45 También se incluyen entre los polipéptidos de la invención los polipéptidos que comprenden una secuencia de aminoácidos que es al menos un 90 % idéntica a la de una tirosil aminoacil-ARNt sintetasa de origen natural (TyrRS) (por ejemplo, SEQ ID NO. :2) y comprende dos o más aminoácidos de los grupos A-E. Por ejemplo, el grupo A incluye valina, isoleucina, leucina, glicina, serina, alanina, o treonina en una posición correspondiente a Tyr37 de TyrRS de *E. coli*; el grupo B incluye aspartato en una posición correspondiente a Asn126 de TyrRS de *E. coli*; el grupo C incluye treonina, serina, arginina, asparagina o glicina en una posición correspondiente a Asp182 de TyrRS de *E. coli*; el grupo D incluye metionina, alanina, valina, o tirosina en una posición correspondiente a Phe183 de TyrRS de *E. coli*; y, el grupo E incluye serina, metionina, valina, cisteína, treonina, o alanina en una posición correspondiente a Leu186 de TyrRS de *E. coli*. Se divulga cualquier subconjunto de combinaciones de estos grupos. Por ejemplo, en una realización, la O-RS tiene dos o más aminoácidos seleccionados de valina, isoleucina, leucina, o treonina aparece en una posición correspondiente a Tyr37 de TyrRS de *E. coli*; treonina, serina, arginina, o glicina en una posición correspondiente a Asp182 de TyrRS de *E. coli*; metionina, o tirosina en una posición correspondiente a Phe183 de TyrRS de *E. coli*; y, serina, o alanina en una posición correspondiente a Leu186 de TyrRS de *E. coli*. En otra realización, la O-RS incluye dos o más aminoácidos seleccionados de glicina, serina, o alanina en una posición correspondiente a Tyr37 de TyrRS de *E. coli*; aspartato en una posición correspondiente a Asn126 de TyrRS de *E. coli*; asparagina en una posición correspondiente a Asp182 de TyrRS de *E. coli*; alanina, o valina, en una posición correspondiente a Phe183 de TyrRS de *E. coli*, y/o metionina, valina, cisteína, o treonina, en una posición correspondiente a Leu186 de TyrRS de *E. coli*. De manera similar, los polipéptidos de la presente divulgación también incluyen un polipéptido que comprende al menos 20 aminoácidos contiguos de SEQ ID NO.: 36-63 (por ejemplo, 36-47, 48-63, o cualquier otro subconjunto de 36-63), y/o 86, y dos o más sustituciones de aminoácido como se indica anteriormente en los grupos A-E. Véanse, también, la Tabla 4, la Tabla 6, y/o la Tabla 8, en el presente documento. Una secuencia de aminoácidos que comprende una variación conservativa de cualquiera de los polipéptidos anteriores también se incluye como un polipéptido de la presente divulgación.

En una realización, una composición incluye un polipéptido de la presente divulgación y un excipiente (por ejemplo, tampón, agua, excipiente farmacéuticamente aceptable, etc.). La presente divulgación también proporciona un anticuerpo o un antisuero específicamente inmunorreactivo con un polipéptido de la presente divulgación.

También se proporcionan polinucleótidos. Los polinucleótidos de la presente divulgación incluyen aquellos que codifican las proteínas o polipéptidos de interés de la presente divulgación, o que incluyen uno o más codones selectores, o ambos. Por ejemplo, los polinucleótidos de la presente divulgación incluyen, por ejemplo, un polinucleótido que comprende una secuencia de nucleótidos como se expone en cualquiera de las SEQ ID NO.: 3-35 (por ejemplo, 3-19, 20-35, o cualquier otro subconjunto de las secuencias 3-35), 64-85; un polinucleótido complementario a o que codifica una secuencia polinucleotídica del mismo; y/o un polinucleótido que codifica un polipéptido que comprende una secuencia de aminoácidos como se expone en una cualquiera de las SEQ ID NO.: 36-63, y/o 86, una variación conservativa de las mismas. Un polinucleótido de la presente divulgación incluye también un polinucleótido que codifica un polipéptido de la presente divulgación. De manera similar, un ácido nucleico que hibrida con un polinucleótido indicado anteriormente en condiciones altamente rigurosas durante sustancialmente toda la longitud del ácido nucleico es un polinucleótido de la presente divulgación.

Un polinucleótido de la presente divulgación también incluye un polinucleótido que codifica un polipéptido que comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la de una tirosil aminoacil-ARNt sintetasa (TyrRS) de origen natural (por ejemplo, SEQ ID NO.: 2) y comprende dos o más mutaciones como se ha indicado anteriormente en los grupos A-E (anteriormente). Un polinucleótido que es al menos un 70 %, (o al menos un 75 %, al menos un 80 %, al menos un 85 %, al menos un 90 %, al menos un 95 %, al menos un 98 %, o al menos un 99 % o más) idéntica a un polinucleótido indicado anteriormente y/o un polinucleótido que comprende una variación conservativa de cualquiera de los polinucleótidos indicados anteriormente también se incluyen entre los polinucleótidos de la presente divulgación. Véanse también, la Tabla 4, la Tabla 6, y/o la Tabla 8, en el presente documento.

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

Un experto apreciará también que se contemplan muchas variantes de las secuencias divulgadas. Por ejemplo, se contemplan variaciones conservativas de las secuencias divulgadas que producen una secuencia funcionalmente idéntica.

Las variantes de las secuencias polinucleotídicas de ácido nucleico, en las que las variantes hibridan con al menos una secuencia divulgada, se incluyen por la presente divulgación. Subsecuencias únicas de las secuencias divulgadas en el presente documento, determinadas por, por ejemplo, técnicas convencionales de comparación de secuencias, también se contemplan.

Variaciones conservativas

Debido a la degeneración del código genético, las "sustituciones silenciosas" (es decir, sustituciones en una secuencia de ácido nucleico que no provocan una alteración en un polipéptido codificado) son una característica implícita de *todas* las secuencias de ácido nucleico que codifican un aminoácido. De manera similar, las "sustituciones conservativas de aminoácidos", en uno o algunos aminoácidos en una secuencia de aminoácidos están sustituidos con diferentes aminoácidos con propiedades muy similares, también se identifican fácilmente como muy similares a una construcción divulgada. Se divulgan dichas variaciones conservativas de cada secuencia divulgada.

Las "variaciones conservativas" de una secuencia de ácido nucleico particular se refieren a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o prácticamente idénticas o, cuando el ácido nucleico no codifica una secuencia de aminoácidos, a secuencias prácticamente idénticas. Un experto reconocerá que las sustituciones, eliminaciones o adicionales individuales que alteran, añaden o eliminan un único aminoácido un pequeño porcentaje de aminoácidos (normalmente menos de un 5 %, más normalmente menos de un 4 %, un 2 % o un 1 %) en una secuencia codificada son "variaciones modificadas de forma conservativa" en las que las alteraciones dan como resultado la eliminación de un aminoácido, la adición de aminoácido, o la sustitución de un aminoácido con un aminoácido químicamente similar. Por lo tanto, las "variaciones conservativas" de una secuencia polipeptídica enumerada de la presente divulgación invención incluyen sustituciones de un pequeño porcentaje, normalmente menos de un 5 %, más normalmente menos de un 2 % o un 1 %, de los aminoácidos de la secuencia polipeptídica, con un aminoácido seleccionado de forma conservativa del mismo grupo de sustituciones conservativas. Por último, la adición de secuencias que no alteran la actividad codificada de una molécula de ácido nucleico, tal como la adición de una secuencia no funcional, es una variación conservativa del ácido nucleico básico.

Las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares se conocen bien en la técnica. Los siguiente expone, por ejemplo, grupos que contienen aminoácidos naturales que incluyen "sustituciones conservativas" para ambos.

5

Grupos de sustitución conservativa

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

Hibridación de ácidos nucleicos

10

Puede utilizarse la hibridación comparativa para identificar ácidos nucleicos de la presente divulgación, incluyendo variaciones conservativas de los ácidos nucleicos de la presente divulgación, y este método de hibridación comparativa es un método para distinguir los ácidos nucleicos de la presente divulgación. Además, los ácidos nucleicos objetivo que hibridan con los ácidos nucleicos representados por la SEQ ID NO: 3-35 (por ejemplo, 3-19, 20-35, o cualquier otro subconjunto de las secuencias 3-35), 64-85 en condiciones de rigurosidad alta, ultra alta y ultra-ultra alta se divulgan.

15

Los ejemplos de dichos ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones de ácido nucleico silenciosas o conservativas en comparación con una secuencia de ácido nucleico dado.

20

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

25

30

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

35

40

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

45

50

55

Las "condiciones de lavado de hibridación rigurosas" en el contexto de experimentos de hibridación de ácidos nucleicos tales como hibridaciones de Southern y Northern son dependientes de la secuencia, y son diferentes en diferentes parámetros ambientales. Se encuentra una amplia guía sobre la hibridación de ácidos nucleicos en Tijssen (1993), *anteriormente*, y en Hames y Higgins, 1 y 2. Las condiciones de hibridación y lavado rigurosas pueden determinarse fácilmente de forma empírica para cualquier ácido nucleico de ensayo. Por ejemplo, en la determinación de las condiciones de hibridación y lavado altamente rigurosas, las condiciones de hibridación y lavado se aumentan gradualmente (por ejemplo, aumentando la temperatura, disminuyendo la concentración salina, aumentando la concentración de detergente y/o aumentando la concentración de disolventes orgánicos tales como formalina en la hibridación o el lavado), hasta que se cumple un conjunto seleccionado de criterios. Por ejemplo, las condiciones de hibridación y lavado se aumentan gradualmente hasta que una sonda se une a una diana complementaria perfectamente coincidente con una relación de señal a ruido que es al menos 5x tan elevada 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 térmica (T_m) para una sonda particular. La T_m es la temperatura (a una fuerza iónica y pH definidos) a la que un 50 % de la secuencia de ensayo hibrida con una sonda perfectamente coincidente. Para los fines de la presente divulgación, generalmente, las condiciones de hibridación y lavado "altamente rigurosas" se seleccionan para que sean aproximadamente 5 °C por debajo de la T_m para la secuencia específica a una fuerza iónica y pH definidos.

Las condiciones de hibridación y lavado "de rigurosidad ultraelevada" son aquellas en que la rigurosidad de las condiciones de hibridación y lavado se aumentan hasta que la relación de señal a ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente coincidente es al menos 10x tan elevada como la observada para la hibridación a cualquiera de los ácidos nucleicos diana no coincidentes. Un ácido nucleico diana que hibrida con una sonda en dichas condiciones, con una relación de señal a ruido de al menos $\frac{1}{2}$ de la del ácido nucleico diana complementario perfectamente coincidente se dice que se une a la sonda en condiciones de rigurosidad ultraelevada.

De manera similar, pueden determinarse niveles incluso mayores de rigurosidad aumentando gradualmente las condiciones de hibridación y/o lavado del ensayo de hibridación relevante. Por ejemplo, aquellos en que la rigurosidad de las condiciones de hibridación y lavado se aumentan hasta que la relación de señal a ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente coincidente es al menos 10x, 20X, 50X, 100x o 500x o más tan elevada como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no coincidentes. Un ácido nucleico diana que hibrida con una sonda en dichas condiciones, con una relación de señal a ruido de al menos $\frac{1}{2}$ de la del ácido nucleico diana complementario perfectamente coincidente se dice que se une a la sonda en condiciones de rigurosidad ultraultraelevada.

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

Subsecuencias únicas

En un aspecto, la presente divulgación proporciona un ácido nucleico que comprende una subsecuencia única en un ácido nucleico seleccionado de las secuencias de O-ARNt y O-RS divulgadas en el presente documento. La subsecuencia única es única en comparación con un ácido nucleico correspondiente a cualquier secuencia de ácido nucleico de O-ARNt u O-RS. La alineación puede realizarse usando, por ejemplo, BLAST configurado a los parámetros por defecto. Cualquier subsecuencia única es útil, por ejemplo, como sonda para identificar los ácidos nucleicos de la presente divulgación.

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

La invención proporciona también ácidos nucleicos objetivo que se hibridan en condiciones rigurosas en un olionucleótido codificante único que codifica una única subsecuencia en un polipéptido seleccionado de las secuencias de O-RS en los que la subsecuencia única es única en comparación con un polipéptido correspondiente a cualquiera de los polipéptidos de control (por ejemplo, secuencias de origen a partir de las cuales se derivaron las sintetasas de la presente divulgación, por ejemplo, mediante mutación). Las secuencias únicas se determinan como se indica anteriormente.

Comparación de secuencias, identidad y homología

Los términos "idéntico" o porcentaje de "identidad", en el contexto de dos o más secuencias de los ácidos nucleicos o polipeptídicas, se refieren a dos o más secuencias o subsecuencias que son iguales o tienen un porcentaje específico de restos aminoacídicos o nucleótidos que son iguales, cuando se comparan y alinean para una correspondencia máxima, medida usando uno de los algoritmos de comparación de secuencias descritos a continuación (u otros algoritmos disponibles para los expertos) o por inspección visual.

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

comparar.

Para la comparación de secuencias y la determinación de la homología, normalmente una secuencia actúa como una secuencia de referencia con la que se comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de prueba y referencia se introducen en un ordenador, se designan las coordenadas de subsecuencia, si fuera necesario, y se designan los parámetros del programa del algoritmo de secuencias. El algoritmo de comparación de secuencias entonces calcula el porcentaje de identidad de secuencia para la o las secuencias de prueba respecto a la secuencia de referencia, basándose en los parámetros del programa designados.

La alineación óptima de secuencias para la comparación puede realizarse, por ejemplo, mediante el algoritmo de homología local de Smith y Waterman, Adv. Appl. Math. 2:482 (1981), mediante el algoritmo de alineación de homología de Needleman y Wunsch, J. Mol. Biol. 48:443 (1970), mediante el procedimiento de búsqueda de similitud de Pearson y Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), mediante implementaciones informatizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Paquete de Software Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, WI), o por inspección visual (véase, generalmente, Ausubel *et al.*, a continuación).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y similitud de secuencia es el algoritmo BLAST, que se describe en Altschul *et al.*, J. Mol. Biol. 215:403-410 (1990). El software para realizar análisis BLAST está disponible públicamente a través del Centro Nacional de Información Biotecnológica (www.ncbi.nlm.nih.gov/). Este algoritmo implica identificar en primer lugar pares de secuencias de alta valoración (HSP) identificando palabras cortas de longitud W en la secuencia de consulta, que coincide o satisface algún valor umbral de valor positivo T cuando se alinea con una palabra de la misma longitud en una secuencia de la base de datos. T se menciona como el umbral del valor de palabra adyacente (Altschul *et al.*, *anteriormente*). Estos aciertos de palabra adyacente iniciales actúan como semillas para iniciar búsquedas para encontrar HSP más largos que los contengan. Los aciertos de palabra entonces se prolongan en ambas direcciones a lo largo de cada secuencia todo el tiempo que pueda aumentarse el valor de alineación acumulativo. Los valores acumulativos se calculan usando, para secuencias nucleotídicas, los parámetros M (valor recompensa para un par de restos coincidentes; siempre > 0) y N (valor de penalización para restos no coincidentes; siempre < 0). Para las secuencias de aminoácidos, se usa una matriz de valores para calcular el valor acumulativo. La extensión de los aciertos de palabra en cada dirección se detiene cuando la puntuación de alineamiento acumulada está dentro de la cantidad X de su valor máximo alcanzado; el valor acumulativo llega a cero o por debajo, debido a la acumulación de una o más alineaciones de restos de valoración negativa; o se alcanza el final de cualquier secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) usa por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un límite de 100, M = 5, N = -4, y una comparación de ambas cadenas. Para las 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 valores BLOSUM62 (véase Henikoff y Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

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

Mutagénesis y otras técnicas de biología molecular

Los textos generales que describen técnicas biológicas moleculares incluyen Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volumen 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al.*, Molecular Cloning - A Laboratory Manual (2.^a Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989 ("Sambrook") y Current Protocols in Molecular Biology, F. M. Ausubel *et al.*, eds., Current Protocols, un proyecto conjunto entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (suplementado hasta 1999) ("Ausubel"). Estos textos describen la mutagénesis, el uso de vectores, promotores y muchos otros aspectos relevantes relacionados, por ejemplo, con la generación de genes que incluyen codones selectores para la producción de proteínas que incluyen aminoácidos artificiales, ARNt ortogonales, sintetisas ortogonales, y pares de los mismos.

En la presente divulgación se utilizan diversos tipos de mutagénesis, por ejemplo, para producir bibliotecas de ARNt, para producir bibliotecas de sintetisas, para insertar codones selectores que codifican aminoácidos artificiales en una proteína o polipéptido de interés. Estas incluyen, pero sin limitación, mutagénesis de sitio dirigido, puntual aleatoria, recombinación homóloga, transposición de ADN u otros métodos recursivos de mutagénesis, construcción química, mutagénesis usando plantillas que contienen uracilo, mutagénesis dirigida por oligonucleótido,

mutagénesis de ADN modificado por fosforotioato, mutagénesis utilizando ADN de doble ranura o similares, o cualquier combinación de las mismas. Los métodos adicionales adecuados incluyen reparación de emparejamiento erróneo puntual, mutagénesis usando cepas hospedadoras de reparación defectuosa, selección de restricción y purificación de restricción, mutagénesis por eliminación, mutagénesis por síntesis génica total, reparación de ruptura bicatenaria, y similares. La mutagénesis, por ejemplo, que implica construcciones químicas, también se incluyen en la presente divulgación. En una realización, la mutagénesis puede guiarse mediante la información conocida de la molécula de origen natural o de la molécula alterada o mutada de origen natural, por ejemplo, secuencia, comparaciones de secuencia, propiedades físicas, estructura cristalina o similares.

Los textos y ejemplos anteriores encontrados en el presente documento describen estos procedimientos. Se encuentra información adicional en las siguientes publicaciones y referencias citadas en: Ling et al., Approaches to DNA mutagenesis: an overview, *Anal Biochem.* 254(2): 157-178 (1997); Dale et al., Oligonucleotide-directed random mutagenesis using the phosphorothioate method, *Methods Mol. Biol.* 57:369-374 (1996); Smith, *In vitro* mutagenesis, *Ann. Rev. Genet.* 19:423-462(1985); Botstein y Shortle, Strategies and applications of *in vitro* mutagenesis, *Science* 229:1193-1201(1985); Carter, Site-directed mutagenesis, *Biochem. J.* 237:1-7 (1986); Kunkel, The efficiency of oligonucleotide directed mutagenesis, en *Nucleic Acids & Molecular Biology* (Eckstein, F. y Lilley, D.M.J. eds., Springer Verlag, Berlín) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. USA* 82:488-492 (1985); Kunkel et al., Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymol.* 154, 367-382 (1987); Bass et al., Mutant Trp repressors with new DNA-binding specificities, *Science* 242:240-245 (1988); *Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Zoller y Smith, Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, *Nucleic Acids Res.* 10:6487-6500 (1982); Zoller y Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Methods in Enzymol.* 100:468-500 (1983); Zoller y Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, *Methods in Enzymol.* 154:329-350 (1987); Taylor et al., The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor et al., The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye y Eckstein, Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 14: 9679-9698 (1986); Sayers et al., Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988); Sayers et al., Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide, (1988) *Nucl. Acids Res.* 16: 803-814; Kramer et al., The gapped duplex DNA approach to oligonucleotide-directed mutation construction, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer & Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, *Methods in Enzymol.* 154:350-367 (1987); Kramer et al., Improved enzymatic *in vitro* reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, *Nucl. Acids Res.* 16: 7207 (1988); Fritz et al., Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions *in vitro*, *Nucl. Acids Res.* 16: 6987-6999 (1988); Kramer et al., Point Mismatch Repair, *Cell* 38:879-887 (1984); Carter et al., Improved oligonucleotide site-directed mutagenesis using M13 vectors, *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter, Improved oligonucleotide-directed mutagenesis using M13 vectors, *Methods in Enzymol.* 154: 382-403 (1987); Eghtedarzadeh y Henikoff, Use of oligonucleotides to generate large deletions, *Nucl. Acids Res.* 14: 5115 (1986); Wells et al., Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986); Nambiar et al., Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223: 1299-1301 (1984); Sakamar y Khorana, Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells et al., Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985); Grundström et al., Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, *Nucl. Acids Res.* 13: 3305-3316 (1985); Mandecki, Oligonucleotide-directed doublestrand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181 (1986); Arnold, Protein engineering for unusual environments, *Current Opinion in Biotechnology* 4:450-455 (1993); Sieber, et al., *Nature Biotechnology*, 19:456-460 (2001). W. P. C. Stemmer, *Nature* 370, 389-91 (1994); y, I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995). Pueden encontrarse detalles adicionales sobre muchos de los métodos anteriores en *Methods in Enzymology* Volumen 154, que también describe controles útiles para resolver problemas con diversos métodos de mutagénesis.

La invención también se refiere a células huésped eucariotas y organismos para la incorporación *in vivo* de un aminoácido artificial a través de pares de ARNt/O-RS ortogonales. Las células huésped están diseñadas genéticamente (por ejemplo, transformadas, transducidas o transfectadas) con los polinucleótidos de la presente divulgación o construcciones que incluyen un polinucleótido de la presente divulgación, por ejemplo, un vector de la presente divulgación, que puede ser, por ejemplo, un vector de clonación o un vector de expresión. El vector puede estar, por ejemplo, en forma de un plásmido, una bacteria, un virus, un polinucleótido desnudo, o un polinucleótido conjugado. Los vectores se introducen en células y/o microorganismos mediante métodos estándar incluyendo electroporación (From et al., *Proc. Natl. Acad. Sci. USA* 82, 5824 (1985), infección por vectores virales, penetración balística a gran velocidad con partículas pequeñas con el ácido nucleico en la matriz de pequeñas perlas o partículas, o sobre la superficie (Klein et al., *Nature* 327, 70-73 (1987)).

Las células huésped diseñadas pueden cultivarse en un medio nutritivo convencional modificado según sea apropiado para dichas actividades como, por ejemplo, etapas de cribado, activación de promotores o selección de transformantes. Estas células pueden cultivarse opcionalmente en organismos transgénicos. Otras referencias útiles, por ejemplo, para el aislamiento y cultivo celular (por ejemplo, para el posterior aislamiento de ácidos nucleicos) incluyen Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, Nueva York y las referencias citadas en el mismo; Payne et al. (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. Nueva York, NY; Gamborg y Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg Nueva York) y Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Ratón, FL.

Se encuentran disponibles varios métodos ya conocidos para introducir ácidos nucleicos diana en células, cualquiera de los cuales se puede usar. Estos incluyen: fusión de las células receptoras con protoplastos bacterianos que contienen el ADN, electroporación, bombardeo de proyectiles, e infección con vectores virales (analizado más en detalle, a continuación), etc. Las células bacterianas pueden utilizarse para amplificar el número de plásmidos que contienen construcciones de ADN de esta divulgación. Las bacterias crecen hasta la fase log y los plásmidos dentro de las bacterias pueden aislarse mediante diversos métodos conocidos en la técnica (véase, por ejemplo, Sambrook). Además, una multitud de kits se encuentran comercialmente disponibles para la purificación de plásmidos de bacterias (véanse, por ejemplo, EasyPrep™, FlexiPrep™, ambos de Pharmacia Biotech; StrataClean™, de Stratagene; y, QIAprep™ de Qiagen). Los plásmidos aislados y purificados se manipulan a continuación adicionalmente para producir otros plásmidos, se utilizan para transfectar células o se incorporan en vectores relacionados para infectar organismos. Los vectores típicos que contienen terminadores de la transcripción y la traducción, y secuencias de inicio de la transcripción y la traducción, y promotores útiles para la regulación de la expresión del ácido nucleico diana particular. Los vectores comprenden opcionalmente casetes de expresión genérica que contienen al menos una secuencia de terminación independiente, secuencias que permiten la replicación del casete en eucariotas, o procariotas, o ambos (por ejemplo, vectores lanzadera) y marcadores de selección para sistemas tanto procariotas como eucariotas. Los vectores son adecuados para la replicación e integración en procariotas, eucariotas, o preferentemente ambos. Véanse, Gilman y Smith, *Gene* 8:81 (1979); Roberts, et al., *Nature*, 328:731 (1987); Schneider, B., et al., *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (todos anteriormente). Se proporciona un catálogo de bacterias y bacteriófagos útiles para clonación, por ejemplo, por la ATCC, por ejemplo, *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna et al. (eds) publicado por la ATCC. También se encuentran procedimientos básicos adicionales para secuenciación, clonación y otros aspectos de biología molecular y consideraciones teóricas subyacentes en Watson et al. (1992) *Recombinant DNA Segunda edición Scientific American Books*, NY. Además, esencialmente cualquier ácido nucleico (y prácticamente cualquier ácido nucleico etiquetado, ya sea estándar o no estándar) puede ordenarse adaptado o estándar de diversas fuentes comerciales, tal como la Midland Certified Reagent Company (Midland, TX mcr.com), The Great American Gene Company (Ramona, CA disponible en la red informática mundial en genco.com), ExpressGen Inc. (Chicago, IL disponible en la red informática mundial en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchos otros.

Ejemplos

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

EJEMPLO 1: MÉTODOS PARA PRODUCIR COMPOSICIONES DE AMINOACIL-ARNt SINTETASAS QUE INCORPORAN AMINOÁCIDOS ARTIFICIALES EN CÉLULAS EUCARIÓTICAS

La expansión del código genético eucariota para incluir aminoácidos artificiales con nuevas propiedades físicas, químicas o biológicas proporcionará potentes herramientas para analizar y controlar la función de las proteínas en estas células. Hacia esta meta, se describe un enfoque general para el aislamiento de aminoacil-ARNt sintetasas que incorporan aminoácidos artificiales con alta fidelidad en proteínas en respuesta a un codón ámbar en *Saccharomyces cerevisiae* (*S. cerevisiae*). El método se basa en la activación de los genes indicadores que responden a GAL4, HIS3, URA3 o LacZ, mediante la supresión de los codones ámbar entre el dominio de unión a ADN y el dominio de activación de transcripción de GAL4. Se describe la optimización de un indicador GAL4 para la selección positiva de variantes de tirosil-ARNt sintetasa de *Escherichia coli* (EcTyrRS). También se ha desarrollado una selección negativa de las variantes de EcTyrRS inactivas con el indicador URA3 mediante el uso de una molécula pequeña (ácido 5-fluorótico (5-FOA)) añadido al medio de crecimiento como un "alelo tóxico". De manera importante, las selecciones tanto positiva como negativa pueden realizarse en una sola célula y con un rango de rigurosidad. Esto puede facilitar el aislamiento de un rango de actividades de la aminoacil-ARNt sintetasa (aaRS) procedente de grandes bibliotecas de sintetasas mutantes. La potencia del método para aislar los fenotipos aaRS deseados se demuestra mediante selecciones de modelo.

La reciente adición de aminoácidos artificiales al código genético de *Escherichia coli* (*E. coli*) proporciona un nuevo procedimiento potente para analizar y manipular la estructura de la proteína y funcionar tanto *in vitro* como *in vivo*. Los aminoácidos con etiquetas de fotoafinidad, átomos pesados, grupos cetó y olefinicos y cromóforos se han incorporado en proteínas en *E. coli* con una eficiencia y fidelidad que rivalizan con la de los veinte aminoácidos

comunes. Véanse, por ejemplo, Chin, et al., (2002), Addition of a Photocrosslinker to the Genetic Code of *Escherichia coli*, Proc. Natl. Acad. Sci. U.S. A. 99:11020-11024; Chin y Schultz, (2002), *In vivo* Photocrosslinking with Unnatural Amino Acid Mutagenesis, Chem BioChem 11:1135-1137; Chin et al., (2002), Addition of p-Azido-L-phenylalanine to the Genetic code of *Escherichia coli*, J. Am. Chem. Soc. 124:9026-9027; Zhang et al., (2002), The selective incorporation of alkenes into proteins in *Escherichia coli*, Angewandte Chemie. International Ed. en Inglés 41:2840-2842; y, Wang y Schultz, (2002), Expanding the Genetic Code, Chem. Comm. 1-10.

Los aminoácidos artificiales se han introducido previamente en el receptor de acetilcolina nicotínico en ovocitos xenopus (por ejemplo, M.W. Nowak, et al. (1998), *In vivo* incorporation of unnatural amino acids into ion channels in *Xenopus* oocyte expression system, Method Enzymol. 293:504-529) mediante microinyección de un ARNt de *Tetrahymena thermophila* químicamente mal acilado (por ejemplo, M.E. Saks, et al. (1996), An engineered *Tetrahymena* tRNA^{Gln} for *in vivo* incorporation of unnatural amino acids into proteins by nonsense suppression, J. Biol. Chem. 271:23169-23175, y el ARNm relevante. Esto ha permitido estudios biofísicos detallados del receptor en ovocitos mediante la introducción de aminoácidos que contienen cadenas laterales con propiedades físicas o químicas únicas. Véase, por ejemplo, D.A. Dougherty (2000), Unnatural amino acids as probes of protein structure and function, Curr. Opin. Chem. Biol. 4:645-652. Desafortunadamente, esta metodología se limita a proteínas en células que pueden microinyectarse, y debido a que el ARNt se aminoacila químicamente *in vitro*, y no puede reaclarse, el rendimiento de la proteína es muy bajo. Esto a su vez requiere técnicas sensibles para ensayar la función de la proteína.

Hay interés en la incorporación genética de aminoácidos artificiales a proteínas en células eucariotas en respuesta a un codón ámbar. Véanse también, H.J. Drabkin et al., (1996), Amber suppression in mammalian cells dependent upon expression of an *Escherichia coli* aminoacyl-tRNA synthetase gene, Molecular & Cellular Biology 16:907-913; A.K. Kowal, et al., (2001), Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria.[comment], Proc. Natl. Acad. Sci. U.S. A. 98:2268-2273; y, K. Sakamoto, et al., (2002), Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, Nucleic Acids Res. 30:4692-4699. Esto tendría ventajas técnicas y prácticas significativas, dado que los ARNt se reaclarían mediante sus sintetasas análogas lo que conduce a grandes cantidades de proteína mutante. Además, las aminoacil-ARNt sintetasas y los ARNt genéticamente codificados son, en principio, heredables, permitiendo que el aminoácido artificial se incorpore en proteínas a través de muchas divisiones celulares sin dilución exponencial.

Se han descrito las etapas necesarias para añadir nuevos aminoácidos al código genético de *E. coli* (véanse, por ejemplo, D.R. Liu, y P.G. Schultz, (1999), Progress toward the evolution of an organism with an expanded genetic code, Proc. Natl. Acad. Sci. U.S. A. 96:4780-4785; y pueden utilizarse principios similares para extender el código genético de eucariotas. En la primera etapa, se identifica el par ortogonal de aminoacil-ARNt sintetasa (aaRS)/ARNt_{CUA}. Este par necesita funcionar con la maquinaria de traducción de las células huésped, pero la aaRS no debe cargar ningún ARNt endógeno con un aminoácido y el ARNt_{CUA} no debe aminoacilarse mediante sintetasas endógenas. Véase, por ejemplo, D.R. Liu, et al., Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins *in vivo*, Proc. Natl. Acad. Sci. U.S. A. 94:10092-10097. En una segunda etapa, los pares de aaRS/ARNt que son capaces de utilizar solo el aminoácido artificial se seleccionan de una biblioteca de aaRS mutantes. En *E. coli*, la selección de los aminoácidos artificiales que utilizan variantes de MjTyrRS se realizó utilizando selecciones de "doble criba" en dos etapas. Véase, por ejemplo, D.R. Liu, y P.G. Schultz, (1999), Progress toward the evolution of an organism with an expanded genetic code, Proc. Natl. Acad. Sci. U.S. A. 96:4780-4785. Se utiliza un método de selección modificado en células eucariotas.

Se seleccionó *Saccharomyces cerevisiae* (*S. cerevisiae*) como el organismo huésped eucariota, dado que es unicelular, tiene un tiempo rápido de generación, así como una genética relativamente bien caracterizada. Véase, por ejemplo, D. Burke, et al., (2000) Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. Además, dado que la maquinaria de traducción de los eucariotas está altamente conservada (véanse, por ejemplo, (1996) Translational Control. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; Y. Kwok, y J.T. Wong, (1980), Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes, Canadian Journal of Biochemistry 58:213-218; y, (2001) The Ribosome. Cold Spring harbor Laboratory Press, Cold Spring Harbor, NY), es probable que los genes aaRS para la incorporación de aminoácidos artificiales descubiertos en *S. cerevisiae* pueden introducirse en organismos eucariotas mayores y utilizarse, en sociedad con los ARNt análogos (véanse, por ejemplo, K. Sakamoto, et al., (2002) Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, Nucleic Acids Res. 30:4692-4699; y, C. Kohrer, et al., (2001), Import of amber and ochre suppressor tRNAs into mammalian cells: a general approach to site-specific insertion of amino acid analogues into proteins, Proc. Natl. Acad. Sci. U.S. A. 98:14310-14315) para incorporar aminoácidos artificiales. Por lo tanto, la expansión del código genético de *S. cerevisiae* es una puerta hacia la expansión del código genético de organismos eucariotas multicelulares complejos. Véanse, por ejemplo, M. Buvoli, et al., (2000), Suppression of nonsense mutations in cell culture and mice by multimerized suppressor tRNA genes, Molecular & Cellular Biology 20:3116-3124. El par tirosil derivado de TyrRS de *Methanococcus jannaschii* (MjTyrRS)/ARNt (véanse, por ejemplo, L. Wang, y P.G. Schultz, (2002), Expanding the Genetic Code, Chem. Comm. 1-10) que se utilizó previamente para expandir el código genético de *E. coli* no es ortogonal en organismos eucariotas (por ejemplo, P. Fechter, et al., (2001), Major tyrosine identity determinants in *Methanococcus jannaschii*

and *Saccharomyces cerevisiae* tRNA(Tyr) are conserved but expressed differently, *Eur. J. Biochem.* 268:761-767) y se requiere un nuevo par ortogonal para expandir el código genético eucariota. Schimmel y colaboradores han mostrado que el par de tirosil-ARNt sintetasa de *E. Coli* (EcTyrRS)/ARNt_{CUA} suprime los codones ámbar en *S. cerevisiae*, y que el ARNt_{CUA} de *E. coli* no se carga por aminoacil-ARNt sintetetas endógenas en el citosol de levadura (**figura 2**). Véanse también, por ejemplo, H. Edwards, et al., (1991), An *Escherichia coli* tyrosine transfer RNA is a leucine-specific transfer RNA in the yeast *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S. A.* 88:1153-1156; y, H. Edwards, y P. Schimmel (1990), A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Molecular & Cellular Biology* 10:1633-1641. Además, la EcTyrRS ha mostrado no cargar ARNt de levadura *in vitro*. Véanse, por ejemplo, Y. Kwok, y J.T. Wong, (1980), Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes, *Canadian Journal of Biochemistry* 58:213-218; B.P. Doctor, et al., (1966), Studies on the species specificity of yeast and *E. coli* tyrosine tRNAs, *Cold Spring Harbor Symp. Quant. Biol.* 31:543-548; y, K. Wakasugi, et al., (1998), Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant, *EMBO Journal* 17:297-305. Por lo tanto, el par de EcTyrRS/ARNt_{CUA} es un candidato para un par ortogonal en *S. cerevisiae*, así como en eucariotas mayores (por ejemplo, A.K. Kowal, et al., (2001), Twenty-first aminoacyl-tRNA synthetase-suppressor tRNA pairs for possible use in site-specific incorporation of amino acid analogues into proteins in eukaryotes and in eubacteria.[comment], *Proc. Natl. Acad. Sci. U.S. A.* 98 (2001) 2268-2273).

Para ampliar la especificidad del sustrato de EcTyrRS en *E. coli*, Nishimura y colaboradores cribaron una biblioteca generada de PCR propensa a error de mutantes de EcTyrRS y descubrieron un mutante con una capacidad mejorada para incorporar 3-azatirosina. Véase, por ejemplo, F. Hamano-Takaku, et al., (2000), A mutant *Escherichia coli* tyrosyl-tRNA synthetase utilizes the unnatural amino acid azatyrosine more efficiently than tyrosine, *J. Biol. Chem.* 275:40324-40328. Sin embargo, este aminoácido se incorpora en todo el proteoma de *E. coli*, y la enzima desarrollada aún prefiere tirosina como sustrato. Yokoyama y colaboradores cribaron una pequeña colección de variantes diseñadas de actiactivas en el sitio de EcTyrRS en un sistema de traducción en germen de trigo y descubrieron una variante de EcTyrRS que utiliza 3-yodotirosina más eficientemente que tirosina. Véase, D. Kiga, et al., (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cellfree system, *Proc. Natl. Acad. Sci. U.S. A.* 99:9715-9720. En contraste con las enzimas se han desarrollado en *E. coli* (por ejemplo, J.W. Chin, et al., (2002), Addition of a Photocrosslinker to the Genetic Code of *Escherichia coli*, *Proc. Natl. Acad. Sci. U.S. A.* 99:11020-11024; J.W. Chin, et al., (2002), Addition of p-Azido-L-phenylalanine to the Genetic code of *Escherichia coli*, *J. Am. Chem. Soc.* 124:9026-9027; L. Wang, et al., (2001), Expanding the genetic code of *Escherichia coli*, *Science* 292:498-500; y, L. Wang, et al., (2002), Adding L-3-(2-naphthyl)alanine to the genetic code of *E. coli*, *J. Am. Chem. Soc.* 124:1836-1837), this enzyme still incorporates tyrosine in the absence of the unnatural amino acid. Véase, por ejemplo, D. Kiga, et al., (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cellfree system, *Proc. Natl. Acad. Sci. U.S. A.* 99:9715-9720. Recientemente, Yokoyama y colaboradores han demostrado también que este mutante de EcTyrRS funciona con un ARNt_{CUA} de *Bacillus stearothermophilus* para suprimir los codones ámbar in células de mamífero. Véase, K. Sakamoto, et al., (2002), Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells, *Nucleic Acids Res.* 30:4692-4699.

Un requisito es que el aminoácido añadido al código genético eucariota se incorpore con una fidelidad similar a la de los veinte aminoácidos comunes. Para lograr esta meta, se ha utilizado un método de selección *in vivo* general para el descubrimiento de variantes de EcTyrRS/ARNt_{CUA} que funcionan en *S. cerevisiae* para incorporar aminoácidos artificiales, pero ninguno de los aminoácidos comunes en respuesta al codón ámbar TAG. Una ventaja principal de una selección es que las enzimas que incorporan selectivamente aminoácidos artificiales pueden seleccionarse rápidamente y enriquecerse de bibliotecas de 10⁸ variantes del sitio activo de EcTyrRS, 6-7 órdenes de magnitud más diversos que los cribados *in vivo*. Véase, por ejemplo, D. Kiga, et al., (2002), An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into proteins in eukaryotic translation and its application in a wheat germ cellfree system, *Proc. Natl. Acad. Sci. U.S. A.* 99:9715-9720. Este aumento en la diversidad aumenta vastamente la probabilidad de aislar las variantes de EcTyrRS para la incorporación de un diverso rango de funcionalidad útil con muy alta fidelidad. Véase, por ejemplo, L. Wang, y P.G. Schultz, (2002), Expanding the Genetic Code, *Chem. Comm.* 1-10.

Para extender el enfoque de selección a *S. cerevisiae*, se utilizó la proteína activadora de transcripción, GAL4 (véase la **figura 1**). Véase, por ejemplo, A. Laughon, et al., (1984), Identification of two proteins encoded by the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:268-275; A. Laughon, y R.F. Gesteland, (1984), Primary structure of the *Saccharomyces cerevisiae* GAL4 gene, *Molecular & Cellular Biology* 4:260-267; L. Keegan, et al., (1986), Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein, *Science* 231:699-704; y, M. Ptashne, (1988), How eukaryotic transcriptional activators work, *Nature* 335:683-689. Los 147 aminoácidos N-terminales de esta proteína de 881 aminoácidos forman un dominio de unión a ADN (DBD) que se une a una secuencia de ADN específicamente. Véase, por ejemplo, M. Carey, et al., (1989), An amino-terminal fragment of GAL4 binds DNA as a dimer, *J. Mol. Biol.* 209:423-432; y, E. Giniger, et al., (1985), Specific DNA binding of GAL4, a positive regulatory protein of yeast, *Cell* 40:767-774. El DBD está unido, por medio de una secuencia de proteína intermedia, a un dominio de activación de 113 aminoácidos C-terminales (AD) que puede

activar la transcripción cuando se une al ADN. Véase, por ejemplo, J. Ma, y M. Ptashne, (1987), Deletion analysis of GAL4 defines two transcriptional activating segments, *Cell* 48:847-853; y, J. Ma, y M. Ptashne, (1987), The carboxy-terminal 30 amino acids of GAL4 are recognized by GAL80, *Cell* 50:137-142. Se consideró que colocando codones ámbar hacia el DBD N-terminal de un solo polipéptido que contenía tanto el DBD de GAL4 y su AD C-terminal, la supresión ámbar por medio del par EcTyrRS/ARNt_{CUA} puede unirse a la activación de la transcripción por GAL4 (**figura 1, Panel A**). Mediante la selección de los genes indicadores activados por GAL4 apropiados pueden realizarse selecciones tanto positivas como negativas con el gen (**figura 1, Panel B**). Aunque pueden utilizarse muchos genes indicadores basándose en la complementación de la auxotrofia aminoacídica de una célula para las selecciones positivas (por ejemplo, URA3, LEU2, HIS3, LYS2), el gen HIS3 es un atractivo gen indicador, dado que la actividad de la proteína que codifica (imidazol glicerol fosfato deshidratasa) puede modularse en una dosis de manera dependiente mediante adición de 3-aminotriazol (3-AT). Véase, por ejemplo, G.M. Kishore, y D.M. Shah, (1988), Amino acid biosynthesis inhibitors as herbicides, *Annual Review of Biochemistry* 57:627-663. En *S. cerevisiae*, se han usado menos genes para las selecciones negativas. Una de varias estrategias de selección negativa (véanse, por ejemplo, A. J. DeMaggio, et al., (2000), The yeast split-hybrid system, *Method Enzymol.* 328:128-137; H.M. Shih, et al., (1996), A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP, *Proc. Natl. Acad. Sci. U.S. A.* 93:13896-13901; M. Vidal, et al., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system.[comentario], *Proc. Natl. Acad. Sci. U.S. A.* 93:10321-10326; y, M. Vidal, et al., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. [comentario], *Proc. Natl. Acad. Sci. U.S. A.* 93:10315-10320) que se ha usado con éxito es la selección negativa de URA3/ácido 5-fluororótico (5-FOA) (por ejemplo, J.D. Boeke, et al., (1984), A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoroorotic acid resistance, *Molecular & General Genetics* 197:345-346), descrito en el sistema de "dos híbridos inversos" desarrollado por Vidal y colaboradores. Véase, M. Vidal, et al., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system.[comentario], *Proc. Natl. Acad. Sci. U.S. A.* 93:10321-10326; y, M. Vidal, et al., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. [comentario], *Proc. Natl. Acad. Sci. U.S. A.* 93:10315-10320). En el sistema de dos híbridos inversos, se coloca un indicador URA3 genómicamente integrado bajo un promotor estrechamente controlado que contiene sitios de unión a ADN de GAL4. Cuando se producen dos proteínas que interactúan como fusiones al GAL4 DBD y GAL4 AD, reconstituyen la actividad de GAL4 y activan la transcripción de URA3. En presencia de 5-FOA, el producto génico de URA3 convierte el 5-FOA en un producto tóxico, destruyendo la célula. Véase, J.D. Boeke, et al., anteriormente. Esta selección se ha utilizado para seleccionar proteínas que alteran una interacción de proteína-proteína y para mutaciones que alteran una interacción de proteína-proteína. Se ha descrito una variante para cribar inhibidores de molécula pequeña de interacción de proteína-proteína. Véase, por ejemplo, J. Huang, y S.L. Schreiber, (1997) A yeast genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets, *Proc. Natl. Acad. Sci. U.S. A.* 94:13396-13401.

La elección apropiada de los codones ámbar en GAL4 de longitud completa permite selecciones positivas eficientes para variantes de EcTyrRS activas utilizando HIS3 o los indicadores activados URA3 GAL4 para complementar la auxotrofia de histidina o uracilo en células de levadura. Además, el indicador URA3 puede utilizarse en las selecciones negativas para variantes de EcTyrRS inactivas en presencia de 5-FOA. Además, pueden utilizarse análisis colorimétricos utilizando lacZ para leer la actividad de aminoacil-ARNt sintetasa en células de levadura.

RESULTADOS Y ANÁLISIS

El gen EcTyrRS se expresó bajo el control del promotor ADH1 constitutivo, y el gen de ARNt_{CUA} se expresó a partir del mismo plásmido de levadura de alto número de copias (pEcTyrRS_{tRNA_{CUA}}, **figura 1, Panel C**). Tras la transformación conjunta de pEcTyrRS_{tRNA_{CUA}} y un indicador de bajo número de copias que contiene una sola mutación ámbar entre el dominio de unión a ADN y el dominio de activación de una construcción quimérica de GAL4 en MaV203, las células crecieron en el medio selectivo careciendo de histidina y conteniendo 3-AT 10-20 mM (**figura 2**). Cuando las células MaV203 se transformaron con la misma construcción GAL4 y un mutante de sintetasa inactiva (A5) o una construcción que carece del gen EcARNt, no se observó ningún crecimiento en 3-AT 10 mM (**figura 2**). Estos experimentos establecen que EcTyrRS puede expresarse constitutivamente en una forma funcional del promotor ADH1, que existe una mínima supresión endógena ámbar en MaV203, y que existe poca carga de EcARNt_{CUA} por sintetasas de levadura en este sistema. Véanse, por ejemplo, H. Edwards, et al., (1991), An *Escherichia coli* tyrosine transfer RNA is a leucine-specific transfer RNA in the yeast *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S. A.* 88:1153-1156; y, H. Edwards, y P. Schimmel, (1990), A bacterial amber suppressor in *Saccharomyces cerevisiae* is selectively recognized by a bacterial aminoacyl-tRNA synthetase, *Molecular & Cellular Biology* 10:1633-1641. Dado que EcTyrRS no carga ARNt de *S. cerevisiae* (por ejemplo, Y. Kwok, y J.T. Wong, (1980), Evolutionary relationship between *Halobacterium cutirubrum* and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes, *Canadian Journal of Biochemistry* 58:213-218; B.P. Doctor, et al., (1966), Studies on the species specificity of yeast and *E. coli* tyrosine tRNAs, *Cold Spring Harbor Symp. Quant. Biol.* 31:543-548; y, K. Wakasugi, et al., (1998), Genetic code in evolution: switching species-specific aminoacylation with a peptide transplant, *EMBO Journal* 17:297-305), estos experimentos confirman que EcTyrRS/Ec_{tRNA_{CUA}} son un par ortogonal en *S. cerevisiae*.

Mientras que la quimera GAL4 de primera generación fue capaz de activar la transcripción del indicador HIS3 débil, fue incapaz de activar la transcripción del indicador URA3 en MaV203 suficientemente para permitir un crecimiento significativo en concentraciones de 3-AT mayores que 20 mM, o en placas -URA (**figura 2**). Para los propósitos de selección de variantes de EcTyrRS, se realizó una segunda generación de construcción GAL4. Este indicador GAL4 se diseñó para ser más activo, para tener un rango dinámico mayor, y para evitar la acumulación de revertientes. Para aumentar la actividad de los indicadores GAL4, se utilizó GAL4 de longitud completa (que tiene una actividad de activación de transcripción de dos veces la de una fusión DBD-AD) (véase, por ejemplo, J. Ma, y M. Ptashne, (1987), Deletion analysis of GAL4 defines two transcriptional activating segments, Cell 48:847-853) bajo el control de un fuerte promotor ADH1, y se utilizó un plásmido de 2 μ m de alto número de copias (con un número de copias de 10-30 veces el del plásmido centromérico de la quimera GAL4 inicial). Un aumento tanto en el número de copias del plásmido como en la actividad de la proteína que codifica debe extender el rango dinámico de los indicadores. Las mutaciones ámbar se dirigieron a la región del gen GAL4 que codifica los residuos de aminoácido 2 y 147 (**figura 3**). Esta región es suficiente para la unión al ADN específica de secuencia (véanse, por ejemplo, M. Carey, et al., (1989), An amino-terminal fragment of GAL4 binds DNA as a dimer, J. Mol. Biol. 209:423-432), y reside en el lado 5' del primer dominio crítico de activación en el gen GAL4 (véase, por ejemplo, J. Ma, y M. Ptashne, (1987) Deletion analysis of GAL4 defines two transcriptional activating segments, Cell 48:847-853), de tal forma que los productos truncados producidos en ausencia de la supresión ámbar no se anticipan para activar la transcripción. La elección de los codones de aminoácido que van a mutarse se guió mediante selecciones de mutagénesis de saturación anteriores en GAL4 (véase, por ejemplo, M. Johnston, y J. Dover, (1988), Mutational analysis of the GAL4-encoded transcriptional activator protein of *Saccharomyces cerevisiae*, Genetics 120:63-74), así como las estructuras de rayos X del dominio de unión a ADN N-terminal de GAL4 (véase, por ejemplo, R. Marmorstein, et al., (1992), DNA recognition by GAL4: structure of a protein-DNA complex.[comentario], Nature 356:408-414; y, J.D. Baleja, et al., (1992), Solution structure of the DNA-binding domain of Cd2-GAL4 from *S. cerevisiae*.[comentario], Nature 356:450-453) y la estructura de RMN de su región de dimerización. Véase, por ejemplo, P. Hidalgo, et al., (2001), Recruitment of the transcriptional machinery through GAL11P: structure and interactions of the GAL4 dimerization domain, Genes & Development 15:1007-1020.

El GAL4 de longitud completa se clonó en un pequeño vector basado en pUC para permitir la rápida construcción de 10 mutantes ámbar únicos (en los codones para los aminoácidos L3, I13, T44, F68, R110, V114, T121, I127, S131, T145) mediante mutagénesis de sitio dirigido. GAL4 y los mutantes ámbar resultantes se subclonaron a continuación en un vector de levadura de 2 μ m bajo el control del promotor ADH1 de longitud completa para crear pGADGAL4 y una serie de mutantes ámbar denominada pGADGAL4 (xxTAG) (**figura 1, Panel C**), donde xx representa el codón de aminoácido en el gen GAL4 que se mutó por el codón ámbar. Cada mutante GAL4 se cotransformó con EcTyrRS/ARNt_{CUA} o A5/ARNt_{CUA} en células MaV203, convirtiendo los transformadores en protrofia de leucina y triptofano. pGADGAL4 en sí transformado con muy baja eficiencia (<10-3 veces la de los mutantes ámbar GAL4) y es presumiblemente dañino para las células MaV203 a tan alto número de copias; no se observó dicho efecto con los mutantes ámbar de GAL4.

Los fenotipos de indicadores GAL4, en presencia de una sintetasa activa o muerta, se ensayaron en placas de -URA, y placas de 5-FOA al 0,1 % (**figura 3, Panel A**). Cinco mutantes GAL4 (L3TAG, I13TAG, T44TAG, F68TAG, S131TAG) crecieron en placas de -URA y no crecieron en 5-FOA al 0,1 % en presencia de EcTyrRS ya sea de tipo silvestre o inactivo. En estos mutantes ámbar, la supresión endógena es aparentemente suficiente para impulsar la supresión mediada por EcTyrRS/ARNt_{CUA} más allá del rango dinámico del indicador URA3 en MaV203. Cinco mutantes ámbar GAL4 únicos (R110TAG, V114TAG, T121TAG, I127TAG, T145TAG) crecieron en ausencia de uracilo y en presencia de EcTyrRS/ARNt_{CUA}, (pero no de A5/ARNt_{CUA}) y mostraron el fenotipo inverso en 5-FOA. Estos mutantes muestran fenotipos dependientes de EcTyrRS que están dentro del rango dinámico del indicador URA3 en MaV203. El fenotipo dependiente de EcTyrRS más limpio tanto en -URA como en 5-FOA al 0,1 % se observó con el mutante R110 TAG de GAL4. Sin embargo, este mutante mostró algo de color azul en ensayos de X-GAL al cotransformarse con A5. Para mejorar adicionalmente el rango dinámico, se preparó una serie de seis mutantes ámbar dobles de GAL4 que contenían R110TAG (**figura 3, Panel B**), (L3TAG, R110TAG; I13TAG, R110TAG; T44TAG, R110TAG; R110TAG, T121TAG; R110TAG, I127TAG; R110TAG, T145TAG). Cuatro de estos mutantes dobles (I13TAG, R110TAG; R110TAG, T121TAG; R110TAG, I127TAG y T145TAG, R110TAG) fueron incapaces de crecer en ausencia de uracilo y crecieron en 5-FOA al 0,1 %. Estos mutantes dobles tienen actividades fuera (por debajo) del rango dinámico de los ensayos de placa. Dos de los mutantes dobles (L3TAG, R110TAG y T44TAG, R110TAG) crecieron en presencia EcTyrRS/ARNt_{CUA} de tipo silvestre, pero no con A5/ARNt_{CUA} en placas de -URA; estos mutantes también mostraron los fenotipos recíprocos esperados en 5-FOA. PGADGAL4 (T44TAG, R110TAG), el más activo de estos dos mutantes GAL4, se seleccionó para una caracterización más detallada (**figura 4**). Las MaV203 que contenían pGADGAL4(T44TAG, R110TAG)/pEcTyrRS-ARNt_{CUA} eran azules en X-GAL, pero la cepa correspondiente que contenía pA5/ARNt_{CUA} no. De manera similar, las MaV203 que contenían pGADGAL4(T44TAG, R110TAG)/pEcTyrRS/ARNt_{CUA} crecieron fuertes en placas con concentraciones de 3-AT de hasta 75 mM, y en placas -URA, pero la cepa correspondiente que contenía pA5/ARNt_{CUA} no creció en 3AT 10 mM o en ausencia de uracilo. Tomados conjuntamente, los fenotipos dependientes de EcTyrRS de pGADGAL4(T44TAG, R110TAG) pueden extender el rango dinámico de los indicadores URA3, HIS3 y lacZ en las MaV203.

Fue de interés determinar la actividad de mutantes GAL4 en los que T44 o R110 se sustituyeron con aminoácidos

diferentes de tirosina, dado que es posible que la capacidad para sustituir aminoácidos variados sin alterar la actividad de GAL4 sea útil para la selección de aminoacil-ARNt sintetasas mutantes que puedan incorporar aminoácidos sintéticos en proteínas. Véase, por ejemplo, M. Pasternak, et al., (2000), A new orthogonal suppressor tRNA/aminoacyl-tRNA synthetase pair for evolving an organism with an expanded genetic code, *Helvetica Chimica Acta* 83:2277. Una serie de cinco mutantes del residuo T44 en GAL4, (T44Y, T44W, T44F, T44D, T44K) se construyeron en pGADGAL4 (R110TAG), ya que pGADGAL4 en sí es tóxico. Se construyó una serie similar de mutantes en la posición R110 en GAL4, (R110Y, R110W, R110F, R110D, R110K) en pGADGAL4(T44TAG). Estos mutantes se desvían hacia las grandes cadenas laterales hidrófobas de aminoácidos de interés para incorporar en proteínas, pero también contienen un residuo cargado positiva y negativamente como prueba rigurosa de permisividad. Cada mutante se cotransformó con pEcTyrRS/ARNt_{CUA} en células MaV203 y los aislados de leu⁺ y trp⁺ se ensayaron para determinar la producción de lacZ mediante la hidrólisis de orto-nitrofenil-β-D-galactopiranosido (ONPG) (**figura 5**). La variación en la actividad entre las células que contienen GAL4 con diferentes aminoácidos sustituidos por T44 o R110 fue menor de 3 veces en todos los casos. Esta mínima variabilidad demuestra la permisividad de estos sitios para una sustitución aminoacídica sin alterar la actividad de transcripción de GAL4. Como se esperaba de la actividad de los mutantes ámbros únicos ensayados en placas selectivas, los mutantes de T44 preparados en la base GAL4 (R110TAG) conducen a una hidrólisis más lenta de ONPG que los mutantes de R110 preparados en la base GAL4(T44TAG).

Se realizaron estudios de enriquecimiento de modelo para examinar la capacidad del sistema para seleccionar una sintetasa activa de un gran exceso de sintetasas inactivas (Tabla 1, Tabla 2, **figura 6**). Esta selección modela la capacidad para seleccionar sintetasas activas de una biblioteca de variantes en presencia de un aminoácido artificial. Las células MaV203 que contenían GAL4(T44, R110) y EcTyrRS/ARNt_{CUA} se mezclaron con un exceso de 10 a 10⁶ veces de GAL4(T44TAG, R110TAG) y A5/tRNA_{CUA} según se determinó tanto por DO₆₆₀, como por la fracción de colonias que se volvió de color azul al colocarse en placas en medio -leu, -trp no selectivo y ensayadas por recubrimiento de X-GAL. Se seleccionaron las células capaces de sobrevivir en 3-AT 50 mM o en ausencia de uracilo. La relación de células que sobrevivieron en 3-AT o -URA que fueron azules en el ensayo X-GAL con respecto a las que fueron blancas, cuando se compararon con la misma relación en ausencia de selección, demuestran claramente que las selecciones positivas pueden enriquecer las sintetasas activas de las sintetasas muertas mediante un factor >10⁵ (Tabla 1). La medición de enriquecimientos precisos para relaciones iniciales mayores de 1:10⁵ generalmente no fue posible, debido a que no pueden colocarse en placas más de 10⁶ células convenientemente sin un cruce significativo entre las células, lo que conduce a fenotipos no fiables.

TABLA 1. SELECCIONES POSITIVAS MODELO PARA EcTyrRS FUNCIONAL.

Relación de partida, EcYRS:A5 ^a	1:10	1:10 ²	1:10 ³	1:10 ⁴	1:10 ⁵
Dilución celular	10 ³	10 ³	10 ² 10 ³	10 10 ³	1 10 ³
-Leu, Trp (#azul ^b)	1360 (81)	1262 (0)	>10 ³ (1) 1774 (0)	>10 ⁴ (-) 1092 (0)	>10 ⁴ (-) 1472 (0)
-Ura (#azul ^b)	152 (152)	9 (9)	8 (8) 0 (-)	5 (5) 0 (-)	16 (14) 0 (-)
-His + 50 mM3AT (#azul ^b)	135 (135)	7 (7)	0 (-) 0 (-)	3 (3) 0 (-)	10 (10) 0 (-)
Factor de enriquecimiento	>10	>10 ²	>10 ³	>10 ⁴	>10 ⁵

a) Determinado por DO₆₆₀
b) En X-GAL

TABLA 2. SELECCIONES NEGATIVAS MODELO PARA EcTyrRS NO FUNCIONALES (A5).

Relación de partida, A5:EcYRS ^a	1:10	1:10 ²	1:10 ³	1:10 ⁴
Dilución celular	10 ³	10 ²	10 ²	10 ² 10
-Leu, Trp (#blanco ^b)	353 (22)	1401 (31)	1336 (2)	1375 (0) >10 ⁴
5-FOA al 0,1 % (#blanco ^b)	16 (16)	41 (41)	4 (4)	0 (-) 2 (2)
Factor de enriquecimiento	>10	>45	>600	>0,67x 10 ⁴

a) Determinado por DO₆₆₀
b) En X-GAL

Después de una selección positiva en presencia de un aminoácido artificial, las células seleccionadas contendrán sintetasas capaces de utilizar aminoácidos naturales y aquellas capaces de utilizar un aminoácido artificial añadido. Para aislar aquellas sintetasas capaces de utilizar solo el aminoácido artificial, las células que codifican sintetasas que utilizan aminoácidos naturales deben suprimirse de los clones seleccionados. Esto puede lograrse con una selección negativa en la que el aminoácido artificial se mantiene y las sintetasas que funcionan con un aminoácido natural se retiran. Una selección negativa de modelo se llevó a cabo de manera análoga a la selección positiva de modelo. EcTyrRS/ARNt_{CUA} se mezcló con un exceso de 10 a 10⁵ veces de A5/ARNt_{CUA} y la selección se realizó en 5-FOA al 0,1 %. La comparación de la relación de células que sobrevivieron en 5-FOA al 0,1 % que fueron blancas en el ensayo X-GAL con las que fueron azules, con la misma relación en condiciones no selectivas (véase la Tabla 2) deja claro que las selecciones negativas pueden enriquecer las sintetasas muertas a partir de sintetasas activas mediante un factor de al menos 0,6 x 10⁴. La medición de enriquecimientos precisos para relaciones iniciales mayores de 1:10⁴ generalmente no fue posible, debido a que no pueden colocarse en placas más de 10⁵ células

convenientemente sin un cruce significativo entre las células, lo que conduce a fenotipos no fiables.

Se desarrolló un enfoque general que permite tanto la selección positiva de aaRS que reconoce aminoácidos artificiales como la selección negativa de aaRS que reconoce aminoácidos naturales. Al variar la rigurosidad de la selección, puede aislarse diversas actividades de sintetasa. La aplicación de este método a una selección de modelo utilizando variantes de EcTyrRS mostró enriquecimientos de más de 10^5 en una sola ronda tanto de selección positiva y de más de $0,6 \times 10^4$ en una sola ronda de selección negativa. Estas observaciones sugieren que este método puede proporcionar un rápido acceso a aminoacil-ARNt sintetasa ortogonales que funcionan para la incorporación específica en el sitio de aminoácidos artificiales con diversas cadenas laterales a proteínas en *S. cerevisiae*. Además, las enzimas desarrolladas en *S. cerevisiae* pueden utilizarse en eucariotas superiores.

Materiales y métodos

Construcción de vectores

El gen ARNt_{CUA} se amplificó por PCR usando los cebadores

ARNt5': GGGGGGACCGGTGGGGGGACCGGTAAGCTTCCCGATAAGGGAGCAGGCCAGTAAAAAGCATTACC
CCGTGGTGGGTTCCCGA (SEQ ID NO:89), y
ARNt3': GGCGGCGCTAGCAAGCTTCCCGATAAGGGAGCAGGCCAGTAAAAAGGGAAGTTCAGGGACTTTTGA
AAAAATGGTGGTGGGGGAAGGAT (SEQ ID NO:90)

de pESCU3URA. Esta, y todas las demás reacciones de PCR se realizaron utilizando el kit Expand PCR de Roche, de acuerdo con las instrucciones del fabricante. Después de la digestión de endonucleasa de restricción con NheI y AgeI, este gen de ARNt se insertó entre los mismos sitios en el vector de 2 µm pESCTrp (Stratagene) para producir pARNt_{CUA}. El promotor ADH1 de longitud completa se amplificó mediante PCR de pDBLeu (Invitrogen) con los iniciadores PADHf: IGGGGGACCGGTGGGGGGACCGGTGGGATCGAAGAAATGATGGTAAATGA
AATAGGAAATCAAGG (SEQ ID NO:91) y pADHR: GGGGGGAATTCAGTTGATTGTATGCTTGGTATAGCTTGAATATTGTGCAGAA AAAGAAAC (SEQ ID NO:92), digerido con AgeI y EcoRI. EcTyrRS se amplificó con los cebadores pESCTrp1: TCATAACGAGAATTCGGGATCGAAGAAATGATGGTAAATGAAATA
GGAAATCTCATAACGAGAATTCATGGCAAGCAGTAACTTG (SEQ ID NO: 93) y pESCTrp2: TTACTACGTGCGGCCGATGGCAAGCAGTAACTTGTTACTACGTGCGGCCGCTTATTTCCAGCAAAT
CAGAC (SEQ ID NO:94). El producto de PCR de EcTyrRS se digirió con EcoRI y Not I. A continuación, el pARNt_{CUA} se digirió con Age I y Not I. Una ligadura triple de estos tres ADN produjo pEcTyrRS-ARNt_{CUA}. El plásmido pA5-ARNt_{CUA} en el que los residuos de aminoácidos (37,126,182,183 y 186 en el sitio activo están mutados en alanina) se creó por PCR de superposición usando los oligonucleótidos F37Afwf: CCGATCGCGCTCGCTTGGCGCTTCGATC (SEQ ID NO:95), N126Afwf: ATCGCGCGAACGCCTATGACTGGTTC (SEQ ID NO:96), 182,183,186A, GTTGCAGGGTTATGCCGCCGCTGTGCGAACAACAGTAC (SEQ ID NO:97) y sus complementos inversos, así como los oligonucleótidos flanqueantes, 4783: GCCGCTTTGCTATCAAGTATAAATAG (SEQ ID NO:98), 3256: CAAGCCGACAACCTTGATTGG (SEQ ID NO:99) y pEcTyrRS-ARNt_{CUA} como plantilla. El producto de PCR de EcTyrRS se digirió con EcoRI y Not I y se ligó en el fragmento grande de pEcTyrRS-ARNt_{CUA} liberado tras la digestión con las mismas enzimas. Para construir la 1.ª generación de indicadores DB-AD, el dominio de unión a ADN de GAL4 se amplificó por PCR de pGADT7 (Clontech) utilizando el cebador directo paADfwf:

GGGGACAAGTTTGTACAAAAAAGCAGGCTACGCCAATTTTAATCAAAGTGG GAATATTGC (SEQ ID NO: 100)
o paADfwf(TAG)
GGGGACAAGTTTGTACAAAAAAGCAGGCTAGGCCAATTTTAATCAAAGTGG GAATATTGC (SEQ ID NO: 101)
y ADrev:

GGGGACCACTTTGTACAAGAAAGCTGGGTTACTCTTTTTTGGGTTTGGTGG GGTATC (SEQ ID NO: 102). Estos productos de PCR se clonaron en el vector pDEST3-2 (Invitrogen) usando el procedimiento Clonase, de acuerdo con las instrucciones del fabricante, produciendo pDB-AD y pDB-(TAG)-AD. Para construir PGADGAL4 y las variantes, el gen GAL4 se amplificó de pCL1 (Clontech) por PCR usando los cebadores ADH1428-1429
AAGCTATACCAAGCATACAATC (SEQ ID NO:103), y GAL4C: ACAAGGCCTTGCTAGCTTACTTTTTTGGGTTTGGTGGGTTATCTTC (SEQ ID NO: 104). Este fragmento se clonó en el vector pCR2.1 TOPO (Invitrogen) de acuerdo con las instrucciones del fabricante. Un clon que contenía el gen GAL4 (pCR2.1 TOPOGAL4) se digirió con Hind III y el gel del fragmento GAL4 de 2,7 kb se purificó y se ligó al fragmento grande de pGADT7 que se había digerido con Hind III, tratado con fosfatasa intestinal de ternera y se purificó con gel. Las variantes del gen GAL4 se crearon por reacciones de Quikchange (Stratagene), realizadas de acuerdo con las instrucciones del fabricante, en pCR2.1 utilizando los cebadores enumerados en la información complementaria. Los mutantes GAL4 se clonaron en pGADT7 de la misma manera que el gen GAL4 de tipo silvestre. Todas las construcciones finales fueron confirmadas por secuenciación de ADN.

Medios de levadura y manipulaciones

La cepa de *S. cerevisiae* MaV203, (Invitrogen) es MAT α ; *leu2-3,112*; *trp1 109*; *his3 Δ 200*; *ade2-101*; *cyh2^R*; *cyh1^R*; *GAL4 Δ* ; *gal80 Δ* ; *GAL1::lacZ*; *HIS3UASGAL1::HIS3@LYS2*; *SPAL10UASGAL1::URA3*. Los medios de levadura se adquirieron en Clontech, 5-FOA y X-GAL eran de Invitrogen, y 3-AT era de BIO 101. YPER (Reactivo de extracción de proteína de levadura) y ONPG se adquirieron en Pierce Chemicals. Las transformaciones de plásmidos se realizaron por el método de PEG/acetato de litio (véase, por ejemplo, D. Burke, et al., (2000) Methods in Yeast Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY) y los transformantes se seleccionaron en el medio de goteo completo artificial apropiado. Para probar los fenotipos conferidos por diversas combinaciones de plásmidos en las colonias de levadura MaV203 de placas sintéticas de goteo completas de cada transformación se resuspendieron en 15 μ l de agua estéril y se extendieron sobre los medios selectivos de interés. Cada fenotipo se confirmó con al menos cinco colonias independientes. Los ensayos X-GAL se realizaron mediante el método de recubrimiento de agarosa. Véase, I.G. Serebriiskii, y E.A. Golemis, (2000), Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, Analytical Biochemistry 285:1-15. En resumen, las colonias o parches celulares se lisaron en placas de agar mediante varias adiciones de cloroformo puro. Después de la evaporación del cloroformo, se aplicó a la superficie de la placa agarosa al 1 % que contenía 0,25 g/l de XGAL y se tamponó con Na₂PO₄ 0,1 M. Una vez que se fijó la agarosa, las placas se incubaron a 37 °C durante 12 h. Los ensayos de ONPG se realizaron mediante la inoculación de 1 ml de SD-leu,-trp en un bloque de 96 pocillos con una sola colonia y se incubaron a 30 °C con agitación. La DO₆₆₀ de 100 μ l de células y varias diluciones de células se registraron en paralelo en una placa de microtitulación de 96 pocillos. Las células (100 μ l) se mezclaron con 100 μ l de YPER:ONPG (1X PBS, YPER al 50 % v/v, MgCl₂ 20 mM, β -mercaptoetanol al 0,25 % v/v y ONPG 3 mM) y se incubaron con agitación a 37 °C. Tras el desarrollo del color, las células se sedimentaron por centrifugación, el sobrenadante se transfirió a una placa limpia de microtitulación de 96 pocillos (Nuncclon, n.º de cat. 167008), y se registró el A420. Todos los datos mostrados son la media de los ensayos de al menos 4 clones independientes y las barras de error que se muestran representan la desviación estándar. La hidrólisis de ONPG se calculó utilizando la ecuación: unidades de beta-galactosidasa = 1000. A420/(V.t.DO₆₆₀), donde V es el volumen de células en mililitros, t es el tiempo de incubación en minutos. Véase, por ejemplo, I.G. Serebriiskii, y E.A. Golemis, (2000), Uses of lacZ to study gene function: evaluation of beta-galactosidase assays employed in the yeast two-hybrid system, Analytical Biochemistry 285:1-15. Una unidad de beta-galactosidasa corresponde a la hidrólisis de 1 μ mol de ONPG por minuto por célula. Véase, Serebriiskii y Golemis, anteriormente. Las lecturas espectrofotométricas se realizaron en un lector de placas SPECTRAMax190.

Selecciones de modelo

Selecciones positivas: Se cultivaron dos cultivos durante una noche en SD -Leu, -Trp. Uno contenía MaV203 que albergaban pEcTyrRS-ARNt_{CUA}/pGADGAL4(T44, R110TAG) y el otro pA5-ARNtSU3/pGADGAL4(T44, R110TAG). Estas células se cosecharon por centrifugación y se resuspendieron en NaCl al 0,9 % mediante agitación vorticial. A continuación, las dos soluciones celulares se diluyeron a DO₆₆₀ idénticas. MaV203 que albergaban pEcTyrRS-ARNt_{CUA}/pGADGAL4(T44,R110TAG) se diluyeron en serie en 7 órdenes de magnitud y cada dilución se mezcló a continuación 1:1 vol:vol con MaV203 sin diluir que albergaban pA5-ARNt_{CUA}/pGADGAL4(T44, R110TAG) para proporcionar relaciones definidas de células que contenían tirosil-ARNt sintetasa activa e inactiva. Para cada relación, se realizó una segunda dilución en serie en la que se disminuyó el número de células pero se mantuvo la relación de células que albergan pEcTyrRS-ARNt_{CUA}/pGADGAL4(T44,R110TAG) y pA5-ARNt_{CUA}/pGADGAL4(T44,R110TAG). Estas diluciones se pusieron en placas en SD -Leu, -trp, SD -Leu,-Trp,- URA y SD - Leu, - Trp, -His + 3-AT 50 mM. Después de 60 h, se contó el número de colonias en cada placa, utilizando una cámara CCD Eagle Eye (Stratagene), y se confirmó el fenotipo de los sobrevivientes con un ensayo de beta-galactosidasa X-GAL. Las células de varias colonias azules o blancas individuales se aislaron y crecieron hasta saturación en SD -leu, -trp y el ADN plasmídico se aisló por métodos estándar. La identidad de la variante EcTyrRS se confirmó mediante secuenciación de ADN.

Selección negativa: La selección negativa de modelo se realizó de manera análoga a la selección positiva, excepto que MaV203 que albergaban pA5-ARNt_{CUA}/pGADGAL4(T44,R110TAG) se diluyeron en serie y se mezclaron con una densidad fija de MaV203 que albergaban pEcTyrRS-ARNt_{CUA}/pGADGAL4(T44,R110TAG). Las células se colocaron en placas en SD -leu, -trp +5-FOA al 0,1 %, se contó el número de colonias después de 48 horas, y las placas se procesaron como se ha descrito anteriormente.

Los siguientes oligonucleótidos (Tabla 3) se usaron en combinación con sus complementos inversos para construir mutantes de sitio dirigido mediante mutagénesis de Quikchange. La posición de la mutación se representa con texto en negrita.

TABLA 3: OLIGONUCLEOTIDOS UTILIZADOS PARA LA CONSTRUCCIÓN DE MUTANTES DE SITIO DIRIGIDO.

Ámbar

Secuencia de mutantes oligo

L3TAG 5'-ATGAAGTAGCTGTCTTCTATCGAACAAAGCATGCG-3' (SEQ ID NO:66)
 I13TAG 5'-CGAACAAAGCATGCGATTAGTGCCGACTTAAAAAG-3' (SEQ ID NO:67)
 T44TAG 5'-CGCTACTCTCCCAAATAGAAAAGGTCTCCGCTG-3' (SEQ ID NO:68)
 F68TAG 5'-CTGGAACAGCTATAGCTACTGATTTTCTCG-3' (SEQ ID NO:69)
 R110TAG 5'-GCCGTCACAGATTAGTTGGCTTCAGTGGAGACTG-3' (SEQ ID NO:70)
 V114TAG 5'-GATTGGCTTCATAGGAGACTGATATGCTCTAAC-3' (SEQ ID NO:71)
 T121TAG 5'-GCCTCTATAGTTGAGACAGCATAGAATAATGCG-3' (SEQ ID NO:72)
 I127TAG 5'-GAGACAGCATAGATAGAGTGCGACATCATCATCGG-3' (SEQ ID NO:73)
 S131TAG 5'-GAATAAGTGCGACATAGTCATCGGAAGAGAGTAGTAG-3' (SEQ ID NO:74)
 T145TAG 5'-GGTCAAAGACAGTTGTAGGTATCGATTGACTCGGC-3' (SEQ ID NO:75)

Permisivo

Secuencia de mutantes oligo en el sitio

T44F 5'-CGCTACTCTCCCAAATTTAAAGGTCTCCGCTG-3' (SEQ ID NO:76)
 T44Y 5'-CGCTACTCTCCCAAATATAAAGGTCTCCGCTG-3' (SEQ ID NO:77)
 T44W 5'-CGCTACTCTCCCAAATGGAAAAGGTCTCCGCTG-3' (SEQ ID NO:78)
 T44D 5'-CGCTACTCTCCCAAAGATAAAGGTCTCCGCTG-3' (SEQ ID NO:79)
 T44K 5'-CGCTACTCTCCCAAATAAAGGTCTCCGCTG-3' (SEQ ID NO:80)
 R110F 5'-GCCGTCACAGATTTTTGGCTTCAGTGGAGACTG-3' (SEQ ID NO:81)
 R110Y 5'-GCCGTCACAGATTATTTGGCTTCAGTGGAGACTG-3' (SEQ ID NO:82)
 R110W 5'-GCCGTCACAGATTGGTTGGCTTCAGTGGAGACTG-3' (SEQ ID NO:83)
 R110D 5'-GCCGTCACAGATGATTTGGCTTCAGTGGAGACTG-3' (SEQ ID NO:84)
 R110K 5'-GCCGTCACAGATAAATTGGCTTCAGTGGAGACTG-3' (SEQ ID NO:85)

EJEMPLO 2: UN CÓDIGO GENÉTICO EUCARIOTA EXPANDIDO

- 5 Se describe una ruta general y rápida para la adición de aminoácidos artificiales al código genético de *Saccharomyces cerevisiae*. Se han incorporado cinco aminoácidos en proteínas de manera eficiente, con alta fidelidad, en respuesta al codón TAG sin sentido. Las cadenas laterales de estos aminoácidos contienen un grupo ceto, que puede modificarse de forma única *in vitro* e *in vivo* con una amplia gama de sondas químicas y reactivos; un aminoácido que contiene átomos pesados para estudios estructurales; y fotorreticuladores para estudios
- 10 celulares de interacciones proteicas. Esta metodología no solo elimina las restricciones impuestas por el código genético sobre nuestra capacidad para manipular la estructura y la función de la proteína en la levadura, sino que proporciona una puerta de entrada a la expansión sistemática de los códigos genéticos de eucariotas multicelulares.

- Aunque los químicos han desarrollado una poderosa variedad de métodos y estrategias para sintetizar y manipular las estructuras de moléculas pequeñas (véase, por ejemplo, E. J. Corey, y X.-M. Cheng, *The Logic of Chemical Synthesis* (Wiley-Interscience, Nueva York, 1995)), la capacidad de controlar racionalmente la estructura y función de la proteína todavía está en su fase temprana. Los métodos de mutagénesis se limitan a los componentes básicos comunes de 20 aminoácidos, aunque en varios casos ha sido posible incorporar competitivamente análogos estructurales cercanos de aminoácidos comunes en todo el proteoma. Véase, por ejemplo, K. Kirshenbaum, et al., (2002), *ChemBioChem* 3:235-7; y, V. Doring et al., (2001), *Science* 292:501-4. La síntesis total (véase, por ejemplo, B. Merrifield, (1986), *Science* 232:341-7 (1986)), y las metodologías semisintéticas (véanse, por ejemplo, D. Y. Jackson et al., (1994) *Science* 266:243-7; y, P. E. Dawson, y S. B. Kent, (2000), *Annual Review of Biochemistry* 69:923-60, han hecho posible sintetizar péptidos y proteínas pequeñas, pero tienen una utilidad más limitada con proteínas de más de 10 kilodaltons (kDa). Los métodos bioartificiales que involucran ARNt ortogonales químicamente acilados (véanse, por ejemplo, D. Mendel, et al., (1995), *Annual Review of Biophysics and Biomolecular Structure* 24:435-462; y, V. W. Cornish, et al. (31 de marzo de 1995), *Angewandte Chemie-International Edition in English* 34:621-633) han permitido que se incorporen aminoácidos artificiales en proteínas más grandes, tanto *in vitro* (véase, por ejemplo, J. A. Ellman, et al., (1992), *Science* 255:197-200) como en células microinyectadas (véase, por ejemplo, véase, por ejemplo, D. A. Dougherty, (2000), *Current Opinion in Chemical Biology* 4:645-52). Sin embargo, la naturaleza estequiométrica de la acilación química limita severamente la cantidad de proteína que se puede generar. Por lo tanto, a pesar de los considerables esfuerzos, las propiedades de las proteínas, y posiblemente de organismos enteros, han sido limitadas a lo largo de la evolución por los veinte aminoácidos codificados genéticamente (con las raras excepciones de la pirrolisina y la selenocisteína (véase, por ejemplo, A. Bock et al., (1991), *Molecular Microbiology* 5:515-20; y, G. Srinivasan, et al., (2002), *Science* 296:1459-

62)).

Para superar esta limitación, se añadieron nuevos componentes a la maquinaria biosintética de las proteínas del procarionta *Escherichia coli* (*E. coli*) (por ejemplo, L. Wang, et al., (2001), *Science* 292:498-500), que hacen posible codificar genéticamente aminoácidos artificiales *in vivo*. Varios aminoácidos nuevos con nuevas propiedades químicas, físicas o biológicas se han incorporado de manera eficiente y selectiva en proteínas en respuesta al codón ámbar, TAG. Véase, por ejemplo, J. W. Chin et al., (2002), *Journal of the American Chemical Society* 124:9026-9027; J. W. Chin, y P. G. Schultz, (2002), *ChemBioChem* 11:1135-1137; J. W. Chin, et al., (2002), *PNAS United States of America* 99:11020-11024; y, L. Wang, y P. G. Schultz, (2002), *Chem. Comm.*, 1:1-10. Sin embargo, debido a que la maquinaria de traducción no está bien conservada entre procariotas y eucariotas, los componentes de la maquinaria biosintética añadida a *E. coli* generalmente no se pueden usar para incorporar de forma específica de sitio aminoácidos artificiales en proteínas para estudiar o manipular procesos celulares en células eucariotas.

Por lo tanto, se crearon componentes traduccionales que expandirán el número de aminoácidos codificados genéticamente en células eucariotas. Se eligió *Saccharomyces cerevisiae* como el organismo huésped eucariota inicial, porque es un modelo eucariota útil, las manipulaciones genéticas son fáciles (véase, por ejemplo, D. Burke, et al., (2000), *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY), y su maquinaria de traducción es muy homóloga a la de los eucariotas superiores (véase, por ejemplo, T. R. Hughes, (2002), *Funct. Integr. Genomics* 2:199-211). La adición de nuevos componentes básicos al código genético de *S. cerevisiae* requiere un codón único, ARNt, y aminoacil-tRNA sintetasa ("aaRS") que no reaccionan de forma cruzada con ningún componente de la maquinaria de traducción de levadura (véase, por ejemplo, Noren et al., (1989) *Science* 244:182; Furter (1998) *Protein Sci.* 7:419; y, Liu et al., (1999) *PNAS USA* 96:4780). Un par ortogonal candidato es el par supresor ámbar tirosil-ARNt sintetasa-ARNt_{CUA} de *E. coli* (véase, por ejemplo, H. M. Goodman, et al., (1968), *Nature* 217:1019-24; y, D. G. Barker, et al., (1982), *FEBS Letters* 150:419-23). La tirosil-ARNt sintetasa de *E. coli* (TyrRS) aminoacila eficientemente ARNt_{CUA} de *E. coli* cuando ambos están codificados genéticamente en *S. cerevisiae* pero no aminoacilan ARNt citoplasmáticos de *S. cerevisiae*. Véase, por ejemplo, H. Edwards, y P. Schimmel, (1990), *Molecular & Cellular Biology* 10:1633-41; y, H. Edwards, et al., (1991), *PNAS United States of America* 88:1153-6. Además, el tirosil ARNt_{CUA} de *E. coli* es un sustrato deficiente para las aminoacil-ARNt sintetisas de *S. cerevisiae* (véase, por ejemplo, V. Trezeguet, et al., (1991), *Molecular & Cellular Biology* 11:2744-51.) pero se procesa y se exporta desde el núcleo al citoplasma (véase, por ejemplo, S. L. Wolin, y A. G. Matera, (1999) *Genes & Development* 13:1-10) y funciona eficientemente en la traducción de proteínas en *S. cerevisiae*. Véase, por ejemplo, H. Edwards, y P. Schimmel, (1990) *Molecular & Cellular Biology* 10:1633-41; H. Edwards, et al., (1991), *PNAS United States of America* 88:1153-6; y, V. Trezeguet, et al., (1991), *Molecular & Cellular Biology* 11:2744-51. Además, TyrRS de *E. coli* no tiene un mecanismo de edición y, por lo tanto, no debe corregir un aminoácido artificial ligado al ARNt.

Para alterar la especificidad de aminoácidos de la TyrRS ortogonal para que aminoacile ARNt_{CUA} con un aminoácido artificial deseado y ninguno de los aminoácidos endógenos, se generó una gran biblioteca de mutantes TyrRS y se sometió a una selección genética. Basándose en la estructura cristalina de la TyrRS homóloga de *Bacillus stearothermophilus* (véase, por ejemplo, P. Brick, et al., (1989), *Journal of Molecular Biology* 208:83) se mutaron cinco residuos ((*B. stearothermophilus*, **figura 7, Panel A**)) en el sitio activo de TyrRS de *E. coli*. que están dentro de 6,5 Å de la posición para el anillo de aro de tirosina unida. Por ejemplo, para crear la biblioteca de EcTyrRS de mutantes, las cinco posiciones seleccionadas para la mutación se convirtieron primero en codones de alanina para producir el gen A5RS. Esto se dividió entre dos plásmidos en un sitio Pst I único en el gen. La biblioteca se creó esencialmente como se describe mediante técnicas conocidas en la técnica (véase, por ejemplo, Stemmer et al., (1993) *Biotechniques* 14:256-265). Un plásmido contiene la mitad 5' del gen A5RS, el otro plásmido contiene la mitad 3' del gen A5RS. La mutagénesis se realizó en cada fragmento por PCR con cebadores oligonucleotídicos para la amplificación del plásmido completo. Los cebadores están dopados, conteniendo NNK (N = A + G + T + C y K = G + T) y sitios de reconocimiento de endonucleasa de restricción Bsa I. La digestión con Bsa I y la ligadura produjeron dos plásmidos circulares, cada uno con copias mutantes de la mitad del gen EcTyrRS. Los dos plásmidos se digirieron a continuación con Pst I y se ensamblaron en un solo plásmido mediante ligadura, lo que condujo al ensamblaje de los genes mutantes de longitud completa. Los genes mutantes EcTyrRS se escindieron de este plásmido y se ligaron en pA5RS/ARNt_{CUA} entre los sitios EcoR I y Not I. La biblioteca se transformó en Mav203 de *S. cerevisiae*: pGADGAL4 (2TAG) usando el método de PEG-acetato de litio produciendo ~10⁸ transformantes independientes.

Una cepa de selección de *S. cerevisiae* [MaV203: pGADGAL4 (2 TAG) (véase, por ejemplo, M. Vidal, et al., (1996), *PNAS United States of America* 93:10321-6; M. Vidal, et al., (1996), *PNAS United States of America* 93:10315-201 y, Chin et al., (2003) *Chem. Biol.* 10:511)] se transformó con la biblioteca para proporcionar 10⁸ transformantes independientes y creció en presencia de 1 mM de aminoácidos artificiales (**figura 8, Panel C**). La supresión de dos codones ámbar permisivos en el activador transcripcional GAL4 conduce a la producción de GAL4 de longitud completa y la activación transcripcional de los genes indicadores HIS3, URA3 y lacZ sensibles a GAL4 (**figura 8, Panel A**). Por ejemplo, los codones permisivos son para T44 y R110 de Gal4. Expresión de HIS3 y URA3 en medios que carecen de uracilo (-ura), o que contienen 3-aminotriazol 20 mM (véase, por ejemplo, G. M. Kishore, y D. M. Shah, (1988), *Annual Review of Biochemistry* 57, 627-63) (3-AT, a competitive inhibitor of the His3 protein) y que carecen de histidina (-his), permite que los clones que expresan pares activos de aaRS-ARNt_{CUA} se seleccionen

positivamente. Si una TyrRS mutante carga el ARNt_{CUA} con un aminoácido, entonces la célula biosintetiza la histidina y el uracilo y sobrevive. Las células supervivientes se amplificaron en ausencia de 3-AT y aminoácidos artificiales para eliminar GAL4 de longitud completa de las células que incorporan selectivamente el aminoácido artificial. Para eliminar los clones que incorporan aminoácidos endógenos en respuesta al codón ámbar, se cultivaron células en medios que contenían ácido 5-fluorootico al 0,1 % (5-FOA) pero que carecían del aminoácido artificial. Las células que expresan URA3, como resultado de la supresión de las mutaciones ámbar GAL4 con aminoácidos naturales, convierten el 5-FOA en un producto tóxico, destruyendo la célula. Véase, por ejemplo, J. D. Boeke, et al., (1984), Molecular & General Genetics 197:345-6. Los clones supervivientes se amplificaron en presencia de aminoácidos artificiales y se volvieron a aplicar a la selección positiva. El indicador lacZ permite que los pares de ARNt de sintetasa activa e inactiva se discriminen colorométricamente (**figura 8, Panel B**).

Con el uso de este enfoque, se añadieron de forma independiente cinco nuevos aminoácidos con distintas propiedades estéricas y electrónicas (**figura 7, Panel B**) al código genético de *S. cerevisiae*. Estos aminoácidos incluyen *p*-acetil-L-fenilalanina (**1**), *p*-benzoil-L-fenilalanina (**2**), *p*-azido-L-fenilalanina (**3**), O-metil-L-tirosina (**4**), y *p*-yodo-L-fenilalanina (**5**) (indicados por los números en la **figura 7, Panel B**). La reactividad única del grupo ceto funcional de *p*-acetil-L-fenilalanina permite la modificación selectiva de proteínas con una serie de reactivos que contienen hidrazina o hidroxilamina *in vitro* e *in vivo* (véase, por ejemplo, V. W. Cornish, et al., (28 de agosto de 1996), Journal of the American Chemical Society 118:8150-8151; y, Zhang, Smith, Wang, Brock, Schultz, en preparación). El átomo pesado de *p*-yodo-L-fenilalanina puede resultar útil para sincronizar los datos de la estructura de rayos X (con el uso de difracción anómala de longitud de onda múltiple). Las cadenas laterales de benzofenona y fenilazida de *p*-benzoil-L-fenilalanina y *p*-azido-L-fenilalanina permiten una eficiente fotorreticulación de proteínas *in vivo* e *in vitro* (véanse, por ejemplo, Chin et al., (2002) J. Am. Chem. Soc., 124:9026; Chin y Schultz, (2002) Chem. Bio.Chem. 11:1135; y, Chin et al., (2002) PNAS, USA 99:11020). El grupo metilo de O-metil-L-tirosina puede sustituirse fácilmente con un grupo metilo marcado isotópicamente como una sonda de estructura y dinámica local con el uso de resonancia magnética nuclear y espectroscopía vibratoria. Después de tres rondas de selección (positivo-negativo-positivo), se aislaron varias colonias cuya supervivencia en medio -ura o en 3-AT 20 mM -his dependía estrictamente de la adición del aminoácido artificial seleccionado. Véase, la **figura 8, Panel D**. Los mismos clones eran azules en x-gal solo en presencia de un aminoácido artificial 1 mM. Estos experimentos demuestran que los fenotipos observados son el resultado de la combinación de los pares de aminoacil-ARNt sintetasa-ARNt_{CUA} desarrollados y sus aminoácidos afines (véase, la Tabla 4).

Por ejemplo, para seleccionar las sintetasas mutantes, se cultivaron células (~10⁹) durante 4 horas en SD líquido -leu, -trp + aminoácido 1 mM. A continuación, las células se cosecharon por centrifugación, se resuspendieron en NaCl al 0,9 % y se colocaron en placas en SD -leu, -trp, -his + 3-AT 20 mM, + aminoácido artificial 1 mM o SD -leu, -trp, -ura, + aminoácido artificial 1 mM. Después de 48 a 60 horas a 30 °C, las células se rasparon de las placas en SD líquido -leu, -trp y se cultivaron durante 15 horas a 30 °C. Las células se cosecharon por centrifugación, se resuspendieron en NaCl al 0,9 % y se colocaron en placas en SD -leu, -trp + 5-FOA al 0,1 %. Después de 48 horas a 30 °C, las células se rasparon en SD líquido -leu, -trp + aminoácido artificial 1 mM y se cultivaron durante 15 horas. A continuación, las células se cosecharon por centrifugación, se resuspendieron en NaCl al 0,9 % y se colocaron en placas en SD -leu, -trp, -his + 3-AT 20 mM, + aminoácido artificial 1 mM o SD -leu, -trp, -ura, + aminoácido artificial 1 mM. Para cribar fenotipos de células seleccionadas, las colonias (192) de cada selección se transfirieron a pocillos de bloques de 96 pocillos que contenían 0,5 ml de SD -leu, -trp y se cultivaron a 30 °C durante 24 horas. Se añadió glicerol (50 % v/v; 0,5 ml) a cada pocillo, y las células se replicaron en placas sobre agar (SD -leu, -trp; SD -leu, -trp, -his, + 3-AT 20 mM; SD -leu, -trp, -ura) en presencia o ausencia de un aminoácido artificial 1 mM. Los ensayos X-Gal se realizaron en placas SD -leu, -trp utilizando el método de recubrimiento de agarosa.

Para demostrar aún más que los fenotipos observados se deben a la incorporación específica de sitio de los aminoácidos artificiales por los pares mutantes ortogonales de TyrRS/ARNt, mutantes de la superóxido dismutasa humana 1 (hSOD) (véase, por ejemplo, H. E. Parge, et al., (1992), PNAS United States of America 89:6109-13) que contenían cada aminoácido artificial se generaron y se caracterizaron.

Por ejemplo, la adición de ADN que codifica una etiqueta de hexahistidina C-terminal, y la mutación del codón para Trp 33 en un codón ámbar en el gen de superóxido dismutasa humana se realizó mediante PCR de superposición usando PS356 (ATCC) como plantilla. hSOD (Trp 33 TAG) HIS se clonó entre el promotor GAL1 y el terminador CYC1 de pYES2.1 (Invitrogen, Carlsbad, CA EE.UU.). Los genes de sintetasa mutante y ARNt en plásmidos derivados de pECTyrRS-ARNt_{CUA} se co-transformaron con pYES2.1 hSOD (Trp 33 TAG) HIS en la cepa InvSc (Invitrogen). Para la expresión de proteínas, las células se cultivaron en SD -trp, -ura + refinosa y la expresión se indujo a una DO₆₆₀ de 0,5 mediante la adición de galactosa. Los mutantes HSOD se purificaron por cromatografía Ni-NTA (Qiagen, Valencia, CA, EE.UU.).

La producción de hSOD etiquetada con hexa-histidina a partir de un gen que contiene un codón ámbar en la posición 33 dependía estrictamente de *p*-acetilPheRS-1-ARNt_{CUA} y *p*-acetil-L-fenilalanina 1 mM (<0,1 % por densitometría, en ausencia de cualquiera de los componentes) (véase la **figura 9**). La *p*-acetil-L-fenilalanina que contenía hSOD de longitud completa se purificó (por ejemplo, mediante cromatografía de afinidad de Ni-NTA) con un rendimiento de 50 ng/ml, comparable al purificado de las células que contenían TyrRSARNt_{CUA} de *E. coli*. A efectos comparativos, el hSODHIS de tipo silvestre podría purificarse con un rendimiento de 250 ng/ml en condiciones idénticas.

La **figura 9** ilustra la expresión de proteínas de hSOD (33TAG)HIS en *S. cerevisiae* que codifica genéticamente aminoácidos artificiales (como se ilustra en la **figura 7, Panel B** e indicados en la **figura 9** por su numeración en la **figura 7, Panel B**). La parte superior de la **figura 9** ilustra la electroforesis en gel de SDS-poliacrilamida de hSOD purificada de levadura en presencia (+) y ausencia (-) del aminoácido artificial indicado por el número, que

corresponde al amino artificial ilustrado en la **figura 7, Panel B** tinción con Coomassie. Las células contienen el par mutante de sintetasa-ARNt seleccionado para el aminoácido indicado. La parte central de la **figura 9** ilustra una transferencia de Western sondeada con un anticuerpo contra hSOD. La parte inferior de la **figura 9** ilustra una transferencia de Western sondeada con un anticuerpo contra la etiqueta His6 C-terminal.

La identidad del aminoácido incorporado se determinó sometiendo una digestión trípica de la proteína mutante a cromatografía líquida y espectrometría de masas en tándem. Por ejemplo, para la espectrometría de masas, se visualizaron bandas de proteínas mediante tinción coloidal de Coomassie. Las bandas de gel correspondientes a SOD de tipo silvestre y mutante se cortaron de geles de poliácridamida, se cortaron en cubos de 1,5 mm, se redujeron y se alquilaron, a continuación se sometieron a hidrólisis de tripsina esencialmente como se describe. Véase, por ejemplo, A. Shevchenko, et al., (1996), *Analytical Chemistry* 68, 850-858. Los péptidos tríplicos que contienen el aminoácido artificial se analizaron por HPLC/μESI/MS de fase inversa de nanoflujo con un espectrómetro de masas con trampa de iones LCQ. El análisis de espectrometría de masas en tándem de cromatografía líquida (LC-MS/MS) se realizó en un espectrómetro de masas con trampa de iones Deca Finnigan LCQ (Thermo Finnigan) equipado con una HPLC Nanospray (serie Agilent 1100). Véase, por ejemplo, **figura 10, Paneles A-H**.

Los iones precursores correspondientes a los iones cargados individual y doblemente del péptido Val-Y*-Gly-Ser-Ile-Lys (SEQ ID NO:87) que contienen los aminoácidos artificiales (representados Y*) se separaron y se fragmentaron con un espectrómetro de masas con trampa de iones. Las masas de iones de fragmentos podrían asignarse sin ambigüedades, confirmando la incorporación específica de sitio de p-acetil-L-fenilalanina (véase, la **figura 10, Panel A**). No se observó indicación de tirosina u otros aminoácidos en lugar de p-acetil-L-fenilalanina, y se obtuvo un mínimo de pureza de incorporación del 99,8 % a partir de la relación señal-ruido de los espectros de péptidos. Se observó una fidelidad y eficiencia similares en la expresión de proteínas cuando se utilizó p-benzoil-PheRS-1, p-azidoPheRS-1, O-meTyrRS-1, o p-yodoPheRS-1 para incorporar p-benzoil-L-fenilalanina, p-azido-L-fenilalanina, O-metil-L-tirosina, o p-yodo-L-fenilalanina en hSOD (véase, la **figura 9**, y la **figura 10, Paneles A-H**). En los experimentos, la p-azido-L-fenilalanina se reduce en p-amino-L-fenilalanina en la preparación de la muestra, y esta última se observa en los espectros de masas. La reducción no tiene lugar *in vivo* por derivación química de SOD purificada que contiene p-azido-L-fenilalanina. En los experimentos de control, se preparó hSOD etiquetada con hexahistidina que contenía triptófano, tirosina y leucina en la posición 33 y se sometió a espectrometría de masas (véase, la **figura 10, Paneles F, G y H**). Los iones que contenían el aminoácido 33 eran claramente visibles en los espectros de masas de estas muestras.

La adición independiente de cinco aminoácidos artificiales al código genético de *S. cerevisiae* demuestra la generalidad de este método y sugiere que puede ser aplicable a otros aminoácidos artificiales que incluyen aminoácidos marcados por espín, de unión a metales, o fotoisomerizables. Esta metodología puede permitir la generación de proteínas con propiedades nuevas o mejoradas, así como facilitar el control de la función de la proteína en la levadura. Además, en las células de mamíferos, la tirosil-ARNt sintetasa de *E. coli* forma un par ortogonal con el ARN^t_{CUA} de *B. stearothermophilus*. Véase, por ejemplo, Sakamoto et al., (2002) *Nucleic Acids Res.* 30:4692. Por lo tanto, se pueden usar las aminoacil-ARNt sintetetas que se han desarrollado en levaduras para añadir aminoácidos artificiales a los códigos genéticos de eucariotas superiores.

TABLA 4. SECUENCIAS DE AMINOACIL-ARNt SINTETASAS SELECCIONADAS

Residuo N.º	37	126	182	183	186	N.º de clones
Ec TyrRS	Tyr	Asn	Asp	Phe	Leu	
p-yodoPheRS-1	Val	Asn	Ser	Tyr	Leu	1/8
p-yodoPheRS-2	Ile	Asn	Ser	Met	Leu	1/8
p-yodoPheRS-3	Val	Asn	Ser	Met	Ala	6/8
p-OMePheRS-1	Val	Asn	Ser	Met	Leu	5/13
p-OMePheRS-2	Thr	Asn	Thr	Met	Leu	1/13
p-OMePheRS-3	Thr	Asn	Thr	Tyr	Leu	1/13
p-OMePheRS-4	Leu	Asn	Ser	Met	Ser	1/13
p-OMePheRS-5	Leu	Asn	Ser	Met	Ala	1/13
p-OMePheRS-6	Thr	Asn	Arg	Met	Leu	4/13
p-acetilPheRS-1 ^a	Ile	Asn	Gly	Met	Ala	10/10
p-benzoilPheRS-1	Gly	Asn	Gly	Phe	Ala	1/2
p-benzoilPheRS-2	Gly	Asn	Gly	Tyr	Met	1/2
p-azidoPheRS-1	Leu	Asn	Ser	Met	Ala	1/6
p-azidoPheRS-2	Val	Asn	Ser	Ala	Ala	1/6
p-azidoPheRS-3	Leu	Asn	Ser	Ala	Ala	1/6
p-azidoPheRS-4	Val	Asn	Ser	Ala	Val	1/6

(continuación)

Residuo N.º	37	126	182	183	186	N.º de clones
<i>p</i> -azidoPheRS-5	Ile	Asp	Asn	Phe	Val	1/6
<i>p</i> -azidoPheRS-6	Thr	Asn	Ser	Ala	Leu	1/6

^a Estos clones también contienen una mutación Asp165Gly**EJEMPLO 3: ADICIÓN DE AMINOÁCIDO CON NUEVA REACTIVIDAD AL CÓDIGO GENÉTICO DE EUCARIOTAS**

- 5 Se demuestra un método específico de sitio, rápido, fiable e irreversible de bioconjugación con proteínas basado en una cicloadición [3+2]. Existe una considerable necesidad de reacciones químicas que modifiquen las proteínas en condiciones fisiológicas de una manera altamente selectiva. Véase, por ejemplo, Lemineux, y Bertozzi, (1996) TIBTECH, 16:506-513. La mayoría de las reacciones utilizadas actualmente para la modificación selectiva de proteínas implican la formación de un enlace covalente entre compañeros de reacción nucleófila o electrófila, por ejemplo, la reacción de α -halocetonas con las cadenas laterales de histidina o cisteína. La selectividad en estos casos se determina por el número y accesibilidad de los residuos nucleófilos en la proteína. En el caso de proteínas sintéticas o semisintéticas, se pueden usar otras reacciones más selectivas, tal como la reacción de un ceto aminoácido artificial con hidrazidas o compuestos aminooxi. Véase, por ejemplo, Cornish, et al., (1996) Am. Chem. Soc., 118:8150-8151; y, Mahal, et al., (1997) Science, 276:1125-1128. Recientemente, ha sido posible codificar genéticamente aminoácidos artificiales (véase, por ejemplo, Wang, et al., (2001) Science 292:498-500; Chin, et al., (2002) Am. Chem. Soc. 124:9026-9027; y, Chin, et al., (2002) Proc. Natl. Acad. Sci., 99:11020-11024), incluidos los aminoácidos que contienen cetona (véase, por ejemplo, Wang, et al., (2003) Proc. Natl. Acad. Sci., 100:56-61; Zhang, et al., (2003) Biochemistry, 42:6735-6746; y, Chin, et al., (2003) Science, en prensa), en bacterias y levaduras usando pares de ARNt-sintetasa ortogonales con especificidades de aminoácidos alteradas. Esta metodología ha hecho posible el etiquetado selectivo de prácticamente cualquier proteína con una gran cantidad de reactivos, incluidos fluoróforos, agentes de reticulación y moléculas citotóxicas.

- Se describe un método altamente eficiente para la modificación selectiva de proteínas, que implica la incorporación genética de azida o acetileno que contiene aminoácidos artificiales en proteínas en respuesta a, por ejemplo, el codón ámbar sin sentido, TAG. Estas cadenas laterales de aminoácidos pueden modificarse a continuación mediante una reacción de cicloadición [3+2] de Huisgen (véase, por ejemplo, Padwa, A. en Comprehensive Organic Synthesis, Vol. 4, (1991) Ed. Trost, B. M., Pergamon, Oxford, pág. 1069-1109; y, Huisgen, R. en 1,3-Dipolar Cycloaddition Chemistry, (1984) Ed. Padwa, A., Wiley, Nueva York, pág. 1-176) con derivados alquínico (acetileno) o azida, respectivamente. Debido a que este método implica la cicloadición en lugar de una sustitución nucleófila, las proteínas pueden modificarse con una selectividad extremadamente alta (otro método que puede usarse es el intercambio de ligando en un compuesto bisarsénico con un motivo de tetracisteína, véase, por ejemplo, Griffin, et al., (1998) Science 281:269-272). Esta reacción puede realizarse a temperatura ambiente en condiciones acuosas con excelente regioselectividad ($1,4 > 1,5$) mediante la adición de cantidades catalíticas de sales de Cu(I) a la mezcla de reacción. Véase, por ejemplo, Tornøe, et al., (2002) Org. Chem. 67:3057-3064; y, Rostovtsev, et al., (2002) Angew. Chem. Int. Ed. 41:2596-2599. De hecho, Finn y colaboradores han demostrado que esta cicloadición [3+2] de azida-alquino puede realizarse en la superficie de un virus de mosaico de caupí intacto. Véase, por ejemplo, Wang, et al., (2003) J. Am. Chem. Soc., 125:3192-3193. Para otro ejemplo reciente de la introducción electrófila de un grupo azido en una proteína y una posterior cicloadición [3+2], véase, por ejemplo, Speers, et al., (2003) J. Am. Chem. Soc., 125:4686-4687.

- Para introducir selectivamente el grupo funcional alquínico (acetileno) o azida en proteínas eucariotas en sitios únicos, se generaron pares de TyrRS/ARNt_{CUA} desarrollados en levadura que codifican genéticamente los aminoácidos acetileno y azido, **figura 11, 1 y 2**, respectivamente. Las proteínas resultantes pueden marcarse eficaz y selectivamente con fluoróforos en una reacción de cicloadición posterior en condiciones fisiológicas.

- Previamente, se demostró que un par de tirosil ARNt-ARNt sintetasa de *E. coli* era ortogonal en levadura, es decir, ni el ARNt ni la sintetasa reaccionan de forma cruzada con el ARNt de levadura endógena o las sintetasas. Véase, por ejemplo, Chin, et al., (2003) Chem. Biol., 10:511-519. Este par de ARNt-sintetasa ortogonal se ha utilizado para incorporar selectiva y eficientemente una cantidad de aminoácidos artificiales en la levadura en respuesta al codón TAG (por ejemplo, Chin, et al., (2003) Science, en prensa). Para alterar la especificidad de aminoácidos de la tirosil-ARNt sintetasa de *E. coli* para aceptar el aminoácido **1** o **2** de la **figura 11**, se generó una biblioteca de $\sim 10^7$ mutantes aleatorizando los codones para Tyr³⁷, Asn¹²⁶, Asp¹⁸², Phe¹⁸³, y Leu¹⁸⁶. Estos cinco residuos fueron elegidos basándose en una estructura cristalina de la sintetasa homóloga de *B. stearrowthenophilus*. Para obtener una sintetasa para la cual el aminoácido particular sirve como sustrato, se usó un esquema de selección en el que los codones para Thr⁴⁴ y Arg¹¹⁰ del gen para el activador transcripcional GAL4 se convirtieron en codones sin sentido ámbar (TAG). Véase, por ejemplo, Chin, et al., (2003) Chem. Biol., 10:511-519. La supresión de estos codones ámbar en la cepa de levadura MaV203:pGADGAL4 (2TAG) conduce a la producción de GAL4 de longitud completa (véase, por ejemplo, Keegan, et al., (1986) Science, 231:699-704; y, Ptashne, (1988) Nature, 335:683-689) que a su vez impulsa la expresión de los genes indicadores HIS3 y URA3. Los últimos productos genéticos complementan la auxotrofia de histidina y de uracilo permitiendo que los clones que albergan mutantes de sintetasa activa se seleccionen en presencia de **1** o **2** de la **figura 11**. Las sintetasas que cargan aminoácidos endógenos se eliminan

por crecimiento en medio que carece de **1** o **2** de la **figura 11** pero que contiene ácido 5-fluoroorótico, que se convierte en un producto tóxico por *URA3*. Al pasar la biblioteca a través de tres rondas de selección (positiva, negativa, positiva), se identificaron sintetasas selectivas para **1** de la **figura 11** (pPR-EcRS1-5) y para **2** de la **figura 11** (pAZ-EcRS1-6) como se muestra en la Tabla 8.

Todas las sintetasas muestran una fuerte similitud de secuencia, incluyendo un Asn¹²⁶ conservado, lo que sugiere un papel funcional importante para este residuo. Sorprendentemente, las sintetasas pPR-EcRS-2 y pAZ-EcRS-6, desarrolladas para unirse a **1** y **2** de la **figura 11** respectivamente, convergieron en la misma secuencia (Tyr³⁷ → Thr³⁷, Asn¹²⁶ → Asn¹²⁶, Asp¹⁸² → Ser¹⁸², y Phe¹⁸³ → Ala¹⁸³, Leu¹⁸⁶ → Leu¹⁸⁶). Ambos enlaces de hidrógeno entre el grupo hidroxifenólico de tirosina unida y Tyr³⁷ y Asp¹⁸² están alterados por mutaciones en Thr y Ser, respectivamente. Phe¹⁸³ se convierte en Ala, posiblemente proporcionando más espacio para el alojamiento del aminoácido artificial. Para confirmar la capacidad de esta sintetasa (y las otras sintetasas) de aceptar cualquiera de los aminoácidos como sustrato, se cultivaron cepas de selección que albergaban los plásmidos de sintetasa en medios carentes de uracilo (se obtuvieron los mismos resultados para los medios carentes de histidina) pero complementados con **1** o **2** de la **figura 11**. Los resultados del crecimiento revelaron que cuatro de las cinco alquino sintetasas pudieron cargar ambos aminoácidos artificiales en su ARNt. Las azido sintetasas parecen ser más selectivas, ya que solo pAZ-EcRS-6 (que es idéntico a pPR-EcRS-2) pudo aminoacilar su ARNt con **1** y **2** de la **figura 11**. El hecho de que no se detectó crecimiento en ausencia de **1** o **2** de la **figura 11** sugiere que las sintetasas no aceptan ninguno de los 20 aminoácidos comunes como sustrato. Véase la **figura 14**.

Para todos los experimentos adicionales se usó pPR-EcRS-2 (pAZ-EcRS-6), lo que permite controlar qué aminoácido artificial se incorpora simplemente añadiendo **1** o **2** de la **figura 11** a los medios que contienen la cepa de expresión. Para la producción de proteínas, el codón para el residuo permisivo Trp³³ de la superóxido dismutasa-1 humana (SOD) fusionada a una etiqueta 6xHis C-terminal se mutó en TAG. Por ejemplo, la superóxido dismutasa humana (Trp³³ TAG) HIS se clonó entre el promotor GAL1 y el terminador CYC1 de pYES2.1 (Invitrogen, Carlsbad, CA EE.UU.). Los genes de sintetasa mutante y ARNt en plásmidos derivados de pECTyrRS-ARNt_{CUA} se co-transformaron con pYES2.1 SOD(Trp³³ TAG) HIS en la cepa InvSc (Invitrogen). Para la expresión de proteínas, las células se cultivaron en SD -trp, -ura + rafinosa y la expresión se indujo a una DO₆₀₀ de 0,5 mediante la adición de galactosa. La proteína se expresó en presencia o ausencia de 1 mM de **1** o **2** de la **figura 11** y se purificó por cromatografía Ni-NTA (Qiagen, Valencia, CA, EE.UU.).

El análisis por SDS-PAGE y transferencia de Western reveló una expresión de proteína dependiente de aminoácidos artificiales con una fidelidad de >99 % según lo juzgado por las comparaciones de densitometría con la expresión de proteína en ausencia de **1** o **2** de la **figura 11**. Véase la **figura 12**. Para confirmar adicionalmente la identidad del aminoácido incorporado, se sometió una digestión triptica a cromatografía líquida y espectrometría de masas en tándem.

Por ejemplo, la hSOD mutante y de tipo silvestre se purificó usando una columna de afinidad de níquel y las bandas de proteínas se visualizaron mediante tinción coloidal de Coomassie. Las bandas de gel correspondientes a SOD de tipo silvestre y mutante se cortaron de geles de poliacrilamida, se cortaron en cubos de 1,5 mm, se redujeron y se alquilaron, a continuación se sometieron a hidrólisis de tripsina esencialmente como se describe. Véase, por ejemplo, Shevchenko, A et al., (1996) Anal. Chem. 68:850-858. Los péptidos tripticos que contienen el aminoácido artificial se analizaron por HPLC/μESI/MS de fase inversa de nanoflujo con un espectrómetro de masas con trampa de iones LCQ. Véase la **figura 15, Panel A y B**. El análisis de espectrometría de masas en tándem de cromatografía líquida (LC-MS/MS) se realizó en un espectrómetro de masas con trampa de iones Deca Finnigan LCQ (Thermo Finnigan) equipado con una HPLC Nanospray (serie Agilent 1100).

Los iones precursores correspondientes a los iones precursores cargados individual y doblemente del péptido VY*GSIK (SEQ ID NO:87) que contienen el aminoácido artificial (representado Y*) se separaron y se fragmentaron con un espectrómetro de masas con trampa de iones. Las masas de iones de fragmentos podrían asignarse sin ambigüedades, confirmando la incorporación específica del sitio de cada aminoácido artificial. El análisis por LC MS/MS no indicó la incorporación de ningún aminoácido natural en esta posición. La relación señal-ruido del péptido para todos los mutantes fue >1000, lo que sugiere una fidelidad de incorporación mejor del 99,8 %. Véase la **figura 15, Panel A y B**.

Para demostrar que las moléculas orgánicas pequeñas se pueden conjugar con proteínas mediante una reacción de cicloadición [3+2] de azida-alquino, los colorantes **3-6** indicados en la **figura 13, Panel A**, que contienen un grupo acetilénico o azido y portan un fluoróforo de dansilo o fluoresceína, se sintetizaron (véase el Ejemplo 5 en el presente documento). La propia cicloadición se realizó con proteína 0,01 mM en tampón fosfato (PB), pH 8, en presencia de 2 mM de **3-6** indicado en la **figura 13, Panel A**, CuSO₄ 1 mM y ~1 mg de alambre de Cu durante 4 horas a 37 °C (véase la **figura 13, Panel B**).

Por ejemplo, a 45 μl de proteína en tampón PB (pH = 8) se le añadió 1 μl de CuSO₄ (50 mM en H₂O), 2 μl de colorante (50 mM en EtOH), 2 μl de tris(1-bencil-1H-[1,2,3]triazol-4-ilmetil)amina (50 mM en DMSO), y alambre de Cu. Después de 4 horas a temperatura ambiente o a 37 °C o durante una noche a 4 °C, se añadieron 450 μl de H₂O y la mezcla se centrifugó a través de una membrana de diálisis (corte de 10 kDa). Después de lavar el sobrenadante

con 2x500 µl por centrifugación, la solución se llevó a un volumen de 50 ml. Se analizó una muestra de 20 ml por SDS-PAGE. Ocasionalmente, el colorante restante podría eliminarse del gel sumergiéndolo en H₂O/MeOH/AcOH (5:5:1) durante una noche. El uso de tris(carboxietil)fosfina como agente reductor generalmente condujo a un etiquetado menos eficiente. En contraste con observaciones anteriores (por ejemplo, Wang, Q. et al., (2003) J. Am. Chem. Soc. 125:3192-3193), la presencia o ausencia del ligando de tris(triazolil)amina no tenía una influencia sustancial en el resultado de la reacción.

Después de la diálisis, las proteínas marcadas se analizaron a continuación por SDS-PAGE y se tomaron imágenes en gel usando un densitómetro en el caso de los colorantes de dansilo **3-4** indicados en la **figura 13, Panel A** ($\lambda_{\text{ex}} = 337 \text{ nm}$, $\lambda_{\text{em}} = 506 \text{ nm}$) o un phosphorimager en el caso de los colorantes de fluoresceína **5-6** indicados en la **figura 13, Panel A** ($\lambda_{\text{ex}} = 483 \text{ nm}$, $\lambda_{\text{em}} = 516 \text{ nm}$). Véanse, *por ejemplo*, Blake, (2001) Curr. Opin. Pharmacol., 1:533-539; Wouters, et al., (2001) Trends in Cell Biology 11:203-211; y, Zacharias, et al., (2000) Curr. Opin. Neurobiol., 10:416-421. Las proteínas marcadas se caracterizaron por análisis por LC MS/MS de digeridos tripticos que mostraban una fijación específica de sitio de los fluoróforos y la conversión fue del 75 % en promedio (por ejemplo, según lo determinado por la comparación de los valores de A_{280}/A_{495} para SOD marcado con **5** o **6** indicados en la **figura 13, Panel A**). La selectividad de esta bioconjugación se verifica por el hecho de que no hubo reacción observable entre **3** indicado en la **figura 13, Panel A** y la proteína alquino, o **4** indicado en la **figura 13, Panel A** y la proteína azido.

TABLA 8 SINTETASAS DESARROLLADAS.

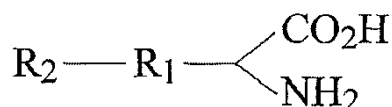
pPR-EcRS seleccionada para 1 y pAZ-EcRS seleccionada para 2 (como se indica en la figura 11)

sintetasa	37	126	182	183	186
tipo silvestre	Tyr	Asn	Asp	Phe	Leu
pPR-EcRS-1	Gly	Asn	Ser	Met	Leu
pPR-EcRS-2	Thr	Asn	Ser	Ala	Leu
pPR-EcRS-3	Ser	Asn	Thr	Met	Val
pPR-EcRS-4	Ala	Asn	Ser	Tyr	Leu
pPR-EcRS-5	Ala	Asn	Thr	Met	Cys
pAZ-EcRS-1	Leu	Asn	Ser	Met	Ala
pAZ-EcRS-2	Val	Asn	Ser	Ala	Ala
pAZ-EcRS-3	Leu	Asn	Ser	Ala	Ala
pAZ-EcRS-4	Val	Asn	Ser	Ala	Val
pAZ-EcRS-5	Ile	Asp	Asn	Phe	Val
pAZ-EcRS-6	Thr	Asn	Ser	Ala	Leu

EJEMPLO 4: SÍNTESIS DE UN AMINOÁCIDO ALQUINO

En un aspecto, la presente divulgación proporciona aminoácidos alquinilo. Un ejemplo de una estructura del aminoácido alquinilo se ilustra en la Fórmula IV:

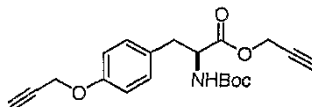
IV



Un aminoácido alquino es normalmente cualquier estructura que tenga la Fórmula IV, donde R₁ es un sustituyente usado en uno de los veinte aminoácidos naturales y R₂ es un sustituyente alquinilo. Por ejemplo, 1 en la **figura 11** ilustra la estructura de *para*-propargiloxifenilalanina. La *p*-propargiloxifenilalanina se puede sintetizar, por ejemplo, como se describe a continuación. En esta realización, la síntesis de *p*-propargiloxifenilalanina se puede completar en tres etapas a partir de la N-Boc-tirosina disponible comercialmente.

Por ejemplo, N-*terc*-butoxicarbonil-tirosina (2 g, 7 mmol, 1 equiv.) y K₂CO₃ (3 g, 21 mmol, 3 equiv.) se suspendieron en DMF anhidra (15 ml). Bromuro de propargilo (2,1 ml, 21 mmol, 3 equiv., solución al 80 % en tolueno) se añadió lentamente y la mezcla de reacción se agitó durante 18 horas a temperatura ambiente. Se añadieron agua (75 ml) y Et₂O (50 ml), las capas se separaron y la fase acuosa se extrajo con Et₂O (2 x 50 ml). Las capas orgánicas combinadas se secaron (MgSO₄) y el disolvente se eliminó a presión reducida. El producto se obtuvo en forma de un aceite de color amarillo (2,3 g, 91 %) y se usó en la siguiente etapa sin más purificación. El producto protegido con Boc se ilustra a continuación como estructura química 8:

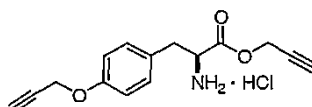
8



Éster propargílico del ácido 2-*tert*-butoxicarbonilamino-3-[4-(prop-2-iniloxy)fenil]-propiónico

- 5 Se añadió cuidadosamente cloruro de acetilo (7 ml) a metanol (60 ml) a 0 °C para dar una solución 5 M de HCl anhidro en MeOH. El producto de la etapa anterior (2 g, 5,6 mmol) se añadió y la reacción se agitó durante 4 horas mientras se dejaba calentar a temperatura ambiente. Después de eliminar los volátiles a presión reducida, un sólido amarillento (1,6 g, 98 %) (véase la estructura química 9) se obtuvo, que se usó directamente en la siguiente etapa.

9



10

Éster propargílico del ácido 2-amino-3-[4-(prop-2-iniloxy)fenil]-propiónico

- 15 El éster propargílico (1,6 g, 5,5 mmol) de la etapa anterior se disolvió en una mezcla de NaOH acuoso 2 N (14 ml) y MeOH (10 ml). Después de la agitación durante 1,5 h a temperatura ambiente, el pH se ajustó a 7 añadiendo HCl conc. Se añadió agua (20 ml) y la mezcla se mantuvo a 4 °C durante una noche. El precipitado se filtró, se lavó con H₂O enfriado con hielo y se secó al vacío, produciendo 1,23 g (90 %) de **1** en la **figura 11** (ácido 2-amino-3-fenilpropiónico (**1**) (también conocido como p-propargiloxifenilalanina) en forma de un sólido de color blanco. ¹H RMN (400 MHz, D₂O) (como la 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); ¹³C RMN (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].
- 20

25 EJEMPLO 5: ADICIÓN DE MOLÉCULAS A PROTEÍNAS CON UNA AMINOÁCIDO ARTIFICIAL A TRAVÉS DE UNA CICLOADICIÓN [3+2]

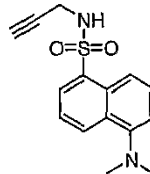
- En un aspecto, la invención proporciona métodos y composiciones relacionadas de proteínas que comprenden aminoácidos artificiales acoplados a moléculas de sustituyentes adicionales. Por ejemplo, se pueden añadir sustituyentes adicionales a un aminoácido artificial a través de una cicloadición [3+2]. Véase, por ejemplo, la **figura**
- 30 **16**. Por ejemplo, la cicloadición [3+2] de una molécula deseada (por ejemplo, que incluye un segundo grupo reactivo, tal como un enlace triple alquino o un grupo azido) a una proteína con un aminoácido artificial (por ejemplo, que tiene un primer grupo reactivo, tal como un grupo azido o enlace triple) puede hacerse siguiendo las condiciones publicadas para la reacción de cicloadición [3+2]. Por ejemplo, una proteína que comprende el aminoácido artificial en el tampón PB (pH = 8) se añade a CuSO₄, la molécula deseada y el alambre de Cu. Después de incubar la mezcla (por ejemplo, aproximadamente 4 horas a temperatura ambiente o 37 °C, o durante una noche a 4 °C), se añade H₂O y la mezcla se filtra a través de una membrana de diálisis. La muestra puede analizarse para la adición, por ejemplo, mediante análisis en gel.
- 35

- Ejemplos de dichas moléculas incluyen, pero sin limitación, por ejemplo, una molécula que tiene un triple enlace o grupo azido, tal como las moléculas tienen la estructura de Fórmula **3**, **4**, **5** y **6** de la **figura 13, Panel A** y similares. Además, se pueden incorporar enlaces triples o grupos azido en las estructuras de otras moléculas de interés, tales como polímeros (por ejemplo, poli(etilenglicol) y derivados), agentes de reticulación, colorantes adicionales, fotorreticuladores, compuestos citotóxicos, etiquetas de afinidad, biotina, sacáridos, resinas, perlas, una segunda proteína o polipéptido, quelatos metálicos, cofactores, ácidos grasos, carbohidratos, polinucleótidos (por ejemplo, ADN, ARN, etc.), y similares, que también pueden usarse en cicloadiciones [3+2].
- 40
- 45

- En un aspecto, las moléculas que tienen la Fórmula **3**, **4**, **5**, o **6** de la **figura 13, Panel A** pueden sintetizarse como se describe a continuación. Por ejemplo, un colorante alquino como se muestra en **3** de la **figura 13, Panel A** y en una estructura química 3 a continuación se sintetizó añadiendo propargilamina (250 µl, 3,71 mmol, 3 equiv.) 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 la agitación durante 1 hora, la mezcla de reacción se calentó a temperatura ambiente y se agitó durante 1 hora más. Los volátiles se eliminaron al vacío y el producto bruto se purificó por cromatografía sobre gel de sílice (Et₂O/hexanos = 1:1), produciendo **3** de la **figura 13, Panel A** (418 mg, 78 %) en forma de un sólido de color amarillo. Los datos analíticos son idénticos a los indicados en la bibliografía. Véanse, por ejemplo, Bolletta, F et al., (1996) Organometallics 15:2415-17. Un ejemplo de una estructura de un colorante alquino que
- 50
- 55

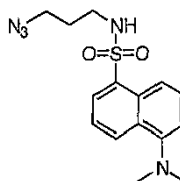
puede usarse en la invención se muestra en la estructura química 3:

3



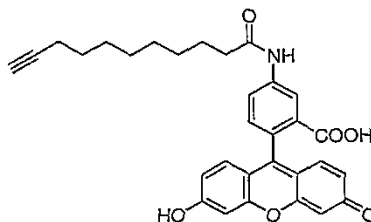
- 5 Un colorante azido como se muestra en 4 de la **figura 13, Panel A** y en la estructura química 4 a continuación se sintetizó añadiendo 3-azidopropilamina (por ejemplo, como se describe por Carboni, B et al., (1993) J. Org. Chem. 58:3736-3741) (371 mg, 3,71 mmol, 3 equiv.) a una solución de cloruro 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 la agitación durante 1 hora, la mezcla de reacción se calentó a temperatura ambiente y se agitó durante 1 hora más. Los volátiles se eliminaron al vacío y el producto bruto se purificó por cromatografía sobre gel de sílice (Et₂O/hexanos = 1:1), produciendo **4** de la **figura 13, Panel A** (548 mg, 89 %) en forma de un aceite de color amarillo. ¹H RMN (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 (s a, 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); ¹³C RMN (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 átomos de carbono cuaternario son visibles en el espectro de ¹³C RMN); HRMS (CI) *m/z* 334,1336 [C₁₅H₂₀N₅O₂S (M+1) requiere 334,1332]. Un ejemplo de una estructura de un colorante azido se muestra en la estructura química 4:

4



- 20 Un colorante alquino como se muestra en **5** de la **figura 13, Panel A** y en la estructura química 5 a continuación se sintetizó añadiendo EDCI (clorhidrato de 1-etil-3-(3-dimetilaminopropil)carbodiimida) (83 mg, 0,43 mmol, 1 equiv.) a una solución de fluoresceinamina (150 mg, 0,43 mmol, 1 equiv.) y ácido 10-undecinoico (79 mg, 0,43, 1 equiv.) en piridina (2 ml) a temperatura ambiente. La suspensión se agitó durante una noche y la mezcla de reacción se vertió en H₂O (15 ml). La solución se acidificó (pH <2) mediante la adición de HCl conc. Después de agitar durante 1 h, el precipitado se retiró por filtración, se lavó con H₂O (5 ml) y se disolvió en una pequeña cantidad de EtOAc. La adición de hexanos condujo a la precipitación de **5** de la **figura 13, Panel A** en forma de cristales de color naranja, que se recogieron y se secaron al vacío (138 mg, 63 %). Los datos analíticos son idénticos a los indicados en la bibliografía. Véase, por ejemplo, Crisp, G. T.; y Gore, J. (1997) Tetrahedron 53:1505-1522. Un ejemplo de una estructura de un colorante alquino se muestra en la estructura química 5:

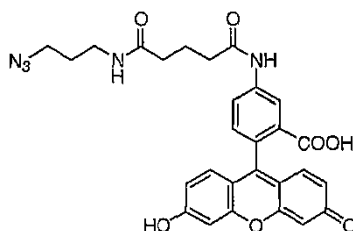
5



- 35 Un colorante azido como se muestra en **6** de la **figura 13, Panel A** y en la estructura química 6 a continuación se sintetizó añadiendo 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)-bútrico (por ejemplo, sintetizado por reacción de 3-azidopropilamina con anhídrido del ácido glutámico) (92 mg, 0,43, 1 equiv.) en piridina (2 ml) a temperatura ambiente. La suspensión se agitó durante una noche y la mezcla de reacción se vertió en H₂O (15 ml). La solución se acidificó (pH <2) mediante la adición de HCl conc. Tras agitar durante 1 hora, el precipitado se retiró por filtración, se lavó con HCl 1 N (3x3 ml)

y se disolvió en una pequeña cantidad de EtOAc. La adición de hexanos condujo a la precipitación de **6** de la **figura 13, Panel A** en forma de cristales de color naranja, que se recogieron y se secaron al vacío (200 mg, 86 %). ^1H RMN (400 MHz, CD_3OD) 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 (s, 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); ^{13}C RMN (100 MHz, CD_3OD) 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 (no todas las señales de átomos de carbono cuaternario son visibles en el espectro de ^{13}C RMN); HRMS (CI) m/z 544,1835 [$\text{C}_{28}\text{H}_{25}\text{N}_5\text{O}_7$ (M+1) requiere 544,1827]. Un ejemplo de una estructura de un colorante azido se muestra en la estructura química 6:

6



En una realización, una molécula de PEG también se puede añadir a una proteína con un aminoácido artificial, por ejemplo, un aminoácido azido o un aminoácido propargilo. Por ejemplo, una propargil amida PEG (por ejemplo, ilustrada en la **figura 17, Panel A**) puede añadirse a una proteína con un aminoácido azido a través de una cicloadición [3+2]. Véase, por ejemplo, la **figura 17, Panel A**. La **figura 17, Panel B** ilustra un análisis en gel de una proteína con un sustituyente PEG añadido.

En un aspecto, una propargil amida PEG (por ejemplo, ilustrada en la **figura 17, Panel A**) puede sintetizarse como se describe a continuación. Por ejemplo, una solución de propargilamina (30 μl) en CH_2Cl_2 (1 ml) se añadió al PEG-hidroxisuccinimida éster de 20 kDa (120 mg, adquirido en Nektar). La reacción se agitó durante 4 horas a temperatura ambiente. Después, se añadió Et_2O (10 ml), el precipitado se eliminó por precipitación, y se recrystalizó dos veces en MeOH (1 ml) mediante la adición de Et_2O (10 ml). El producto se secó al vacío proporcionando un sólido de color blanco (105 mg, 88 % de rendimiento). Véase, por ejemplo, la **figura 17, Panel C**.

EJEMPLO 6: O-RS Y O-ARNt A MODO DE EJEMPLO.

Un O-ARNt a modo de ejemplo comprende la SEQ ID NO.:65 (Véase la Tabla 5). Las O-RS a modo de ejemplo incluyen las SEQ ID NOs.: 36-63, 86 (Véase la Tabla 5). Los ejemplos de polinucleótidos que codifican las O-RS o porciones de las mismas (por ejemplo, el sitio activo) incluyen las SEQ ID NOs.: 3-35. Además, los cambios de aminoácidos a modo de ejemplo de O-RS se indican en la Tabla 6.

Tabla 6: Variantes de EcTyrRS desarrolladas

Residuo N.º	37	126	182	183	186	Representación
Ec TyrRS	Tyr	Asn	Asp	Phe	Leu	
p-yodoPheRS-1	Val	Asn	Ser	Tyr	Leu	1/8
p-yodoPheRS-2	Ile	Asn	Ser	Met	Leu	1/8
p-yodoPheRS-3	Val	Asn	Ser	Met	Ala	6/8
OMeTyrRS-1	Val	Asn	Ser	Met	Leu	5/13
OMeTyrRS-2	Thr	Asn	Thr	Met	Leu	1/13
OMeTyrRS-3	Thr	Asn	Thr	Tyr	Leu	1/13
OMeTyrRS-4	Leu	Asn	Ser	Met	Ser	1/13
OMeTyrRS-5	Leu	Asn	Ser	Met	Ala	1/13
OMeTyrRS-6	Thr	Asn	Arg	Met	Leu	4/13
p-acetilPheRS-1	Ile	Asn	Gly	Met	Ala	4/4
p-acetilPheRS-1 ^a	Ile	Asn	Gly	Met	Ala	10/10
p-benzoilPheRS-1	Gly	Asn	Gly	Phe	Ala	1/2
p-benzoilPheRS-2	Gly	Asn	Gly	Tyr	Met	1/2
p-azidoPheRS-1	Leu	Asn	Ser	Met	Ala	1/6
p-azidoPheRS-2	Val	Asn	Ser	Ala	Ala	1/6
p-azidoPheRS-3	Leu	Asn	Ser	Ala	Ala	1/6
p-azidoPheRS-4	Val	Asn	Ser	Ala	Val	1/6
p-azidoPheRS-5	Ile	Asp	Asn	Phe	Val	1/6
p-azidoPheRS-6	Thr	Asn	Ser	Ala	Leu	1/6
p-PR-EcRS-1	Gly	Asn	Ser	Met	Leu	1/10
p-PR-EcRS-2	Thr	Asn	Ser	Ala	Leu	1/10

(continuación)						
Residuo N.º	37	126	182	183	186	Representación
p-PR-EcRS-3	Ser	Asn	Thr	Met	Val	1/10
p-PR-EcRS-4	Ala	Asn	Ser	Tyr	Leu	1/10
p-PR-EcRS-5	Ala	Asn	Thr	Met	Cys	1/10
p-PR-EcRS-6	Thr	Asn	Thr	Phe	Met	1/10
p-PR-EcRS-7	Thr	Asn	Ser	Val	Leu	1/10
p-PR-EcRS-8	Val	Asn	Ser	Met	Thr	1/10
p-PR-EcRS-9	Ser	Asn	Ser	Phe	Leu	1/10
p-PR-EcRS-10	Thr	Asn	Thr	Phe	Thr	1/10
* Estos clones también contienen una mutación Asp165Gly						

TABLA 5

SEQ ID NO.:	Etiqueta	SECUENCIA
SEQ ID NO.:1	Polinucleótido de TyrRS (sintetasa) de tipo silvestre de <i>E. coli</i>	ATGGCAAGCAGTAACCTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGG CCCGATCGCGCTCTATTGCGGCTTCGATCCCTACCGCTGACAGCTTGCAT TTGGGGCATCTTGTTCATTGTTATGCCTGAAACGCTTCCAGCAGCGGG GCCACAAGCCGGTTGCGCTGGTAGCGGCGCGACGGGTCTGATTGGCG ACCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTG TTCAGGAGTGGTGGACAAAATCCGTAAGCAGGTTGCCCGTTCCCTCG ATTTGCACTGTGGAGAAAACCTCTGCTATCGCGGCGAAACAATATGACT GGTTCGGCAATATGAATGTGCTGACCTTCTCTGCGGATATTGGCAACA CTCTCCGTTAACAGATGATCAACAAGAAGCGGTTAAGCAGCGTCT CAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTCTTACAACTG TTGCAGGGTTATGACTTCGCTGTCTGAAACAACAAGTACGGTGTGGTGC TGCAAAATTGGTGGTTCTGACCAGTGGGTAACTCACTTCTGGTATCGA CCTGACCCGTCGTCGATCAGAAATCAGGTGTTTGGCTGACCGTTCCG CTGATCACTAAAGCAGATGGCACCAAAATTGTTAAAACTGAAGGCGGC GCAGTCTGGTTGGATCCGAAGAAACCAAGCCGTACAAATTTCTACCA TTCTGGATCAACACTCGGGATGCCGACGTTTACCGCTTCTGAAATCT TCACCTTTATGAGCATTGAAGAGATCAACGCCCTGGAAAGAAGATA AAAACAGCGGTAAGCACCGCGCCAGTATGTACTTGGCGGAGCAG GTGACTCGTCTGGTTACGGTGAAGAAGGTTTACAGCGGCAAAACGT ATTACCGAATGCCCTGTTCAGCGGTTCTTTGAGTGCCTGAGTGAAGCGG ACTTCGAACAGCTGGCGCAGGACGGCTACCGATGGTTGAGATGGAAA AGGCGCAGACCTGATGCAGGCACTGGTCGATTCTGAACTGCAACCTT CCCGTGGTCAAGCAGTAAAACATATCGCCTCCAATGCCATCACCAATA CGGTGAAAAACAGTCCGATCCTGAATACTTCTTTAAAGAAGAAGATCG TCTGTTTGGTCGTTTACCTTACTGCGTCGCGGTAAAAAGAATTACTGT CTGATTGCTGGAAATAA

(continuación)	
SEQ ID NO.:	Etiqueta
SEQ ID NO.: 2	Aminoácido (aa) de TyrRS (sintetasa) de tipo silvestre de <i>E. coli</i>
	SECUENCIA MASSNLJKQLQERGLVAQVTDEEALAEERLAQGPIALYCGFDPTADSLHLGH LVPLCLCKRFQAGHKPVAVGGATGLIGDPSFKAERKLNTEETVQEWV DKRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYDFACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALLEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAARKITECLFSGSLSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRQARKTASNAITNGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLJCWK

SEQ ID NO.:	Etiqueta	(continuación)
SEQ ID NO.: 3	Polinucleótido pOMe-1 sintetasa	SECUENCIA ATGGCAAGCAGTAACCTTGATTAAACAATTGCAAGAGCGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCGTGTGGCTTCGATCCTACCGCTGACAGCTTGCAATT GGGCATCTTGTTCATTGTTATGCCTGAAACGCTTCAGCAGCGGGC CACAAGCCGGTTCGCTGGTAGCGCGCGCAGCGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCCGAAGAACTGTT CAGGAGTGGTGGAACAAATCCGTAGCAGGTGGCCCGTTCCTCGATT TCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATAATTATGACTGGTT CGGCAATATGAATGTGCTGACCTTCTCGCGGATATTGGCAAAACACTTC TCCGTTAACAGATGATCAACAAGAACGCGTTAAGCAGCTCTCAAC CGTGAAGATCAGGGGATTTCTGCTCACTGAGTTTCTCTACAACTGCTGC AGGTTATAGTATGGCTGTTTGAACAACAAGTACGGTGTGGTGTGCA AATTGGTGGTCTGACCAAGTGGGTAAACATCATTCTGGTATCGACCTG ACCGTCTGCTGCATCAGATCAGGTGTTTGGCTGACCGTTCGCTGA TCACTAAAGCAGATGGCACCATAATTGGTAAACTGAAGCGCGCGCAG TCTGGTTGGATCCGAAGAAAACAGCCCGTACAAATTCTACAGTTCTG GATCAACACTGCGGATGCGGACGTTTACCGCTTCTGAACTCTTCACC TTTATGAGCATTGAAGAGATCAACGCCCTGGAGAAGAGATAAAAAC AGCGTAAAGCACCGCGCGCCAGTATGTACTGGCGAGCAGGTGACT CGTCTGGTTACCGTGAAAGAGTTTACAGCGCGCAAAACGTATTACC GAATGCCCTGTTACGCGGTTCCTTGTAGTGCCTGAGTGAAGCGGACTTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACCTGCAACCTTCCGTGG TCAGGCACGTAAAACATATCGCCTCCAAATGCCATCACCATTACCGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAAGAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTCGCGGTAAAAAGAAATTACTGTCTGATTGC TGGAAATAA

SEQ ID NO.:	Etiqueta	(continuación)
SEQ ID NO.: 4	Polinucleótido pOMe-2 sintetasa	SECUENCIA ATGGCAAGCAGTAACCTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA gCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTACTTGTGGCTTCGATCCTACCGCTGACAGCTTGCAATTT GGGGCATCTTGTTCATTTGTTATGCTGAACCGCTCCAGCAGCGGGC CACAAAGCCGGTTGCGCTGGTAGGGCGCGACGGGTCTGATTGGCGAC CCGAGCTTCAAAAGCTGCCGAGCTAAGCTGAACACCCGAAGAACTGTT CAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCCCGCTTCCTCGATT TCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATAATTATGACTGGTT CAGCAATATGAATGTGCTGACCTTCCCTGCGCGATATTGGCAACACTTC TCCGTTAACCCAGATGATCAACAAAGAACGGTTAAGCAGCGTCTCAAC CGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTTCCTACAACCTGCTGC AGGTTATACGTATGCTGTCTGAACAAACAGTACGGTGTGGTGTCTGCA AATTGGTGGTCTGACCAAGTGGGTAAACATCACTTCTGTTATCGACCTG ACCGTCTGCTGTCATCAGATCAAGGTGTTTGGCCCTGACCGTTCGGCTGA TCACTAAAGCAGATGGCACCATAATTGGTAAACTGAAAGCGGGCGCAG TCTGGTTGGATCCGAAGAAAACAGCCCGTACAAATCTACCACTCTG GATCAACACTGGGATGCCGACGTTTACCGCTTCCCTGAAGTCTTCCACC TTTATGAGCATTGAAGAGATCAACGCCCTGGAAAGAAAGATAAAAAC AGCGTAAAGCACCGCGCGCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTCACGGTGAAGAGGTTTACAGCGGCAAAACGTAATTACC GAATGCCCTGTTACGCGGTTCTTTGAGTGCCTGAGTGAAAGCGGACTTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAAAGGGCG CAGACCTGATGCAGGCACCTGGTCGATTCTGAACCTGCAACCTTCCCGTGG TCAGGCACGTAAAAACTATCGCCTCCAATGCCATCACCAATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAAGAGATCGTCTGTTTG GTCGTTTTACCTTACTGCGTGGCGGTAAAAAGAAATTACTGTCTGATTGC TGGAATAA

SEQ ID NO.:	Etiqueta	(continuación) SECUENCIA
SEQ ID NO.: 5	Polinucleótido pOMe-3 sintetasa	ATGGCAAGCAGTAACCTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA GCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCAGCTGGCGCAAGGC CCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTGACAGCTTGCAATTT GGGGCATCTTGTTCATTTGTTATGCCTGAAACGCTTCCAGCAGCGGGGC CACAAAGCCGTTGCGCTGGTAGGGCGCGGACGGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAAGCTGAACACCGAAGAAACTGTT CAGGAGTGGGTGGACAAAATCCGTAAGCAGTTGCCCGGTTCCCTCGATT TCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATAATTATGACTGGTT CGCAATATGAATGTGCTGACCTTCCGCGCGATAATTGGCAAACTTC TCCGTTAACCCAGATGATCAACAAAGAAAGCGGTTAAGCAGCTCTCAAC CGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTTCCTACAACTGCTGC AGGCTATAGTATGGCCTGTTTGAACAACAGTACGGTGTGTGCTGCA AATTGGTGGTCTGACCAAGTGGGTAAACATCACCTTCTGGTATCGACCTG ACCCGTCTCTGCATCAGAAATCAGGTGTTTGGCCTGACCCGTTCCGCTGA TCACTAAAGCAGATGGCACCATAATTTGGTAAACTGAAAGCGGGCGCAG TCTGGTTGGATCCGAAGAAACCAGCCCGTACAAATCTACCACTCTG GATCAACACTGGGGATGCCGACGTTTACCGCTTCCGAAAGTCTTCAAC TTTATGAGCATTGAAGAGATCAACGCCCTGGAAGAAGATAAAAAC AGCGGTAAAGCACCGCGCGCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTCACGGTGAAGAAAGTTTACAGCGGCAAAACGTATTACC GAATGCCCTGTTACGGGTTCTTTGAGTGCCTGAGTGAAGCGGACTCG AACAGCTGGCGCAGGACGGCGTACCGATGGTTGAGATGGAAGAGGCG CAGACCTGATGCAAGGCACTGGTCTGATTCTGAACCTGCAACCTTCCCGTGG TCAGGCACGTAAACCTATCGCCTCCAATGCCATCACCATTACGGTGAA AAACAGTCCGATCCCTGAATACTTCTTTAAAGAAAGATCGTCTGTTG GTCGTTTACCTTACTGCGTCCGGTAAAAAGAAATTACTGTCTGATTGCTG TGGAATAA

SEQ ID NO.: 6	Etiqueta	(continuación)
SEQ ID NO.: 6	Polinucleótido pOMe-4 sintetasa	SECUENCIA ATGGCAAGCAGTAACCTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA GCCAGGTGACGACGAGGAGCGTLAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCGTGTGGCTTCGATCCTACCGCTGACAGCTTGCAATT GGGCATCTTGTTCATTGTTATGCCTGAAACGGTTCCAGAGCGGGC CACAAAGCCGGTTGCGTGGTAGGGCGCGACGGGTCTGATTGGCGAC CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTT CAGGAGTGGGTGGACAAATCCGTAAAGCAGGTTGCCCGGTTCTCGATT TCGACTGTGGAGAAACTCTGCTATCGCGGCCAATAATTATGACTGGTT CGGCAATATGAATGTGCTGACCTTCCTGCGGATATTGGCAACACATTC TCCGTTAACCCAGATGATCAACAAGAAAGCGGTTAAGCAGCGTCTCAAC CGTGAAGATCAGGGGATTTCTGTTCACTGAGTTTCTACAACTGCTGC AGGTTATAGTATGGCCTGTTTGAACAACAGTACGGTGTGGTGTGCA AATTGGTGGTCTGACCAGTGGGTAAACATCACTTCTGGTATCGACCTG ACCGTCTGCTGCATCAGAAATCAGGTGTTTGGCCCTGACCGTCCGCTGA TCACTAAAGCAGATGGCACCAAAATTTGGTAAACTGAAGGGCGGCGAG TCTGGTTGGATCCGAAGAAACCCAGCCGTACAAATCTACAGTCTG GATCAACACTGCGGATGCCGACGTTTACCGCTTCTGAACTCTTCAAC TTATGAGCAITGAAGAGATCAACGCCCTGGAAGAAGAGATAAAAC AGCGTAAAGCACCGCGGCCAGTATGTACTGGCGGAGCAGGTGACT CGTCTGGTTACGGTGAAAGGTTTACAGGGCGGCAAAACGTATTACC GAATGCCTGTTACGGGTCTTGTGATGCGCTGAGTGAAGCGGACTTCG AACAGCTGGCGCAGGACGGGTACCGATGGTTGAGATGGAAAGGGCG CAGACCTGATGCAGGCATGGTCTGATTCGAACTGCAACCTTCCCGTGG TCAGGCACGTAAACTATCGCCTCCAAATGCCATCACCATTACCGTGAA AAACAGTCCGATCCTGAATACTTCTTTAAAGAAAGATCGTCTGTTTG GTCGTTTACCTTACTGCGTCCGCGTAAAGAAATTAATGCTCTGATTGC TGGAAATAA

(continuación)		Etiqueta	SECUENCIA
SEQ ID NO.: 7	Polinucleótido pOMe-5 sintetasa		ATGGCAAGCAGTAACCTTGATTAAACAATTGCAAGAGCGGGGGCTGGTA gCCAGGTGACGGACGAGGAAAGCGTTAGCAGAGCGACTGGCGCAAGGC CCGATCGCACTCACGTGTGGCTTCGATCCTACCGCTGACAGCTTGCATT TGGGCACTCTTGTCCATTGTTATGCCCTGAAACGCTTCCAGCAGCGGG CCACAAGCCGGTGGCTGGTAGCGCGGACGGGTCTGATTGGCGA CCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCCGAAGAAACTGT TCAGGAGTGGTGGACAAAATCCGTAAGCAGGTTGCCCGTTCTCGAT TTCGACTGTGGAGAAAACCTCTGCTATCGCGCCAAATAATTATGACTGGT TCGGCAATATGAATGTGTGACCTTCTCGCGGATATTGGCAACACATT CTCCGTTAACAGATGATCAACAAGAAAGCGGTTAAGCAGCGTCTCAA CCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTCTCTACAGCTGCTG CAGGGTTATACGATGGCTGTCTGAACAAACAGTACGGTGTGTGCTGC AAATTGGTGGTCTGACCAAGTGGGTAACATCACTTCTGGTATCGACCT GACCGTCTGTGCTGATCAGAATCAGGTGTTTGGCCTGACCGTTCCGCTG ATCACTAAAGCAGATGGCACCAAAATTTGGTAAACTGAAGCGCGGCA GTCTGTTGGATCCGAAGAAACCAAGCCGTACAAATCTACCAATTCT GGATCAACACTGCGGATGCCGACGTTACCGCTCTCTGAAAGTTCTTAC CTTTATGAGCAATTGAAGAGATCAACGCCCTGGAAGAAGAGATAAAA CAGCGGTAAGCACCGCGGCCAGTATGTACTGGCGGAGCAGGTGAC TCGCTGTGTTCAAGGTGAAGAAGTTTACAGCGGCAAAACGTTATTACC GAATGCCTGTTCAAGCGGTTCTTTGAGTGCCTGAGTGAAGCGGACTTCG AACAGTGGCGCAGGACGGGTACCGATGGTTGAGATGGAAAGGGCG CAGACCTGATGCAGGCACTGGTCGATTCTGAACCTGCAACCTTCCCGTGG TCAGGCACGTAAAACCTATCGCCTCCAATGCCATCACCAATTAACGGTGAA AAACAGTCCGATCCTGAATACTTCTTAAAGAAGAAGATCGTCTGTGTG GTCGTTTACCTTACTGCGTGGCGGTAAAAAGAAATTACTGTCTGATTGCTG TGGAAATAA

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 8	Polinucleótido pOMe-6 (sitio activo) sintetasa		CGGGGGCTGGTAGCCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGGCATCTTGTTCATTGTTATGCCTGAAACGCTTC CAGCAGCGGGCCACAAAGCGGTTCGCTGGTAGCGGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAACTGTTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCTCGATTTTCGACTGTGGAGAAAATCTGTCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTCTGACCTTCTCGCGGATAT TGGCAAAACACTTCTCCGTTAACCAAGATGATCAACAAAAGAACGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG
SEQ ID NO.: 9	Polinucleótido pOMe-7 (sitio activo) sintetasa		CGGGGGCTGGTACCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGGCATCTTGTTCATTGTTATGCCTGAAACGCTTC CAGCAGCGGGGCCACAAAGCGGTTCGCTGGTAGCGGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAACTGTTTCAGGAGTGGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCTCGATTTTCGACTGTGGAGAAAATCTGTCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTCTGACCTTCTCGCGGATAT TGGCAAAACACTTCTCCGTTAACCAAGATGATCAACAAAAGAACGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 10	Polinucleótido pOMe-8 (sitio activo) sintetasa	CGGGGGCTGGTAGCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTGCTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTGGGGCATCTTGTTCATTGTTATGCCTGAAACGGCTTC CAGCAGGGGGCCACAAAGCGGTTGGCTGGTAGGGGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAAGAGTGGTGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTGATTTCGACTGTGGAGAAAATCTGCTATCGCGGCCAATA ATTATGACTGGTTCAGCAATATGAATGTGCTGACCTTCCTGCGGATAT TGCAAAACACTTCTCCGTTAACAGATGATCAACAAAGAACGGGTTAA GCAGCTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTCC TACAAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAAACAGTACG GTGTG	
SEQ ID NO.: 11	Polinucleótido pOMe-9 (sitio activo) sintetasa	CGGGGGCTGGTAGCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTGCTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTGGGGCATCTTGTTCATTGTTATGCCTGAAACGGCTTC CAGCAGGGGGCCACAAAGCGGTTGGCTGGTAGGGGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAAGAGTGGTGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTGATTTCGACTGTGGAGAAAATCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTGCGGATAT TGCAAAACACTTCTCCGTTAACAGATGATCAACAAAGAACGGGTTAA GCAGCTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTCC TACAAACCTGCTGCAGGGTTATTCGTATGCCTGTGCGAACAACAGTACG GTGTG	
SEQ ID NO.: 12	Polinucleótido pOMe-10 (sitio activo) sintetasa	CGGGGGCTGGTAGCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTGCTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTGGGGCATCTTGTTCATTGTTATGCCTGAAACGGCTTC CAGCAGGGGGCCACAAAGCGGTTGGCTGGTAGGGGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAAGAGTGGTGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTGATTTCGACTGTGGAGAAAATCTGCTATCGCGGCCAATA	

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 13	Polinucleótido pOMe-11 (sitio activo) sintetasa	ATTATGACTGGTTCAGCAATATGAATGTGTGACCTTCCTGCGCGATAI TGGAACAACACTTCTCCGTTAACAGATGATCAACAAGAAGCGGTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATACGTATGCCTGTCTGAACAACAAGTACG GTGTG CGGGGGCTGGTACCcCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCCTTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGGCATCTGTCCATTGTATGCCGTGAACGCTTC CAGCAGCGGGCCACAAGCCGTTGCGTGGTAGGGCGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGATTCGACTGTGGAGAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGTGACCTTCCTGCGCGATAT TGGCAAAACACTTCTCCGTTAACAGATGATCAACAAGAAGCGGTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATCTATTGCCGTTCGAACAACAAGTACG GTGTG	
SEQ ID NO.: 14	Polinucleótido pOMe-12 (sitio activo) sintetasa	CGGGGGCTGGTAGCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCGTGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGGCATCTGTCCATTGTATGCCGTGAACGCTTC CAGCAGCGGGCCACAAGCCGTTGCGTGGTAGGGCGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGATTCGACTGTGGAGAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGTGACCTTCCTGCGCGATAT TGGCAAAACACTTCTCCGTTAACAGATGATCAACAAGAAGCGGTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATAGTATTGCCGTTCGAACAACAAGTACG GTGTG	

(continuación)			
SEQ ID NO.:	Etiqueta	SECuencia	
SEQ ID NO.: 15	Polinucleótido pOMe-13 (sitio activo) sintetasa	CGGGGGCTGTTACCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCGTGTGTGGCTTGGATCTACCGCTG ACAGCTTGCAITTTGGGGCATCTTGTTCCATTGTTATGCCCTGAAACGCTTC CAGCAGGGGGCCACAAGCCGGTTGCGTGGTAGGGGGCGGACGGGT CTGATTGGCGACCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGAITTCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCTCGCGGATAT TGGCAACACTTCTCCGTTAACAGATGATCAACAAGACGGGTAA GCAGGCTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATAGTATTGCCTGTTTGAACAAACAGTACG GTGTG	
SEQ ID NO.: 16	Polinucleótido pOMe-14 (sitio activo) sintetasa	CGGGGGCTGTTAGCCACAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCTGTGTGGCTTGGATCTACCGCTG ACAGCTTGCAITTTGGGGCATCTTGTTCCATTGTTATGCCCTGAAACGCTTC CAGCAGGGGGCCACAAGCCGGTTGCGTGGTAGGGGGCGGACGGGT CTGATTGGCGACCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGAITTCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATT GTTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCTCGCGGATAT TGGCAACACTTCTCCGTTAACAGATGATCAACAAGACGGGTAA GCAGGCTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATAGTATTGCCTGCTGTGAGAAACAACAGTACG GTGTG	
SEQ ID NO.: 17	Polinucleótido p-acetilPhe-1 (sitio activo) sintetasa	CGGGGGCTGTTAGCCACAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCATTTGTGGCTTGGATCTACCGCTG ACAGCTTGCAITTTGGGGCATCTTGTTCCATTGTTATGCCCTGAAACGCTTC CAGCAGGGGGCCACAAGCCGGTTGCGTGGTAGGGGGCGGACGGGT CTGATTGGCGACCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC	

SEQ ID NO.:	Etiqueta	(continuación) SECUENCIA
		CCGTTCTCTCGATTTCGACTGTGGAGAAACTCTGCTATCGCGGCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCTCGCGGATAT TGGCAACACACTTCTCCGTTAACAGATGATCAACAAGCGGTTAA GCAGCGTCTCAACCGTGAAGTCAAGGGATTTCTGCTACAGTGTTC TACAACCTGCTGACAGGTTATGGTATGGCTGTGTAACAACAGTACG GTGTGTGCTGCAAAATGGTGTCTGACCAATGGGTAAACATCACTTC TGGTATCGAAGCTGACCCGTCGTCGATCAGAAATCAGGTG
SEQ ID NO.: 18	Polinucleótido pBenzofenona-1 (sitio activo) sintetasa	CAGGTGACGGACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCG ATCGCACTCGGTTGTGGCTTCGATCCTACCGCTGACAGCTTGCATTTGG GGCATCTGTTCATGTTATGCTGAACGCTTCCAGCAGCGGGCCA CAAGCCGTTGCGCTGTTAGGCGCGCGACGGGTCTGATTGGCGACCC GAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACCGAAGAACTGTTCA GGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCCCGTTCCTCGATTTC GACTGTGGAGAAAACCTCTGCTATCGCGGCAATAATATGACTGGTTTCG GCAATATGAATGTGCTGACCTTCTCGCGGATATTGGCAAAACACTTCTC CGTTAACCAGATGATCAACAAAGAGCGGTTAAGCAGCGTCTCAACCG TGAAAGATCAGGGGATTTCTGTTCACTGAGTTTCTCAACACCTGCTGACG GGTATGTTTGTGCTGTTTGAACAAACAGTACGGTGTGTTGCTGCAAA TTGGTGGTCTGACCAAGTGGGTAAACATCACTTCTGCTATCGACCTGAC CCGTCGTCGATCAGAAATCAGGTG
SEQ ID NO.: 19	Polinucleótido pBenzofenona-2 (sitio activo) sintetasa	GCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTCGGGTGTGGC TTCGATCCTACCGCTGACAGCTTGCAATTGGGGCATCTTGTTCCATTGTT ATGCTGAAACGCTTCCAGCAGCGGGGCCAACAGCCGGTTGCGCTGGT AGCGGCGCGACGGGCTGATTGGCGACCCGAGCTTCAAAGCTGCCGA GCGTAAGCTGAACACCGAAGAACTGTTCAAGGAGTGGGTGGACAAAAT CCGTAAGCAGGTTGCCCGTCTCGATTTCGACTGTGGAGAAAACCTCT GCTATCGCGGCAATAATATGACTGGTTCGGCAATATGAATGTGCTGA CCTTCTGCGGATATTGGCAACACTTCTCCGTTAACAGATGATCAA CAAAGAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTC GTTCACTGAGTTTCTTCAACCTGCTGCAAGGTTATGGTTATGCCTGTA TGAACAAACAGTACGGTGTGGTGTGCAAAATGGTGGTCTGACCAAGTG GGGTAAACATCACTTCTGTTATCGAACCTGACCCGTCGTCTGATCAGAAAT CAGGTG

(continuación)		SECUENCIA	
SEQ ID NO.: 20	Etiqueta Polinucleótido pAzidoPhe-1 (sitio activo) sintetasa	GGGTGTTAGCCAGGTGACGGACGNAGAAAGCGTTAGCAGAGCGACTG GCGAAGGCCGATCGACTCTTTGTGGCTTCGATCCTACCGCTGACA GCTTGCAATTTGGGCACTTTGTTCATTGTTATGCCTGAAACGCTTCCAG CAGCGGGCCACAAGCCGTTGCGTGTAGCGGGCGACGGGTCTG ATTGGGACCCGAGCTTCAAAAGCTGCCGAGCTAAGCTGAACACCGAA GAAACTGTTACAGGAGTGGGTGGACAAAATCCGTAAAGAGTTGCCCG TTCTCGATTTCCGACTGTGGAGAAAACCTGCTATCGCGGCCAATAATT ATGACTGGTTCGGCAATATGAATGTGCTGAACCTTCTCGCGGATATTGG CAAAACACTTCTCCGTTAACCAAGATGATCAACAAAGAGCGGTTAAGCA GCGTCTCAACCGTGAAGATCAGGGGATTTGCTTCACTGAGTTTCTCTAC AACCTGCTGCAGGGTTATTCTATGGCCTGTGCGAACAACAGTACGGTG TGGTGTGCAAAATTGGTGGTTCTGACCAAGTGGGTAACATCAGTTCTGG TATCGACCTGACCCCGTCTGCTGCATCANAAATCANGTG	
SEQ ID NO.: 21	Polinucleótido pAzidoPhe-2 (sitio activo) sintetasa	TTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTCGTTTGTGGCTTCG ATCCTACCGCTGACAGCTTGCAATTTGGGGCATCTTGTTCCATTGTTATGC CTGAAACGCTTCCAGCAGCGGGCCACAAGCCGGTTGCGTGGTAGGC GGCGGACGCGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGT AAGCTGAACACCGAAGAAACTGTTTCAGGAGTGGTGGACAAAATCCGT AAGCAGGTTGCCCGTTCTCGATTTTCGACTGTGGAGAAAACCTCTGCTA TCGCGGCCAATAATTATGACTGGTTCGGCAATATGAATGTGCTGACCTT CCTGCGGATATTGGCAAAACACTTCTCCGTTAACCCAGATGATCAACAAA GAAGCGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTCGTTT ACTGAGTTTCTTCAACACCTGCTGCAGGGTTATTCTCGCGCCTGTGCGA ACAAACAGTACGGTGTGCTGCAAAATTGGTGGTTCTGACCAAGTGGG GTAACATCACTTCTTGGTATCGACCTGACCCCGTCTGCTGCATCAGAAATCA GGTG	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 22	Polinucleótido pAzidoPhe-3 (sitio activo) sintetasa	GACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTC CTGTGTGGCTTCGATCCTACCGCTGACAGCTTGCAATTTGGGGCATCTTGT TCCATTGTATGCCTGAAACGCTTCCAGCAGCGGGCCACAAGCCGGTT GGCTGTGTAGCGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAA GCTGCCGAGCGTAAGCTGAACACCGAAGAAACTGTTCAAGGAGTGGGTG GACAAAATCCGTAGCAGGTTGCCCGTTCCCTCGATTTCCGACTGTGGAG AAAACCTCTGCTATCGCGGCCAATAATATGACTGTTTCGGCAATATGAA TGTGCTGACCTTCTGCGGATATTGGCAACACACTTCTCCGTTAACCCAG ATGATCAACAAANAAGCGGTTAAGCAGCGTCTCAACCGTGAAAGATCAG GGGATTTTCGTTCACTGAGTTTTCCTACAACTGCTGCAGGGTTATTCGGC TGCTGTGCGAAACAAACAGTACGGNGNGNGCTGCAAAATTGGNGGTTTC TGACCAAGGGGGTAACATCACTTCTGGTATOGACCTGACCCCGTCGTCTG CATCAAAATCAGGTG	
SEQ ID NO.: 23	Polinucleótido pAzidoPhe-4 (sitio activo) sintetasa	GC GTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTCGTTTGTGGCT TCGATCCTACCGCTGACAGCTTGCAATTTGGGGCATCTTGTCCATTGTGTG TGCTGAAACGCTTCCAGCAGCGGGCCACAAGCCGGTTGCGCTGGTA GGCGCGCGACGGGTCTGATTGGCGACCCGAGCTTCAAAGCTGCGGAG CGTAAGCTGAACACCGAAGAAACTGTTCAAGGAGTGGTGGACAAAATC CGTAAGCAGGTTGCCCGTTCTCGATTTCCGACTGTGGAGAAACTCTG CTATCGCGGCCAATAATTATGACTGGTTCGGCAATATGAATGTGCTGAC CTTCTGCGCGATATTGGCAACACACTTCTCCGTTAACCCAGATGATCAAC AAAGAAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCG TTCACTGAGTTTTCCTACAACTGCTGCAGGGTTATAGTGGGGCTGTGT TAACAAACAGTACGGTGTGTTGCTGCAAAATTGGTGGTTCTGACCAAGTGG GGTAACATCACTTCTGTTATCGACCTGACCCCGTCGTCTGCATCAGAATC ANGTG	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 24	Polinucleótido pAzidoPhe-5 (sitio activo) sintetasa	GACGAGGAAGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATCGCACTC ATTTGTGGCTTCGATCCTACCGCTGACAGCTTGCAATTTGGGCACTCTTGT TCCATTGTTATGCCTGAAACGCTTCCAGCAGCGGGGCCACAAGCCGGTT GCGCTGGTAGCGCGCGACGGGCTCTGATTGGCGACCCGAGCTTCAAA GCTGCCGAGCGTAAGCTGAACACCCGAAGAAACTGTTCAAGAGTGGTG GACAAAATCCGTAAGCAGGTTGCCCGTTCTTCGATTTCCGACTGTGGAG AAAACTCTGCTATCGGGCAATGATATATGACTGGTTCGGCAATATGAA TGTGCTGACCTTCTCGCGGATATTGGCAACACTTCTCCGTTAACCCAG ATGATCAACAAAGAGCGGTTAAGCAGCGTCTCAACCGTGAAGATCAG GGGATTTCTGTTCACTGAGTTTCTCAACCTGCTGCAGGGTTATAATTI TGCTGTGTGAACAAACAGTACGGTGTGTGCTGCAAAATTGGTGGTCT GACCAAGTGGGTAAACATCACTTCTGTGTAATCGACCTGACCCGTCGTCTGC ATCAGAAATCAGGTG	
SEQ ID NO.: 25	Polinucleótido pAzidoPhe-6 (sitio activo) sintetasa	CGACTGGCGCAAGGCCGATCGCACTCACGTGTGGCTTCGATCCTACCG CTGACAGCTTGCAATTTGGGCACTCTTGTTCATTTGTTATGCCTGAAACGCG TTCCAGCAGGGGGCCACAAGCCGGTTGCGCTGGTAGCGGGCGCGACG GGTCTGATTGGCGACCCGAGCTTCAAAAGCTGCCGAGCGTAAGCTGAAC ACCGAAGAAACTGTTCAGGAGTGGTGGACAAATCCGTAAGCAGGT GCCCGTTCTCGATTTCGACTGTGGAGAAAACCTCTGCTATCGCGGCCA ATAATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCTGCGCGA TATTGGCAACACTTCTCCGTTAACAGATGATCAACAAAGAAGCGGTT AAGCAGCGTCTCAACCGTGAAGATCAGGGGATTTCGTTCACTGAGTTTT CCTACAATCTGCTGCAGGGTTATTCGGCTGCCTGTCTTAAACAAACAGTA CGGTGTGTGCTGCAAAATTGGTGGTTCTGACCAATGGGGTAACATCACT TCGTGTATCGACCTGACCCGTCGTCTGCAATCAGAAATCAGGTG	
SEQ ID NO.: 26	Polinucleótido pPR-EcRS-1 (propargiloxi fenilalanina sintetasa) (sitio activo) sintetasa	CGGGGGCTGGTANCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCGGGTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGCACTCTTGTTCATTGTTATGCCTGAAACGCTTC CAGCAGCGCGGCCACAAGCCGGTTGCGCTGGTAGCGGGCGCGAGCGGT CTGATTGGCGACCCGAGCTTCAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGTGGACAAATCCGTAAGCAGGTGACC CCGTTCTCGATTTCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATA	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 27	Polinucleótido pPR-EcRS -2 (sitio activo) sintetasa	ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCCTGCGCGATAT TGGCAAAACACTTCTCCGTTAACCCAGATGATCAACAAAGACGGTTAA GCAGCTCTCAACCCGTGAAGATCAGGGGATTTCTGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATTTCTATGGCTGTTTGAACAAACAGTACG GTGTGTGCTGCAAAATTGGTGTCTGACCAAGTGGGTAAACATCACTTC TGGTATCGACCTGANCCGTCGCTGTCATCAGAAATCAGGTG CGGGGGCTGTGAGCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCTCTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTGGGGCATCTTGTTCATTGTTATGCCTGAAACGCTTC CAGCAGCGGGGCCACAAGCCGGTTGCGCTGTTAGCGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAAGGAGTGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCTCGATTTCTGACTGTGGAGAAACTCTGCTATCGCGGCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCTCGCGGATAT TGGCAAAACACTTCTCCGTTAACCCAGATGATCAACAAAGACGGTTAA GCAGCTCTCAACCCGTGAAGATCAGGGGATTTCTGTTCACTGAGTTTCC TACAATCTGCTGCAGGGTTATTCGGCTGCTCTTAAACAAACAGTACG GTGTGTGCTGCAAAATTGGTGTCTGACCAAGTGGGTAAACATCACTTC TGGTATCGAACCTGANCCGTCGCTGTCATCAAAATCAAGTG	
SEQ ID NO.: 28	Polinucleótido pPR-EcRS -3 (sitio activo) sintetasa	CGGGGGCTGTGATACCCCAAGTGACGGACGAGGAACGTTAGCAGAGCGA CTGGCGCAAGGCCGATCGCACTCTCTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTGGGGCATCTTGTTCATTGTTATGCCTGAAACGCTTC CAGCAGGCAGGCCACAAGCCGGTTGCGCTGTTAGCGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAAGGAGTGGTGGACAAAATCCGTAAGCAGGTTGCC CCGTTCTCGATTTCTGACTGTGGAGAAACTCTGCTATCGCGGCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCTCGCGGATAT TGGCAAAACACTTCTCCGTTAACCCAGATGATCAACAAAGACGGTTAA GCAGCTCTCAACCCGTGAAGATCAGGGGATTTCTGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATACGATGGCTGTGACCAAGTGGGTAAACATCACTTC GTGTGTGCTGCAAAATTGGTGTCTGACCAAGTGGGTAAACATCACTTC TGGTATCGACCTGACCCGTCGCTGTCATCAGAAATCAGGTG	

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 29	Polinucleótido pPR-EcRS -4 (sitio activo) sintetasa	CGGGGGCTGGTAGCCAGGTGACGGACGAGGAAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGCGTGGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGCATCTTGTCCATTGTTATGCCTGAAACGGTTC CAGCAGCGGGGCCACAAGCCGTTGCGTGGTAGGGCGCGCAGCGGT CTGATTGGCAGCCGAGCTTCAAGGCTGCCAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGATTTCCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATA ATTATGACTGTTTCGGCAATATGAATGTGCTGACCTTCCCTGCGCGATAT TGGCAACACACTCTCCGTTAACCCAGATGATCAACAAAGAGCGGTAA GCAGCTCTCAACCGTGAAGATCAGGGGATTTCTGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATTCTTATGCCTGTCTTAACAAACAGTACG GTGTGGTGTGCAAAATTGGTGGTTCTGACCAAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCCGTCTGTCATCAGAAATCAGGTG	
SEQ ID NO.: 30	Polinucleótido pPR-EcRS-5 (sitio activo) sintetasa	CGGGGGCTGGTAGCCAGGTGACGGACGAGGAAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCGCGTGGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGCATCTTGTCCATTGTTATGCCTGAAACGGTTC CAGCAGCGGGGCCACAAGCCGTTGCGTGGTAGGGCGCGCAGCGGT CTGATTGGCAGCCGAGCTTCAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGATTTCCGACTGTGGAGAAAACCTCTGCTATCGCGCCAATA ATTATGACTGTTTCGGCAATATGAATGTGCTGACCTTCCCTGCGCGATAT TGGCAACACACTCTCCGTTAACCCAGATGATCAACAAAGAGCGGTAA GCAGCTCTCAACCGTGAAGATCAGGGGATTTCTGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATACGATGGCCTGTGTAACAAACAGTACG GTGTGGTGTGCAAAATTGGTGGTTCTGACCAAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCCGTCTGTCATCAGAAATCAGGTG	
SEQ ID NO.: 31	Polinucleótido pPR-EcRS-6 (sitio activo) sintetasa	CGGGGGCTGGTAGCCCAAGTGACGGACGAGGAAAGCGTTAGCAGAGCGA	

SEQ ID NO.:	Etiqueta	(continuación) SECUENCIA
		CTGGCGCAAGGCCGATGCACTACGCTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGCACTTTGTTCCATTTGTTATGCTGAAACGCTTC CAGCAGGGGGCCACAAGCGGTTGGCTGGTAGGGGGCGGACGGGT CTGATTTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTCAGGAGTGGTGACAAAATCCGTAAGCAGGTTGCC CCGTTCTCGAATTCGACTGTGGAGAAAATCTGCTATCGCGGCCAATA ATTATGACTGGTTTCGGCAATATGAATGTGCTGACCTTCCTGCGCGATAT TGGCAAAACACTTCTCCGTTAACAGATGATCAACAAAAGAGCGGTTAA GCAGGTTCTCAACCGTGAAGATCAGGGGATTCGTTGCTGAGTTTCC TACAACTGCTGCAGGGTTATACGTTTGCCTGTATGAACAAACAGTACG GTGTGGTGTGCAAAATTTGGTGGTTCTGACCAGTGGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCGATCAGATCAGGTG
SEQ ID NO.: 32	Polinucleótido pPR-EcRS-7 (sitio activo) sintelasa	GTGACGGAAGGAGGCGTTAGCAGAGCGACTGGCGCAAGGCCCGATC GCACTCACGTGTGGCTTGGATCTACCGCTGACAGCTTGCATTTGGGGC ATCTTGTTCATTTGTTATGCTGAAACGCTTCCAGCAGGGGGCCACAA GCCGTTGGCTGGTAGGGGGCGCGACGGGTCTGATTTGGCGACCCGAG CTCAAAAGCTGCCGAGCGTAAGCTGAACACCCGAAAGAACTGTTCAAG GTGGGTGGACAAAATCCGTAAGCAGGTTGCCCGCTTCCTCGATTTTCGAC TGTGGAGAAAACCTCTGCTATCGCGGCCAATATATGACTGGTTCGGCA ATATGAATGTGCTGACCTTCTGCGCGATATTGGCAAAACACTTCTCCGT TAACCAAGATGATCAACAAAGAGCGGTTAAGCAGCGCTCAACCCGTGA AGATCAGGGGATTTTCGTTCACTGAGTTTTCCTACAACTGCTGCAAGGT TATTCGGCTGCCTGTCTTAAACAACAGTACGGTGTGGTGTGCTGCAAAATG GTGGTTCTGACCAAGTGGGGTAACATCACCTTCTGGTATCGACCTGACCCG TCGTTGCAATCAGAAATCAGGTG

(continuación)		SECUCENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 33	Polinucleótido pPR-EcRS-8 (sitio activo) sintetasa	CGGGGGCTGGTAGCCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGGA CTGGCGCAAGGCCCGATCGCACTCGTTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGGCATCTTGTTCATTTCCATTTGTTATGCCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGTGGTGGCGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGAATTCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTCTGACCTTCCTGCGCGATAT TGGCAAAACACTTCTCCGTTAACAGATGATCAACAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATTCGATGGCTGTACGAACAACAACAGTACG GTGTGGTGTCTGCAAAATTGGTGGTTCTGACCAAGTGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAAATCAGGTG	
SEQ ID NO.: 34	Polinucleótido pPR-EcRS-9 (sitio activo) sintetasa	CGGGGGCTGGTANCCCCAAGTGACGGACGGGGAAGCGTTAGCAGAGCGGA CTGGCGCAAGGCCCGATCGCACTCAGTTTGTGGCTTCGATCCTACCGCTG ACAGCTTGCAATTTGGGGCATCTTGTTCATTTCCATTTGTTATGCCCTGAAACGCTTC CAGCAGGCGGGCCACAAGCCGGTTGCGTGGTGGCGGCGGACGGGT CTGATTGGCGACCCGAGCTTCAAAGCTGCCGAGCGTAAGCTGAACACC GAAGAACTGTTCAGGAGTGGGTGGACAAAATCCGTAAAGCAGGTTGCC CCGTTCCCTCGAATTCGACTGTGGAGAAAACCTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTCTGACCTTCCTGCGCGATAT TGGCAAAACACTTCTCCGTTAACAGATGATCAACAAGAAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTTCGTTCACTGAGTTTCC TACAACCTGCTGCAGGGTTATAGTTTGGCTGTCTGAACAACAACAGTACG GTGTGGTGTCTGCAAAATTGGTGGTTCTGACCAAGTGGGTAACATCACTTC TGGTATCGACCTGACCCGTCGTCTGCATCAGAAATCAGGTG	

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 35	Polinucleótido pPR-EcRS-10 (sitio activo) sintetasa	CGGGGGCTGGTAGCCCCAGGTGACGGACGAGGAAGCGTTAGCAGAGCGA CTGGCGCAAGGCCCGATCGCACTCACGTGTGGCTTCGATCTACCGCTG ACAGCTTGCAATTTGGGGCATCTTGTTCATTGTTATGCTGAAACGCTTC CAGAGCGGGGCCACAAGCGGTTGCGTGGTAGGGCGCGCGACGGGT CTGATTGGGACCCGAGCTTCAAGCTGCCGAGCGTAAGCTGAACACC GAAGAAACTGTTGAGGAGTGGGTGGACAAATCCGTAAGCAGGTTGCC CCGTTCTCGATTTCGACTGTGGAGAAACTCTGCTATCGCGGCCAATA ATTATGACTGGTTCGGCAATATGAATGTGCTGACCTTCCTCGCGGATAT TGGCAACACTTCTCCGTTAACAGATGATCAACAAAGAGCGGTTAA GCAGCGTCTCAACCGTGAAGATCAGGGGATTTCTGTCAGTGTTTCC TACAACTGCTGCAGGGTTATACGTTTGCCTGTACTAACAAACAGTACG GTGTTGCTGCTGCAAAATTGGTGGTCTGACCACTGGGTAAACATCACTTC TGGTATCGACCTGACCCGCTGCTGCATCAGAAATCAGGTG	
SEQ ID NO.: 36	Aminoácido (aa) p-yodoPheRS-1 sintetasa	MASSNLKQLQERGLVAQVTDEEALAEFLAQGPALVCGFDPTADSLHLGH LVPLLCIKRFQQAQGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYSAACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTGGAVWLDPKKTS PYKPYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAYVL AEQVTRLVHGEGLQAAKRITCECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYHFKKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 37	Aminoácido (aa) p-yodoPheRS-2 sintetasa	MASSNLKQLQERGLVAQVTDEEALAEFLAQGPALVCGFDPTADSLHLGH LVPLLCIKRFQQAQGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYSMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTGGAVWLDPKKTS PYKPYQFWINTADADVYRFLKFFTFMSIEEINALEEEDKNSGKAPRAYVL AEQVTRLVHGEGLQAAKRITCECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYHFKKEEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 38	Aminoácido (aa) p-yodoPheRS-3 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFPQAAGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTESYNLLQGYSMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKFGTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGGARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 39	Aminoácido (aa) OMeTyrRS-1 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLLCLKRFPQAAGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTESYNLLQGYSMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKFGTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGGARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 40	Aminoácido (aa) OMeTyrRS-2 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLLCLKRFPQAAGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTESYNLLQGYTMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKFGTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGGARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)

SEQ ID NO.:	Etiqueta	SECUENCIA
SEQ ID NO.: 41	Aminoácido (aa) OMeTyrRS-3 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRFQQAQHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFVSNQM INKEAVKQRLNREDQGISFTEFSYNLLQGTYACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGAVWLDPKKTS PYKQYQFWINTADADVYRFLKFFTFMSIEEINALLEEEDKNSGKAPRAQYVL AEQVTRLVHGEGLQAARKITECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITINGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 42	Aminoácido (aa) OMeTyrRS-4 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRFQQAQHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFVSNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACSNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGAVWLDPKKTS PYKQYQFWINTADADVYRFLKFFTFMSIEEINALLEEEDKNSGKAPRAQYVL AEQVTRLVHGEGLQAARKITECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITINGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 43	Aminoácido (aa) OMeTyrRS-5 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRFQQAQHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFVSNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACSNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGAVWLDPKKTS PYKQYQFWINTADADVYRFLKFFTFMSIEEINALLEEEDKNSGKAPRAQYVL AEQVTRLVHGEGLQAARKITECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITINGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLICWK

(continuación)		SECUENCIA	
SEQ ID NO.: 44	Etiqueta Aminoácido (aa) OMeTyrRS-6 sintetasa	MASSNLKQLQERGLVAQVTDEEALAERLAQGPALTCGFDPADSLHLGH LVPLCLKRFQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYMACLNKQYGVVLIQGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALLEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 45	Aminoácido (aa) p-acetilPheRS-1 sintetasa	MASSNLKQLQERGLVAQVTDEEALAERLAQGPALTCGFDPADSLHLGH LVPLCLKRFQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYMACANKQYGVVLIQGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALLEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 46	Aminoácido (aa) p-benzoilPheRS -1 sintetasa	MASSNLKQLQERGLVAQVTDEEALAERLAQGPALTCGFDPADSLHLGH LVPLCLKRFQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYGFACANKQYGVVLIQGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALLEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)			
SEQ ID NO.:	Etiqueta	SECuencia	
SEQ ID NO.: 47	Aminoácido (aa) p-benzoilPheRS -2 sintetasa	MASSNLIKQLQERGLVAQVTDDEEALAERLAQGPIALCGGFDPTADSLHLGH LVPLCLKRFQQAQGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYGACMNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGAVWLDPKKT PYKPYQFWINTADADVYRFLKHFTEFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 48	Aminoácido (aa) p-azidoPheRS-1 sintetasa	MASSNLIKQLQERGLVAQVTDDEEALAERLAQGPIALCGGFDPTADSLHLGH LVPLCLKRFQQAQGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYSMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGAVWLDPKKT PYKPYQFWINTADADVYRFLKHFTEFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 49	Aminoácido (aa) p-azidoPheRS-2 sintetasa	MASSNLIKQLQERGLVAQVTDDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLCLKRFQQAQGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISTEFSYNLLQGYSAACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGAVWLDPKKT PYKPYQFWINTADADVYRFLKHFTEFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGSLALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)

SEQ ID NO.:	Etiqueta	SECUENCIA
SEQ ID NO.: 50	Aminoácido (aa) p-azidoPheRS-3 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAERLAQGPIALVCGFDPTADSLHLGH LVPLCLCRFQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGNSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALFEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKFEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 51	Aminoácido (aa) p-azidoPheRS-4 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAERLAQGPIALVCGFDPTADSLHLGH LVPLCLCRFQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGNSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACVKNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALFEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKFEEDRLF GRFTLLRRGKKNYCLICWK
SEQ ID NO.: 52	Aminoácido (aa) p-azidoPheRS-5 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAERLAQGPIALICGFDPTADSLHLGH LVPLCLCRFQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGNSAIAANDYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYNFACVKNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTTS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALFEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITINGEKQSDPEYFFKFEEDRLF GRFTLLRRGKKNYCLICWK

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 53	Aminoácido (aa) p-azidoPheRS-6 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRQQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTIS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALSEADFEQLAQDGVPMVEM AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 54	Aminoácido (aa) pPR-EcRS-1 sintetasa <i>p-propargiloxifenilalanina sintetasa</i>	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRQQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSMACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTIS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALSEADFEQLAQDGVPMVEM AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 55	Aminoácido (aa) pPR-EcRS-2 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRQQAGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTEFSYNLLQGYSAACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTIS PYKFYQFWINTADADVYRFLKFFTFMSIEEINALSEADFEQLAQDGVPMVEM AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKKEEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 56	Aminoácido (aa) pPR-EcRS-3 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAERLAQGPIALACGFDPTADSLHLGH LVPLLCIKRFQQA GHKPV ALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYTMACVNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 57	Aminoácido (aa) pPR-EcRS-4 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAERLAQGPIALACGFDPTADSLHLGH LVPLLCIKRFQQA GHKPV ALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYSYACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 58	Aminoácido (aa) pPR-EcRS-5 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAERLAQGPIALACGFDPTADSLHLGH LVPLLCIKRFQQA GHKPV ALVGGATGLIGDPSFKAAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSEAIAANNYDWFGNMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYTMACCNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGKTEGGAVWLDPKKTS PYKFYQFWINTADADVYRFLKFTFMSIEEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 59	Aminoácido (aa) pPR-EcRS-6 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRFQQAQGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYTFACMNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGTEGGA VWLDPKKT PYKFYQFWINTADADVYRFLKFFTFMSIEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEGLQA AKRITTECLFSGSLSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITINGEKQSDPEYFFKEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 60	Aminoácido (aa) pPR-EcRS-7 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALTCGFDPTADSLHLGH LVPLCLKRFQQAQGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYSVACLNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGTEGGA VWLDPKKT PYKFYQFWINTADADVYRFLKFFTFMSIEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEGLQA AKRITTECLFSGSLSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITINGEKQSDPEYFFKEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 61	Aminoácido (aa) pPR-EcRS-8 sintetasa	MASSNLIKQLQERGLVAQVTDEEALAERLAQGPIALVCGFDPTADSLHLGH LVPLCLKRFQQAQGHKPVVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREDQGISFTFSYNLLQGYSMACNTNKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNVFGLTVPLITKADGTFKGTEGGA VWLDPKKT PYKFYQFWINTADADVYRFLKFFTFMSIEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEGLQA AKRITTECLFSGSLSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTASNAITINGEKQSDPEYFFKEDRLF GRFTLLRRGKKNYCLICWK	

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 62	Aminoácido (aa) pPR-EcRS-9 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAEERLAQGPIALSCGFDPDPTADSLHLGH LVPLCLKRFQAGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGGENSAIAANNYDWFNGMNVLTFLRDIGKHFVSQNM INKEAVKQRLNREDQGISFTEFSYNLLQGYSFACLNKQYGVVLQIGGSDQW GNTSGIDLTRRLHQNQVFGTLVPLTKADGTGKGTGGAVWLDPKKTS YKPYQFWINTADADVYRFLKFFTFMSIEEINALEEDKNSGKAPRAQYVLA EQVTRLVHGEELQAARKITECLFSGSLSEADFEQLAQDGVPMVEME KGADLMQALVDSLEQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 63	Aminoácido (aa) pPR-EcRS-10 sintetasa	MASSNLIKQLQERGLVAQVTDDEALAEERLAQGPIALTCGFDPDPTADSLHLGH LVPLCLKRFQAGHKPVALVGGATGLIGDPSFKAERKLNTEETVQEWV DKIRKQVAPFLDFDCGGENSAIAANNYDWFNGMNVLTFLRDIGKHFVSQNM INKEAVKQRLNREDQGISFTEFSYNLLQGYTFACNKKYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGTLVPLTKADGTGKGTGGAVWLDPKKTS PYKPYQFWINTADADVYRFLKFFTFMSIEEINALEEDKNSGKAPRAQYVLA AEQVTRLVHGEELQAARKITECLFSGSLSEADFEQLAQDGVPMVEME EKGADLMQALVDSLEQPSRGQARKTIASNAITINGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
SEQ ID NO.: 64	polinucleótido ARN ^h /Tyr	AGCTTCCCGATAAGGAGCAGGCCAGTAAAAAGCATTACCCCGTGGTG GGGTTCCCGAGCGGCCAAAGGGAGCAGACTCTAAATCTGCCGTCATCG ACCTCGAAGGTTCGAATCCTTCCCCCACCACCA	
SEQ ID NO.: 65	ARN ^h /Tyr	AGCUUCCCCGAUAAGGGAGCAGGCCAGUAAAAAGCAUUAACCCCGUGGU GGGGUUCCCCGAGCGGCCAAAGGGAGCAGACUCUAAAUUCUGCCGUCAU CGACCUUGAAGGUUCGAAUCCUUCUCCCCACCA	
SEQ ID NO.: 66	Mutantes ámbar L3TAG	5'-ATGAAGTAGCTGTCTTCTATCGAACAAGCATGCG-3'	
SEQ ID NO.: 67	Mutantes ámbar I13TAG	5'-CGAACAAGCATGCGATTAGTGCCCGACTTAAAAAG-3'	
SEQ ID NO.: 68	Mutantes ámbar T44TAG	5'-CGCTACTCTCCCAAAATAGAAAAAGGTCTCCGCTG-3'	

(continuación)			
SEQ ID NO.:	Etiqueta	SECUENCIA	
SEQ ID NO.: 69	Mutantes ámbar F68TAG	5'-CTGGAACAGCTATAGCTACTGATTTTTCCTCG-3'	
SEQ ID NO.: 70	Mutantes ámbar R110TAG	5'-GCCGTCACAGATTAGTTGGCTTCAGTGGAGACTG-3'	
SEQ ID NO.: 71	Mutantes ámbar V114TAG	5'-GATTGGCTTCATAGGAGACTGATGCTCTAAC-3'	
SEQ ID NO.: 72	Mutantes ámbar T121TAG	5'-GCCCTCTATAGTTGAGACAGCATAGAATAATGCG-3'	
SEQ ID NO.: 73	Mutantes ámbar I127TAG	5'-GAGACAGCATAGATAGAGTGGACATCATCATCGG-3'	
SEQ ID NO.: 74	Mutantes ámbar S131TAG	5'-GAATAAGTGCACATAGTCATCGGAAGAGAGTAGTAG-3'	
SEQ ID NO.: 75	Mutantes ámbar T145TAG	5'-GGTCAAAAGACAGTTGTAGGTATCGATTGACTCGGC-3'	
SEQ ID NO.: 76	Mutantes sitio permisivo T44F	5'-CGCTACTCTCCCCAAATTTAAAAGGTCTCCGCTG-3'	
SEQ ID NO.: 77	Mutantes sitio permisivo T44Y	5'-CGCTACTCTCCCCAAATATAAAAGGTCTCCGCTG-3'	
SEQ ID NO.: 78	Mutantes sitio permisivo T44W	5'-CGCTACTCTCCCCAAATGGAAGGTCTCCGCTG-3'	
SEQ ID NO.: 79	Mutantes sitio permisivo T44D	5'-CGCTACTCTCCCCAAAGATAAAAGGTCTCCGCTG-3'	
SEQ ID NO.: 80	Mutantes sitio permisivo T44K	5'-CGCTACTCTCCCCAAAAAAGGTCTCCGCTG-3'	
SEQ ID NO.: 81	Mutantes sitio permisivo R110F	5'-GCCGTCACAGATTTTGGCTTCAGTGGAGACTG-3'	
SEQ ID NO.: 82	Mutantes sitio permisivo R110Y	5'-GCCGTCACAGATTATTGGCTTCAGTGGAGACTG-3'	

(continuación)		SECUENCIA	
SEQ ID NO.:	Etiqueta		
SEQ ID NO.: 83	Mutantes sitio permisivo R110W	5'-GCCGTCACAGATTGGTTGGCTTCAGTGGAGACTG-3'	
SEQ ID NO.: 84	Mutantes sitio permisivo R110D	5'-GCCGTCACAGATGATTTGGCTTCAGTGGAGACTG-3'	
SEQ ID NO.: 85	Mutantes sitio permisivo R110K	5'-GCCGTCACAGATAAAATTGGCTTCAGTGGAGACTG-3'	
SEQ ID NO.: 86	Aminoácido (aa) p-acetilPheRS-1 sintetasa ^a	MASSNLKQLQERGLVAQVTDEEALAEERLAQQPIALICGFDPTADSLHLGH LVPLCLKRFQAGHKPVVALVGGATGLIGDPSFKAARLKLNTETVQEWV DKIRKQVAPFLDFDCGGENSAIAANNYDWFNGMNVLTFLRDIGKHFSVNQM INKEAVKQRLNREGQGHSFTFSYNLLQGYGMACANKQYGVVLQIGGSDQ WGNITSGIDLTRRLHQNQVFGLTVPLITKADGTFKGTEGGA VWLDPKKTIS PYKFYQFWINTADADVYRFLKFFTFMSIEINALEEEDKNSGKAPRAQYVL AEQVTRLVHGEEGLQAAKRITTECLFSGLSALSEADFEQLAQDGVPMVEM EKGADLMQALVDSELQPSRGQARKTIASNAITNGEKQSDPEYFFKEEDRLF GRFTLLRRGKKNYCLICWK	
^a Estos clones también contienen una mutación Asp165Gly			

LISTA DE SECUENCIAS

	<110> The Scripps Research Institute	
5	<120> ADICIONES AL CÓDIGO GENÉTICO DE AMINOÁCIDOS REACTIVOS ARTIFICIALES	
	<130> CSC/FP6973622	
	<140> EP TBC	
10	<141> 16-04-2004	
	<150> EP 11163742.7	
	<151> 16-04-2004	
15	<150> EP 04785787.5	
	<151> 16-04-2004	
	<150> PCT/US2004/011833	
20	<151> 16-04-2004	
	<150> US 60/479.931	
	<151> 18-06-2003	
	<150> US 60/493.014	
25	<151> 05-08-2003	
	<150> US 60/496.548	
	<151> 19-08-2003	
30	<160> 104	
	<170> PatentIn versión 3.3	
	<210> 1	
35	<211> 1275	
	<212> ADN	
	<213> Escherichia coli	
40	<400> 1	
	atggcaagca gtaacttgat taaacaattg caagagcggg ggctggtagc ccaggtgacg	60
	gacgaggaag cgtagcaga gcgactggcg caaggcccg tgcgctcta ttgcggcttc	120
	gacctaaccg ctgacagctt gcatttgggg catcttgctt cattgttatg cctgaaacgc	180
	ttccagcagg cgggccacaa gccggttgcg ctggtaggcg gcgcgacggg tctgattggc	240
	gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg	300
	gtggacaaaa tccgtaagca ggttgccccg ttcctcgatt tcgactgtgg agaaaactct	360
	gctatcgcgg cgaacaacta tgactggctt ggcaatatga atgtgctgac cttcctgcgc	420
	gatattggca aacacttctc cgtaaccag atgatcaaca aagaagcggg taagcagcgt	480
	ctcaaccgtg aagatcaggg gatttcgctt actgagtttt cctacaacct gttgcagggt	540
	tatgacttcg cctgtctgaa caaacagtac ggtgtgggtg tgcaaattgg tggttctgac	600
	cagtggggta acatcacttc tggatcgcac ctgacccgct gtctgcatca gaatcagggt	660
	tttggcctga ccgttccgct gatcactaaa gcagatggca ccaaatttgg taaaactgaa	720

ES 2 773 664 T3

```

ggcggcgcag tctggttgga tccgaagaaa accagcccgt acaaattcta ccagttcttg 780
atcaacactg cggatgccga cgtttaccgc ttcctgaagt tcttcacett tatgagcatt 840
gaagagatca acgccctgga agaagaagat aaaaacagcg gttaaagcacc gcgcgcccag 900
tatgtactgg cggagcaggt gactcgtctg gttcacggtg aagaagggtt acaggcggca 960
aaacgtatta ccgaatgcct gttcagcggg tctttgagtg cgctgagtga agcggacttc 1020
gaacagctgg cgcaggacgg cgtaccgatg gttgagatgg aaaagggcgc agacctgatg 1080
caggcactgg tcgattctga actgcaacct tcccgtgggc aggcacgtaa aactatcgcc 1140
tccaatgcca tcaccattaa cggtgaaaaa cagtccgatc ctgaatactt ctttaaagaa 1200
gaagatcgtc tgtttgggtc ttttacctta ctgcgtcgcg gtaaaaagaa ttactgtctg 1260
atttgctgga aataa 1275

```

<210> 2
 <211> 424
 <212> PRT
 <213> Escherichia coli
 <400> 2

5

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1		5		10				15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly		20		25					30
Pro	Ile	Ala	Leu	Tyr	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His		35		40				45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala		50		55			60		
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65			70			75		80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr			85		90				95
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu		100			105			110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp		115		120			125		
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys		130		135			140		

ES 2 773 664 T3

His 145	Phe	Ser	Val	Asn 150	Gln	Met	Ile	Asn	Lys	Glu 155	Ala	Val	Lys	Gln	Arg 160
Leu	Asn	Arg	Glu	Asp 165	Gln	Gly	Ile	Ser	Phe 170	Thr	Glu	Phe	Ser	Tyr 175	Asn
Leu	Leu	Gln	Gly 180	Tyr	Asp	Phe	Ala	Cys 185	Leu	Asn	Lys	Gln	Tyr 190	Gly	Val
Val	Leu	Gln 195	Ile	Gly	Gly	Ser	Asp 200	Gln	Trp	Gly	Asn	Ile 205	Thr	Ser	Gly
Ile	Asp 210	Leu	Thr	Arg	Arg	Leu 215	His	Gln	Asn	Gln	Val 220	Phe	Gly	Leu	Thr
Val 225	Pro	Leu	Ile	Thr	Lys 230	Ala	Asp	Gly	Thr	Lys 235	Phe	Gly	Lys	Thr	Glu 240
Gly	Gly	Ala	Val	Trp 245	Leu	Asp	Pro	Lys	Lys 250	Thr	Ser	Pro	Tyr	Lys 255	Phe
Tyr	Gln	Phe	Trp 260	Ile	Asn	Thr	Ala	Asp 265	Ala	Asp	Val	Tyr	Arg 270	Phe	Leu
Lys	Phe	Phe 275	Thr	Phe	Met	Ser	Ile 280	Glu	Glu	Ile	Asn	Ala 285	Leu	Glu	Glu
Glu	Asp 290	Lys	Asn	Ser	Gly	Lys 295	Ala	Pro	Arg	Ala	Gln 300	Tyr	Val	Leu	Ala
Glu 305	Gln	Val	Thr	Arg	Leu 310	Val	His	Gly	Glu	Glu 315	Gly	Leu	Gln	Ala	Ala 320
Lys	Arg	Ile	Thr	Glu 325	Cys	Leu	Phe	Ser	Gly 330	Ser	Leu	Ser	Ala	Leu 335	Ser
Glu	Ala	Asp	Phe 340	Glu	Gln	Leu	Ala	Gln 345	Asp	Gly	Val	Pro	Met 350	Val	Glu
Met	Glu	Lys 355	Gly	Ala	Asp	Leu	Met 360	Gln	Ala	Leu	Val	Asp 365	Ser	Glu	Leu
Gln	Pro 370	Ser	Arg	Gly	Gln	Ala 375	Arg	Lys	Thr	Ile	Ala 380	Ser	Asn	Ala	Ile
Thr 385	Ile	Asn	Gly	Glu	Lys 390	Gln	Ser	Asp	Pro	Glu 395	Tyr	Phe	Phe	Lys	Glu 400

ES 2 773 664 T3

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 3
<211> 1275
<212> ADN
<213> Artificial

<220>
<223> sintetasa artificial

<400> 3

```
atggcaagca gtaacttgat taaacaattg caagagcggg ggctggtagc ccaggtgacg      60
gacgaggaag cgtagcaga gcgactggcg caaggcccga tcgcactcgt gtgtggcttc      120
gacacctaccg ctgacagctt gcatttgggg catcttgctt cattgttatg cctgaaacgc      180
ttccagcagg cgggccacaa gccgggttgcg ctggtaggcg gcgcgacggg tctgattggc      240
gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg      300
gtggacaaaa tccgtaagca ggttgccccg ttctctcgatt tcgactgtgg agaaaactct      360
gctatcgcgg ccaataatta tgactggctt ggcaatatga atgtgctgac cttcctgcgc      420
gatattggca aacacttctc cgtaaccag atgatcaaca aagaagcggg taagcagcgt      480
ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacaacct gctgcagggt      540
tatagtatgg cctgtttgaa caaacagtac ggtgtggtgc tgcaaattgg tggttctgac      600
cagtggggta acatcacttc tggatcgac ctgaccgcgc gtctgcatca gaatcagggtg      660
tttggcctga ccgttccgct gatcactaaa gcagatggca ccaaatttgg taaaactgaa      720
ggcggcgagc tctggttgga tccgaagaaa accagcccgt acaaattcta ccagttctgg      780
atcaacactg cggatgccga cgtttaccgc ttctgaagt tcttcacctt tatgagcatt      840
gaagagatca acgccctgga agaagaagat aaaaacagcg gtaaagcacc gcgcgcccag      900
tatgtactgg cggagcaggg gactcgtctg gttcacgggtg aagaagggtt acaggcggca      960
aaacgtatta ccgaatgcct gttcagcggg tctttgagtg cgctgagtga agcggacttc     1020
gaacagctgg cgcaggacgg cgtaccgatg gttgagatgg aaaagggcgc agacctgatg     1080
caggcactgg tcgattctga actgcaacct tcccgtgggc aggcacgtaa aactatcgcc     1140
tccaatgcc aaccattaa cggtgaaaaa cagtccgatc ctgaatactt ctttaaagaa     1200
gaagatcgtc tgtttggctg ttttacctta ctgcgtcgcg gtaaaaagaa ttactgtctg     1260
atttgctgga aataa                                     1275
```

<210> 4

<211> 1275
<212> ADN
<213> artificial

5 <220>
<223> sintetasa artificial

<400> 4

```
atggcaagca gtaacttgat taaacaattg caagagcggg ggctggtagc ccaggtgacg      60
gacgaggaag cgtttagcaga gcgactggcg caaggcccga tcgcactcac ttgtggcttc      120
gacccagcag cgttagcaga gcgactggcg caaggcccga tcgcactcac ttgtggcttc      180
ttccagcagg cgggccacaa gccgggttgcg ctggtaggcg gcgcgacggg tctgattggc      240
gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg      300
gtggacaaaa tccgtaagca ggttgccccg ttccctcgatt tcgactgtgg agaaaactct      360
gctatcgcgg ccaataatta tgactgggtc agcaatatga atgtgctgac cttcctgcgc      420
gatattggca aacactttctc cgtaaccag atgatcaaca aagaagcggg taagcagcgt      480
ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacaacct gctgcagggt      540
tatacgtatg cctgtctgaa caaacagtac ggtgtggtgc tgcaaattgg tggttctgac      600
cagtggggta acatcacttc tggatcgac ctgaccgcgt gtctgcatca gaatcagggtg      660
tttggcctga ccgttcgct gatcactaaa gcagatggca ccaaatttgg taaaactgaa      720
ggcggcgag tctggttgga tccgaagaaa accagcccggt acaaattcta ccagttctgg      780
atcaaacactg cggatgccga cgtttaccgc ttccctgaagt tcttcacctt tatgagcatt      840
gaagagatca acgccctgga agaagaagat aaaaacagcg gtaaagcacc gcgcgcccag      900
tatgtactgg cggagcaggg gactcgtctg gttcacgggtg aagaagggtt acaggcggca      960
aaacgtatta ccgaatgcct gttcagcggg tctttgagtg cgctgagtga agcggacttc     1020
gaacagctgg cgcaggacgg cgtaccgatg gttgagatgg aaaagggcgc agacctgatg     1080
caggcactgg tcgattctga actgcaacct tcccgtgggtc aggcacgtaa aactatcgcc     1140
tccaatgcc aaccatttaa cggtgaaaaa cagtccgatc ctgaatactt ctttaaagaa     1200
gaagatcgtc tgtttggtcg ttttacctta ctgcgtcgcg gtaaaaagaa ttactgtctg     1260
atttgctgga aataa                                         1275
```

10

<210> 5
<211> 1275
<212> ADN
<213> artificial

15

<220>
<223> sintetasa artificial

20 <400> 5

ES 2 773 664 T3

atggcaagca gtaacttgat taaacaattg caagagcggg ggctggtagc ccaggtgacg 60
gacgaggaag cgtagcaga gcgactggcg caaggcccga tcgcactcgt gtgtggcttc 120
gatcctaccg ctgacagctt gcatttgggg catcttggtc cattgttatg cctgaaacgc 180
ttccagcagg cgggccacaa gccggttgcg ctggtaggcg gcgcgacggg tctgattggc 240
gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg 300
gtggacaaaa tccgtaagca ggttgccccg ttctctgatt tcgactgtgg agaaaactct 360
gctatcgcg ccaataatta tgactgggtc ggcaatatga atgtgctgac cttcctgcgc 420
gatattggca aacacttctc cgtaaccag atgatcaaca aagaagcggg taagcagcgt 480
ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacaacct gctgcagggt 540
tatagtatgg cctgtttgaa caaacagtac ggtgtggtgc tgcaaattgg tggttctgac 600
cagtggggta acatcacttc tggatcgcac ctgacccgtc gtctgcatca gaatcagggtg 660
tttggcctga ccgttccgct gatcactaaa gcagatggca ccaaatttgg taaaactgaa 720
ggcggcgag tctggttga tccgaagaaa accagcccg acaaattcta ccagttctgg 780
atcaacactg cggatgccga cgtttaccgc ttctgaagt tcttcacctt tatgagcatt 840
gaagagatca acgccctgga agaagaagat aaaaacagcg gtaaagcacc gcgcgcccag 900
tatgtactgg cggagcagg gactcgtctg gttcacggtg aagaaggttt acaggcggca 960
aaacgtatta ccgaatgcct gttcagcggg tctttgagtg cgctgagtga agcggacttc 1020
gaacagctgg cgcaggacgg cgtaccgatg gttgagatgg aaaagggcgc agacctgatg 1080
caggcactgg tcgattctga actgcaacct tcccggtggtc aggcacgtaa aactatcgcc 1140
tccaatgcc taccattaa cggtgaaaaa cagtccgatc ctgaatactt ctttaaagaa 1200
gaagatcgtc tgtttggtcg ttttacctta ctgcgtcgcg gtaaaaagaa ttactgtctg 1260
atttgctgga aataa 1275

<210> 6
<211> 1275
<212> ADN
<213> artificial

<220>
<223> sintetasa artificial

<400> 6

atggcaagca gtaacttgat taaacaattg caagagcggg ggctggtagc ccaggtgacg 60
gacgaggaag cgtagcaga gcgactggcg caaggcccga tcgcactcgt gtgtggcttc 120
gatcctaccg ctgacagctt gcatttgggg catcttggtc cattgttatg cctgaaacgc 180
ttccagcagg cgggccacaa gccggttgcg ctggtaggcg gcgcgacggg tctgattggc 240

ES 2 773 664 T3

gacccgagct tcaaagctgc cgagcgtaag ctgaacacog aagaaactgt tcaggagtgg 300
 gtggacaaaa tccgtaagca ggttgccccg ttccctcgatt tcgactgtgg agaaaactct 360
 gctatcgcgg ccaataatta tgactgggtc ggcaatatga atgtgctgac cttcctgcgc 420
 gatattggca aacactttctc cgttaaccag atgatcaaca aagaagcggg taagcagcgt 480
 ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacaacct gctgcagggg 540
 tatagtatgg cctgtttgaa caaacagtac ggtgtggtgc tgcaaattgg tggttctgac 600
 cagtggggta acatcacttc tggatcgac ctgacccgtc gtctgcatca gaatcaggtg 660
 tttggcctga ccgttccgct gatcactaaa gcagatggca ccaaatttgg taaaactgaa 720
 ggcggcgag tctggttgga tccgaagaaa accagcccggt acaaattcta ccagttctgg 780
 atcaacactg cggatgccga cgtttaccgc ttccctgaagt tcttcacctt tatgagcatt 840
 gaagagatca acgcccctgga agaagaagat aaaaacagcg gtaaagcacc gcgcgcccag 900
 tatgtactgg cggagcaggt gactcgtctg gttcacgggtg aagaaggttt acaggcggca 960
 aaacgtatta ccgaatgcct gttcagcggg tctttgagtg cgctgagtga agcggacttc 1020
 gaacagctgg cgcaggacgg cgtaccgatg gttgagatgg aaaagggcgc agacctgatg 1080
 caggcactgg tcgattctga actgcaacct tcccgtgggtc aggcacgtaa aactatcgcc 1140
 tccaatgcca tcaccattaa cggtgaaaaa cagtccgatc ctgaatactt ctttaaagaa 1200
 gaagatcgtc tgtttggtcg ttttacctta ctgcgtcgcg gtaaaaagaa ttactgtctg 1260
 atttgctgga aataa 1275

<210> 7
 <211> 1275
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 7

ES 2 773 664 T3

atggcaagca gtaacttgat taaacaattg caagagcggg ggctggtagc ccaggtgacg 60
 gacgaggaag cgtagcaga gcgactggcg caaggcccga tcgactcac gtgtggcttc 120
 gatcctaccg ctgacagctt gcatttgggg catcttggtc cattgttatg cctgaaacgc 180
 ttccagcagg cgggccacaa gccggttgcg ctggtaggcg gcgcgacggg tctgattggc 240
 gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg 300
 gtggacaaaa tccgtaagca ggttgccccg ttcctcgatt tcgactgtgg agaaaactct 360
 gctatcgcg ccaataatta tgactggttc ggcaatatga atgtgctgac cttcctgcgc 420
 gatattggca aacacttctc cgtaaccag atgatcaaca aagaagcggg taagcagcgt 480
 ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacagcct gctgcagggt 540

 tatacgatgg cctgtctgaa caaacagtac ggtgtggtgc tgcaaattgg tggttctgac 600
 cagtggggta acatcacttc tggtagcgc ctgacccgtc gtctgcatca gaatcagggtg 660
 tttggcctga ccgttccgct gatcactaaa gcagatggca ccaaatttgg taaaactgaa 720
 ggcggcgag tctggttggg tccgaagaaa accagcccgt acaaattcta ccagttctgg 780
 atcaacactg cggatgccga cgtttaccgc ttcctgaagt tcttcacctt tatgagcatt 840
 gaagagatca acgcccgtga agaagaagat aaaaacagcg gtaaagcacc gcgcgcccag 900
 tatgtactgg cggagcaggg gactcgtctg gttcacgggtg aagaagggtt acaggcggca 960
 aaacgtatta ccgaatgcct gttcagcggg tctttgagtg cgctgagtga agcggacttc 1020
 gaacagctgg cgcaggacgg cgtaccgatg gttgagatgg aaaagggcgc agacctgatg 1080
 caggcactgg tcgattctga actgcaacct tcccgtggtc aggcacgtaa aactatcgcc 1140
 tccaatgcca tcaccattaa cggtgaaaaa cagtccgatc ctgaatactt ctttaaagaa 1200
 gaagatcgtc tgtttggctg ttttacctta ctgcgtcgcg gtaaaaagaa ttactgtctg 1260
 atttgctgga aataa 1275

- 5 <210> 8
 <211> 540
 <212> ADN
 <213> artificial

 10 <220>
 <223> sintetasa artificial

 <400> 8

ES 2 773 664 T3

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcacttgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcagcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatacg tatgcctgtc tgaacaaaca gtacggtgtg 540

<210> 9
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 9

cgggggctgg taccacaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcacttgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcagcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatacg tatgcctgtc tgaacaaaca gtacggtgtg 540

<210> 10
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 10

ES 2 773 664 T3

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcacttgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcagcaat 360
 atgaatgtgc tgaccttctt gcgcgatatt ggcaaact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatacg tatgcctgtc tgaacaaaca gtacggtgtg 540

<210> 11
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 11

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcacttgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttctt gcgcgatatt ggcaaact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttattcg tatgcctgtg cgaacaaaca gtacggtgtg 540

<210> 12
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 12

ES 2 773 664 T3

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcacttgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcagcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatag tatgcctgtc tgaacaaaca gtacggtgtg 540

<210> 13
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 13

cgggggctgg taccacaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcctttgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttattct attgcctgtt cgaacaaaca gtacggtgtg 540

<210> 14
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 14

ES 2 773 664 T3

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcgtgtgtgg ctctgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcagggtgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgtaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatagt attgcctgtt tgaacaaaca gtacggtgtg 540

<210> 15
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 15

cgggggctgg taccacaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcgtgtgtgg ctctgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcagggtgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgtaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatagt attgcctgtt tgaacaaaca gtacggtgtg 540

<210> 16
 <211> 540
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 16

ES 2 773 664 T3

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tctggtgtgg ctccgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcca cgggtctgat tggcgacccg agcttcaagg ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaatt gttatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatatg cgtgcctgtg agaacaaaca gtacggtgtg 540

<210> 17
 <211> 624
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 17

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tcatttgtgg ctccgatcct accgctgaca gcttgcattt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
 ggcggcgcca cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgttaa ccagatgac 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaaggtc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatggt atggcctgtg ctaacaaaca gtacggtgtg 540
 gtgctgcaaa ttggtgggtc tgaccaatgg ggtaacatca cttctggtat cgacctgacc 600
 cgtcgtctgc atcagaatca ggtg 624

<210> 18
 <211> 609
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 18

ES 2 773 664 T3

cagggtgacgg acgaggaagc gttagcagag cgactggcgc aaggcccgat cgcactcggg 60
 tgtggcttcg atcctaccgc tgacagcttg catttggggc atcttggtcc attggttatgc 120
 ctgaaacgct tccagcaggc gggccacaag ccggttgccg tggtaggcgg cgcgacgggt 180
 ctgattggcg acccgagctt caaagctgcc gagcgttaagc tgaacaccga agaaactggt 240
 caggagtggg tggacaaaat ccgtaagcag gttgccccgt tcctcgattt cgactgtgga 300
 gaaaactctg ctatcgcggc caataattat gactggttcg gcaatatgaa tgtgctgacc 360
 ttctcgcgcg atattggcaa acacttctcc gttaaccaga tgatcaacaa agaagcgggt 420
 aagcagcgct tcaaccgtga agatcagggg atttcgttca ctgagttttc ctacaacctg 480
 ctgcagggtt atggttttgc ctgtttgaac aaacagtacg gtgtggtgct gcaaattggt 540
 ggttctgacc agtggggtaa catcacttct ggtatcgacc tgacccgctg tctgcatcag 600
 aatcagggtg 609

<210> 19
 <211> 591
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 19

gcgttagcag agcgactggc gcaaggcccg atcgactcgc ggtgtggctt cgatcctacc 60
 gctgacagct tgcatttggg gcatcttggt ccattgttat gcctgaaacg cttccagcag 120
 gcggggccaca agccggttgc gctggtaggc ggcgcgacgg gtctgattgg cgacccgagc 180
 ttcaaagctg ccgagcgtaa gctgaacacc gaagaaactg ttcaggagtg ggtggacaaa 240
 atccgtaagc aggttgcccc gttcctcgat ttcgactgtg gagaaaactc tgctatcgcg 300
 gccataatt atgactgggt cggcaatatg aatgtgctga ccttcctgcg cgatattggc 360
 aaacacttct ccgttaacca gatgatcaac aaagaagcgg ttaagcagcg tctcaaccgt 420
 gaagatcagg ggatttcggt cactgagttt tcctacaacc tgctgcaggg ttatggttat 480
 gcctgtatga acaaacagta cgggtgtgtg ctgcaaattg gtggttctga ccagtggggg 540
 aacatcactt ctggtatcga cctgacccgt cgtctgcatc agaatcagggt g 591

<210> 20
 <211> 621
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<220>

<221> misc_feature
<222> (26)..(26)
<223> n es a, c, g, o t

5 <220>
<221> misc_feature
<222> (612)..(612)
<223> n es a, c, g, o t

10 <220>
<221> misc_feature
<222> (618)..(618)
<223> n es a, c, g, o t

15 <400> 20

gggctggtag	cccaggtgac	ggacgnagaa	gcgttagcag	agcgactggc	gcaaggcccg	60
atcgcaactcc	tttgtggctt	cgatcctacc	gctgacagct	tgcatttggg	gcatcttggt	120
ccattgttat	gcctgaaacg	cttccagcag	gcgggccaca	agccggttgc	gctggtaggc	180
ggcgcgacgg	gtctgattgg	cgacccgagc	ttcaaagctg	ccgagcgtaa	gctgaacacc	240
gaagaaactg	ttcaggagtg	ggtggacaaa	atccgtaagc	aggttgcccc	gttcctcgat	300
ttcgactgtg	gagaaaactc	tgctatcgcg	gccaataatt	atgactgggt	cggcaatatg	360
aatgtgctga	ccttcctgcg	cgatattggc	aaacacttct	ccgttaacca	gatgatcaac	420
aaagaagcgg	ttaagcagcg	tctcaaccgt	gaagatcagg	ggatttcgtt	cactgagttt	480
tcctacaacc	tgctgcaggg	ttattctatg	gcctgtgcga	acaaacagta	cgggtgtggtg	540
ctgcaaattg	gtggttctga	ccagtggggg	aacatcactt	ctggtatcga	cctgacccgt	600
cgtctgcac	anaatcangt	g				621

20 <210> 21
<211> 588
<212> ADN
<213> artificial

25 <220>
<223> sintetasa artificial

<400> 21

ttagcagagc	gactggcgca	aggcccgatc	gcactcgttt	gtggcttcga	tcctaccgct	60
gacagcttgc	atttggggca	tcttgttcca	ttgttatgcc	tgaaacgctt	ccagcaggcg	120
ggccacaagc	cggttgcgct	ggtaggcggc	gcgacgggtc	tgattggcga	cccgagcttc	180
aaagctgccg	agcgtaagct	gaacaccgaa	gaaactgttc	aggagtgggt	ggacaaaatc	240
cgtaaagcagg	ttgccccgtt	cctcgatttc	gactgtggag	aaaactctgc	tatcgcgggc	300

30

ES 2 773 664 T3

aataattatg actggttcgg caatatgaat gtgctgacct tcctgcgcga tattggcaaa 360
cacttctccg ttaaccagat gatcaacaaa gaagcgggta agcagcgtct caaccgtgaa 420
gatcagggga tttcgttcac tgagttttcc tacaacctgc tgcaggggta ttctgcggcc 480
tgtgcgaaca aacagtacgg tgtgggtgctg caaattgggtg gttctgacca gtggggtaac 540
atcacttctg gtatcgacct gacccgtcgt ctgcatcaga atcagggtg 588

<210> 22
<211> 600
<212> ADN
<213> artificial

<220>
<223> sintetasa artificial

<220>
<221> misc_feature
<222> (403)..(403)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (513).. (513)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (515).. (515)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (518)..(518)
<223> n es a, c, g, o t

<220>
<221> misc_feature
<222> (531)..(531)
<223> n es a, c, g, o t

<400> 22

gacgaggaag cgtagcaga gcgactggcg caaggcccg tgcactcct gtgtggcttc 60
gatcctaccg ctgacagctt gcatttgggg catcttggtc cattgttatg cctgaaacgc 120
ttccagcagg cgggccacaa gccggttgcg ctggtaggcg gcgcgacggg tctgattggc 180
gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg 240
gtggacaaaa tccgtaagca ggttgccccg ttcctcgatt tcgactgtgg agaaaactct 300
gctatcgcgg ccaataatta tgactgggtc ggcaatatga atgtgctgac ctctctgcgc 360
gatattggca aacacttctc cgtaaccag atgatcaaca aanaagcggg taagcagcgt 420
ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacaacct gctgcagggt 480
tattcggctg cctgtgcgaa caaacagtac gngngngngc tgcaaattgg nggttctgac 540

	caggggggta acatcacttc tggatcgcac ctgaccgcgc gtctgcatca aaatcagggtg	600
5	<210> 23 <211> 591 <212> ADN <213> artificial	
10	<220> <223> sintetasa artificial	
15	<220> <221> misc_feature <222> (588)..(588) <223> n es a, c, g, o t	
	<400> 23	
	gcgttagcag agcgactggc gcaaggcccg atcgcaactcg tttgtggctt cgatcctacc	60
	gctgacagct tgcatttggg gcatcttggt ccattgttgt gcctgaaacg cttccagcag	120
	gcggggccaca agccggttgc gctggtaggc ggcgcgacgg gtctgattgg cgacccgagc	180
	ttcaaagctg ccgagcgtaa gctgaacacc gaagaaactg ttcaggagtg ggtggacaaa	240
	atccgtaagc aggttgcccc gttcctcgat ttcgactgtg gagaaaactc tgctatcgcg	300
	gccaataatt atgactggtt cggcaatatg aatgtgctga ccttcctgcg cgatattggc	360
	aaacacttct ccgttaacca gatgatcaac aaagaagcgg ttaagcagcg tctcaaccgt	420
	gaagatcagg ggatttcgtt cactgagttt tcctacaacc tgctgcaggg ttatagtgcg	480
	gcctgtgtta acaaacagta cgggtgtggtg ctgcaaattg gtggttctga ccagtggggt	540
	aacatcactt ctggtatcga cctgaccggt cgtctgcacg agaatcangt g	591
20	<210> 24 <211> 600 <212> ADN <213> artificial	
25	<220> <223> sintetasa artificial	
	<400> 24	
	gacgaggaag cgtagcaga gcgactggcg caaggcccgga tcgcactcat ttgtggcttc	60
	gatcctaccg ctgacagctt gcatttgggg catcttggtc cattgttatg cctgaaacgc	120
	ttccagcagg cggggccacaa gccggttgcg ctggtaggcg gcgcgacggg tctgattggc	180
	gacccgagct tcaaagctgc cgagcgtaag ctgaacaccg aagaaactgt tcaggagtgg	240
	gtggacaaaa tccgtaagca ggttgccccg ttcctcgatt tcgactgtgg agaaaactct	300
	gctatcgcgg ccaatgatta tgactggttc ggcaatatga atgtgctgac cttcctgcgc	360
30	gatattggca aacacttctc cgtaaccag atgatcaaca aagaagcggg taagcagcgt	420

ES 2 773 664 T3

	ctcaaccgtg aagatcaggg gatttcgttc actgagtttt cctacaacct gctgcagggg	480
	tataattttg cctgtgtgaa caaacagtac ggtgtggtgc tgcaaattgg tggttctgac	540
	cagtggggta acatcacttc tggatatcac ctgaccctgc gtctgcatca gaatcaggtg	600
5	<210> 25 <211> 579 <212> ADN <213> artificial	
10	<220> <223> sintetasa artificial <400> 25	
	cgactggcgc aaggcccgat cgcactcacg tgtggcttcg atcctaccgc tgacagcttg	60
	catttggggc atcttggtcc attgttatgc ctgaaacgct tccagcaggc gggccacaag	120
	ccggttgccg tggtaggcgg cgcgacgggt ctgattggcg acccgagctt caaagctgcc	180
	gagcgtaacg tgaacaccga agaaactggt caggagtggg tggacaaaat ccgtaagcag	240
	gttgccccgt tcctcgattt cgactgtgga gaaaactctg ctatcgcggc caataattat	300
	gactgggttcg gcaatatgaa tgtgctgacc ttcctgcgcg atattggcaa acacttctcc	360
	gttaaccaga tgatcaacaa agaagcgggt aagcagcgtc tcaaccgtga agatcagggg	420
	atttcgttca ctgagttttc ctacaatctg ctgcaggggt attcggctgc ctgtcttaac	480
	aaacagtacg gtgtggtgct gcaaattggt ggttctgacc agtggggtaa catcacttct	540
	ggtatcgacc tgaccctgcg tctgcatcag aatcaggtg	579
15	<210> 26 <211> 624 <212> ADN <213> artificial	
20	<220> <223> sintetasa artificial	
25	<220> <221> misc_feature <222> (13)..(13) <223> n es a, c, g, o t	
30	<220> <221> misc_feature <222> (599)..(599) <223> n es a, c, g, o t <400> 26	
	cgggggcttg tancccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc	60
	ccgatcgcac tcgggtgtgg cttcgatcct accgctgaca gcttgcattt ggggcattct	120
35	gttccattgt tatgcctgaa acgcttccag caggcgggcc acaagccggt tgcgctggta	180

ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgtaa ccagatgac 420
aacaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
ttttcctaca acctgctgca gggttattct atggcctgtt tgaacaaaca gtacgggtgtg 540
gtgctgcaaa ttggtggttc tgaccagtgg ggtaacatca cttctggtat cgacctganc 600
cgtcgtctgc atcagaatca ggtg 624

<210> 27
<211> 625
<212> ADN
<213> artificial

<220>
<223> sintetasa artificial

<220>
<221> misc_feature
<222> (600).. (600)
<223> n es a, c, g, o t

<400> 27

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc 60
ccgatcgac tcacgtgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt 120
gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta 180
ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc 300
gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgtaa ccagatgac 420
aacaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
ttttcctaca atctgctgca gggttattcg gctgcctgtc ttaacaaaca gtacgggtgtg 540
gtgctgcaaa ttggtggttc tgaccagtgg ggtaacatca cttctggtat cgaacctgan 600
ccgtcgtctg catcaaaatc aagtg 625

<210> 28
<211> 624
<212> ADN
<213> artificial

<220>
<223> sintetasa artificial

<400> 28

ES 2 773 664 T3

cgggggctgg taccccaagt gacggacgag gaaacgtag cagagcgact ggcgcaaggc 60
 ccgatcgac tctcttgagg ctctgatcct accgctgaca gcttgcatctt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcaggcc acaagccggg tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcagggttg cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaact tctccgttaa ccagatgatc 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttatacg atggcctgtg tgaacaaaca gtacggtgtg 540
 gtgctgcaaa ttgggtggtc tgaccagtgg ggtaacatca cttctggtat cgacctgacc 600
 cgtcgtctgc atcagaatca ggtg 624

<210> 29
 <211> 624
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 29

cgggggctgg tagcccagggt gacggacgag gaagcgttag cagagcgact ggcgcaaggc 60
 ccgatcgac tcgctgagg ctctgatcct accgctgaca gcttgcatctt ggggcatctt 120
 gttccattgt tatgcctgaa acgcttccag caggcgggccc acaagccggg tgcgctggta 180
 ggcggcgcgga cgggtctgat tggcgacccg agcttcaagg ctgccgagcg taagctgaac 240
 accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcagggttg cccgttcctc 300
 gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat 360
 atgaatgtgc tgaccttcct gcgcgatatt ggcaaact tctccgttaa ccagatgatc 420
 aacaaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag 480
 ttttcctaca acctgctgca gggttattct tatgcctgtc ttaacaaaca gtacggtgtg 540
 gtgctgcaaa ttgggtggtc tgaccagtgg ggtaacatca cttctggtat cgacctgacc 600
 cgtcgtctgc atcagaatca ggtg 624

<210> 30
 <211> 624
 <212> ADN
 <213> artificial

<220>
 <223> sintetasa artificial

ES 2 773 664 T3

<400> 30

```

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc      60
ccgatcgcac tcgcgtgtgg cttecatcct accgctgaca gcttgcattt ggggcatctt      120
gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta      180
ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac      240
accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc      300
gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat      360
atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgtaa ccagatgac      420
aacaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag      480
ttttcctaca acctgctgca gggttatacg atggcctgtt gtaacaaaca gtacggtgtg      540
gtgctgcaaa ttggtggttc tgaccagtgg ggtaacatca cttctggtat cgacctgacc      600
cgtcgtctgc atcagaatca ggtg                                           624

```

5 <210> 31
 <211> 624
 <212> ADN
 <213> artificial

10 <220>
 <223> sintetasa artificial
 <400> 31

```

cgggggctgg taccccaagt gacggacgag gaagcgtag cagagcgact ggcgcaaggc      60
ccgatcgcac tcacgtgtgg cttecatcct accgctgaca gcttgcattt ggggcatctt      120
gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta      180
ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac      240
accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc      300
gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat      360
atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgtaa ccagatgac      420
aacaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcgtgag      480
ttttcctaca acctgctgca gggttatacg tttgcctgta tgaacaaaca gtacggtgtg      540
gtgctgcaaa ttggtggttc tgaccagtgg ggtaacatca cttctggtat cgacctgacc      600
cgtcgtctgc atcagaatca ggtg                                           624

```

15
 20 <210> 32
 <211> 606
 <212> ADN
 <213> artificial

<220>

<223> sintetasa artificial

<400> 32

gtgacggacg aggaagcgtt agcagagcga ctggcgcaag gcccgatcgc actcacgtgt	60
ggcttcgatc ctaccgctga cagcttgcat ttggggcatc ttgttccatt gttatgcctg	120
aaacgcttcc agcaggcggg ccacaagccg gttgcgctgg taggcggcgc gacgggtctg	180
attggcgacc cgagcttcaa agctgccgag cgtaagctga acaccgaaga aactgttcag	240
gagtgggtgg acaaaatccg taagcagggt gccccgttcc tcgatttcga ctgtggagaa	300
aactctgcta tcgcggccaa taattatgac tggttcggca atatgaatgt gctgaccttc	360
ctgcgcgata ttggcaaaca cttctccgtt aaccagatga tcaacaaaga agcggttaag	420
cagcgtctca accgtgaaga tcaggggatt tcgttcactg agttttccta caatctgctg	480
cagggttatt cggtgcctg tcttaacaaa cagtacggtg tggtgctgca aattggtggt	540
tctgaccagt ggggtaacat cacttctggt atcgacctga cccgtcgtct gcatcagaat	600
caggtg	606

5

<210> 33

<211> 624

<212> ADN

<213> artificial

10

<220>

<223> sintetasa artificial

<400> 33

15

cgggggctgg tagcccaggt gacggacgag gaagcgttag cagagcgact ggcgcaaggc	60
ccgatcgac tcgtttgtgg cttcgatcct accgctgaca gcttgcattht ggggcatctt	120
gttccattgt tatgcctgaa acgcttccag caggcgggccc acaagccggt tgcgctggta	180
ggcggcgcca cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac	240
accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcagggtgc cccgttcctc	300
gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat	360
atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgttaa ccagatgatc	420
aacaaagaag cgggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag	480
ttttcctaca acctgctgca ggggtattcg atggcctgta cgaacaaaca gtacgggtgtg	540
gtgctgcaaa ttggtgggtc tgaccagtgg ggtaacatca cttctggtat cgacctgacc	600
cgtcgtctgc atcagaatca ggtg	624

<210> 34

<211> 624

<212> ADN

<213> artificial

20

<220>
<223> sintetasa artificial

5 <220>
<221> misc_feature
<222> (13)..(13)
<223> n es a, c, g, o t

10 <400> 34

cgggggctgg	tancccaagt	gacggacggg	gaagcgtag	cagagcgact	ggcgcaaggc	60
ccgatcgac	tcagttgtgg	cttcgatcct	accgctgaca	gcttgcattt	ggggcatctt	120
gttcatttgt	tatgcctgaa	acgcttcag	caggcgggcc	acaagccggt	tgcgctggta	180
ggcggcgcg	cgggtctgat	tggcgacccg	agcttcaaag	ctgccgagcg	taagctgaac	240
accgaagaaa	ctgttcagga	gtgggtggac	aaaatccgta	agcaggttgc	cccgttcctc	300
gatctcgact	gtggagaaaa	ctctgctatc	gcggccaata	attatgactg	gttcggcaat	360
atgaatgtgc	tgaccttcct	gcgcgatatt	ggcaaacact	tctccgttaa	ccagatgatc	420
aacaaagaag	cggttaagca	gcgtctcaac	cgtgaagatc	aggggatttc	gttcaactgag	480
ttttcctaca	acctgctgca	gggttatagt	tttgctgtgc	tgaacaaaca	gtacgggtgtg	540
gtgctgcaaa	ttggtggttc	tgaccagtgg	ggtaacatca	cttctgggat	cgacctgacc	600
cgtcgtctgc	atcagaatca	ggtg				624

15 <210> 35
<211> 624
<212> ADN
<213> artificial

20 <220>
<223> sintetasa artificial

<400> 35

ES 2 773 664 T3

```

cgggggctgg tagcccaggt gacggacgag gaagcgtag cagagcgact ggcgcaaggc      60
ccgatcgcac tcacgtgtgg cttcgatcct accgctgaca gcttgcattt ggggcatctt      120
gttccattgt tatgcctgaa acgcttccag caggcggggc acaagccggt tgcgctggta      180
ggcggcgcgga cgggtctgat tggcgacccg agcttcaaag ctgccgagcg taagctgaac      240
accgaagaaa ctgttcagga gtgggtggac aaaatccgta agcaggttgc cccgttcctc      300
gatttcgact gtggagaaaa ctctgctatc gcggccaata attatgactg gttcggcaat      360
atgaatgtgc tgaccttcct gcgcgatatt ggcaaacact tctccgttaa ccagatgac      420
aacaagaag cggttaagca gcgtctcaac cgtgaagatc aggggatttc gttcactgag      480
ttttcctaca acctgctgca gggttatacg ttgcctgta ctaacaaaca gtacggtgtg      540
gtgctgcaaa ttggtgggtc tgaccagtgg ggtaacatca cttctggtat cgacctgacc      600
cgtcgtctgc atcagaatca ggtg                                           624

```

<210> 36
 <211> 424
 <212> PRT
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 36

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Val	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155	160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175	
Leu	Leu	Gln	Gly	Tyr	Ser	Tyr	Ala	Cys	Leu	Asn	Lys	Gln	Tyr	Gly	Val	180	185	190	
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	195	200	205	

ES 2 773 664 T3

Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220

Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240

Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255

Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
 370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
 385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
 405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
 420

<210> 37
 <211> 424
 <212> PRT
 <213> artificial

<220>
 <223> sintetasa artificial

ES 2 773 664 T3

<400> 37

```

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1          5          10          15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20          25          30

Pro Ile Ala Leu Ile Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35          40          45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50          55          60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65          70          75          80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
85          90          95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100         105         110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
115         120         125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130         135         140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145         150         155         160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
165         170         175

Leu Leu Gln Gly Tyr Ser Met Ala Cys Leu Asn Lys Gln Tyr Gly Val
180         185         190

Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
195         200         205

Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
210         215         220

Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
225         230         235         240

```

ES 2 773 664 T3

Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
245 250 255

Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
260 265 270

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 38
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 38

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

ES 2 773 664 T3

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
 20 25 30
 Pro Ile Ala Leu Val Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
 35 40 45
 Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50 55 60
 Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65 70 75 80
 Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
 85 90 95
 Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
 100 105 110
 Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
 115 120 125
 Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
 130 135 140
 His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
 145 150 155 160
 Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175
 Leu Leu Gln Gly Tyr Ser Met Ala Cys Ala Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270

ES 2 773 664 T3

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 39
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 39

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Val Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

ES 2 773 664 T3

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65 70 75 80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
 85 90 95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
 100 105 110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
 115 120 125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
 130 135 140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
 145 150 155 160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175

Leu Leu Gln Gly Tyr Ser Met Ala Cys Leu Asn Lys Gln Tyr Gly Val
 180 185 190

Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205

Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220

Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240

Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255

Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300

ES 2 773 664 T3

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 40
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 40

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65 70 75 80

ES 2 773 664 T3

Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175
Leu	Leu	Gln	Gly	Tyr	Thr	Met	Ala	Cys	Leu	Asn	Lys	Gln	Tyr	Gly	Val	180	185	190
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	195	200	205
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr	210	215	220
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu	225	230	235
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe	245	250	255
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu	260	265	270
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu	275	280	285
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala	290	295	300
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala	305	310	315
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser	325	330	335

ES 2 773 664 T3

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 41
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 41

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65 70 75 80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
85 90 95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100 105 110

ES 2 773 664 T3

Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175
Leu	Leu	Gln	Gly	Tyr	Thr	Tyr	Ala	Cys	Leu	Asn	Lys	Gln	Tyr	Gly	Val	180	185	190
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	195	200	205
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr	210	215	220
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu	225	230	235
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe	245	250	255
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu	260	265	270
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu	275	280	285
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala	290	295	300
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala	305	310	315
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser	325	330	335
Glu	Ala	Asp	Phe	Glu	Gln	Leu	Ala	Gln	Asp	Gly	Val	Pro	Met	Val	Glu	340	345	350
Met	Glu	Lys	Gly	Ala	Asp	Leu	Met	Gln	Ala	Leu	Val	Asp	Ser	Glu	Leu	355	360	365

ES 2 773 664 T3

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 42
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 42

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Leu Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65 70 75 80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
85 90 95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100 105 110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
115 120 125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130 135 140

ES 2 773 664 T3

His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg		
145					150					155					160		
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn		
				165					170					175			
Leu	Leu	Gln	Gly	Tyr	Ser	Met	Ala	Cys	Ser	Asn	Lys	Gln	Tyr	Gly	Val		
			180					185					190				
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly		
		195					200					205					
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr		
	210					215					220						
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu		
225					230					235					240		
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe		
				245					250					255			
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu		
			260					265					270				
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu		
		275					280					285					
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala		
	290					295					300						
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala		
305					310					315					320		
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser		
				325					330					335			
Glu	Ala	Asp	Phe	Glu	Gln	Leu	Ala	Gln	Asp	Gly	Val	Pro	Met	Val	Glu		
			340					345					350				
Met	Glu	Lys	Gly	Ala	Asp	Leu	Met	Gln	Ala	Leu	Val	Asp	Ser	Glu	Leu		
		355					360					365					
Gln	Pro	Ser	Arg	Gly	Gln	Ala	Arg	Lys	Thr	Ile	Ala	Ser	Asn	Ala	Ile		
	370					375					380						
Thr	Ile	Asn	Gly	Glu	Lys	Gln	Ser	Asp	Pro	Glu	Tyr	Phe	Phe	Lys	Glu		
385					390					395					400		

ES 2 773 664 T3

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 43
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 43

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Leu Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65 70 75 80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
85 90 95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100 105 110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
115 120 125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130 135 140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145 150 155 160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
165 170 175

ES 2 773 664 T3

Leu Leu Gln Gly Tyr Ser Met Ala Cys Ala Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285
 Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300
 Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320
 Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335
 Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350
 Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365
 Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
 370 375 380
 Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
 385 390 395 400
 Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
 405 410 415
 Asn Tyr Cys Leu Ile Cys Trp Lys
 420

ES 2 773 664 T3

<210> 44
 <211> 424
 <212> PRT
 <213> artificial

5

<220>
 <223> sintetasa artificial

10

<400> 44

```

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1           5           10           15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
                20           25           30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
                35           40           45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50           55           60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65           70           75           80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
                85           90           95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
                100          105          110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
                115          120          125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130          135          140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145          150          155          160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
                165          170          175

Leu Leu Gln Gly Tyr Arg Met Ala Cys Leu Asn Lys Gln Tyr Gly Val
                180          185          190

Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
195          200          205
  
```

ES 2 773 664 T3

Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
210 215 220

Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
225 230 235 240

Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
245 250 255

Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
260 265 270

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 45
<211> 424
<212> PRT
<213> artificial
<220>

<223> sintetasa artificial

<400> 45

```

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1          5          10          15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20          25          30

Pro Ile Ala Leu Ile Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35          40          45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50          55          60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65          70          75          80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
85          90          95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100         105         110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
115         120         125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130         135         140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145         150         155         160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
165         170         175

Leu Leu Gln Gly Tyr Gly Met Ala Cys Ala Asn Lys Gln Tyr Gly Val
180         185         190

Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
195         200         205

Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
210         215         220

Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
225         230         235         240

```

ES 2 773 664 T3

Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285
 Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300
 Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320
 Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335
 Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350
 Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365
 Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
 370 375 380
 Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
 385 390 395 400
 Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
 405 410 415
 Asn Tyr Cys Leu Ile Cys Trp Lys
 420

5 <210> 46
 <211> 424
 <212> PRT
 <213> artificial

10 <220>
 <223> sintetasa artificial

<400> 46

ES 2 773 664 T3

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly

ES 2 773 664 T3

20					25					30					
Pro	Ile	Ala	Leu	Gly	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His
		35					40					45			
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala
	50					55					60				
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly
65					70					75					80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr
				85					90					95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu
			100					105					110		
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp
		115					120					125			
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys
	130					135					140				
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg
145					150					155					160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn
				165					170					175	
Leu	Leu	Gln	Gly	Tyr	Gly	Phe	Ala	Cys	Ala	Asn	Lys	Gln	Tyr	Gly	Val
			180					185					190		
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly
		195					200					205			
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr
	210					215					220				
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu
225					230					235					240
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe
				245					250					255	
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu
			260					265					270		

ES 2 773 664 T3

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 47
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 47

ES 2 773 664 T3

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Gly Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala

ES 2 773 664 T3

50	55	60
Gly 65	His Lys Pro Val Ala 70	Leu Val Gly Gly Ala 75 Thr Gly Leu Ile Gly 80
Asp	Pro Ser Phe 85	Lys Ala Ala Glu Arg Lys 90 Leu Asn Thr Glu Glu 95 Thr
Val	Gln Glu Trp 100	Val Asp Lys Ile Arg Lys 105 Gln Val Ala Pro Phe Leu 110
Asp	Phe Asp 115	Cys Gly Glu Asn Ser Ala 120 Ile Ala Ala Asn Asn Tyr Asp 125
Trp	Phe Gly Asn Met Asn 130	Val Leu Thr Phe Leu Arg Asp 140 Ile Gly Lys
His 145	Phe Ser Val Asn 150	Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg 160
Leu	Asn Arg Glu Asp 165	Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn 175
Leu	Leu Gln Gly Tyr 180	Gly Tyr Ala Cys Met Asn Lys Gln Tyr Gly Val 190
Val	Leu Gln Ile Gly Gly Ser 200	Asp Gln Trp Gly Asn Ile Thr Ser Gly 205
Ile	Asp Leu Thr Arg Arg 210	Leu His Gln Asn Gln Val Phe Gly Leu Thr 220
Val	Pro Leu Ile Thr Lys 230	Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu 240
Gly	Gly Ala Val Trp 245	Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe 255
Tyr	Gln Phe Trp 260	Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu 270
Lys	Phe Phe Thr Phe Met Ser 275	Ile Glu Glu Ile Asn Ala Leu Glu Glu 285
Glu	Asp Lys Asn Ser Gly 290	Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala 300

ES 2 773 664 T3

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 48
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 48

ES 2 773 664 T3

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Leu Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65 70 75 80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr

ES 2 773 664 T3

				85					90					95					
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu				
			100					105					110						
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp				
		115					120					125							
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys				
		130					135				140								
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg				
145					150					155					160				
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn				
				165					170					175					
Leu	Leu	Gln	Gly	Tyr	Ser	Met	Ala	Cys	Ala	Asn	Lys	Gln	Tyr	Gly	Val				
			180					185					190						
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly				
		195					200					205							
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr				
	210					215					220								
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu				
225					230					235					240				
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe				
				245					250					255					
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu				
			260					265					270						
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu				
		275					280					285							
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala				
	290					295					300								
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala				
305					310					315					320				
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser				
				325					330					335					

ES 2 773 664 T3

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 49
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 49

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Val	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp				

ES 2 773 664 T3

115		120		125
Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys				
130		135		140
His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg				
145		150		155
160				
Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn				
		165		170
				175
Leu Leu Gln Gly Tyr Ser Ala Ala Cys Ala Asn Lys Gln Tyr Gly Val				
		180		185
				190
Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly				
		195		200
				205
Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr				
		210		215
				220
Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu				
		225		230
				235
Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe				
		245		250
				255
Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu				
		260		265
				270
Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu				
		275		280
				285
Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala				
		290		295
				300
Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala				
		305		310
				315
Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser				
		325		330
				335
Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu				
		340		345
				350
Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu				
		355		360
				365

ES 2 773 664 T3

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 50
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 50

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Leu	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg				

ES 2 773 664 T3

145					150						155					160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	
				165					170					175		
Leu	Leu	Gln	Gly	Tyr	Ser	Ala	Ala	Cys	Ala	Asn	Lys	Gln	Tyr	Gly	Val	
			180					185					190			
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	
		195					200					205				
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr	
	210					215					220					
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu	
225					230					235					240	
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe	
				245					250					255		
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu	
			260					265					270			
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu	
		275					280					285				
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala	
	290					295					300					
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala	
305					310					315					320	
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser	
				325					330					335		
Glu	Ala	Asp	Phe	Glu	Gln	Leu	Ala	Gln	Asp	Gly	Val	Pro	Met	Val	Glu	
			340					345					350			
Met	Glu	Lys	Gly	Ala	Asp	Leu	Met	Gln	Ala	Leu	Val	Asp	Ser	Glu	Leu	
		355					360					365				
Gln	Pro	Ser	Arg	Gly	Gln	Ala	Arg	Lys	Thr	Ile	Ala	Ser	Asn	Ala	Ile	
	370					375					380					
Thr	Ile	Asn	Gly	Glu	Lys	Gln	Ser	Asp	Pro	Glu	Tyr	Phe	Phe	Lys	Glu	
385					390					395					400	

ES 2 773 664 T3

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
 405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
 420

5 <210> 51
 <211> 424
 <212> PRT
 <213> artificial

10 <220>
 <223> sintetasa artificial
 <400> 51

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Val	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155	160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175	
Leu	Leu	Gln	Gly	Tyr	Ser	Ala	Ala	Cys	Val	Asn	Lys	Gln	Tyr	Gly	Val				

ES 2 773 664 T3

180								185				190			
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly
		195					200					205			
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr
	210					215					220				
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu
225					230					235					240
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe
				245					250					255	
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu
			260					265					270		
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu
		275					280					285			
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala
	290					295					300				
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala
305					310					315					320
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser
				325					330					335	
Glu	Ala	Asp	Phe	Glu	Gln	Leu	Ala	Gln	Asp	Gly	Val	Pro	Met	Val	Glu
			340					345					350		
Met	Glu	Lys	Gly	Ala	Asp	Leu	Met	Gln	Ala	Leu	Val	Asp	Ser	Glu	Leu
		355					360					365			
Gln	Pro	Ser	Arg	Gly	Gln	Ala	Arg	Lys	Thr	Ile	Ala	Ser	Asn	Ala	Ile
	370					375					380				
Thr	Ile	Asn	Gly	Glu	Lys	Gln	Ser	Asp	Pro	Glu	Tyr	Phe	Phe	Lys	Glu
385					390					395					400
Glu	Asp	Arg	Leu	Phe	Gly	Arg	Phe	Thr	Leu	Leu	Arg	Arg	Gly	Lys	Lys
				405					410					415	
Asn	Tyr	Cys	Leu	Ile	Cys	Trp	Lys								
			420												

ES 2 773 664 T3

	<210> 52
	<211> 424
	<212> PRT
	<213> artificial
5	<220>
	<223> sintetasa artificial
10	<400> 52

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Ile	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asp	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155	160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175	
Leu	Leu	Gln	Gly	Tyr	Asn	Phe	Ala	Cys	Val	Asn	Lys	Gln	Tyr	Gly	Val	180	185	190	
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	195	200	205	
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr				

ES 2 773 664 T3

210	215	220
Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu 225 230 235 240		
Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe 245 250 255		
Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu 260 265 270		
Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu 275 280 285		
Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala 290 295 300		
Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala 305 310 315 320		
Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser 325 330 335		
Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu 340 345 350		
Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu 355 360 365		
Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile 370 375 380		
Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu 385 390 395 400		
Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys 405 410 415		
Asn Tyr Cys Leu Ile Cys Trp Lys 420		

<210> 53
 <211> 424
 <212> PRT
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 53

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Thr	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155	160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175	
Leu	Leu	Gln	Gly	Tyr	Ser	Ala	Ala	Cys	Leu	Asn	Lys	Gln	Tyr	Gly	Val	180	185	190	
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	195	200	205	
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr	210	215	220	
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu	225	230	235	240
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe				

ES 2 773 664 T3

				245				250						255	
Tyr	Gln	Phe	Trp 260	Ile	Asn	Thr	Ala	Asp 265	Ala	Asp	Val	Tyr	Arg 270	Phe	Leu
Lys	Phe	Phe 275	Thr	Phe	Met	Ser	Ile 280	Glu	Glu	Ile	Asn	Ala 285	Leu	Glu	Glu
Glu	Asp 290	Lys	Asn	Ser	Gly	Lys 295	Ala	Pro	Arg	Ala	Gln 300	Tyr	Val	Leu	Ala
Glu 305	Gln	Val	Thr	Arg	Leu 310	Val	His	Gly	Glu	Glu 315	Gly	Leu	Gln	Ala	Ala 320
Lys	Arg	Ile	Thr	Glu 325	Cys	Leu	Phe	Ser	Gly 330	Ser	Leu	Ser	Ala	Leu 335	Ser
Glu	Ala	Asp	Phe 340	Glu	Gln	Leu	Ala	Gln 345	Asp	Gly	Val	Pro	Met 350	Val	Glu
Met	Glu	Lys 355	Gly	Ala	Asp	Leu	Met 360	Gln	Ala	Leu	Val	Asp 365	Ser	Glu	Leu
Gln	Pro 370	Ser	Arg	Gly	Gln	Ala 375	Arg	Lys	Thr	Ile	Ala 380	Ser	Asn	Ala	Ile
Thr 385	Ile	Asn	Gly	Glu	Lys 390	Gln	Ser	Asp	Pro	Glu 395	Tyr	Phe	Phe	Lys	Glu 400
Glu	Asp	Arg	Leu	Phe 405	Gly	Arg	Phe	Thr	Leu 410	Leu	Arg	Arg	Gly	Lys 415	Lys
Asn	Tyr	Cys	Leu 420	Ile	Cys	Trp	Lys								

5 <210> 54
 <211> 424
 <212> PRT
 <213> artificial

10 $\langle 220 \rangle$
 $\langle 223 \rangle$ sintetasa artificial

<400> 54

ES 2 773 664 T3

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
 1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
 20 25 30

ES 2 773 664 T3

Pro Ile Ala Leu Gly Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
 35 40 45
 Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50 55 60
 Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65 70 75 80
 Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
 85 90 95
 Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
 100 105 110
 Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
 115 120 125
 Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
 130 135 140
 His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
 145 150 155 160
 Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175
 Leu Leu Gln Gly Tyr Ser Met Ala Cys Leu Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu

ES 2 773 664 T3

275	280	285
Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala		
290	295	300
Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala		
305	310	315
Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser		
325	330	335
Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu		
340	345	350
Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu		
355	360	365
Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile		
370	375	380
Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu		
385	390	395
Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys		
405	410	415
Asn Tyr Cys Leu Ile Cys Trp Lys		
420		

<210> 55
 <211> 424
 <212> PRT
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 55

ES 2 773 664 T3

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

ES 2 773 664 T3

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65 70 75 80
 Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
 85 90 95
 Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
 100 105 110
 Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
 115 120 125
 Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
 130 135 140
 His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
 145 150 155 160
 Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175
 Leu Leu Gln Gly Tyr Ser Ala Ala Cys Leu Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285
 Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300
 Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala

ES 2 773 664 T3

305		310		315		320
Lys Arg Ile Thr	Glu Cys Leu Phe Ser Gly	Ser Leu Ser Ala Leu Ser				
	325	330				335
Glu Ala Asp Phe	Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu					
	340	345				350
Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu						
	355	360				365
Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile						
	370	375				380
Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu						
	385	390				395
Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys						
	405	410				415
Asn Tyr Cys Leu Ile Cys Trp Lys						
	420					

<210> 56
 <211> 424
 <212> PRT
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 56

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val
1				5					10					15	
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly
			20					25					30		
Pro	Ile	Ala	Leu	Ser	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His
		35					40					45			
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala
	50					55					60				
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly
65					70					75				80	
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr
				85					90					95	

ES 2 773 664 T3

Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu			
			100					105					110					
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp			
		115					120					125						
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys			
	130					135					140							
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg			
145					150					155					160			
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn			
				165					170					175				
Leu	Leu	Gln	Gly	Tyr	Thr	Met	Ala	Cys	Val	Asn	Lys	Gln	Tyr	Gly	Val			
			180					185					190					
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly			
		195					200					205						
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr			
	210					215					220							
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu			
225					230					235					240			
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe			
				245					250					255				
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu			
			260					265					270					
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu			
		275					280					285						
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala			
	290					295					300							
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala			
305					310					315					320			
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser			
				325					330					335				
Glu	Ala	Asp	Phe	Glu	Gln	Leu	Ala	Gln	Asp	Gly	Val	Pro	Met	Val	Glu			

ES 2 773 664 T3

340

345

350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 57
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 57

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val
1				5					10					15	
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly
			20					25					30		
Pro	Ile	Ala	Leu	Ala	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His
		35					40					45			
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala
	50					55					60				
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly
65					70					75				80	
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr
				85					90					95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu
			100					105					110		
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp
		115					120					125			

ES 2 773 664 T3

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
 130 135 140
 His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
 145 150 155 160
 Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175
 Leu Leu Gln Gly Tyr Ser Tyr Ala Cys Leu Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285
 Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300
 Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320
 Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335
 Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350
 Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365
 Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile

ES 2 773 664 T3

370

375

380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 58

<211> 424

<212> PRT

<213> artificial

<220>

<223> sintetasa artificial

<400> 58

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Ala	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155	160

ES 2 773 664 T3

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175
 Leu Leu Gln Gly Tyr Thr Met Ala Cys Cys Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285
 Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300
 Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320
 Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335
 Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350
 Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365
 Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
 370 375 380
 Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
 385 390 395 400
 Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys

ES 2 773 664 T3

405

410

415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 59
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 59

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65 70 75 80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
85 90 95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
100 105 110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
115 120 125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
130 135 140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145 150 155 160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
165 170 175

Leu Leu Gln Gly Tyr Thr Phe Ala Cys Met Asn Lys Gln Tyr Gly Val
180 185 190

ES 2 773 664 T3

Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly
	195						200					205			
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr
	210					215					220				
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu
225					230					235					240
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe
				245					250					255	
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu
			260					265					270		
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu
		275					280					285			
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala
	290					295					300				
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala
305					310					315					320
Lys	Arg	Ile	Thr	Glu	Cys	Leu	Phe	Ser	Gly	Ser	Leu	Ser	Ala	Leu	Ser
				325					330					335	
Glu	Ala	Asp	Phe	Glu	Gln	Leu	Ala	Gln	Asp	Gly	Val	Pro	Met	Val	Glu
			340					345					350		
Met	Glu	Lys	Gly	Ala	Asp	Leu	Met	Gln	Ala	Leu	Val	Asp	Ser	Glu	Leu
		355					360					365			
Gln	Pro	Ser	Arg	Gly	Gln	Ala	Arg	Lys	Thr	Ile	Ala	Ser	Asn	Ala	Ile
	370					375					380				
Thr	Ile	Asn	Gly	Glu	Lys	Gln	Ser	Asp	Pro	Glu	Tyr	Phe	Phe	Lys	Glu
385					390					395					400
Glu	Asp	Arg	Leu	Phe	Gly	Arg	Phe	Thr	Leu	Leu	Arg	Arg	Gly	Lys	Lys
				405					410					415	
Asn	Tyr	Cys	Leu	Ile	Cys	Trp	Lys								
			420												

<210> 60

<211> 424
<212> PRT
<213> artificial

5 <220>
<223> sintetasa artificial

<400> 60

```

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
 1          5          10          15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
          20          25          30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
          35          40          45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50          55          60

Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65          70          75          80

Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
          85          90          95

Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
          100          105          110

Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
          115          120          125

Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
          130          135          140

His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
          145          150          155          160

Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
          165          170          175

Leu Leu Gln Gly Tyr Ser Val Ala Cys Leu Asn Lys Gln Tyr Gly Val
          180          185          190

Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
          195          200          205

Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
          210          215          220

```

ES 2 773 664 T3

Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240

Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255

Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
 370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
 385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
 405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
 420

<210> 61
 <211> 424
 <212> PRT
 <213> artificial

<220>
 <223> sintetasa artificial

<400> 61

ES 2 773 664 T3

Met	Ala	Ser	Ser	Asn	Leu	Ile	Lys	Gln	Leu	Gln	Glu	Arg	Gly	Leu	Val	1	5	10	15
Ala	Gln	Val	Thr	Asp	Glu	Glu	Ala	Leu	Ala	Glu	Arg	Leu	Ala	Gln	Gly	20	25	30	
Pro	Ile	Ala	Leu	Val	Cys	Gly	Phe	Asp	Pro	Thr	Ala	Asp	Ser	Leu	His	35	40	45	
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala	50	55	60	
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	65	70	75	80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	85	90	95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	100	105	110	
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	115	120	125	
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	130	135	140	
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	145	150	155	160
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	165	170	175	
Leu	Leu	Gln	Gly	Tyr	Ser	Met	Ala	Cys	Thr	Asn	Lys	Gln	Tyr	Gly	Val	180	185	190	
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	195	200	205	
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr	210	215	220	
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu	225	230	235	240
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe	245	250	255	

ES 2 773 664 T3

Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
260 265 270

Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
275 280 285

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 62
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 62

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

ES 2 773 664 T3

Pro Ile Ala Leu Ser Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
 35 40 45
 Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50 55 60
 Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
 65 70 75 80
 Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
 85 90 95
 Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
 100 105 110
 Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
 115 120 125
 Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
 130 135 140
 His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
 145 150 155 160
 Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
 165 170 175
 Leu Leu Gln Gly Tyr Ser Phe Ala Cys Leu Asn Lys Gln Tyr Gly Val
 180 185 190
 Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
 195 200 205
 Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
 210 215 220
 Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
 225 230 235 240
 Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
 245 250 255
 Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
 260 265 270
 Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu
 275 280 285

ES 2 773 664 T3

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 63
<211> 424
<212> PRT
<213> artificial

<220>
<223> sintetasa artificial

<400> 63

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
1 5 10 15

Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
20 25 30

Pro Ile Ala Leu Thr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
35 40 45

Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
50 55 60

ES 2 773 664 T3

Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly	
65					70					75					80	
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr	
				85					90					95		
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu	
			100					105					110			
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp	
		115					120					125				
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys	
	130					135					140					
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg	
145				150						155					160	
Leu	Asn	Arg	Glu	Asp	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn	
			165						170					175		
Leu	Leu	Gln	Gly	Tyr	Thr	Phe	Ala	Cys	Thr	Asn	Lys	Gln	Tyr	Gly	Val	
		180						185					190			
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly	
		195					200					205				
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr	
	210					215					220					
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu	
225					230					235					240	
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe	
			245						250					255		
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu	
		260						265					270			
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu	
		275					280					285				
Glu	Asp	Lys	Asn	Ser	Gly	Lys	Ala	Pro	Arg	Ala	Gln	Tyr	Val	Leu	Ala	
	290					295					300					
Glu	Gln	Val	Thr	Arg	Leu	Val	His	Gly	Glu	Glu	Gly	Leu	Gln	Ala	Ala	
305					310					315					320	

ES 2 773 664 T3

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
420

<210> 64
<211> 129
<212> ADN
<213> Escherichia coli

<400> 64

agcttcccgga taagggagca ggccagtaaa aagcattacc ccgtggtggg gttcccgagc 60
ggccaaaggg agcagactct aaatctgccg tcatcgacct cgaagggtcg aatccttccc 120
ccaccacca 129

<210> 65
<211> 129
<212> RNA
<213> Escherichia coli

<400> 65

agcuucccgga uaagggagca ggccaguaaa aagcauuacc ccgugguggg guucccgagc 60
ggccaaaggg agcagacucu aaucugccg ucaucgaccu cgaagguucg aauccuuccc 120
ccaccacca 129

<210> 66
<211> 34
<212> ADN
<213> artificial

<220>
<223> Cebador oligonucleotídico

	<400> 66 atgaagtagc tgtcttctat cgaacaagca tgcg	34
5	<210> 67 <211> 34 <212> ADN <213> artificial	
10	<220> <223> Cebador oligonucleotídico	
	<400> 67 cgaacaagca tgcgattagt gccgacttaa aaag	34
15	<210> 68 <211> 33 <212> ADN <213> artificial	
20	<220> <223> Cebador oligonucleotídico	
25	<400> 68 cgctactctc ccaaataagaa aaggctccg ctg	33
30	<210> 69 <211> 32 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
35	<400> 69 ctggaacagc tatagctact gatttttct ctg	32
40	<210> 70 <211> 34 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
45	<400> 70 gccgtcacag attagtggc ttcagtggag actg	34
50	<210> 71 <211> 33 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
55	<400> 71 gattggcttc ataggagact gatatgctct aac	33
60	<210> 72 <211> 33 <212> ADN <213> artificial	
65	<220> <223> Cebador oligonucleotídico	

	<400> 72 gcctctatag ttgagacagc atagaataat gcg	33
5	<210> 73 <211> 35 <212> ADN <213> artificial	
10	<220> <223> Cebador oligonucleotídico	
	<400> 73 gagacagcat agatagagtg cgacatcatc atcgg	35
15	<210> 74 <211> 37 <212> ADN <213> artificial	
20	<220> <223> Cebador oligonucleotídico	
25	<400> 74 gaataagtgc gacatagtca tcggaagaga gtagtag	37
30	<210> 75 <211> 35 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
35	<400> 75 ggtcaaagac agttgtaggt atcgattgac tcggc	35
40	<210> 76 <211> 34 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
45	<400> 76 cgctactctc cccaaattta aaaggtctcc gctg	34
50	<210> 77 <211> 34 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
55	<400> 77 cgctactctc cccaaataata aaaggtctcc gctg	34
60	<210> 78 <211> 34 <212> ADN <213> artificial	
65	<220> <223> Cebador oligonucleotídico	

	<400> 78 cgctactctc cccaaatgga aaaggtctcc gctg	34
5	<210> 79 <211> 34 <212> ADN <213> artificial	
10	<220> <223> Cebador oligonucleotídico	
	<400> 79 cgctactctc cccaaagata aaaggtctcc gctg	34
15	<210> 80 <211> 34 <212> ADN <213> artificial	
20	<220> <223> Cebador oligonucleotídico	
25	<400> 80 cgctactctc cccaaaaaaa aaaggtctcc gctg	34
30	<210> 81 <211> 34 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
35	<400> 81 gccgtcacag atttttggc ttcagtggag actg	34
40	<210> 82 <211> 34 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
45	<400> 82 gccgtcacag attatttggc ttcagtggag actg	34
50	<210> 83 <211> 34 <212> ADN <213> artificial	
	<220> <223> Cebador oligonucleotídico	
55	<400> 83 gccgtcacag attggttggc ttcagtggag actg	34
60	<210> 84 <211> 34 <212> ADN <213> artificial	
65	<220> <223> Cebador oligonucleotídico	

ES 2 773 664 T3

<400> 84
 gccgtcacag atgattggc ttcagtggag actg 34

 5 <210> 85
 <211> 34
 <212> ADN
 <213> artificial

 10 <220>
 <223> Cebador oligonucleotídico

 <400> 85
 gccgtcacag ataaattggc ttcagtggag actg 34

 15 <210> 86
 <211> 424
 <212> PRT
 <213> artificial

 20 <220>
 <223> sintetasa artificial

 <400> 86

 Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
 1 5 10 15

 Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
 20 25 30

 Pro Ile Ala Leu Ile Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
 25

ES 2 773 664 T3

35					40					45					
Leu	Gly	His	Leu	Val	Pro	Leu	Leu	Cys	Leu	Lys	Arg	Phe	Gln	Gln	Ala
50					55					60					
Gly	His	Lys	Pro	Val	Ala	Leu	Val	Gly	Gly	Ala	Thr	Gly	Leu	Ile	Gly
65					70					75					80
Asp	Pro	Ser	Phe	Lys	Ala	Ala	Glu	Arg	Lys	Leu	Asn	Thr	Glu	Glu	Thr
				85					90					95	
Val	Gln	Glu	Trp	Val	Asp	Lys	Ile	Arg	Lys	Gln	Val	Ala	Pro	Phe	Leu
			100					105					110		
Asp	Phe	Asp	Cys	Gly	Glu	Asn	Ser	Ala	Ile	Ala	Ala	Asn	Asn	Tyr	Asp
		115					120					125			
Trp	Phe	Gly	Asn	Met	Asn	Val	Leu	Thr	Phe	Leu	Arg	Asp	Ile	Gly	Lys
		130				135					140				
His	Phe	Ser	Val	Asn	Gln	Met	Ile	Asn	Lys	Glu	Ala	Val	Lys	Gln	Arg
145					150					155					160
Leu	Asn	Arg	Glu	Gly	Gln	Gly	Ile	Ser	Phe	Thr	Glu	Phe	Ser	Tyr	Asn
				165					170					175	
Leu	Leu	Gln	Gly	Tyr	Gly	Met	Ala	Cys	Ala	Asn	Lys	Gln	Tyr	Gly	Val
			180					185					190		
Val	Leu	Gln	Ile	Gly	Gly	Ser	Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ser	Gly
		195					200					205			
Ile	Asp	Leu	Thr	Arg	Arg	Leu	His	Gln	Asn	Gln	Val	Phe	Gly	Leu	Thr
	210					215					220				
Val	Pro	Leu	Ile	Thr	Lys	Ala	Asp	Gly	Thr	Lys	Phe	Gly	Lys	Thr	Glu
225					230					235					240
Gly	Gly	Ala	Val	Trp	Leu	Asp	Pro	Lys	Lys	Thr	Ser	Pro	Tyr	Lys	Phe
				245					250					255	
Tyr	Gln	Phe	Trp	Ile	Asn	Thr	Ala	Asp	Ala	Asp	Val	Tyr	Arg	Phe	Leu
			260					265					270		
Lys	Phe	Phe	Thr	Phe	Met	Ser	Ile	Glu	Glu	Ile	Asn	Ala	Leu	Glu	Glu
		275					280					285			

Glu Asp Lys Asn Ser Gly Lys Ala Pro Arg Ala Gln Tyr Val Leu Ala
 290 295 300

Glu Gln Val Thr Arg Leu Val His Gly Glu Glu Gly Leu Gln Ala Ala
 305 310 315 320

Lys Arg Ile Thr Glu Cys Leu Phe Ser Gly Ser Leu Ser Ala Leu Ser
 325 330 335

Glu Ala Asp Phe Glu Gln Leu Ala Gln Asp Gly Val Pro Met Val Glu
 340 345 350

Met Glu Lys Gly Ala Asp Leu Met Gln Ala Leu Val Asp Ser Glu Leu
 355 360 365

Gln Pro Ser Arg Gly Gln Ala Arg Lys Thr Ile Ala Ser Asn Ala Ile
 370 375 380

Thr Ile Asn Gly Glu Lys Gln Ser Asp Pro Glu Tyr Phe Phe Lys Glu
 385 390 395 400

Glu Asp Arg Leu Phe Gly Arg Phe Thr Leu Leu Arg Arg Gly Lys Lys
 405 410 415

Asn Tyr Cys Leu Ile Cys Trp Lys
 420

<210> 87
 <211> 6
 <212> PRT
 <213> artificial

<220>
 <223> péptido triptico que incluye aminoácidos artificiales

<220>
 <221> MISC_FEATURE
 <222> (2)..(2)

<223> X es un aminoácido artificial (p-acetil-L-fenilalanina, p-benzoil-L-fenilalanina, p-azido-L-fenilalanina, O-metil-L-tirosina, o p-yodo-L-fenilalanina) o triptófano, tirosina, o leucina

<400> 87

Val Xaa Gly Ser Ile Lys
 1 5

<210> 88
 <211> 11
 <212> ADN
 <213> artificial

<220>
 <223> Caja B

5	<220> <221> misc_feature <222> (8)..(8) <223> n es a, c, g, o t	
	<400> 88 gggtcgantc c	11
10	<210> 89 <211> 82 <212> ADN <213> artificial	
15	<220> <223> Cebador oligonucleotídico	
	<400> 89	
	gggggggaccg gtgggggggac cggtaagctt cccgataagg gagcaggcca gtaaaaagca	60
20	ttaccccgtg gtgggttccc ga	82
	<210> 90 <211> 90 <212> ADN <213> artificial	
25	<220> <223> Cebador oligonucleotídico	
30	<400> 90	
	ggcggcgcta gcaagcttcc cgataaggga gcaggccagt aaaaaggga gttcaggga	60
	ttttgaaaaa aatggtggtg ggggaaggat	90
35	<210> 91 <211> 68 <212> ADN <213> artificial	
40	<220> <223> Cebador oligonucleotídico	
	<220> <221> misc_feature <222> (1)..(1) <223> n = l	
45	<220> <221> misc_feature <222> (14)..(14) <223> n = l	
50	<400> 91	
	nggggggacc ggtngggggg accggtcggg atcgaagaaa tgatggtaaa tgaaatagga	60
55	aatcaagg	68
60	<210> 92 <211> 62 <212> ADN <213> artificial	

	<220>		
	<223> Cebador oligonucleotídico		
5	<400> 92		
	ggggggggaat tcagttgatt gtatgcttgg tatagcttga aatattgtgc agaaaaagaa	60	
	ac	62	
10	<210> 93 <211> 86 <212> ADN <213> artificial		
15	<220> <223> Cebador oligonucleotídico		
	<400> 93		
	tcataacgag aattccggga tcgaagaaat gatggtaaat gaaataggaa atctcataac	60	
	gagaattcat ggcaagcagt aacttg	86	
20	<210> 94 <211> 72 <212> ADN <213> artificial		
25	<220> <223> Cebador oligonucleotídico		
30	<400> 94		
	ttactacgtg cggccgcatg gcaagcagta acttgttact acgtgcggcc gcttatttcc	60	
	agcaaatacag ac	72	
35	<210> 95 <211> 28 <212> ADN <213> artificial		
40	<220> <223> Cebador oligonucleotídico		
	<400> 95		
	ccgatcgcg tcgcttgcgg ctgcgac	28	
45	<210> 96 <211> 27 <212> ADN <213> artificial		
50	<220> <223> Cebador oligonucleotídico		
	<400> 96		
	atcgcgcgga acgcctatga ctggttc	27	
55	<210> 97 <211> 40 <212> ADN <213> artificial		

	<220> <223> Cebador oligonucleotídico	
5	<400> 97 gttgagggt tatgccgccc cctgtgcgaa caaacagtac	40
10	<210> 98 <211> 26 <212> ADN <213> artificial	
15	<220> <223> Cebador oligonucleotídico	
15	<400> 98 gccgcttgc tatcaagtat aaatag	26
20	<210> 99 <211> 21 <212> ADN <213> artificial	
25	<220> <223> Cebador oligonucleotídico	
25	<400> 99 caagccgaca acctgattg g	21
30	<210> 100 <211> 60 <212> ADN <213> artificial	
35	<220> <223> Cebador oligonucleotídico	
40	<400> 100 ggggacaagt ttgtacaaaa aagcaggcta cgccaatttt aatcaaagt ggaatattgc	60
40	<210> 101 <211> 60 <212> ADN <213> artificial	
45	<220> <223> Cebador oligonucleotídico	
50	<400> 101 ggggacaagt ttgtacaaaa aagcaggcta ggccaatttt aatcaaagt ggaatattgc	60
55	<210> 102 <211> 58 <212> ADN <213> artificial	
55	<220> <223> Cebador oligonucleotídico	
60	<400> 102 ggggaccact ttgtacaaga aagctgggtt actctttttt tgggttggt ggggtatc	58
65	<210> 103 <211> 22 <212> ADN <213> artificial	

ES 2 773 664 T3

	<220>	
	<223> Cebador oligonucleotídico	
5	<400> 103 aagctatacc aagcatacaa tc	22
	<210> 104	
	<211> 49	
10	<212> ADN	
	<213> artificial	
	<220>	
	<223> Cebador oligonucleotídico	
15	<400> 104 acaaggcctt gctagcttac tcttttttg ggtttggtgg ggtatcttc	49

REIVINDICACIONES

1. Una composición que comprende una célula eucariota, célula que comprende un anticuerpo o un fragmento de anticuerpo, en la que:

(i) el anticuerpo o el fragmento de anticuerpo comprenden al menos un aminoácido artificial que comprende un grupo reactivo para la unión de una molécula mediante una reacción de cicloadición [3 + 2] y en donde el al menos un aminoácido artificial se incorpora al anticuerpo o al anticuerpo fragmento *in vivo* en la célula eucariota mediante una aminoacil ARNt sintetasa ortogonal (O-RS) y un par de ARNt ortogonal; o

(ii) el anticuerpo o el fragmento de anticuerpo comprende al menos un aminoácido artificial que comprende un primer grupo reactivo incorporado mediante una O-RS y un par de ARNt ortogonal y al menos una modificación postraducciona, en donde la al menos una modificación postraducciona comprende la unión de una molécula que comprende un segundo grupo reactivo mediante una reacción de cicloadición [3 + 2] al al menos un aminoácido artificial que comprende un primer grupo reactivo, y en donde:

(a) la O-RS, o una porción de la misma están codificada por una secuencia polinucleotídica como se expone en una cualquiera de las SEQ ID NO: 20-35;

(b) la O-RS comprende una secuencia de aminoácidos como se expone en una cualquiera de las SEQ ID NO: 48-63; o

(c) la O-RS comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la de la tirosilaminoacil-ARNt sintetasa (TyrRS) natural de SEQ ID NO: 2 y comprende las siguientes sustituciones aminoacídicas:

valina, isoleucina, leucina, glicina, serina, alanina o treonina en una posición correspondiente a Tyr37 de SEQ ID NO: 2; y

treonina, serina, arginina, asparagina o glicina en una posición correspondiente a Asp182 de SEQ ID NO: 2,

y opcionalmente comprende además una o más de las siguientes sustituciones aminoacídicas:

aspartato en una posición correspondiente a Asn126 de SEQ ID NO: 2;

metionina, alanina, valina o tirosina en una posición correspondiente a Phe183 de SEQ ID NO: 2; y

serina, metionina, valina, cisteína, treonina o alanina en una posición correspondiente a Leu186 de SEQ ID NO: 2.

2. La composición de la reivindicación 1 (i), en la que la molécula se selecciona del grupo que consiste en una molécula orgánica pequeña, un compuesto citotóxico y una proteína o un polipéptido.

3. La composición de la reivindicación 1 (i), en la que el al menos un aminoácido artificial es p-azido-L-fenilalanina, p-benzoil-L-fenilalanina o p-propargiloxifenilalanina.

4. La composición de la reivindicación 1 (ii), en la que el primer grupo reactivo es un resto alquínico o azido y el segundo grupo reactivo es un resto azido o alquínico.

5. La composición de la reivindicación 1 (ii), en la que la al menos una modificación postraducciona se realiza *in vivo* en una célula eucariota.

6. Un método para producir en una célula eucariota al menos un anticuerpo o un fragmento de anticuerpo que comprende al menos un aminoácido artificial, comprendiendo el método:

cultivar, en un medio apropiado, una célula eucariota que comprende un ácido nucleico que comprende al menos un codón selector y codifica el anticuerpo o el fragmento de anticuerpo; en donde el medio comprende un aminoácido artificial y la célula eucariota 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 aminoácido artificial, en donde la O-RS comprende una secuencia de aminoácidos que corresponde a las SEQ ID NO.: 48-53.

7. Un método para producir un conjugado que comprende al menos un anticuerpo o un fragmento de anticuerpo que comprenden al menos un aminoácido artificial, en donde dichos anticuerpo o fragmento de anticuerpo se producen en una célula eucariota, comprendiendo el método:

cultivar, en un medio apropiado, una célula eucariota que comprende un ácido nucleico que comprende al menos un codón selector y codifica el anticuerpo o el fragmento de anticuerpo; en donde el medio comprende un aminoácido artificial y la célula eucariota 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 aminoácido artificial;

incorporar en el anticuerpo o el fragmento de anticuerpo el aminoácido artificial en la célula eucariota, en donde el aminoácido artificial comprende un primer grupo reactivo; y poner en contacto el anticuerpo o el fragmento de anticuerpo con una molécula que comprende un segundo grupo reactivo; en donde el primer grupo reactivo reacciona con el segundo grupo reactivo para unir la molécula al aminoácido artificial a través de una reacción de cicloadición, y en donde:

(a) la O-RS, o una porción de la misma, está codificada por una secuencia polinucleotídica como se expone en una cualquiera de las SEQ ID NO: 20-35;

(b) la O-RS comprende una secuencia de aminoácidos como se expone en una cualquiera de las SEQ ID NO: 48-63; o

(c) la O-RS comprende una secuencia de aminoácidos que es al menos un 90 % idéntica a la de la tirosilaminoacil-ARNt sintetasa (TyrRS) natural de SEQ ID NO: 2 y comprende las siguientes sustituciones aminoacídicas:

valina, isoleucina, leucina, glicina, serina, alanina o treonina en una posición correspondiente a Tyr37 de SEQ ID NO: 2; y

treonina, serina, arginina, asparagina o glicina en una posición correspondiente a Asp182 de SEQ ID NO: 2,

y opcionalmente comprende además una o más de las siguientes sustituciones aminoacídicas:

aspartato en una posición correspondiente a Asn126 de SEQ ID NO: 2;

metionina, alanina, valina o tirosina en una posición correspondiente a Phe183 de SEQ ID NO: 2; y

serina, metionina, valina, cisteína, treonina o alanina en una posición correspondiente a Leu186 de SEQ ID NO: 2.

8. El método de la reivindicación 7, en el que la reacción de cicloadición es una reacción de cicloadición [3 + 2].

9. El método de la reivindicación 7, en el que la molécula se selecciona del grupo que consiste en una molécula orgánica pequeña, un compuesto citotóxico y una proteína o un polipéptido.

10. El método de la reivindicación 7, en el que el primer grupo reactivo es un resto alquinilo o azido y el segundo grupo reactivo es un resto azido o alquinilo.

11. El método de la reivindicación 7, en el que el al menos un aminoácido artificial es p-azido-L-fenilalanina, p-benzoyl-L-fenilalanina o p-propargiloxifenilalanina.

12. El método de la reivindicación 7, en el que la reacción de cicloadición comprende al menos una modificación postraducciona realizada *in vivo* en la célula eucariota.

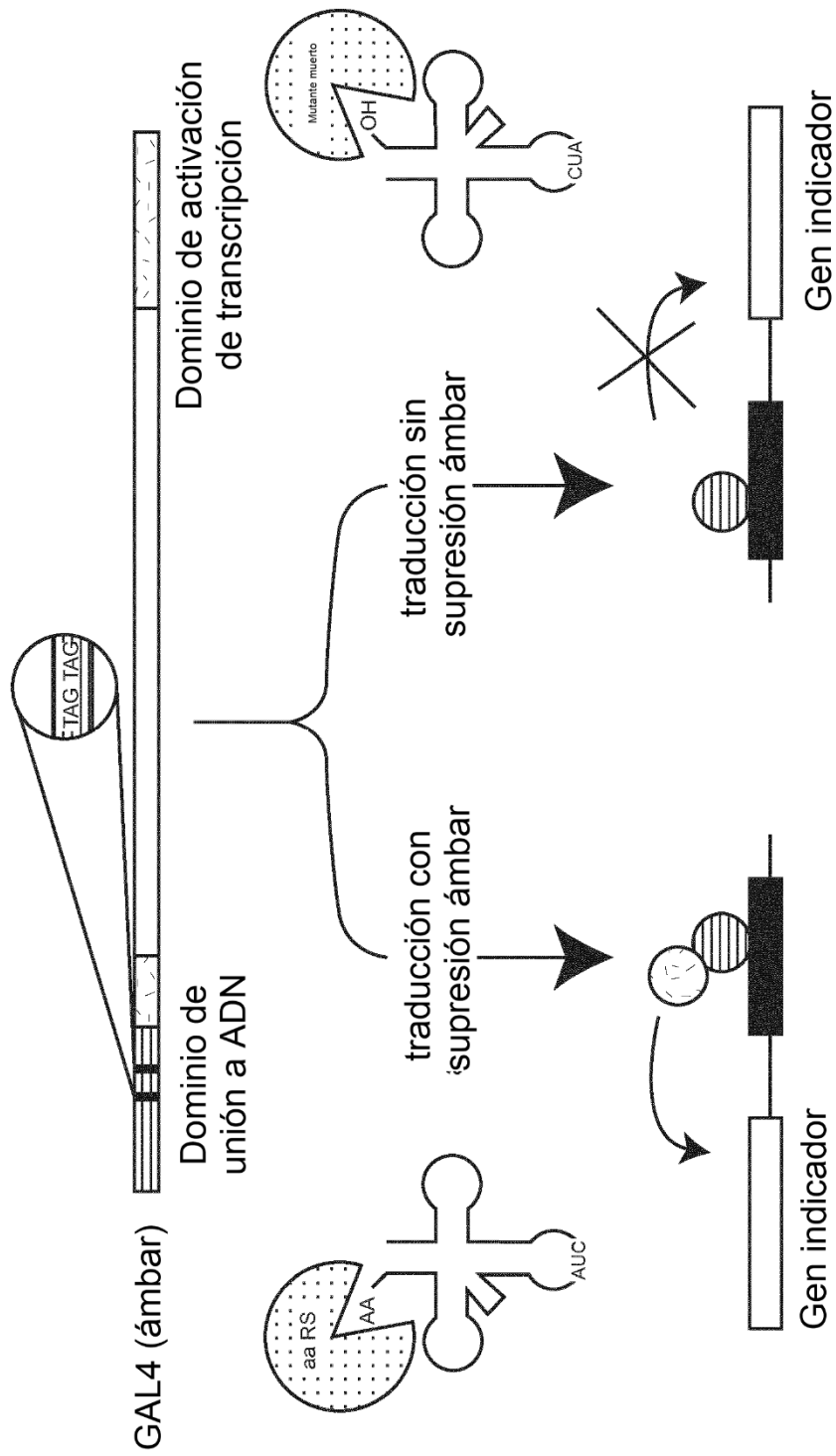


Fig. 1A

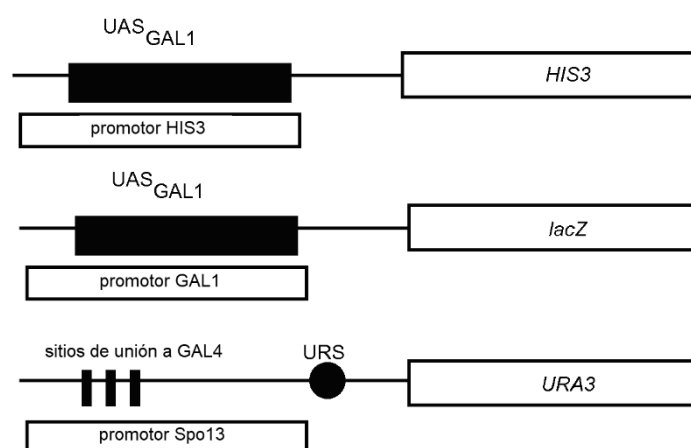


Fig. 1B

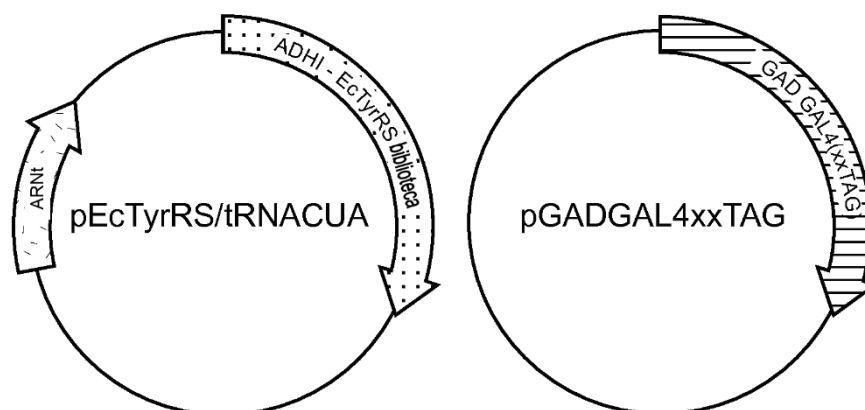


Fig. 1C

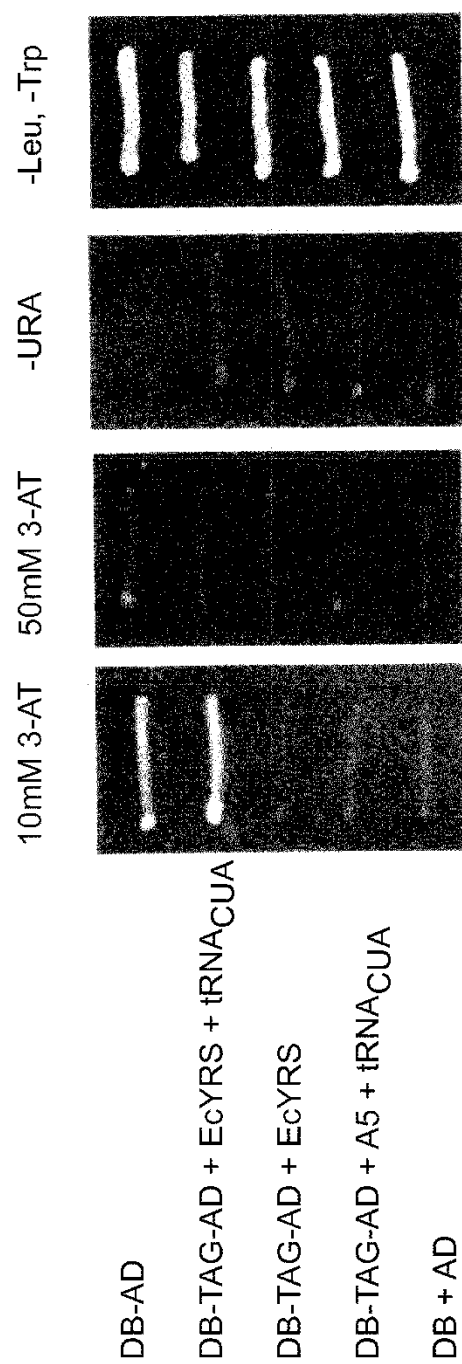


Fig. 2

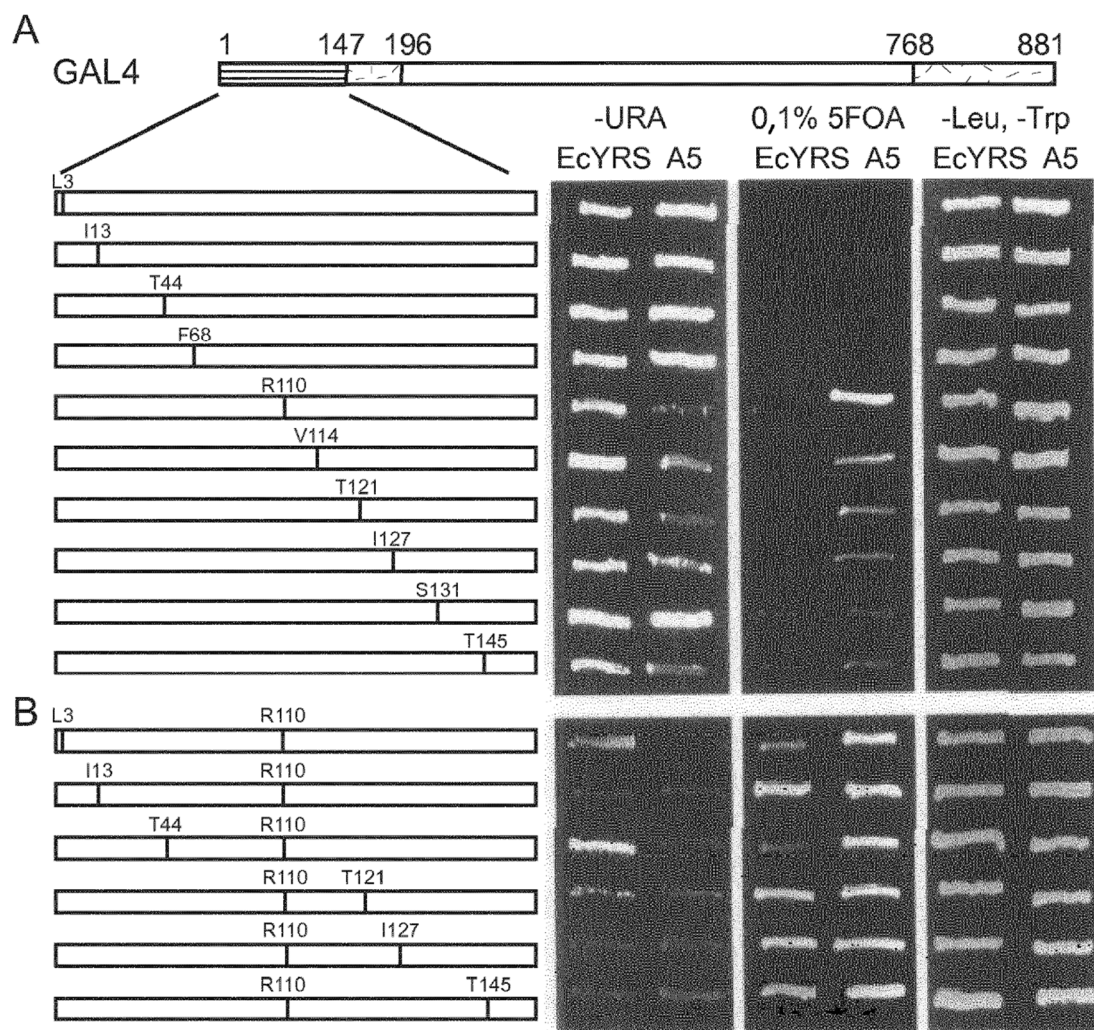


Fig. 3

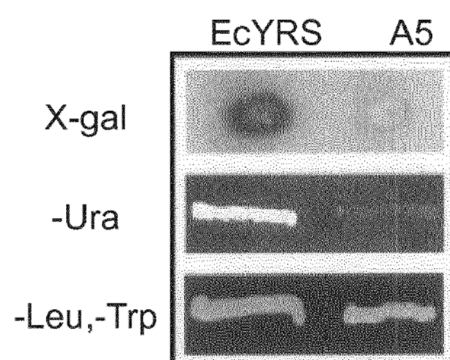


Fig. 4A

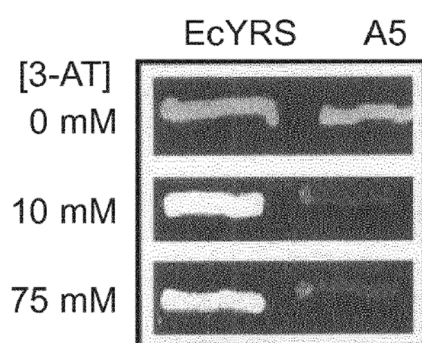


Fig. 4B

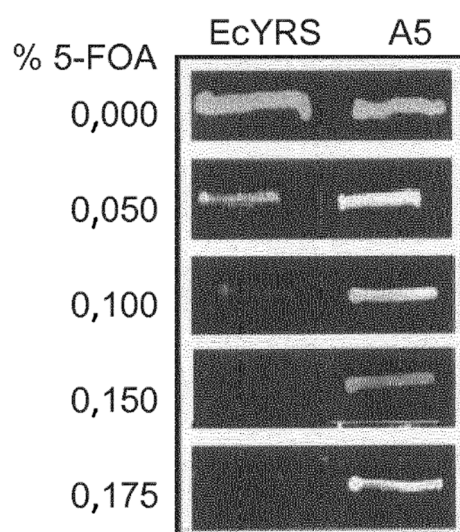


Fig. 4C

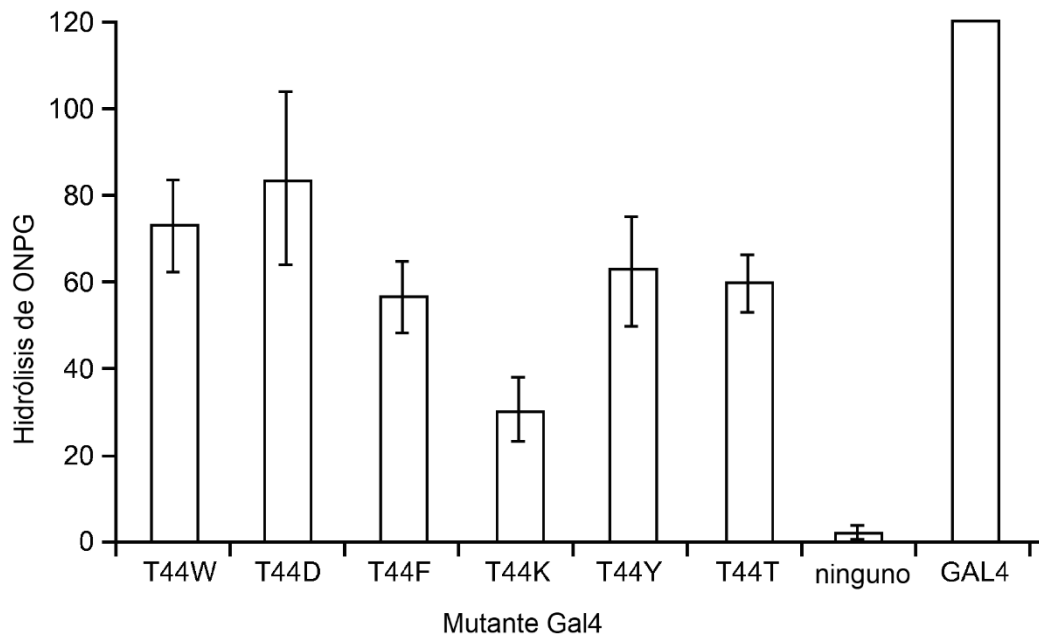


Fig. 5A

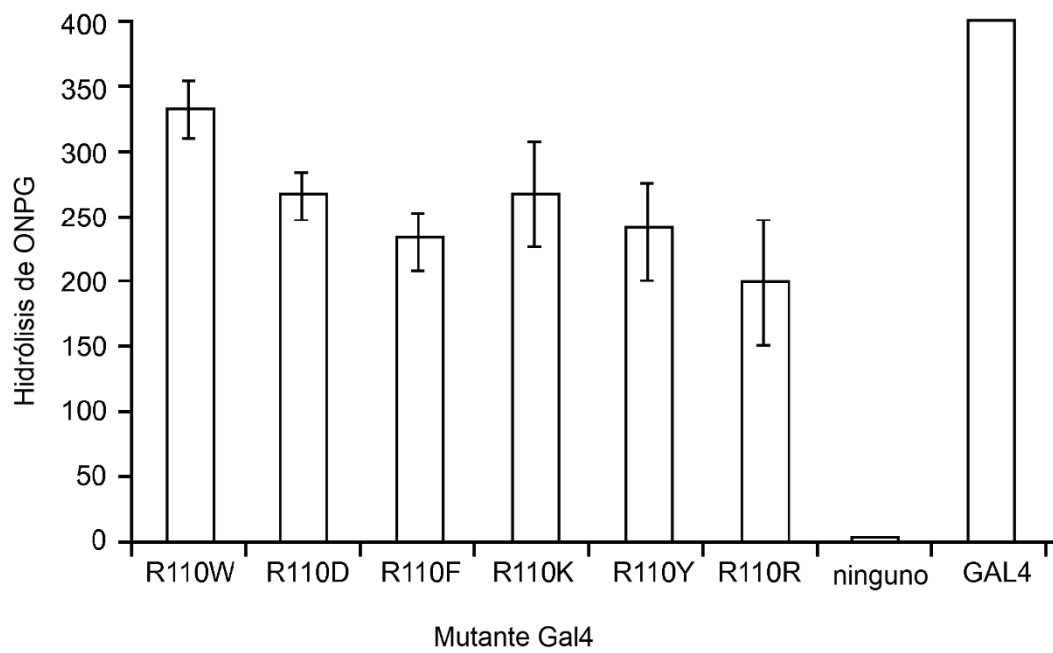


Fig. 5B

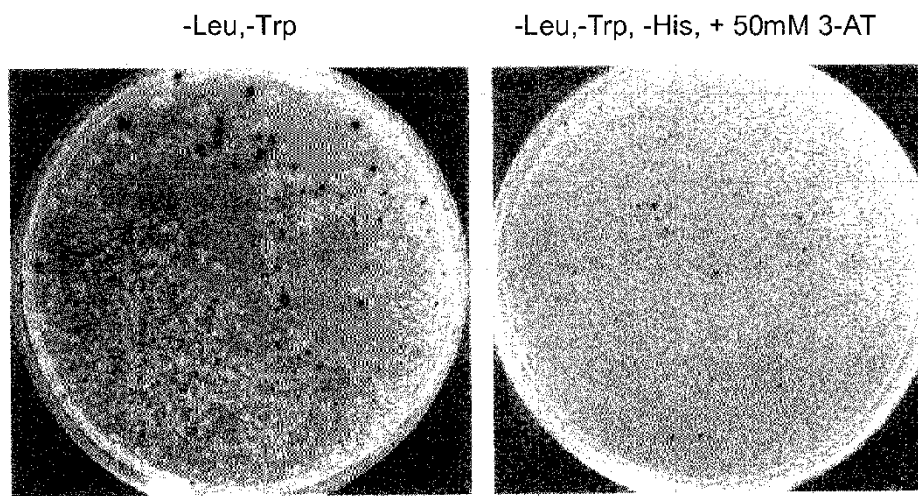


Fig. 6

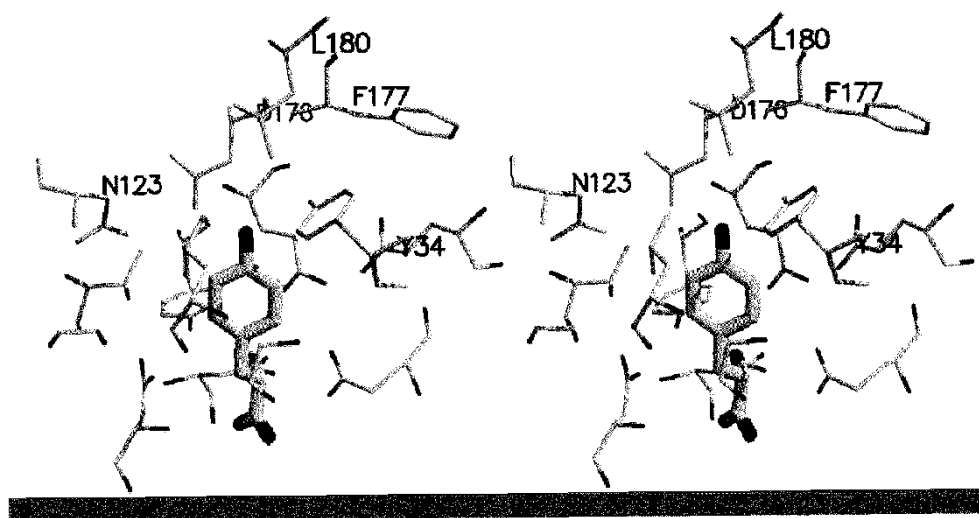


Fig. 7A

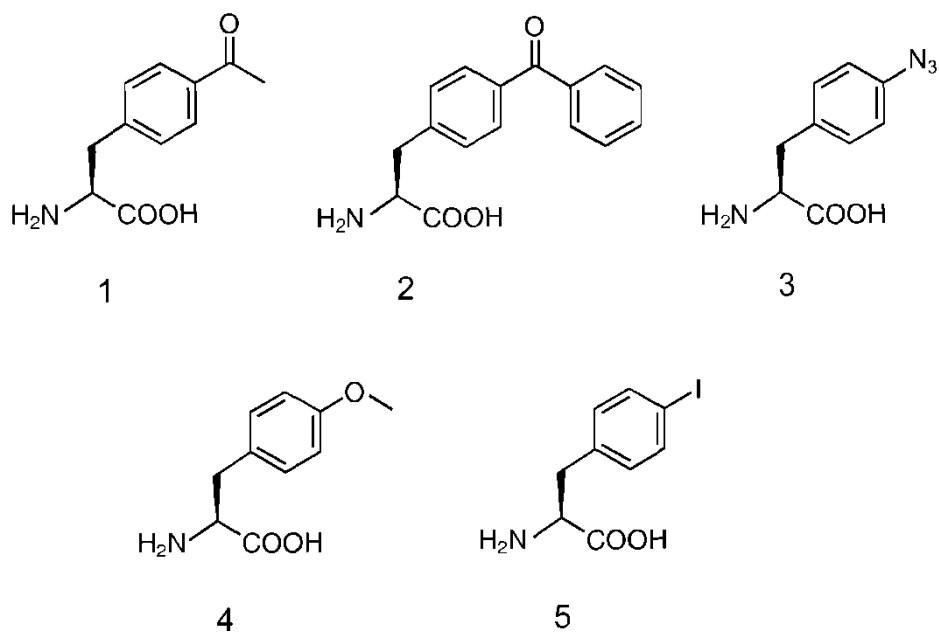


Fig. 7B

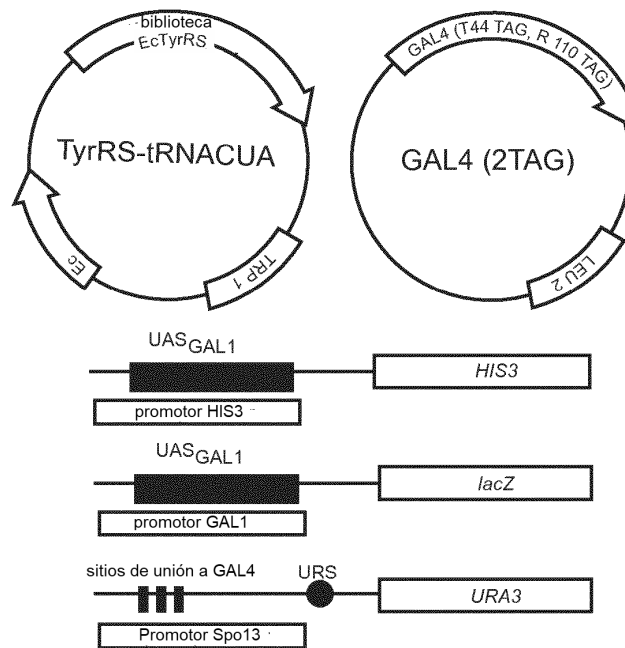


Fig. 8A

aaRS	TyrRS	A5
ARNtCUA	+	+
GAL4 (2TAG)	+	+

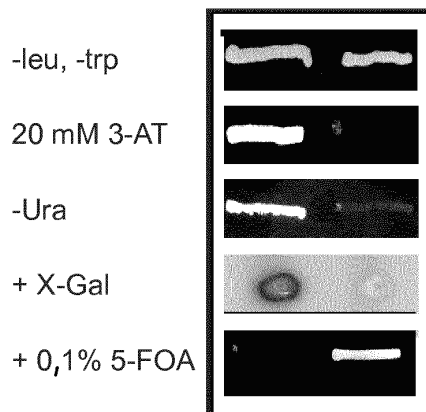


Fig. 8B

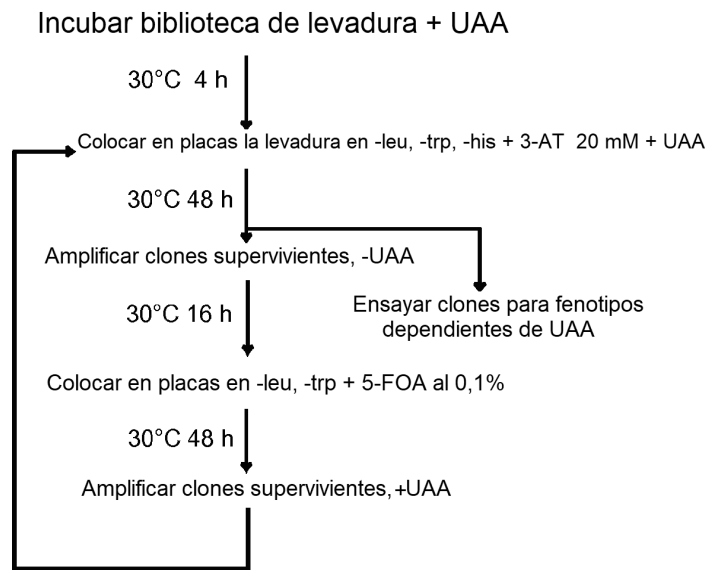


Fig. 8C

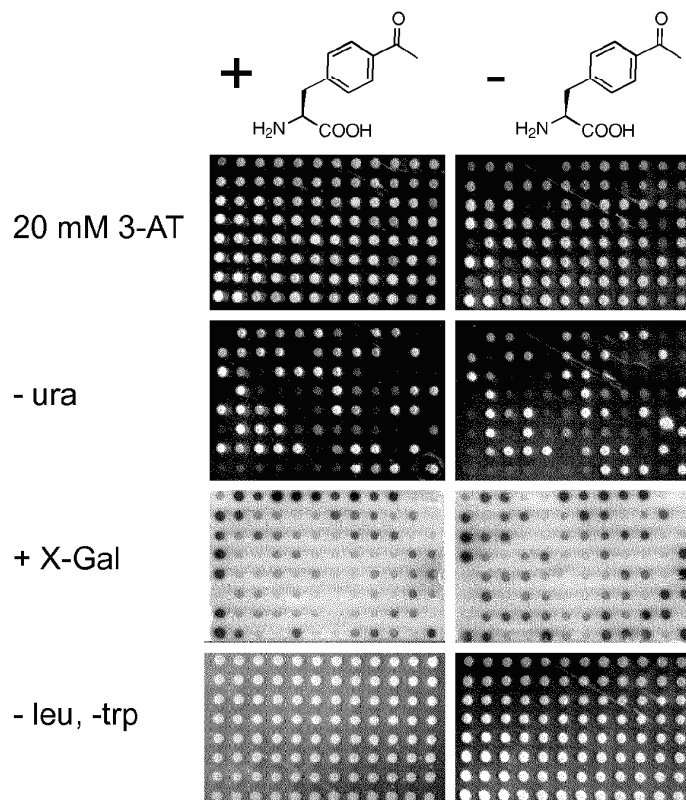


Fig. 8D

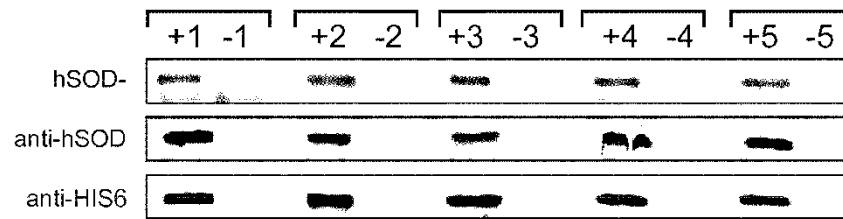


Fig. 9

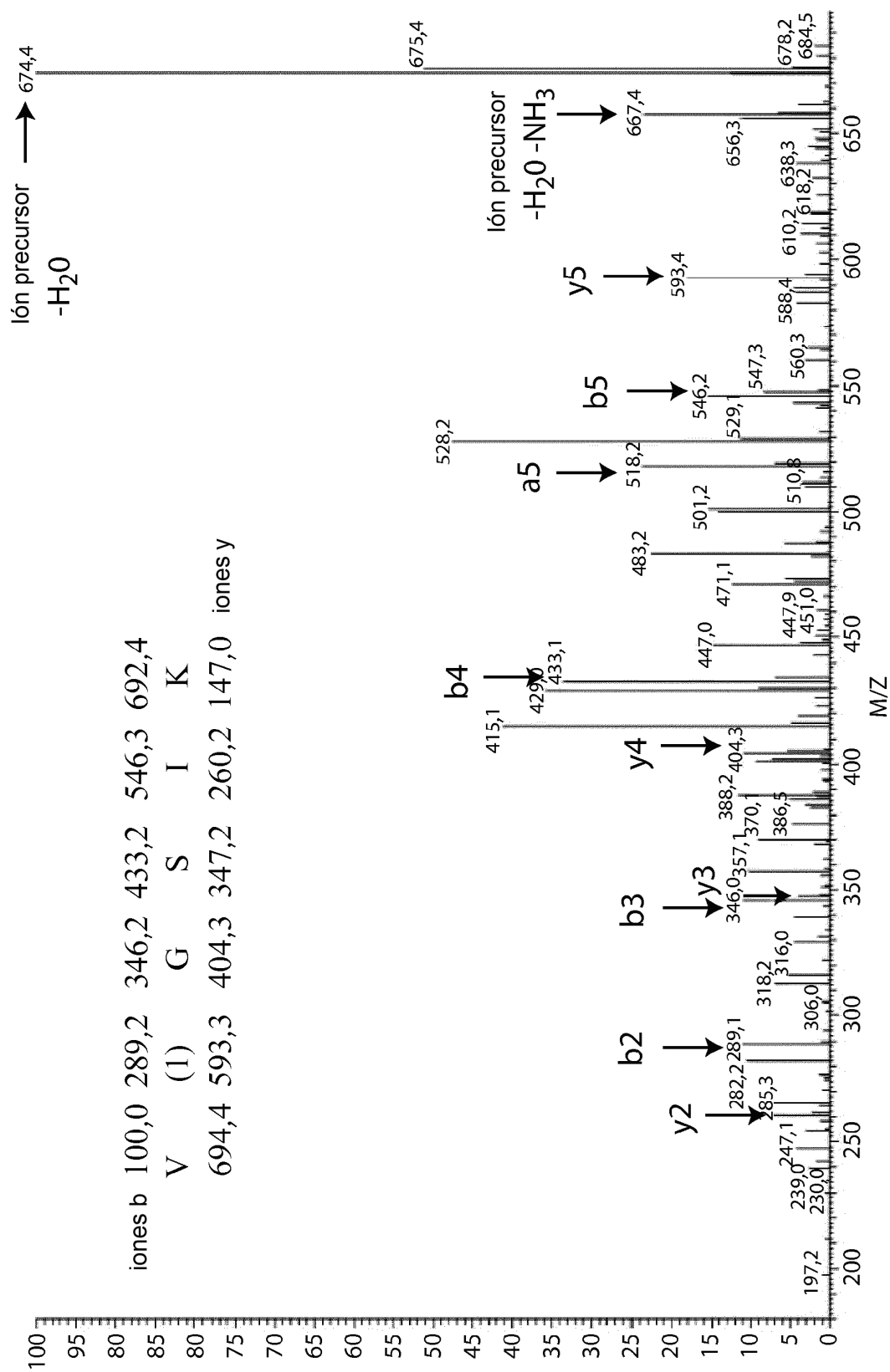


Fig. 10A

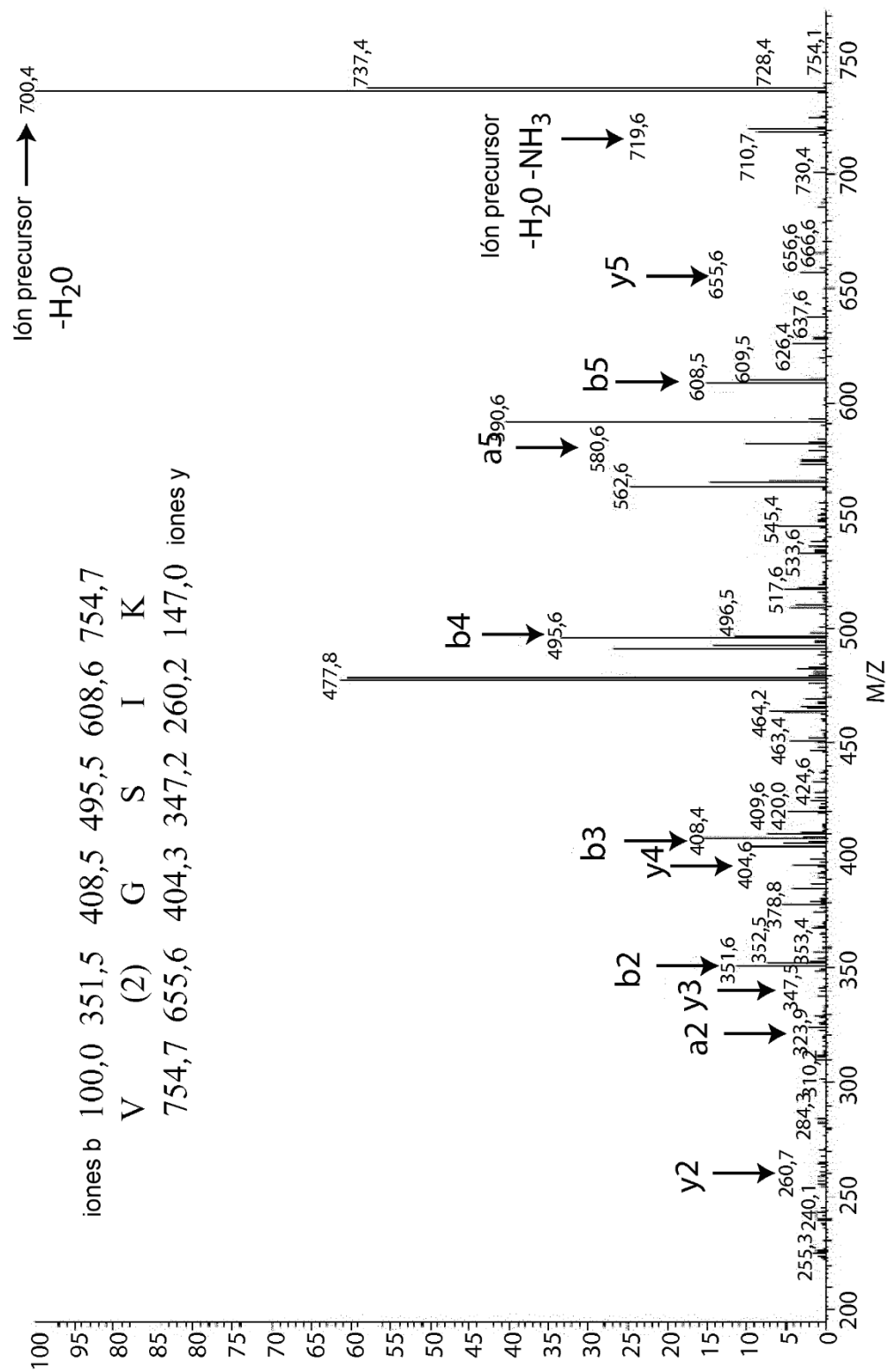


Fig. 10B

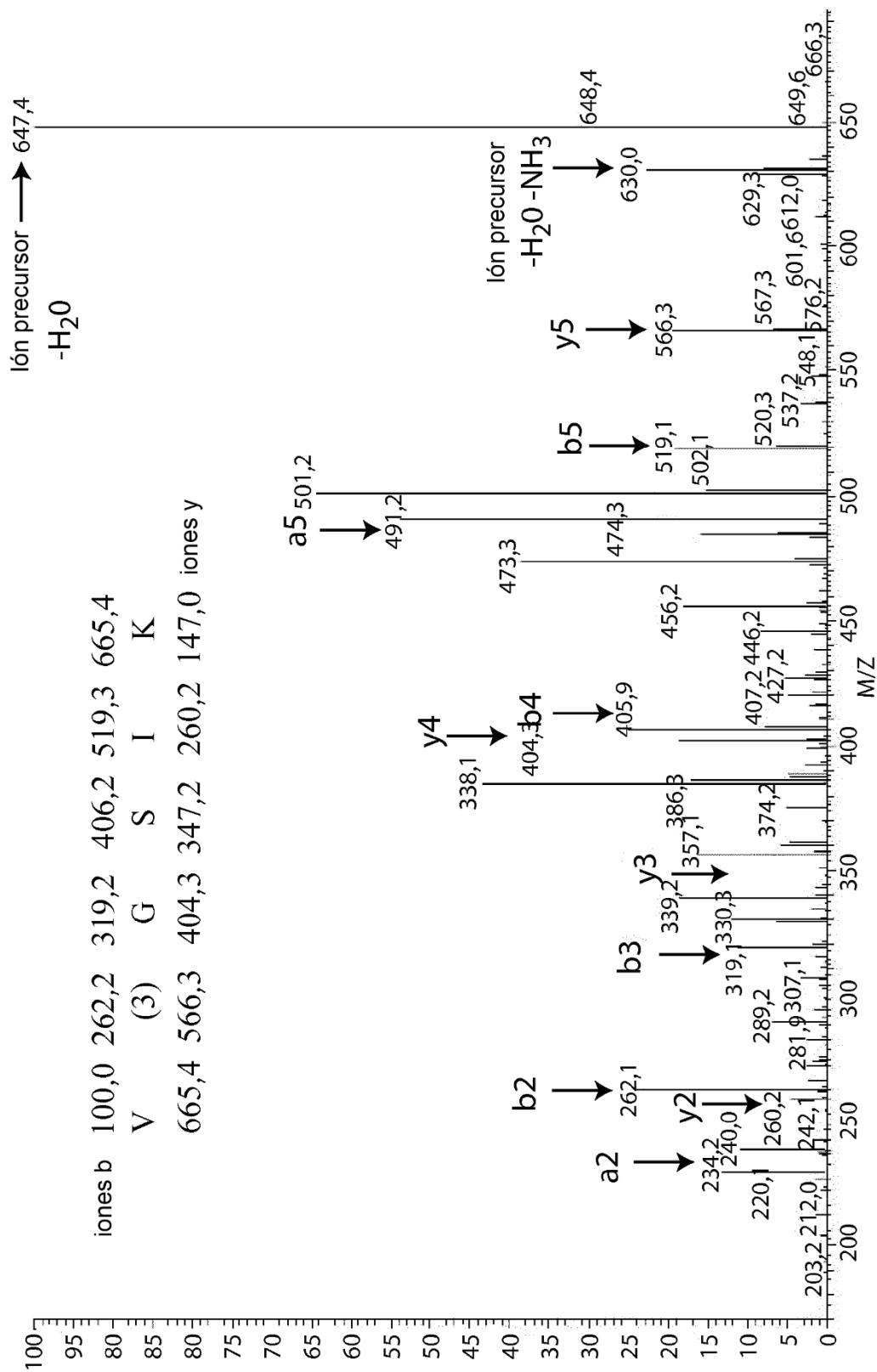


Fig. 10C

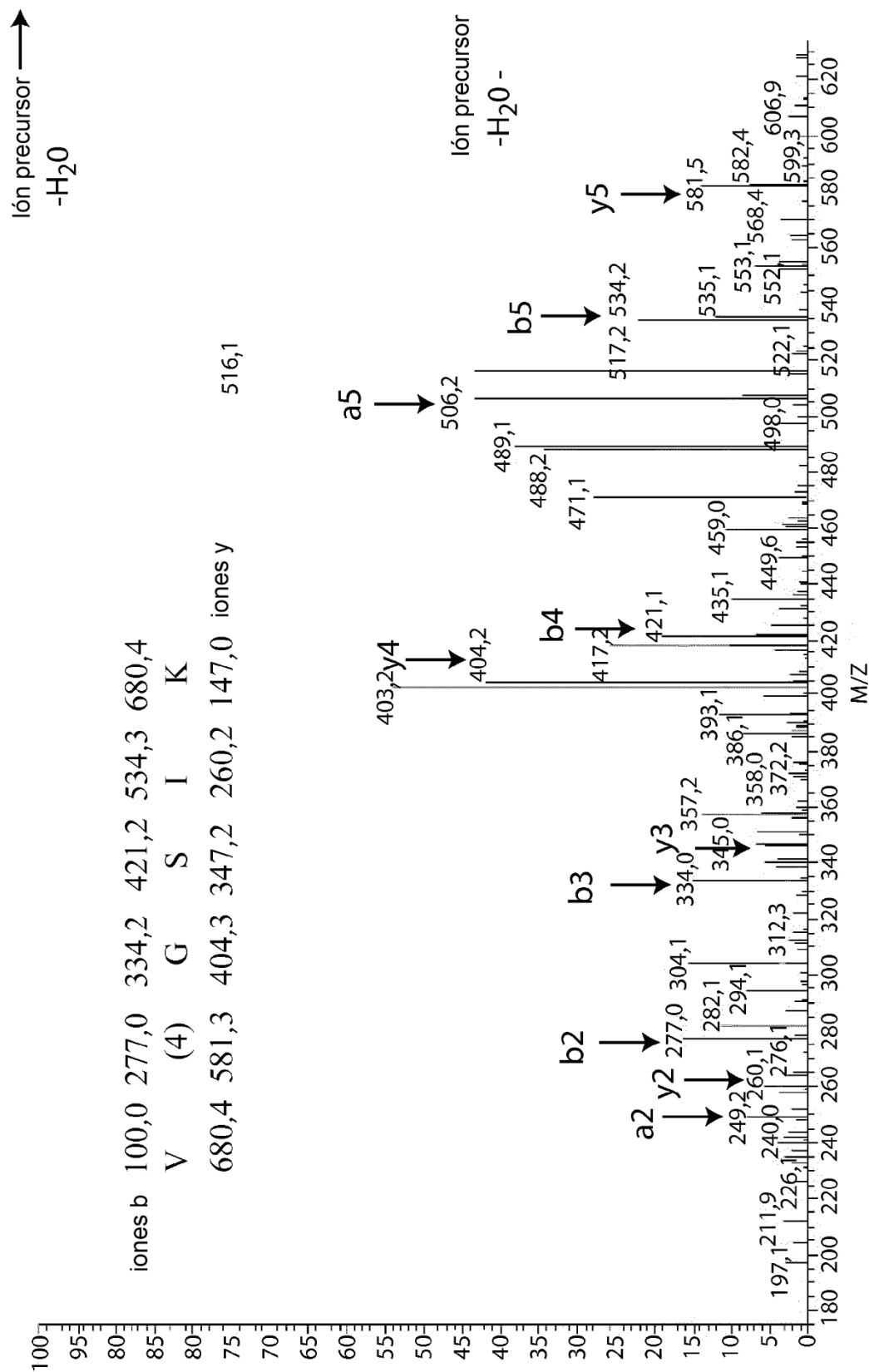


Fig. 10D

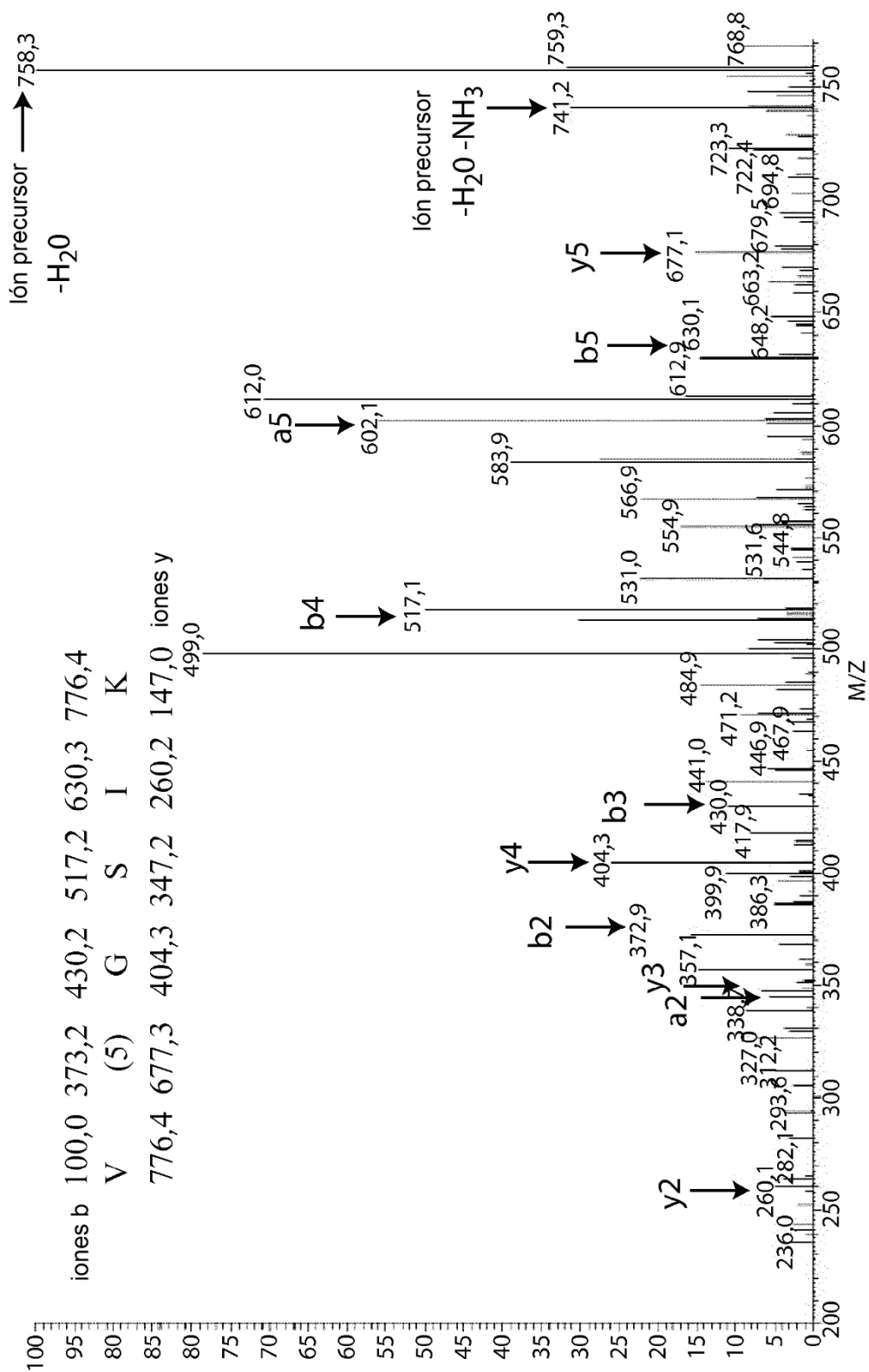


Fig. 10E

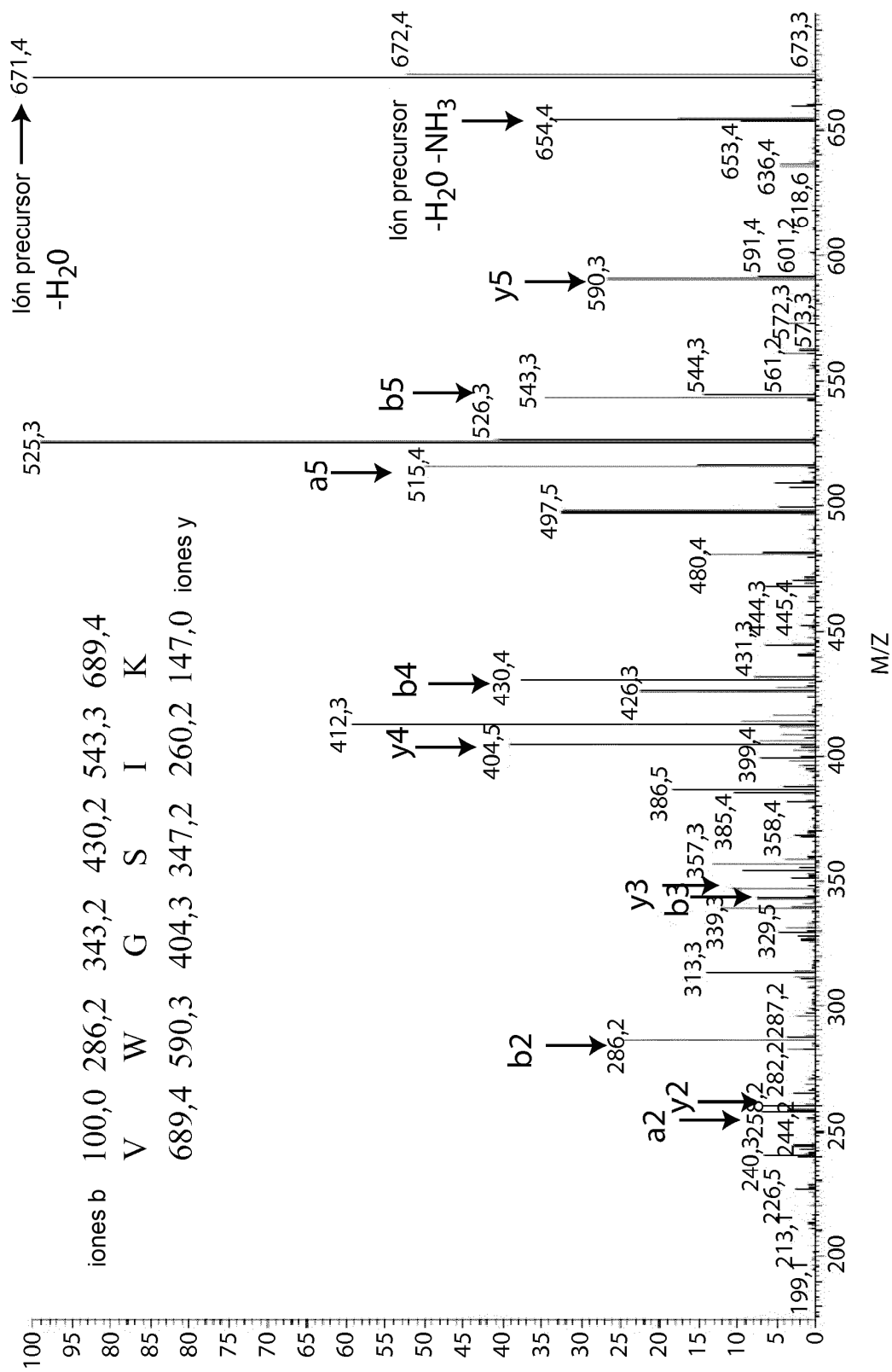


Fig. 10F

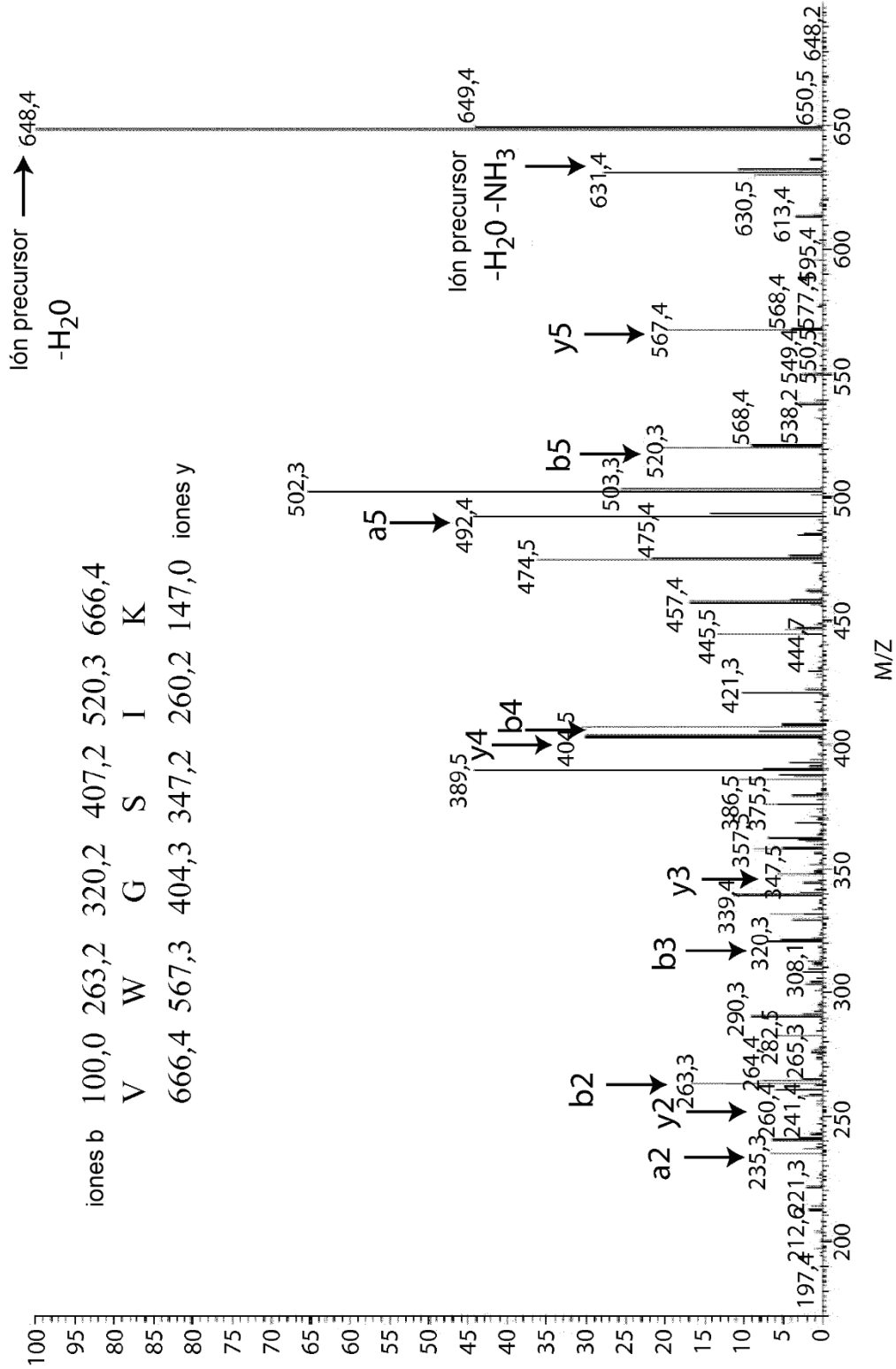


Fig. 10G

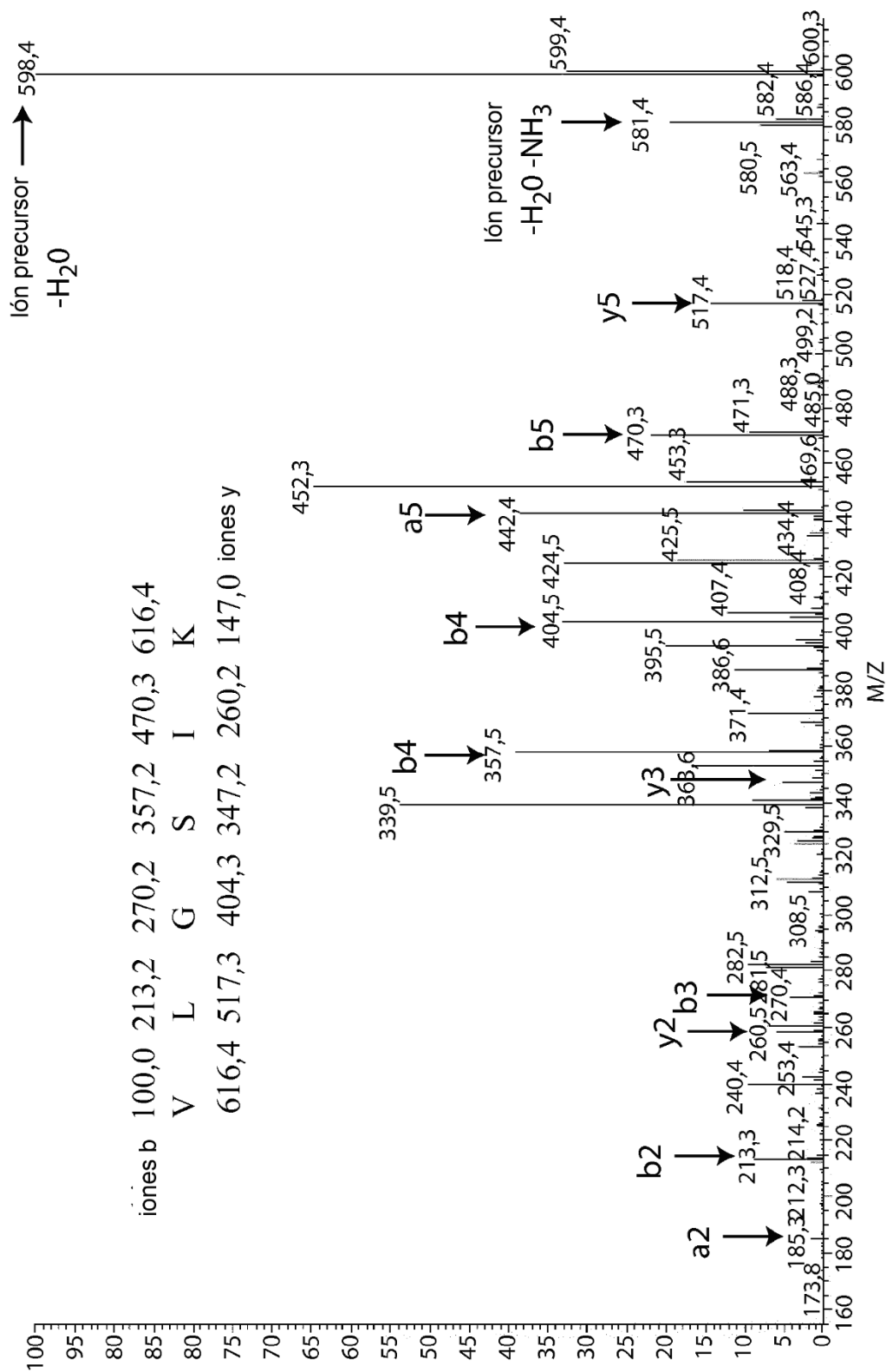


Fig. 10H

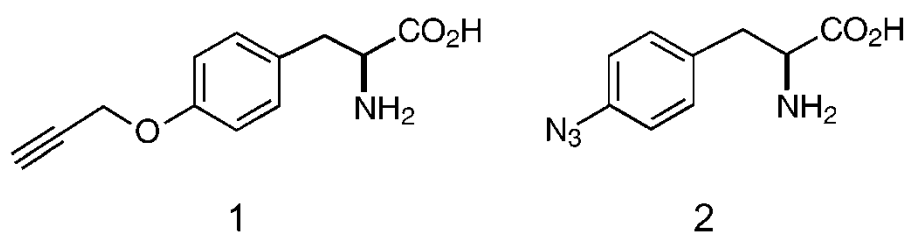


Fig. 11

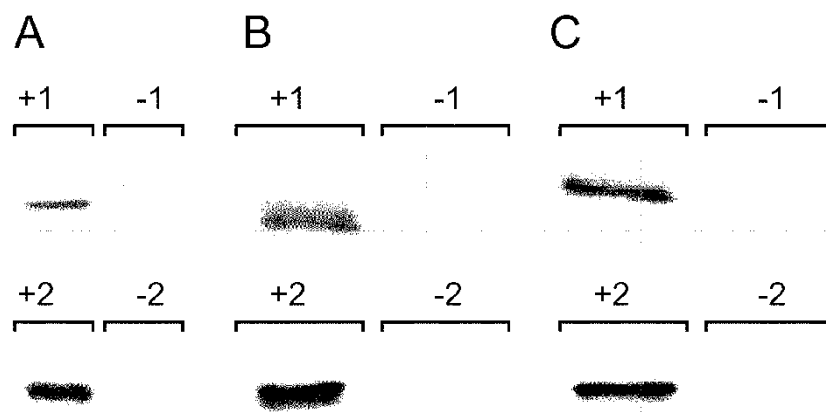


Fig. 12

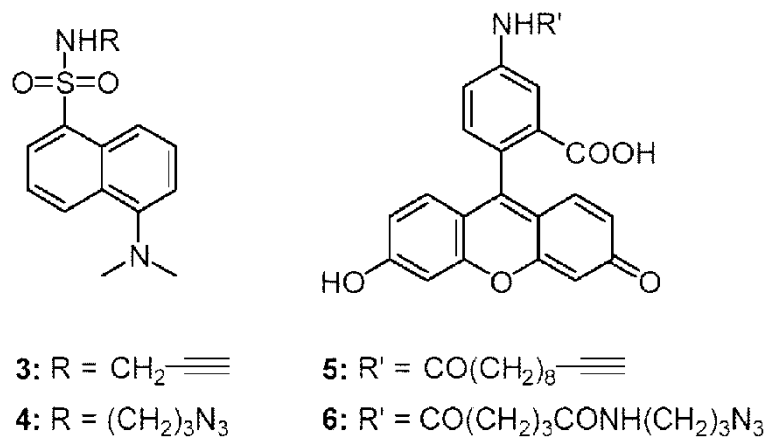


Fig. 13A

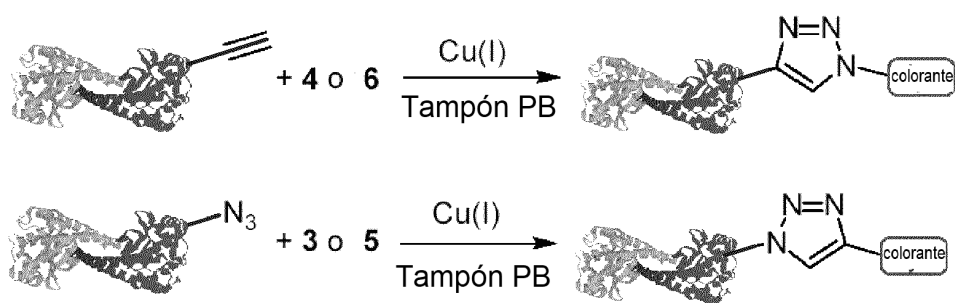


Fig. 13B

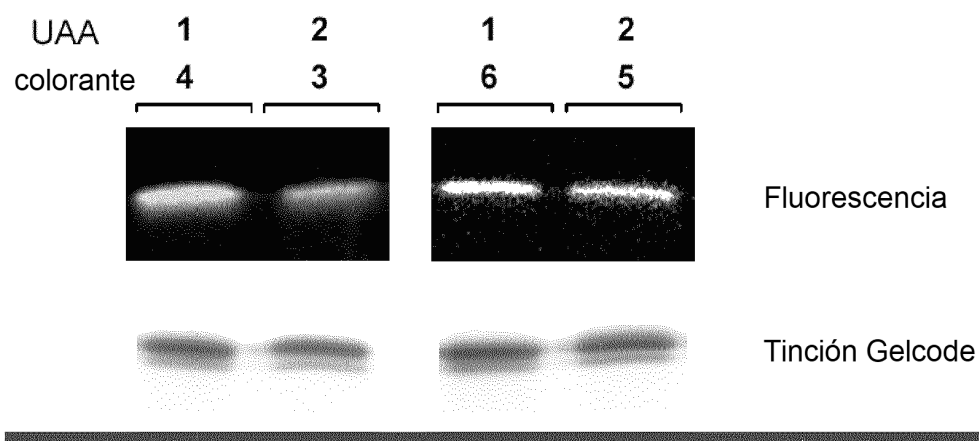


Fig. 13C

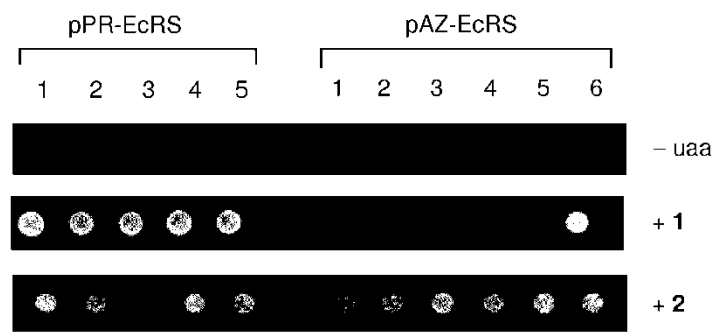


Fig. 14

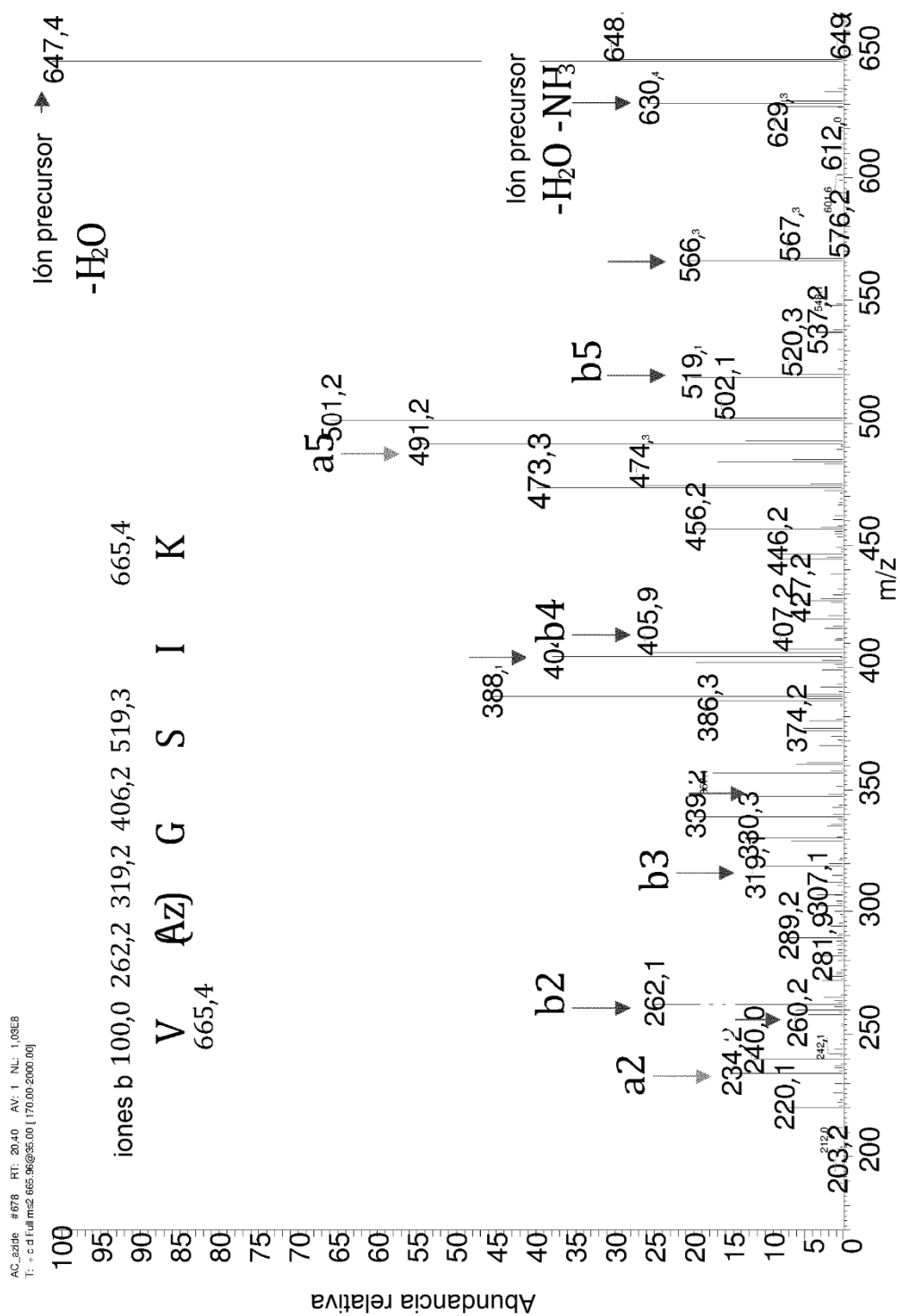


Fig. 15A

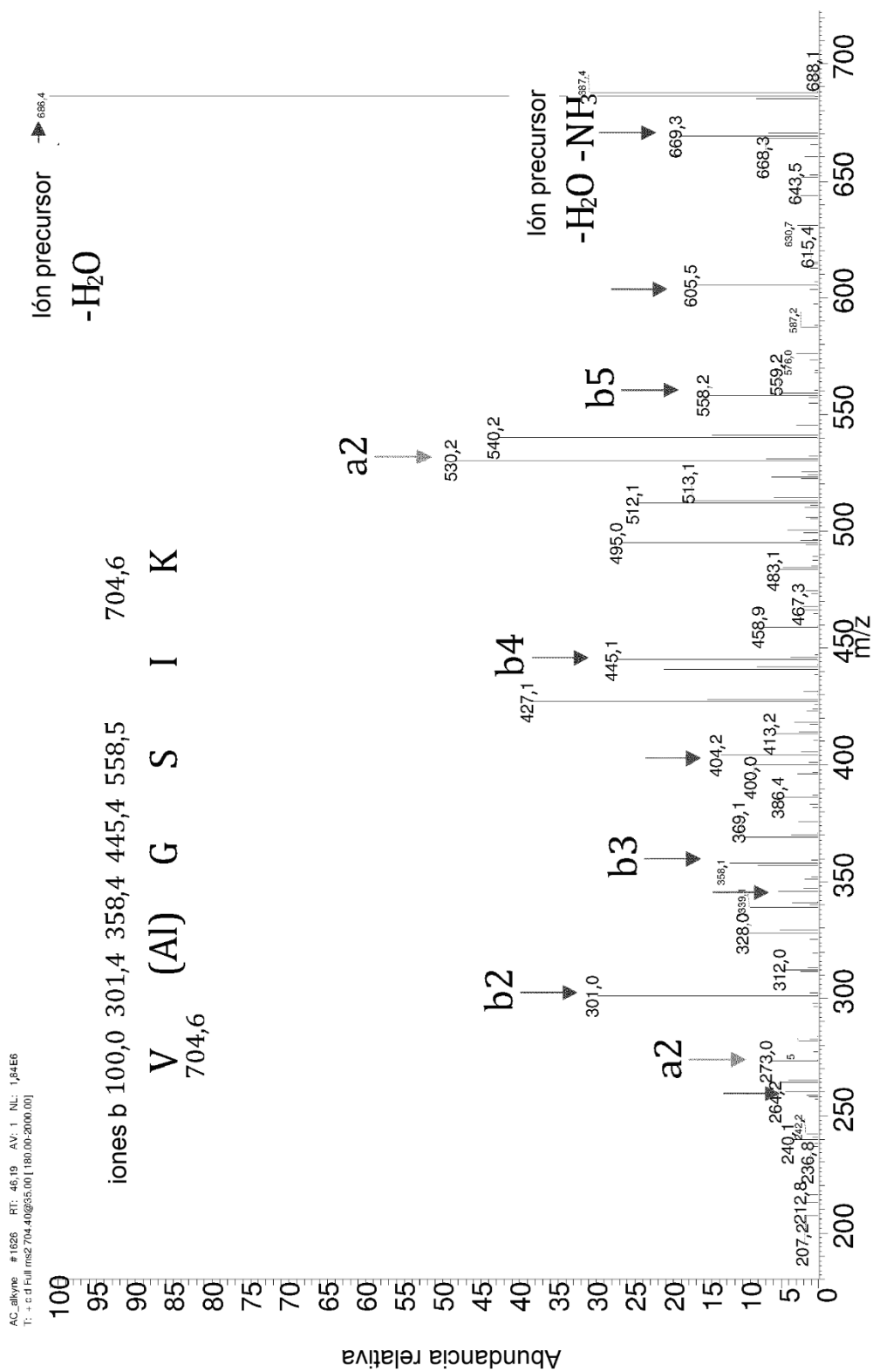


Fig. 15B

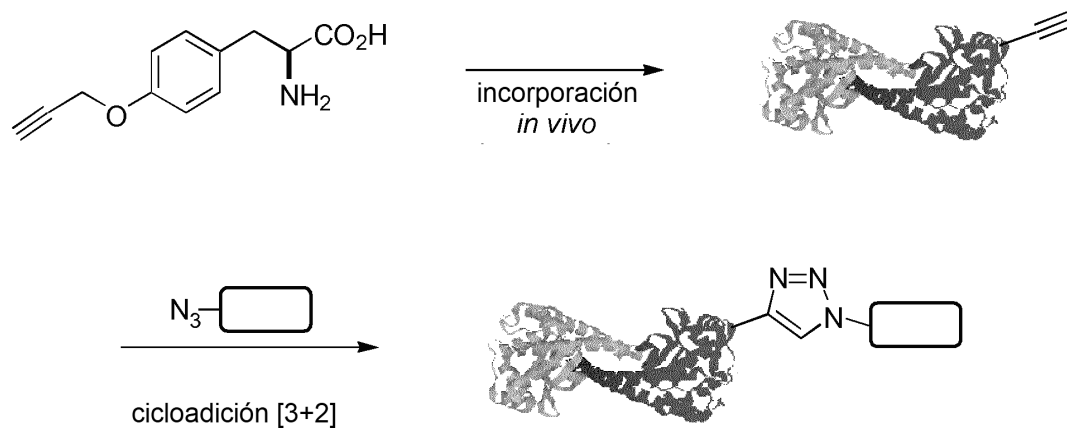


Fig. 16

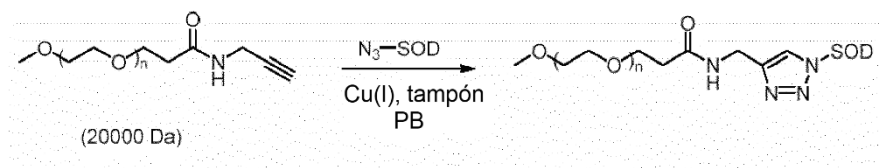


Fig. 17A



Fig. 17B

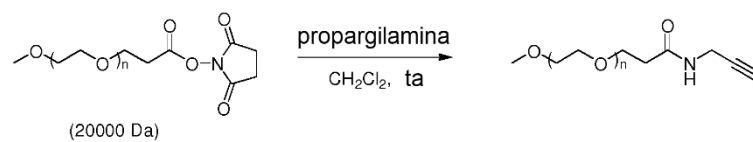


Fig. 17C