

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 660 000**

51 Int. Cl.:

C12N 1/36 (2006.01)

C12P 21/00 (2006.01)

C12N 9/00 (2006.01)

A61K 39/108 (2006.01)

C12N 15/70 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **28.09.2009 PCT/US2009/058668**

87 Fecha y número de publicación internacional: **01.04.2010 WO10037062**

96 Fecha de presentación y número de la solicitud europea: **28.09.2009 E 09817017 (8)**

97 Fecha y número de publicación de la concesión europea: **17.01.2018 EP 2337846**

54 Título: **Vacunas y microorganismos dependientes de la replicación de aminoácidos no naturales**

30 Prioridad:

26.09.2008 US 100688 P

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

20.03.2018

73 Titular/es:

**AMBRX, INC. (100.0%)
10975 North Torrey Pines Road, Suite 100
La Jolla, CA 92037, US**

72 Inventor/es:

**TIAN, FENG y
HEHLI, BRAD**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 660 000 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

Vacunas y microorganismos dependientes de la replicación de aminoácidos no naturales

Campo de la presente invención

5 La invención se refiere a células y a las vacunas que comprenden células. En algunas realizaciones, la presente invención se refiere a composiciones, incluyendo vacunas de organismos completos, con capacidades limitadas o sin replicación mediante el uso de aminoácidos no naturales, artificiales, o codificados de forma no natural.

Antecedentes de la presente invención

10 Hasta ahora, ha habido un progreso técnico relativamente lento e incluso más de 100 años después de la muerte de Louis Pasteur, su protocolo "III" (aislamiento, inactivación, inyección) continua siendo aplicable. Con el advenimiento de la biología molecular y la tecnología de proteínas recombinantes, sin embargo, las personas han comenzado a desarrollar vacunas de subunidades y vacunas similares a organismos completos diseñadas mediante ingeniería genética. Más recientemente, se han hecho avances en inmunología sobre algunas clases de inmunopotenciadores disponibles, incluyendo los ligandos de los receptores de tipo Toll (TLR). En general, se han ido definiendo cada vez más química y genéticamente vacunas de nueva generación y sus adyuvantes.

15 Está en marcha el desarrollo de agentes terapéuticos, y en concreto, vacunas dirigidas contra patógenos tales como virus, bacterias, protozoos, y hongos. Dicha investigación ha demostrado tener un valor incalculable en la prevención de la diseminación de la enfermedad en animales, incluyendo seres humanos. De hecho, en la medicina moderna, la inmunoterapia que incluye la vacunación ha erradicado la viruela y existen enfermedades virtualmente erradicadas tales como la polio, el tétanos, la tuberculosis, la viruela aviar, y el sarampión.

20 En general, las vacunas ideales tienen una larga vida útil, son capaces de inducir inmunidad duradera frente a un patógeno seleccionado y todas las variantes fenotípicas, son incapaces de dar lugar a la enfermedad contra la cual se dirige la vacuna, son terapéutica y profilácticamente eficaces, se preparan fácilmente utilizando metodologías económicas normalizadas y se pueden administrar fácilmente en el campo.

25 Existen cuatro clases principales de vacunas comercialmente disponibles. Incluyen vacunas de organismos completos no vivos, vacunas atenuadas vivas, vacunas de vectores, y vacunas de subunidades. La vacunación con materiales no vivos tales como proteínas conduce generalmente a una respuesta de anticuerpos o linfocitos T auxiliares CD4+, mientras que, la vacunación con materiales vivos (por ejemplo, virus infecciosos) conduce generalmente a una respuesta de linfocitos T citotóxicos (LTC) CD8+. una respuesta de LTC es crucial para la protección contra patógenos del tipo de virus y bacterias infecciosos. Esto plantea un problema práctico, ya que la única forma segura de lograr una respuesta de LTC es usar agentes vivos que sean por sí mismos patógenos. El problema se elude generalmente utilizando cepas víricas y bacterianas atenuadas o destruyendo células completas que se pueden usar para la vacunación. Estas estrategias han funcionado bien, pero el uso de cepas atenuadas conlleva siempre el riesgo de que el agente atenuado pueda recombinarse genéticamente en el hospedador y convertirse en una cepa virulenta. Por lo tanto, existe una necesidad de agentes y procedimientos terapéuticos que puedan conducir a una respuesta LTC CD8+ mediante la vacunación con materiales no vivos, tales como proteínas, de una manera específica.

30 Las vacunas de subunidades han proporcionado un medio de lidiar con alguno de estos problemas. Dichas vacunas comprenden generalmente un componente subcelular derivado de un patógeno de interés. Un componente de subunidad puede producirse tanto a partir de una fracción subcelular del patógeno, como de una proteína, ácido nucleico o polisacáridos purificado. Todos estos elementos tienen un determinante antigénico capaz de estimular una respuesta inmunitaria frente al patógeno de interés. En general, el componente subcelular de la vacuna de subunidad se obtiene tanto purificando una preparación de un patógeno perturbado como sintetizándolo utilizando procedimientos bien conocidos.

35 Existen, sin embargo, algunas limitaciones asociadas con las vacunas de subunidades. En primer lugar, un requisito para la producción de dicha vacuna es el de que el(los) determinante(s) antigénico(s) deben(n) caracterizarse e identificarse. Esto impone limitaciones sobre su uso, particularmente frente a determinantes antigénicos muy variables. En segundo lugar, las vacunas de subunidades son generalmente ineficaces en estimular las respuestas de los linfocitos T citotóxicos. En tercer lugar, la inmunidad conferida por las vacunas de subunidades a menudo es de corta duración y requiere por tanto inyecciones de refuerzo continuas. Muy pocas vacunas de subunidades expresadas de forma recombinante han mostrado inducir una inmunidad fuerte y duradera en animales vacunados (incluyendo el hombre). Una excepción notable es el antígeno de superficie recombinante de la vacuna de la hepatitis B utilizado en el hombre. Uno de los problemas asociados con el uso de dichas vacunas parece ser presentar correctamente los antígenos al sistema inmunitario de tal manera que se inducen una inmunidad humoral fuerte y una inmunidad mediada por célula fuerte. En particular, las vacunas recombinantes existentes (subunidades) no parecen dar como resultado respuestas de 'memoria' fuertes, de tal manera que los animales vacunados reaccionan muy rápidamente cuando se exponen a las infecciones naturales producidas por un patógeno.

Únicamente a modo de ejemplo, Se han notificado extensamente deficiencias en las vacunas de subunidades actuales preparadas a partir del virus de la diarrea vírica de bovino similar a pestivirus (BVDV). Estos estudios han mostrado que incluso aunque se utilizaron grandes cantidades de proteínas recombinantes en las vacunas, se observaron pobres tasas de protección que muestran que las vacunas fracasaron en proteger del estímulo con aislados de BVDV vivos (tanto protección homóloga como protección heteróloga).

Existen muchas enfermedades infecciosas para las cuales no se ha desarrollado una vacuna eficaz, y muchas de las vacunas actualmente disponibles proporcionan solo protección parcial frente a la enfermedad. Además, existen vacíos en el campo de las vacunas. Las vacunas vivas producen una inmunidad más fuerte, más amplia, y más duradera que los otros tipos de vacunas. Existe una necesidad de un vehículo de vacunas vivas más seguro, que sea incapaz de producir enfermedad incluso en individuos inmunosuprimidos. Existe también una necesidad de vacunas que induzcan la inmunidad mediada por células y no exactamente la inmunidad basada en anticuerpos. Y, existe una necesidad de inducir respuestas inmunoprotectoras directamente en las superficies mucosales del cuerpo, donde la mayoría de los patógenos ganan la entrada. Por lo tanto, existe una necesidad de vacunas mejoradas. La presente invención busca proporcionar una vacuna terapéutica mejorada que mejore al menos alguna de las desventajas sobre la técnica anterior existente.

El documento WO 01/00234 describe procedimientos de generar células atenuadas que se pueden utilizar para formar vacunas. dichas células requieren ARNt supresor para replicarse, pero no requieren la presencia de un aminoácido no natural para replicarse.

Sumario de la presente invención

La presente divulgación proporciona organismos completos dependientes de la replicación modulados químicamente. En algunas realizaciones, la presente divulgación proporciona un organismo completo modulado químicamente con una modificación, por lo cual, su capacidad para replicarse es dependiente de la presencia de un aminoácido no natural. En algunas realizaciones, la presente divulgación proporciona un organismo completo con una o más modificaciones, por lo cual, la capacidad del organismo para replicarse depende de que se cultive en presencia de un aminoácido no natural. En algunas realizaciones, la presente divulgación proporciona un organismo completo con una o más modificaciones, por lo cual, la capacidad del organismo de replicarse depende de que se cultive en presencia de uno o más aminoácido(s) no naturales. En algunas realizaciones, la presente divulgación proporciona un organismo completo con dos o más modificaciones, por lo cual, la capacidad del organismo de replicarse depende de que se cultive en presencia de uno o más aminoácido(s) no naturales. En algunas realizaciones de la presente divulgación, el organismo completo es un organismo mitigado - una vacuna modulada por aminoácidos no naturales incorporados específicamente en uno o más sitios, en el que la replicación o la expresión del agente inmunizante está modulada por la presencia o ausencia de un aminoácido no natural.

En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de una única modificación genética. En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de una o más modificaciones genéticas. En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de dos modificaciones genéticas. En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de tres modificaciones genéticas. En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de cuatro modificaciones genéticas. En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de cinco modificaciones genéticas. En algunas realizaciones de la presente divulgación, la replicación o la expresión de un organismo completo está modulada por la presencia o ausencia de un aminoácido no natural a través de cuatro o más modificaciones genéticas.

La presente invención proporciona una célula en la que se ha incorporado un aminoácido no natural específicamente en el sitio en un producto génico requerido para la replicación, en la que dicha célula es capaz de replicación en presencia de dicho aminoácido no natural, y tiene una capacidad limitada o sin replicación en ausencia de dicho aminoácido no natural. La presente invención proporciona también una vacuna que comprende dicha célula. La presente invención proporciona también una formulación de vacuna que comprende dicha célula, un transportador farmacéuticamente aceptable, y opcionalmente uno o más agentes biológicamente activos diferentes. La célula es un ejemplo de un organismo completo,

La capacidad de inducir selectivamente una respuesta inmunitaria frente a organismos, antígenos, proteínas, o autoproteínas o de aumentar la inmunogenicidad de epítopos específicos de antígenos extraños, es significativa en la producción de vacunas para numerosos estados de enfermedad (incluyendo, aunque no de forma limitativa, enfermedades cancerosas y de plegamiento de proteínas) y enfermedades infecciosas (por ejemplo, infecciones bacterianas o víricas). La invención actual utiliza la incorporación de aminoácidos no naturales en organismos para producir inmunógenos dependientes de la replicación y/o deficientes en la replicación que se van a usar en vacunaciones de organismos completos. Se desvela también la incorporación de aminoácidos no naturales en

organismos para producir anticuerpos que se van a usar en la inmunización pasiva. La invención actual utiliza también la incorporación de aminoácidos no naturales en proteínas, antígenos, y/o polipéptidos para producir inmunógenos que se van a usar en vacunaciones. En la invención, los inmunógenos a los cuales se añaden los aminoácidos no naturales corresponden a restos diana (por ejemplo, restos relacionados con una enfermedad) en el sujeto que se va a vacunar/inmunizar o corresponden a restos diana (por ejemplo, restos relacionados con una enfermedad) que son capaces de estar en el sujeto. En las realizaciones donde el inmunógeno con el aminoácido no natural se administra a un sujeto, la presencia del aminoácido no natural estimula una respuesta inmunológica contra el inmunógeno que tiene reactividad cruzada frente al resto diana (por ejemplo, relacionado con la enfermedad).

Se proporcionan procedimientos para producir o potenciar una respuesta inmunológica, por ejemplo, una respuesta mediada por linfocitos B y/o una respuesta mediada por linfocitos T, en un sujeto contra un resto diana, por ejemplo, un polipéptido, un hidrato de carbono, o una combinación de ambos, que está en el sujeto o que es capaz de estar en el sujeto. Los procedimientos incluyen proporcionar un organismo dependiente de un aminoácido no natural modificado genéticamente y administrar el organismo al sujeto. Los procedimientos incluyen también proporcionar un inmunógeno no natural modificado genéticamente que comprende uno o más aminoácidos no naturales, y administrar el inmunógeno no natural al sujeto. El sujeto (por ejemplo, un ser humano, un mono, un ratón, una rata, un animal de ganadería, un cerdo, una vaca, un pollo, un ave ornamental, un pájaro aviario, una mascota doméstica, un perro, un gato, un reptil, y/o un anfibio) produce uno o más anticuerpos contra el inmunógeno natural, cuyos anticuerpos tienen reactividad cruzada contra el resto diana (produciendo o potenciando por tanto la respuesta inmunógena frente a la diana).

En determinadas realizaciones, el organismo completo modificado genéticamente puede ser cualquier organismo completo frente al cual es deseable inmunizar el sujeto, por ejemplo, una bacteria, un virus, un hongo, un Mycoplasma, un protozoo, un helminto, o un prión. Una vacuna de organismo completo puede incluir opcionalmente uno o más de: un antígeno bacteriano, un antígeno vírico, un antígeno fúngico, un antígeno de micoplasma, un antígeno de protozoo, un antígeno de helminto, un antígeno de prión, un antígeno de VIH, gpl20 de VIH, gp41 del VIH, gag de VIH, pol de VIH, env de VIH, tat de VIH, nef de VIH, rev de VIH, un antígeno de la cápsida de calicivirus, un antígeno del núcleo de la hepatitis B, un antígeno de superficie de la hepatitis B, el agente delta de la hepatitis, una glicoproteína del virus del herpes simple, una glicoproteína del virus de la varicela zóster, una hemaglutinina del virus de la gripe, una neuraminidasa del virus de la gripe, una nucleoproteína del virus de la gripe, una proteína de la cápsida del virus VPH, una hemaglutinina/neuraminidasa del virus paragripal, un polipéptido de la cápsida de un poliovirus, un antígeno de Hep A, un polipéptido del virus vaccinia, una glicoproteína G del virus de la rabia, B. burgdorferi OspA, proteína de la membrana externa de H. influenzae de tipo b, lipoarabinomano de Mycobacterium, mAPG de Mycobacterium, proteína M de S. pyogenes, polisacárido capsular de S. pneumoniae, FI de Y. pestis, antígeno V de Y. pestis, circunsporozoito de P. falciparum (PfCSP), proteína 2 de la superficie del esporozoito de P. falciparum (PfSSP2), extremo carboxilo del antígeno 1 de estado hepático de P. falciparum (extremo de PflSAI), proteína 1 exportada de P. falciparum (PfExp-1), Pfs 48/45, Pfs 28, Pfs 25, o Pfs 230.

El organismo dependiente de la replicación de aminoácidos no naturales modificado genéticamente puede ser uno o más de: una bacteria, un virus, un hongo, un Mycoplasma, un protozoo, un helminto, un prión, un Actinomyces, un Bacillus, un Bacteroides, una Bordetella, una Bartonella, una Borrelia, una Brucella, un Campylobacter, un Capnocytophaga, una Chlamydia, un Clostridium, a Corynebacterium, una Coxiella, un Dermatophilus, un Enterococcus, una Ehrlichia, una Escherichia, una Francisella, un Fusobacterium, una Haemobartonella, un Haemophilus, un Helicobacter, una Klebsiella, una bacteria de forma en L, una Leptospira, una Listeria, un Mycobacterium, un Mycoplasma, una Neisseria, una Neorickettsia, una Nocardia, una Pasteurella, a Peptococcus, un Peptostreptococcus, un Pneumococcus, un Proteus, una Pseudomonas, una Rickettsia, una Rochalimaea, una Salmonella, una Shigella, un Staphylococcus, un Streptococcus, un Treponema, una Yersinia, un adenovirus, un alfavirus, un calicivirus, un coronavirus, un CMV, un virus del moquillo, un virus del Ébola, un enterovirus, un VEB, un flavivirus, un Hep C, un hepadnavirus, un Hep B, un agente delta de la hepatitis, un Hep E o virus F, un GBV-C, un herpesvirus, un virus del herpes simple, un virus de la varicela zóster, un virus de la inmunodeficiencia, un VIH, un virus de la peritonitis infecciosa, un virus de la gripe, un virus de la gripe A, un virus de la leucemia, un virus de Marburg, un ortomixovirus, un virus del papiloma, un VPH, un virus paragripal, un paramixovirus, un VSR, un parvovirus, un pestivirus, un picornavirus, un poliovirus, un poxvirus, un virus vaccinia, un virus de la rabia, un reovirus, un retrovirus, un rotavirus, una Absidia, un Acremonium, una Alternaria, un Aspergillus, un Basidiobolus, un Bipolaris, un Blastomyces, una Candida, un Coccidioides, un Conidiobolus, un Cryptococcus, una Curvalaria, un Epidermophyton, una Exophiala, un Geotrichum, un Histoplasma, una Madurella, una Malassezia, un Microsporium, una Moniliella, una Mortierella, un Mucor, un Paecilomyces, un Penicillium, un Phialemonium, una Phialophora, una Prototheca, una Pseudallescheria, un Pseudomicrdochium, un Pythium, un Rhinosporidium, un Rhizopus, un Scolecobasidium, un Sporothrix, un Stemphylium, un Trichophyton, un Trichosporon, una Xylohypha, una Babesia, un Balantidium, una Besnoitia, un Cryptosporidium, una Eimeria, un Encephalitozoon, una Entamoeba, una Giardia, una Hammondia, un Hepatozoon, una Isospora, una Leishmania, una Microsporidia, una Neospora, una Nosema, un Pentatrichomonas, un Plasmodium, un P. falciparum, un Pneumocystis, un Sarcocystis, un Schistosoma, una Theileria, un Toxoplasma, un Trypanosoma, un Acanthocheilonema, un Aelurostrongylus, un Ancylostoma, un Angiostrongylus, un Ascaris, una Brugia, un Bunostomum, una Capillaria, una Chabertia, una Cooperia, un Crenosoma, un Dictyocaulus, un Dioctophyme, un Dipetalonema, un Diphylobothrium, un Diplydium, una Dirofilaria, un Dracunculus, un Enterobius, un Filaroides, un Haemonchus, un Lagochilascaris, un polipéptido de Loa, una

Mansonella, un Muellerius, un Nanophyetus, un Necator, un Nematodirus, un Oesophagostomum, un Onchocerca, un Opisthorchis, una Ostertagia, una Parafilaria, un Paragonimus, un Parascaris, un Physaloptera, un Protostrongylus, una Setaria, una Spirocerca, una Spirometra, una Stephanofilaria, un Strongyloides, un Strongylus, una Thelazia, un Toxascaris, una Toxocara, una Trichinella, un Trichostrongylus, un Trichuris, una Uncinaria, o una Wuchereria.

Se desvelan procedimientos de tratar profiláctica o terapéuticamente una patología en un sujeto, por ejemplo, produciendo una respuesta mediada por linfocitos B y/o linfocitos T en el sujeto. En diversas realizaciones, la patología puede ser, aunque no de forma limitativa, una o más de: una infección bacteriana, una infección vírica, una infección fúngica, una infección por Mycoplasma, una infección por priones, una infección de protozoos, o una infección de helmintos. Un conjunto de procedimientos del aspecto incluye administrar a un sujeto, por ejemplo, un ser humano, un mono, un ratón, una rata, un cerdo, una vaca, un pollo, un ave ornamental, un pájaro aviario, un reptil, o un anfibio, un organismo completo modificado genéticamente. El organismo completo modificado genéticamente estimula por tanto la producción de anticuerpos en el sujeto. Un segundo conjunto de procedimientos comprende tratar profiláctica o terapéuticamente una patología en un sujeto produciendo un anticuerpo contra una o más enfermedades, dolencias u organismos, que implica producir una respuesta de anticuerpos y aislar el anticuerpo o los anticuerpos, que se administran a continuación a un sujeto.

con la mejor herramienta de ingeniería de proteínas, la tecnología Ambrx, la incorporación específica de sitio de aminoácidos no naturales, presenta una gran oportunidad en el desarrollo de vacunas. Puede crear nanoestructuras químicamente bien definidas combinando antígenos con inmunopotenciadores. Estas nanoestructuras, que se pueden usar para el desarrollo de vacunas de subunidades, poseen información del antígeno y funcionalidad de inmunopotenciación.

Además, Se puede usar también la tecnología Ambrx para crear novedosas vacunas análogas a organismos completos atenuados en las que los aminoácidos no naturales servirán como interruptores para la expresión de genes esenciales en organismos patógenos, en vez de vehículos para introducir grupos funcionales no naturales en la proteína. Los organismos resultantes emularán el organismo natural, pero solo serán capaces de replicación en la presencia de aminoácidos no naturales Ambrx, que no existen en la naturaleza.

Vacuna de subunidades:

debido a la baja eficacia de las vacunas de subunidades, se requieren adyuvantes para estimular una respuesta inmunitaria. Tradicionalmente, los adyuvantes se formulan como mezclas con antígenos. Tras la administración, los antígenos serán capturados por las células dendríticas (CD) y presentados a los linfocitos T. Los adyuvantes estimularán las células dendríticas para la liberación de las citoquinas, la presentación mejorada de los antígenos y la maduración de las células dendríticas. se conseguirá una respuesta inmunitaria eficaz específica de antígeno con estos dos procesos independientes, la presentación del antígeno y la inmunoactivación, convergen sobre la misma CD. Se requiere una dosificación relativamente grande. Sin embargo, la dosificación grande de adyuvante evocará la activación inmunitaria independiente de antígeno, un efecto secundario no deseado. Por lo tanto, intrínsecamente esta solución no es ideal y transmite tras la vacuna una baja eficacia y una elevada toxicidad.

A diferencia de los adyuvantes tradicionales, tal como alumbre, muchos adyuvantes descubiertos recientemente están molecularmente bien definidos. Con su única incorporación de aminoácido no natural específica de sitio y sus capacidades de conjugación específicas de sitio, La tecnología Ambrx proporciona a los inventores las herramientas necesarias para crear la siguiente generación de vacunas con elevada potencia y baja toxicidad. La característica clave de las vacunas de subunidades Ambrx será la combinación del antígeno y el inmunopotenciador en una nanoestructura químicamente definida. Cuando es capturada por la CD, los dos procesos, la presentación del antígeno y la activación inmunitaria, convergen sobre la misma CD.

La incorporación específica de sitio de aminoácidos no naturales es una tecnología que permite a los inventores incorporar aminoácidos no naturales con restos muy inmunógenos tales como mononitrofenilo, dinitrofenilo, directamente en antígenos de proteínas. Recientemente, los inmunopotenciadores de moléculas pequeñas, tales como la imidazoquinolina del ligando del receptor 7 de tipo Toll (TLR7), están poniendo a disposición. Es factible diseñar aminoácidos no naturales con estos inmunopotenciadores de moléculas pequeñas como sus cadenas secundarias e incorporarlos sobre un antígeno de proteína.

La tecnología de conjugación Ambrx específica de sitio proporciona incluso más flexibilidad. Esto permite a los inventores conjugar no solo las moléculas pequeñas anteriormente mencionadas en la superficie de los antígenos de proteínas, sino también una variedad de otros inmunopotenciadores (por ejemplo, lípidos, lipopéptidos, polisacáridos, ADN, ARN y nanopartículas).

La conjugación específica de sitio de un ADN monocatenario sobre superficies de proteínas crea otra dimensión de libertad para el diseño de la vacuna. Los ADN de Cpg sin metilar son ligandos de TLR-9, que, de manera interesante, no están presentes sobre la superficie celular sino en el interior de las células. El ADN conjugado con el antígeno puede servir no solo como inmunopotenciador, sino también como un bloque de construcción para crear una a tres dimensiones y nanoestructuras multiantígeno-valentes mediante procesos de hibridación específicos de

secuencias. Se puede usar también este esquema general para combinar antígenos con otros elementos, tales como reactivos de direccionamiento de APC y otros ligandos de TLR. Se ha mostrado que el direccionamiento de APC (anticuerpo o péptido) puede potenciar la eficacia de la vacuna y promover la presentación cruzada. La combinación de dos ligandos de TLR diferentes tendrá mayor efecto sinérgico.

- 5 La conjugación Ambrx específica de sitio permite el diseño preciso y la optimización de las vacunas. Sin la tecnología Ambrx, un proceso de conjugación no específica tiene limitados o sin control los sitios de modificación en el antígeno de la proteína. La conjugación no específica podría alterar epítomos T y B. Esto podría modificar también la conformación 3D del antígeno que es crítica para el reconocimiento del anticuerpo. Adicionalmente, algunas modificaciones podrían obstaculizar el procesamiento del antígeno. Todos estos factores vuelven la vacuna
10 conjugada no específicamente menos eficaz.

Vacunas de organismos completos:

- 15 La vacunología comenzó con vacunas de organismos completos y aún hoy en día estas son las vacunas más satisfactorias, clasificadas tanto como atenuadas vivas como muertas. La esencia de las vacunas de organismos completos es que la vacuna debe estar tan cercana como sea posible al propio organismo patógeno a fin de provocar una respuesta inmunoprotectora pero con capacidades de replicación muy limitadas o sin replicación en el hospedador.

- 20 Por lo tanto, si se limita o no es deseable la replicación, ¿es posible utilizar moléculas pequeñas para controlar el ciclo de vida de los microorganismos para crear vacunas mejores y más seguras? La respuesta, utilizando la tecnología Ambrx, es ahora sí. Si se incorpora el aminoácido no natural en un gen en una bacteria, virus o incluso parásito que es esencial para la replicación del microorganismo, la vida de estos organismos será dependiente de la presencia del aminoácido no natural. Debido a que el aminoácido no natural incorporado no existe en la naturaleza, el resultado serán organismos, tales como virus y bacterias, que tienen componentes y estructuras víricas y celulares prácticamente exactas, sin la capacidad de replicarse en el hospedador. Además, el virus o la bacteria modificada mediante la tecnología Ambrx puede servir no solo como la propia vacuna, sino también como vectores
25 para la administración de genes y antígenos.

- 30 La plataforma de vacunas de subunidades Ambrx tiene una amplia variedad de aplicaciones en áreas de vacunas para enfermedades infecciosas y terapéuticas contra el cáncer. Las vacunas contra el cáncer actuales son generalmente ineficaces. Se han propuesto soluciones de activación de linfocitos T sistémicos e inhibición de las rutas reguladoras negativas y algunas de ellas están en ensayos clínicos. El desacoplamiento de la activación de linfocitos T y el reconocimiento del antígeno pueden producir efectos secundarios graves. La ineficacia de las vacunas contra el cáncer es debida a la ausencia de tecnología para inducir fuertes respuestas tumorales inmuno-específicas. Ambrx proporciona la tecnología para crear nanoestructuras que puedan invocar fuertes respuestas inmuno-específicas contra el cáncer, incluyendo respuestas de linfocitos t y linfocitos B, y al hacerlo abre un nuevo mundo de posibilidades ventajosas, tanto terapéuticas como preventivas.

- 35 La plataforma de vacunas de organismos completos puede encontrar innumerables aplicaciones en el área de las enfermedades infecciosas. Los virus, bacterias, o incluso parásitos que se pueden cultivar y tienen un genoma capaz de mutación son buenos candidatos. Las vacunas de la presente invención mejorarán la gestión de riesgos en el desarrollo de vacunas y las vacunas estarán más bien definidas, tanto química como genéticamente, y proporcionarán eficacias sin precedentes.

- 40 Para expandir el código genético, se proporcionan composiciones y procedimientos para producir ARNt ortogonales. Los aminoacilato ARNt de las aminoacil ARNt sintetetas que se desvelan con un aminoácido codificado no naturalmente. Estos componentes de la traducción se pueden usar para incorporar un aminoácido seleccionado en una posición específica en una cadena polipeptídica en crecimiento (durante la traducción del ácido nucleico) en respuesta a un codón selector que es reconocido por el ARNt.

- 45 Se desvelan también procedimientos para producir una proteína en una célula con un aminoácido seleccionado en una posición especificada. Por ejemplo, un procedimiento incluye crecer, en un medio adecuado, una célula, donde la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína; y, proporcionar el aminoácido seleccionado. La célula comprende además: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector; y, una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente el O-ARNt con el aminoácido seleccionado. Normalmente, el O-ARNt comprende la actividad de supresión en presencia de una sintetasa análoga. Se desvela también una proteína producida mediante este
50 procedimiento.

Breve descripción de los dibujos

- 55 Figura 1 - Se muestra la estructura de tipo hoja de trébol del ARNt de J17 con los sitios de mutación en el tallo TΨC.
Figura 2 - Se muestra la supresión de una mutación ámbar en la hormona de crecimiento humana utilizando J17 o el mutante J17 (F12, F13, F14). Se analizaron los lisados de células totales de cada muestra mediante SDS PAGE.

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

Figura 4 - Muestra un diagrama del vector pVK10-camR que contiene el Sistema de supresión Ambrx.

Figura 5 - Muestra un diagrama del vector pKD46 que contienen los componentes Lambda Red (bet, gam, exonucleasa).

Figura 6 - Muestra un esquema de cebadores de solapamiento y los procedimientos utilizados en el Ejemplo 5. Los cebadores se indican por flechas y un identificador numérico, y por ejemplo. el 1^{er} cebador de los clones Y31 y N51 sería la SEQ ID NO: 18 y 26, respectivamente; el 2^o cebador de los clones Y31 y N51 sería la SEQ ID NO: 19 y 27, respectivamente; el 3^o cebador de los clones Y31 y N51 sería la SEQ ID NO: 20 y 28, respectivamente; el 4^o cebador de los clones Y31 y N51 sería la SEQ ID NO: 21 y 29, respectivamente; el 5^o cebador de los clones Y31 y N51 sería la SEQ ID NO: 22 y 30, respectivamente; el 6^o cebador de los clones Y31 y N51 sería la SEQ ID NO: 23 y 31, respectivamente; el 7^o cebador de los clones Y31 y N51 sería la SEQ ID NO: 24 y 32, respectivamente; el 8^{er} cebador de los clones Y31 y N51 sería la SEQ ID NO: 25 y 33, respectivamente. Los productos de la PCR se identifican por una letra. Se generó el sitio ámbar que contenía la metionina aminopeptidasa solapando la PCR sobre la mutación ámbar (A y B), por lo que es condicionalmente letal para la *E. coli*, como un organismo completo. La metionina aminopeptidasa mutante se fusionó a continuación en un gen completo (C) y se combinó con un marcador seleccionable (D) y una parte parcial de la metionina aminopeptidasa mutante (E) para la selección en la dirección 3'.

Figura 7 - Muestra un esquema de cribado para la dependencia del aminoácido artificial. Se muestran las tres etapas 1) repicando los transformantes, tales como KanR, los transformantes AmpR positivos, tratando las placas; 2) cultivando los supervivientes en KanR, AmpR, + el aminoácido no natural; y 3) haciendo crecer colonias individuales en las placas con y sin el aminoácido no natural y seleccionando aquellas colonias que crecen en presencia del aminoácido no natural y mueren cuando se cultivan sin el aminoácido no natural.

Figura 8 - Muestra las dos placas de cultivo celular de *E. coli* que se están cribando para una mutación ámbar en el gen de la metionina aminopeptidasa en Y31, revestidas cada una con agar LB y con medios que contienen 50 µg/ml de ampicilina y 50 µg/ml de kanamicina, la placa de la izquierda contiene adicionalmente 2 mM de paraacetilfenilalanina y la placa de la derecha no contiene ninguno. A12, C5 y E3 son ejemplos de *E. coli* dependiente de la replicación modulada químicamente.

Figura 9 - Muestra las dos placas de cultivo celular de *E. coli* que se están cribando para una mutación ámbar en el gen de la metionina aminopeptidasa en N51, revestidas cada una con agar LB y con medios que contienen 50 µg/ml de ampicilina y 50 µg/ml de kanamicina, la placa de la izquierda contiene adicionalmente 2 mM de paraacetilfenilalanina y la placa de la derecha no contiene ninguno. A12, C5 y E3 son ejemplos de *E. coli* dependiente de la replicación modulada químicamente.

Figura 10A y 10B - muestran el crecimiento en medios permisivos y no permisivos. Los cultivos durante la noche se inocularon en medio de glucosa (rombos) o arabinosa (círculos) a las relaciones de dilución óptimas que se muestran en la esquina derecha superior de cada panel. Se vigiló la DO600 en una placa de 96 pocillo en un lector de placas Spectramax con agitación a 37°C con una retirada periódica de muestras para el microscopio y sembrado de placas en placas de arabinosa. Se calculó el cambio en la viabilidad, dado en el eje derecho (triángulos), utilizando la concentración de UFC normalizada dividiendo la DO600. Para cada punto temporal, se dividió este valor por el valor en T = 0 y el log de este valor se presenta aquí. Los números de UFC por mililitro por unidad de DO en T = 0 medidos a partir del inóculo saturado y corregido para la dilución fueron del siguiente modo: WT, $4,6 \times 10^9$; *frr* mutante, $3,2 \times 10^9$; *gcpE* mutante, $3,4 \times 10^9$; *lpxC* mutante, $2,2 \times 10^9$; *map* mutante, $2,9 \times 10^9$; *murA* mutante, $8,7 \times 10^8$; *ppa* mutante, $1,5 \times 10^9$; *rpsA* mutante, $1,2 \times 10^9$. La DO se muestra en una escala lineal más bien que en una escala logarítmica para mostrar mejor la dinámica de los cultivos fluctuantes, y la escala del eje de la derecha difiere de *lpxC* y se cartografió para cartografiar las líneas de sesgo procedentes del solapamiento. Para los dos últimos puntos temporales de la medición de *murA*, no se hicieron crecer colonias. De Herring y Blattner JOURNAL OF BACTERIOLOGY, 2004, págs. 2673-2681.

Figura 11 - Muestra un esquema de la caracterización de las etapas de reversión en el que las células dependientes de la replicación de aminoácidos no naturales modificados genéticamente se hacen crecer a una densidad óptica de 1,0 en medios + aminoácidos no naturales, a continuación, aquellas células se sembraron en placas en diluciones en serie en medios con aminoácidos no naturales añadidos para calcular las unidades de formación de colonias, y las células restantes se cultivaron sin la presencia de los aminoácidos no naturales para detectar revertientes.

Figura 12 - Muestra un esquema de una estrategia de mutagénesis. Se produjo un fragmento de ADN lineal a partir del ADN genómico del molde WT mediante solapamiento de la PCR de extensión. Las posiciones de los cebadores se indican por flechas de un solo lado. El producto de fusión de la PCR se sometió a electroporación en células, y seleccionaron las integraciones resultantes de un evento de recombinación doble. Tras la identificación de un clon que transporta el codón de terminación ámbar acompañante, Se indujo el gen I-SceI, dando como resultado la eliminación del gen que codifica Camr. La recombinación en una región duplicada corta conduce a la generación de un mutante ámbar que no tiene, de otra manera, cicatrices.

Figura 13 - muestra los plásmidos utilizados en la mutagénesis acompañante.

Figura 14 - Muestra un esquema de las etapas en la creación de un mutante letal condicional en *E. coli* que incluye, comenzando en la parte superior, 1) crear un molde de mutante condicional; 2) transformar en lambda RED recombinante, la cepa competente de supresión ámbar; y 3) inducir la recombinación lambda RED.

Figura 15 - Muestra un esquema de las etapas de caracterización de un mutante letal condicional en *E. coli* que

incluyen, comenzando en la parte superior, 1) caracterizar el crecimiento en la presencia y ausencia de un aminoácido no natural tras el cual sería dependiente un organismo transformado satisfactoriamente; 2) caracterizar la mutación como bactericida o bacteriostática, y 3) caracterizar la frecuencia revertiente.

Figura 16 - Muestra una vista estructural de los picos de restos MAP.

5 Figura 17 - Muestra una vista estructural de los picos de restos MAP.

Definiciones

Tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un", "uno/a" y "el/la", incluyen las referencias plurales salvo que el contexto indique claramente otra cosa. Por lo tanto, por ejemplo, la referencia a "una célula" incluye una combinación de dos o más células e incluye equivalentes de las mismas conocidos por las personas normalmente expertas en la materia, y así sucesivamente. La referencia a "bacterias" incluye mezclas de bacterias, y similares.

Salvo que se defina en el presente documento y a continuación en el resto de la memoria descriptiva, Todos los términos técnicos y científicos utilizados en el presente documento tienen el mismo significado que el que entiende comúnmente una persona normalmente experta en la técnica a la cual pertenece la invención.

15 Las proteínas y/o las secuencias de proteínas son "homólogas" cuando se derivan, natural o artificialmente, de una proteína o secuencia de proteínas común. De manera similar, Los ácidos nucleicos y/o las secuencias de ácidos nucleicos son homólogas cuando se derivan, natural o artificialmente, de un ácido nucleico ancestral común o una secuencia de ácidos nucleicos. Por ejemplo, cualquier ácido nucleico que se produzca naturalmente puede modificarse mediante cualquier procedimiento de mutagénesis disponible para incluir uno o más codones selectores.

20 Cuando se expresa, este ácido nucleico mutagenizado codifica un polipéptido que comprende uno o más aminoácidos seleccionados, por ejemplo, aminoácidos no naturales. El proceso de mutación puede, por supuesto, alterar adicionalmente uno o más codones de inicio, cambiando también por tanto uno o más aminoácidos normalizados en la proteína mutante resultante. El uno o más aminoácidos normalizados puede cambiarse por un aminoácido no natural o un aminoácido natural. La homología se infiere generalmente a partir de la similitud de la secuencia entre dos o más ácidos nucleicos o proteínas (o secuencias de los mismos). El porcentaje preciso de similitud entre secuencias que es útil para establecer la homología varía con el ácido nucleico y la proteína en cuestión, pero tan poco como un 25% de similitud de la secuencia se usa para establecer la homología. Se pueden utilizar niveles de similitud de la secuencia, por ejemplo, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% o 99% o más, para establecer la homología. Se describen en el presente documento los procedimientos para determinar los porcentajes de similitud de las secuencias (por ejemplo, BLASTP y BLASTN utilizando parámetros por defecto) y están generalmente disponibles.

Tal como se usa en el presente documento, el término "ortogonal" se refiere a una molécula (por ejemplo, un ARNt ortogonal (O-ARNt) y una aminoacil ARNt sintetasa ortogonal (O-RS)) que se usa con eficacia reducida por un sistema de interés (por ejemplo, un sistema de traducción, por ejemplo, una célula). Ortogonal se refiere a la incapacidad o a una eficacia reducida, por ejemplo, menos del 20% de eficacia, menos del 10% de eficacia, menos del 5% de eficacia, o, por ejemplo, menos del 1% de eficacia, de un ARNt ortogonal y/o de una RS ortogonal para funcionar en el sistema de traducción de interés. Por ejemplo, un ARNt ortogonal en un sistema de traducción de interés está aminoacilado por cualquier RS endógena de un sistema de traducción de interés con eficacia reducida o incluso cero, cuando se compara con la aminoacilación de un ARNt endógeno mediante una RS endógena. En otro ejemplo, una RS ortogonal aminoacila cualquier ARNt endógeno en el sistema de traducción de interés con eficacia reducida o incluso cero, en comparación con la aminoacilación del ARNt endógeno mediante una RS endógena. se puede introducir una segunda molécula ortogonal en la célula que funciona con la primera molécula ortogonal. Por ejemplo, una pareja de ARNt/RS ortogonal incluye componentes complementarios introducidos que funcionan juntos en la célula con una eficacia (por ejemplo, aproximadamente un 50% de eficacia, aproximadamente un 60% de eficacia, aproximadamente un 70% de eficacia, aproximadamente un 75% de eficacia, aproximadamente un 80% de eficacia, aproximadamente un 85% de eficacia, aproximadamente un 90% de eficacia, aproximadamente un 95% de eficacia o aproximadamente un 99% o más de eficacia) que la de una pareja de ARNt/RS correspondiente. "Mejora en la ortogonalidad" se refiere a una ortogonalidad potenciada en comparación con un material de partida i un ARNt u una RS que se produce naturalmente.

50 El término "análogo" se refiere a componentes que funcionan juntos, por ejemplo, un ARNt y una aminoacil ARNt sintetasa. Los componentes se pueden denominar también como complementarios.

La expresión "aminoacila preferentemente" se refiere a una eficacia, por ejemplo, aproximadamente un 70% de eficacia, aproximadamente un 75% de eficacia, aproximadamente un 80% de eficacia, aproximadamente un 85% de eficacia, aproximadamente un 90% de eficacia, aproximadamente un 95 % de eficacia, o aproximadamente un 99% o más de eficacia, a la cual una O-RS aminoacila un O-ARNt con un aminoácido seleccionado, por ejemplo, un aminoácido artificial, en comparación con la O-RS que aminoacila un ARNt que se produce naturalmente o un material de partida utilizado para general el O-ARNt. El aminoácido no natural se incorpora a continuación en una cadena polipeptídica en crecimiento con una alta fidelidad, por ejemplo, de más de aproximadamente un 70% de eficacia para un codón selector dado, de más de aproximadamente un 75% de eficacia para un codón selector dado, de más de aproximadamente un 80% de eficacia para un codón selector dado, de más de aproximadamente un 85%

de eficacia para un codón selector dado, de más de aproximadamente un 90% de eficacia para un codón selector dado, más de aproximadamente un 95% de eficacia para un codón selector dado, o más de aproximadamente un 99% de eficacia para un codón selector dado.

5 Un "tratamiento profiláctico" es un tratamiento administrado a un sujeto que no presenta signos o síntomas de una enfermedad, patología, o trastorno médico, o presenta solo signo o síntomas tempranos de una enfermedad, patología, o trastorno, de tal manera que el tratamiento se administra con el objetivo de disminuir, prevenir, o disminuir el riesgo de desarrollar la enfermedad, patología o trastorno médico. Un tratamiento profiláctico funciona como un tratamiento preventivo frente a una enfermedad o trastorno. Una "actividad profiláctica" es una actividad de un agente, tal como un inmunógeno no natural y/o un anticuerpo, o composición del mismo, que, cuando se
10 administra a un sujeto que no presenta signos o síntomas de una patología, enfermedad, o trastorno (o que presenta solo signos o síntomas tempranos de las mismas) reduce, previene, o disminuye el riesgo del sujeto de desarrollar la patología, enfermedad o trastorno. Un agente o compuesto "profilácticamente útil" (por ejemplo, un inmunógeno no natural y/o un anticuerpo de la descripción, se refiere a un agente o compuesto que es útil para reducir, prevenir, tratar, o disminuir el desarrollo de una patología, enfermedad o trastorno.

15 Los tratamientos, vacunas, y/o tratamientos profilácticos se pueden administrar a un paciente que lo necesita. Los tratamientos, vacunas, y/o tratamientos profilácticos pueden administrarse también a una variedad de animales incluyendo, aunque no de forma limitativa, ganadería doméstica, tal como vacas, cerdos, cabras, ovejas, pollos, y/u otros animales de granja comunes y mascotas domésticas comunes, por ejemplo, gatos, perros, loros, periquitos, etc.

20 La expresión "codón selector" se refiere a codones reconocidos por el O-ARNt en el proceso de traducción y no reconocidos por un ARNt endógeno. El bucle anticodón del O-ARNt reconoce el codón selector en el ARNm e incorpora su aminoácido, por ejemplo, un aminoácido seleccionado, tal como un aminoácido artificial, en este sitio en el polipéptido. Los codones selectores pueden incluir, aunque no de forma limitativa, por ejemplo, codones finalizadores, tales como, codones de terminación, incluyendo, aunque no de forma limitativa, codones ámbar, ocre, y ópalo; cuatro o más codones base; codones raros; codones derivados de parejas de bases naturales o no naturales y/o similares. Para un sistema dado, un codón selector puede incluir también uno de los tres codones base naturales, en el que el sistema endógeno no utiliza (o utiliza raramente) dicho codón de tres bases natural. Por
25 ejemplo, esto incluye un sistema que carece de ARNt que reconoce el codón de tres bases natural, y/o un sistema en el que el codón de tres bases natural es un codón raro.

30 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 que puede leer a través de un codón incluyendo, aunque no de forma limitativa, un codón de terminación, un codón de cuatro bases, o un codón raro.

35 La expresión "actividad de supresión" se refiere a la capacidad de un ARNt, por ejemplo, un ARNt supresor, de leer a través de un codón selector. La actividad puede expresarse como un porcentaje de actividad observado en comparación con un control (por ejemplo, careciendo de una sintetasa análoga).

40 La expresión "sistema de traducción" se refiere a los componentes necesarios para incorporar un aminoácido que se produce naturalmente en una cadena polipeptídica en crecimiento (proteína). Los componentes de un sistema de traducción pueden incluir, por ejemplo, ribosomas, ARNt, sintetetasas, ARNm y similares. Los componentes de la presente divulgación pueden añadirse a un sistema de traducción in vitro o in vivo. Los ejemplos de sistemas de traducción incluyen, aunque no de forma limitativa, una célula no eucariota, por ejemplo, una bacteria (tal como *E. coli*), una célula eucariota, por ejemplo, una célula de levadura, una célula de mamífero, una célula vegetal, una célula de alga, una célula de hongo, una célula de insecto, un sistema de traducción exento de células, por ejemplo, un lisado celular, y/o similares.

45 Los sistemas de traducción pueden ser celulares o exentos de células, y pueden ser procariotas o eucariotas. Los sistemas de traducción celulares incluyen, aunque no de forma limitativa, preparaciones de organismos completos tales como células o cultivos de células permeabilizadas en los que una secuencia de ácido nucleico deseada se puede transcribir en ARNm y el ARNm traducirse. Los sistemas de traducción exentos de células están comercialmente disponibles y se conocen bien muchos tipos y sistemas diferentes. Los ejemplos de sistemas
50 exentos de células incluyen, aunque no de forma limitativa, lisados procarionóticos tales como lisados de *Escherichia coli*, y lisados eucarióticos tales como extractos de germen de trigo, lisados de células de insectos, lisados de reticulocitos de conejos, lisados de oocitos de conejos y lisados de células humanas. Pueden preferirse extractos o lisados eucarióticos cuando la proteína resultante está glicosilada, fosforilada o modificada de otra forma debido a que muchas de dichas modificaciones son solo posibles en sistemas eucarióticos. Algunos de estos extractos y lisados están disponibles comercialmente (Promega; Madison, Wis.; Stratagene; La Jolla, Calif.; Amersham; Arlington Heights, Ill.; GIBCO/BRL; Grand Island, N.Y.). Están también disponibles extractos membranosos, tales como extractos de páncreas caninos que contienen membranas ribosómicas, que son útiles para traducir proteínas secretoras.

Se pueden utilizar también sistemas de traducción reconstituidos. Se han utilizado también satisfactoriamente

mezclas de factores de traducción purificados para traducir ARNm en una proteína así como combinaciones de lisados o lisados suplementados con factores de traducción purificados tales como el factor-1 de inicio (IF-1), IF-2, IF-3 (α o β), factor T de alargamiento (EF-Tu), o factores de terminación. Los sistemas exentos de células pueden también acoplarse a los sistemas de transcripción/traducción en los que el ADN se introduce en el sistema, se transcribe en ARNm y el ARNm se traduce como se describe en Current Protocols in Molecular Biology (F. M. Ausubel y col. editores, Wiley Interscience, 1993). El ARN transcrito en el sistema de transcripción eucariota puede estar en la forma de un ARN heteronuclear (ARNhn) o protecciones en extremo 5' (7-metil guanósina) y ARNm madura con una cola poli A en el extremo 3', que puede ser una ventaja en determinados sistemas de traducción. Por ejemplo, los ARNm protegidos se traducen con una alta eficacia en el sistema de lisados de reticulocitos.

La expresión "aminoácido seleccionado" se refiere a cualquier aminoácido que se produce naturalmente o aminoácido no natural deseado. Tal como se usa en el presente documento, la expresión "aminoácido artificial" o "aminoácido codificado de forma no natural" se refiere a cualquier aminoácido, aminoácido modificado, y/o análogo de aminoácido que no es uno de los 20 aminoácidos que se producen naturalmente o selenocisteína o pirrolisina. Otros términos que se pueden usar de forma sinónima con la expresión "aminoácido codificado de forma no natural" y "aminoácido artificial" son "aminoácido no natural", y aminoácido que no se produce naturalmente", y versiones con varios guiones y sin guiones de los mismos. La expresión "aminoácido codificado no naturalmente" incluye también, aunque no de forma limitativa, aminoácidos que se producen mediante modificación (por ejemplo, modificaciones posteriores a la traducción) de un aminoácido codificado naturalmente (incluyendo, aunque no de forma limitativa, los 20 aminoácidos comunes o pirrolisina y selenocisteína) pero no se incorporan por sí mismos en una cadena polipeptídica en crecimiento mediante el complejo de traducción. Los ejemplos de dichos aminoácidos que no se producen naturalmente incluyen, aunque no de forma limitativa, *N*-acetilglucosaminil-L-serina, *N*-acetilglucosaminil-L-treonina, y O-fosfotirosina.

Aminoácido no natural: Tal como se usa en el presente documento, un aminoácido no natural o un análogo de aminoácido codificado se refiere a cualquier aminoácido, aminoácido modificado, o análogo de aminoácidos diferente que la selenocisteína y/o pirrolisina y los siguiente veinte alfa-aminoácidos clásicos codificados genéticamente: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. En diversas realizaciones de la invención, el uno o más aminoácido artificial que se incorpora en el inmunógeno artificial puede ser cualquier aminoácido artificial. Por lo tanto, se apreciará que la citación de aminoácidos artificiales específicos en el presente documento no debe tomarse como limitante sobre la invención. Se ha incorporado una amplia variedad de aminoácidos artificiales en las proteínas codificándolas in vivo, por ejemplo, utilizando los sistemas de traducción que comprenden elementos ortogonales. Véanse, por ejemplo, Liu, y col. (2007) "Genetic incorporation of unnatural amino acids into proteins in mammalian cells" *Nat Methods* 4:239-244; Wang, y col. (2006) "Expanding the genetic code" *Annu Rev Biophys Biomol Struct* 35:225-249; Xie & Schultz (2006) "A chemical toolkit for proteins-an expanded genetic code" *Nat Rev Mol Cell Biol* 7:775-782; Wang y Schultz "Expanding the Genetic Code," *Angewandte Chemie Int. Ed.*, 44(1):34-66 (2005) y Chin, y col. (2003) "An expanded eukaryotic genetic code" *Science* 301:964-967 para una revisión.

En algunas realizaciones de la presente invención es deseable utilizar aminoácidos no naturales que no son uno de los 20 aminoácidos comunes que se producen naturalmente o los aminoácidos raros que se producen naturalmente, por ejemplo, selenocisteína o pirrolisina. Por ejemplo, los aminoácidos no naturales, p-nitrofenilalanina, p-sulfotirosina, y p-carboxifenilalanina encuentran uso en diversas realizaciones en el presente documento. En algunas realizaciones, los aminoácidos no naturales pueden incluir, aunque no de forma limitativa: p-nitrofenilalanina; una o-nitrofenilalanina; una m-nitrofenilalanina; una p-boronilfenilalanina; una o-boronilfenilalanina; una m-boronilfenilalanina; una p-aminofenilalanina; una o-aminofenilalanina; una m-aminofenilalanina; una p-acilfenilalanina; una o-acilfenilalanina; una m-acilfenilalanina; una p-OMe fenilalanina; una o-OMe fenilalanina; una m-OMe fenilalanina; una p-sulfofenilalanina; una o-sulfofenilalanina; una m-sulfofenilalanina; una 5-nitro His; una 3-nitro Tyr; una 2-nitro Tyr; una Leu nitrosustituida; una His nitrosustituida; un De nitrosustituido; un Trp nitrosustituido; un 2-nitro Trp; un 4-nitro Trp; un 5-nitro Trp; un 6-nitro Trp; un 7-nitro Trp; 3-aminotirosina, 2-aminotirosina, O-sulfotirosina, 2-sulfoxifenilalanina, 3-sulfoxioxifenilalanina o p-carboxifenilalanina, o-carboxifenilalanina, y m-carboxifenilalanina. Igualmente, Se apreciara que la invención no se limita a aminoácidos no naturales concretos.

Además, en diversas realizaciones de la presente invención, se pueden incorporar aminoácidos artificiales en inmunógenos in vitro, por ejemplo, utilizando procedimientos biosintéticos en los que un ARNt supresor se acila químicamente con un aminoácido artificial deseado y se añade a un extracto in vitro capaz de soportar la biosíntesis del inmunógeno. Para una descripción de dichos procedimientos sintéticos in vitro, véanse, por ejemplo, V. W. Cornish, D. Mendel y P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* 1995, 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, "A general method for site-specific incorporation of unnatural amino acids into proteins," *Science* 244 182-188 (1989); y, J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Dials, "Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide," *J. Am. Chem. Soc.* 111 8013-8014 (1989). Se pueden añadir también aminoácidos artificiales a proteínas producidas de forma natural o sintética mediante las químicas de péptidos sintéticos disponibles (o se pueden convertir aminoácidos naturales en aminoácidos artificiales mediante dichos procedimientos), o mediante procesamiento posterior a la traducción. Igualmente, sin embargo, se apreciará que dichas modificaciones posteriores a la traducción y químicas se llevan a cabo normalmente junto con, o además de, la incorporación de uno o más aminoácidos artificiales durante la síntesis de una molécula (por

- ejemplo, incorporación directa tal como traducción ortogonal, síntesis en fase sólida, etc.). Por lo tanto, la adición posterior a la traducción o la modificación química de aminoácidos se llevan a cabo normalmente, si es que es posible, solo en moléculas que ya tienen aminoácidos artificiales que se añadieron durante la síntesis de la molécula. Se proporciona a continuación información adicional sobre la incorporación no ortogonal de aminoácidos artificiales en inmunógenos.
- 5 Tal como se usa en el presente documento, la expresión "derivado de" se refiere a un componente que está aislado o preparado utilizando información de una molécula u organismo especificado.
- Una "célula hospedadora recombinante" o "célula hospedadora" se refiere a una célula que incluye un polinucleótido exógeno, sin tener en cuenta el procedimiento utilizado para la inserción, por ejemplo, captación directa, transducción, emparejamiento f, u otros procedimientos conocidos en la técnica para crear células hospedadoras recombinantes. El polinucleótido exógeno puede mantenerse como un vector no integrado, por ejemplo, un plásmido, o, de forma alternativa, puede integrarse en el genoma del hospedador.
- 10 Tal como se usa en el presente documento, el término "medio" o "medios" incluye cualquier medio de cultivo, solución, soporte sólido, semisólido, o rígido que pueda soportar o contener cualquier célula hospedadora, incluyendo células hospedadoras bacterianas, células hospedadoras de levaduras, células hospedadoras de insectos, células hospedadoras vegetales, células hospedadoras eucariotas, células hospedadoras de mamíferos, células CHO, células hospedadoras procariotas, E. coli, o células hospedadoras de Pseudomonas, y los contenidos celulares. De esta manera, el término puede abarcar el medio en el que se ha hecho crecer la célula hospedadora, por ejemplo, el medio en el que se hace crecer un cultivo u organismo completo o célula, y el medio puede tener un aminoácido no natural incluido para soportar el crecimiento de organismos o células deficientes en la replicación, incluyendo cualquier medio antes o después de una etapa de proliferación. El término puede abarcar también tampones o reactivos que contienen lisados de células hospedadoras, tales como en el caso donde los antígenos se producen intracelularmente y las células hospedadoras se lisan o perturban para liberar los antígenos o antígenos recombinantes.
- 15 "Agente reductor", tal como se usa en el presente documento con respecto al replegado de una proteína, se define como cualquier compuesto o material que mantiene grupos sulfhidrilo en el estado reducido y reduce los enlaces disulfuro intramoleculares o intermoleculares. Los agentes reductores adecuados incluyen, aunque no de forma limitativa, ditioneitol (DTT), 2-mercaptoetanol, ditioneitol, cisteína, cisteamina (2-aminoetanol), y glutatión reducido. Es fácilmente evidente para las personas normalmente expertas en la materia que una amplia variedad de agentes reductores son adecuados para el uso en las composiciones desveladas en el presente documento.
- 20 "Agente oxidante", tal como se usa en el presente documento con respecto al replegado de una proteína, se define como cualquier compuesto o material que es capaz de eliminar un electrón de un compuesto que se está oxidando. Los agentes oxidantes adecuados incluyen, aunque no de forma limitativa, glutatión oxidado, cistina, cistamina, ditioneitol oxidado, eritreitol oxidado, y oxígeno. Es fácilmente evidente para las personas normalmente expertas en la materia que es adecuada una amplia variedad de agentes oxidantes para el uso en los procedimientos desvelados en el presente documento.
- 25 "Replegado", tal como se usa en el presente documento describe cualquier proceso, reacción o procedimiento que transforma los polipéptidos que contienen enlaces disulfuro desde un estado plegado o desplegado inadecuadamente a una conformación natural o adecuadamente plegada con respecto a los enlaces disulfuro.
- 30 "Plegado simultáneo", tal como se usa en el presente documento, se refiere específicamente a procesos, reacciones o procedimientos de replegado que emplean al menos dos polipéptidos que interactúan entre sí y dan como resultado la transformación de polipéptidos desplegados o plegados inadecuadamente en polipéptidos naturales, plegados adecuadamente.
- 35 Un "grupo modificador en el extremo amino" se refiere a cualquier moléculas que se puede unir al extremo amino de un polipéptido. De manera similar, un "grupo modificador en el extremo carboxi" se refiere a cualquier molécula que se puede unir al extremo carboxi de un polipéptido. Los grupos modificadores de extremos incluyen, aunque no de forma limitativa, diversos polímeros, péptidos o proteínas solubles en agua tales como albúmina de suero, u otros restos que aumentan la semivida en suero de los péptidos.
- 40 Las expresiones "grupo funcional", "resto activo", "grupo activador", "grupo saliente", "sitio reactivo", "grupo químicamente reactivo" y "resto químicamente reactivo" se utilizan en la técnica y en el presente documento para referirse a partes o unidades distintas, definibles de una molécula. Los términos son algunas veces sinónimos en las técnicas químicas y se usan en el presente documento para indicar las partes de moléculas que llevan a cabo alguna función o actividad y son reactivas con otras moléculas.
- 45 El término "enlace" o "enlazador" se usa en el presente documento para referirse a grupos o enlaces que normalmente se forman como el resultado de una reacción química y normalmente son enlaces covalentes. Los enlaces hidrolíticamente estables significan que los enlaces son sustancialmente estables en agua y no reaccionan con agua a valores de pH útiles, incluyendo, aunque no de forma limitativa, en condiciones fisiológicas durante un periodo extendido de tiempo, quizás incluso de forma indefinida. Los enlaces hidrolíticamente inestables o
- 50
- 55

el término "alquileo" por sí mismo o como parte de otro sustituyente significa un radical divalente derivado de un alcano, como se ilustra, aunque no de forma limitativa, por las estructuras $-\text{CH}_2\text{CH}_2-$ y $-\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2-$, e incluye además aquellos grupos descritos a continuación como "heteroalquileo". Normalmente, un grupo alquilo (o alquileo) tendrá de 1 a 24 átomos de carbono, siendo aquellos grupos que tienen 10 o unos pocos átomos de carbono capaces de ser una descripción concreta de los procedimientos y composiciones descritos en el presente documento. Un "alquilo inferior" o "alquileo inferior" es un grupo alquilo de cadena más corta o un grupo alquileo, teniendo generalmente ocho o unos pocos átomos de carbono.

Los términos "-alcoxi", "alquilamino" y "alquiltio" (o tioalcoxi) se utilizan en su sentido convencional, y se refieren a aquellos grupos alquilo unidos al resto de la molécula mediante un átomo de oxígeno, un grupo amino o un átomo de azufre, respectivamente.

El término "heteroalquilo", por sí mismo o en combinación con otro término, significa, salvo que se indique otra cosa, una cadena lineal o ramificada estable, o un radical hidrocarburo cíclico, o combinaciones de los mismos, que consiste en un número indicado de átomos de carbono y al menos un heteroátomo seleccionado entre el grupo que consiste en O, N, Si y S, y en el que los átomos de nitrógeno y azufre pueden opcionalmente oxidarse y el heteroátomo de nitrógeno puede opcionalmente cuaternizarse. Los heteroátomo(s) O, N y S y Si pueden colocarse en cualquier posición inferior del grupo heteroalquilo o en la posición en la cual el grupo alquilo se une al resto de la molécula. Los ejemplos incluyen, aunque no de forma limitativa, $-\text{CH}_2\text{-CH}_2\text{-O-CH}_3$, $-\text{CH}_2\text{-CH}_2\text{-NH-CH}_3$, $-\text{CH}_2\text{-CH}_2\text{-N(CH}_3\text{)-CH}_3$, $-\text{CH}_2\text{-S-CH}_2\text{-CH}_3$, $-\text{CH}_2\text{-CH}_2\text{-S(O)-CH}_3$, $-\text{CH}_2\text{-CH}_2\text{-S(O)}_2\text{-CH}_3$, $-\text{CH=CH-O-CH}_3$, $-\text{Si(CH}_3\text{)}_3$, $-\text{CH}_2\text{-CH=N-OCH}_3$, y $-\text{CH=CH-N(CH}_3\text{)-CH}_3$. Hasta dos heteroátomos pueden ser consecutivos, tales como, por ejemplo, $-\text{CH}_2\text{-NH-OCH}_3$ y $-\text{CH}_2\text{-O-Si(CH}_3\text{)}_3$. De manera similar, el término "heteroalquileo" por sí mismo o como parte de otro sustituyente significa un radical divalente derivado a partir de heteroalquilo, como se ilustra, aunque no de forma limitativa por, $-\text{CH}_2\text{-CH}_2\text{-S-CH}_2$ CH_2- y $-\text{CH}_2\text{-S-CH}_2\text{-CH}_2\text{-NH-CH}_2-$. Para grupos heteroalquileo, el mismo o diferentes heteroátomos pueden ocupar cualquiera o ambos de los términos de la cadena (incluyendo, aunque no de forma limitativa, oxialquileo, dioalquileo, aminoalquileo, diaminoalquileo, aminoalquileo, y similares). Aún más, para los grupos de enlace de alquileo y heteroalquileo, ninguna orientación del grupo de enlace está implícita en la dirección en la que se escribe la fórmula del grupo de enlace. Por ejemplo, la fórmula $-\text{C(O)}_2\text{R}'$ representa $-\text{C(O)}_2\text{R}'$ y $-\text{R}'\text{C(O)}_2$.

Los términos "cicloalquilo" y "heterocicloalquilo", por sí mismo o en combinación con otros términos, representa, salvo que se indique otra cosa, versiones cíclicas de "alquilo" y "heteroalquilo", respectivamente. De esta manera, un cicloalquilo o heterocicloalquilo incluye enlaces de anillo saturados, parcialmente insaturados y completamente insaturados. Además, para heterocicloalquilo, un heteroátomo puede ocupar la posición en la que el grupo alquilo se une al resto de la molécula. Los ejemplos de cicloalquilo incluyen, aunque no de forma limitativa, ciclopentilo, ciclohexilo, 1-ciclohexenilo, 3-ciclohexenilo, cicloheptilo y similares. Los ejemplos de heterocicloalquilo incluyen, aunque no de forma limitativa, 1-(1,2,5,6-tetrahidropiridilo), 1-piperidinilo, 2-piperidinilo, 3-piperidinilo, 4-morfolinilo, 3-morfolinilo, tetrahidrofuran-2-ilo, tetrahidrofuran-3-ilo, tetrahidrotien-2-ilo, tetrahidrotien-3-ilo, 1-piperazinilo, 2-piperazinilo y similares. Además, el término abarca estructuras de anillos bicíclicos y tricíclicos. De manera similar, el término "heterocicloalquileo" por sí mismo o como parte de otro sustituyente significa un radical divalente derivado de heterocicloalquilo, y el término "cicloalquileo" por sí mismo o como parte de otro sustituyente significa un radical divalente derivado de cicloalquilo.

El término "aminoácido" se refiere a aminoácidos que se producen naturalmente y aminoácidos que no se producen naturalmente, así como análogos de aminoácidos y miméticos de aminoácidos que funcionan de un modo similar a los aminoácidos que se producen naturalmente. Los aminoácidos codificados naturalmente son los 20 aminoácidos comunes (alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, y valina) y pirrolisina y selenocisteína. Los análogos de aminoácidos se refiere a compuestos que tienen la misma estructura química básica que un aminoácido que se produce naturalmente, es decir, un carbono α que está unido a hidrógeno, un grupo carboxilo, un grupo amino, y un grupo R, tales como, homoserina, norleucina, sulfóxido de metionina, metilsulfonio de metionina. Dichos análogos tiene grupos R modificados (tal como, norleucina) o las estructuras principales de los péptidos modificadas, pero retienen la misma estructura química básica que un aminoácido que se produce naturalmente. La referencia a un aminoácido incluye, por ejemplo, L-aminoácidos proteogénicos que se producen naturalmente; D-aminoácidos, aminoácidos químicamente modificados tales como variantes y derivados de aminoácidos; aminoácidos no proteogénicos que se producen naturalmente tales como α -alanina, ornitina, etc.; y compuestos sintetizados químicamente que tienen propiedades conocidas en la materia por ser características de los aminoácidos. Los ejemplos de aminoácidos que no se producen naturalmente incluyen, aunque no de forma limitativa, α -metil aminoácidos (por ejemplo, α -metil alanina), D-aminoácidos, aminoácidos similares a histidina (por ejemplo, 2-amino histidina, α -hidroxi-histidina, homohistidina, α -fluorometil-histidina y α -metil-histidina), aminoácidos que tiene un metileno extra en la cadena secundaria ("homo" aminoácidos), y aminoácidos en los que está sustituido un grupo funcional de ácido carboxílico en la cadena secundaria con un grupo de ácido sulfónico (por ejemplo, ácido cisteico).

En el presente documento se hace referencia a los aminoácidos bien por su símbolos de tres letras conocido o por el símbolo de una letra recomendados por la IUPAC-IUB Biochemical Nomenclature Commission. Los nucleótidos, igualmente, pueden denominarse por sus códigos de una sola letra comúnmente aceptados.

"Variantes modificadas conservativamente" se aplica tanto a secuencias de aminoácidos como de ácidos nucleicos. Con respecto a secuencias de ácido nucleico concretas, "variantes modificadas conservativamente" se refiere a los ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o en las que el ácido nucleico no codifica una secuencia de aminoácidos, en secuencias esencialmente idénticas. Debido a la degeneración del código genético, un gran número de ácidos nucleicos funcionalmente idénticos codifica cualquier proteína dada. Por ejemplo, los codones GCA, GCC, GCG y GCU codifican, todos ellos, el aminoácido alanina. De esta manera, en cada posición en la que una alanina se especifica por un codón, el codón puede alterarse para cualquiera de los codones correspondientes descritos sin alterar el polipéptido codificado. Dichas variaciones de ácidos nucleicos son "variaciones silenciosas", que son una especie de variaciones modificadas conservativamente. En el presente documento, toda secuencia de ácido nucleico que codifica un polipéptido también describe toda posible variación silente del ácido nucleico. Una persona normalmente experta en la materia reconocerá que cada codón en un ácido nucleico (excepto AUG, que es, de forma ordinaria, el único codón para metionina, y TGG, que es, de forma ordinaria, el único codón para triptófano) puede modificarse para dar como resultado una molécula funcionalmente idéntica. Por consiguiente, cada variación silenciosa de un ácido nucleico que codifica un polipéptido está implícita en cada secuencia descrita.

Como para las secuencias de aminoácidos, una persona normalmente experta en la materia reconocerá que las sustituciones, deleciones o adiciones individuales en una secuencia de un ácido nucleico, péptido, polipéptido, o proteína que alteran, añaden o eliminan un único aminoácido o un pequeños porcentaje de aminoácidos en la secuencia codificada es una "variante modificada conservativamente" en la que la alteración da como resultado de deleción de un aminoácido, la adición de un aminoácido, o la sustitución de un aminoácido con un aminoácido químicamente similar. Las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares son bien conocidas por las personas normalmente expertas en la materia. Dichas variantes modificadas conservativamente se suman y no excluyen las variantes polimórficas, homólogos entre especies y alelos de la invención.

Las tablas de sustituciones conservativas que proporcionan aminoácidos funcionalmente similares son bien conocidas por las personas normalmente expertas en la materia. Los siguientes ocho grupos contienen aminoácidos que son sustituciones conservativas unos de otros:

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

(véase, por ejemplo, Creighton, Proteins: Structures and Molecular Properties (W H Freeman & Co.; 2ª edición (Diciembre 1993)

Los términos "idéntico" o porcentaje de "identidad" en el contexto de dos o más secuencias de ácidos nucleicos o polipéptidos, se refieren a dos o más secuencias o subsecuencias que son iguales. Las secuencias son "sustancialmente idénticas" si tienen un porcentaje de restos de aminoácidos o nucleótidos que son iguales (es decir, aproximadamente un 60% de identidad, aproximadamente 65%, aproximadamente 70%, aproximadamente 75%, aproximadamente 80%, aproximadamente 85%, aproximadamente 90% o aproximadamente 95% de identidad sobre una región especificada), cuando se comparan y alinean para correspondencia máxima sobre una ventana de comparación, o una región designada medida usando uno de los siguientes algoritmos de comparación de secuencias (u otros algoritmos disponibles para las personas normalmente expertas en la materia) o mediante alineamiento manual e inspección visual. Esta definición también se refiere a la complementaria de una secuencia de ensayo. La identidad puede existir sobre una región que tiene una longitud de al menos aproximadamente 50 aminoácidos o nucleótidos, o, sobre una región que tiene una longitud de 75-100 aminoácidos o nucleótidos, o, donde no se especifica, a través de la secuencia completa de un polinucleótido o polipéptido. Se puede obtener un polinucleótido que codifica un polipéptido de la presente divulgación, incluyendo los homólogos de especies diferentes del ser humano, mediante un proceso que comprende las etapas de cribar una biblioteca en condiciones de hibridación rigurosas con una sonda marcada que tiene una secuencia de polinucleótidos de la descripción o un fragmento de la misma, y aislar el ADNc de longitud completa y los clones genómicos que contienen dicha secuencia de polinucleótidos. Dichas técnicas de hibridación son bien conocidas el técnico experto.

Para la comparación de la secuencia, normalmente una secuencia actúa como una secuencia de referencia, con la

cual se comparan las secuencias de prueba. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de ensayo y de referencia se introducen en un ordenador, se designan las coordenadas de la subsecuencia, en caso necesario, y se designan los parámetros del programa del algoritmo de secuencias. Se pueden usar los parámetros por defecto del programa o parámetros alternativos. A continuación, el algoritmo de comparación de secuencias calcula el porcentaje de identidades de secuencia para las secuencias problema con respecto a la secuencia de referencia, según los parámetros del programa.

Una "ventana de comparación", tal como se usa en el presente documento, incluye referencias a un segmento de una cualquiera de las numerosas posiciones contiguas seleccionadas entre el grupo que consiste en 20 a 600, normalmente aproximadamente 50 a aproximadamente 200, más normalmente aproximadamente 100 a aproximadamente 150 en el que una secuencia se puede comparar con una secuencia de referencia del mismo número de posiciones contiguas después de que las dos secuencias están alineadas de forma óptima. Los procedimientos de alineación de secuencias para comparación son bien conocidos por las personas normalmente expertas en la materia. La alineación óptima de las secuencias para comparación puede llevarse a cabo, incluyendo, aunque no de forma limitativa, mediante el algoritmo de homología local de Smith y Waterman (1970) *Adv. Appl. Math.* 2:482c, mediante el algoritmo de alineación de homología de Needleman y Wunsch (1970) *J. Mol. Biol.* 48:443, mediante la búsqueda del procedimiento de similitud de Pearson y Lipman (1988) *Proc. Nat'l. Acad. Sci. USA* 85:2444, mediante las 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 mediante alineación manual e inspección visual (véase, por ejemplo, Ausubel y col., *Current Protocols in Molecular Biology* (suplemento de 1995)).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul y col. (1997) *Nuc. Acids Res.* 25:3389-3402, y Altschul y col. (1990) *J. Mol. Biol.* 215:403-410, respectivamente. El software para realizar los análisis BLAST está disponible para el público a través del National Center for Biotechnology Information en el sitio web en ncbi.nlm.nih.gov. Los parámetros W, T, y X del algoritmo BLAST determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para las secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, M=5, N=-4 y una comparación de ambas hebras. Para las secuencias de aminoácidos, el programa BLASTP utiliza como valores predeterminados una longitud de palabra de 3 y una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, (1992) *Proc. Natl. Acad. Sci. USA* 89:10915) alineaciones (B) de 50, una expectativa (E) de 10, M=5, N=-4, y una comparación de ambas hebras. El algoritmo BLAST se lleva a cabo normalmente con el filtro de "baja complejidad" apagado.

El algoritmo BLAST lleva a cabo también un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787). Una medida de la similitud proporcionada por el algoritmo BLAST es la menor probabilidad de la suma (P(N)), que proporciona una indicación de la probabilidad por la cual se produciría un emparejamiento entre las dos secuencias de nucleótidos o aminoácidos por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la menor probabilidad de la suma en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es inferior a aproximadamente 0,2, o inferior a aproximadamente 0,01, o inferior a aproximadamente 0,001.

La frase "se hibrida selectivamente (o específicamente) a" se refiere a la unión, formación de dupletes o hibridación de una molécula solo con una secuencia nucleotídica concreta en condiciones rigurosa de hibridación cuando esta secuencia está presente en una mezcla compleja (incluyendo, aunque no de forma limitativa, ADN o ARN celular total o de biblioteca).

La frase "condiciones rigurosas de hibridación" se refiere a la hibridación de secuencias de ADN, ARN, PNA, u otras imitaciones de ácidos nucleicos, o combinaciones de las mismas en condiciones de baja fuerza iónica y alta temperatura, como se conoce en la técnica. Normalmente, en condiciones rigurosas, una sonda se hibridará con su diana posterior en una mezcla compleja de ácido nucleico (que incluye, aunque no forma limitativa, ADN o ARN celular total o de biblioteca) pero no se hibrida con otras secuencias en la mezcla compleja. Las condiciones rigurosas dependen de la secuencia y serán diferentes en circunstancias diferentes. Las secuencias más largas se hibridan específicamente a temperaturas más altas. Se encuentra una extensa guía de la hibridación de ácidos nucleicos en Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology* Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). En general, las condiciones rigurosas se seleccionan para que sean aproximadamente 5-10 °C menos que el punto de fusión térmica (T_f) para la secuencia específica a un pH y fuerza iónica definidos. La T_f es la temperatura (para una fuerza iónica, pH, y concentración de ácido nucleico definidas) a la cual el 50% de las sondas complementarias con la diana se hibridan a la secuencia diana en equilibrio) ya que las secuencias diana están presentes en exceso, a T_f, el 50% de las sondas están ocupadas en equilibrio). Las condiciones rigurosas pueden ser aquellas en las que la concentración de sal es menor de aproximadamente 1,0 M de ion sodio, normalmente aproximadamente 0,01 a 1,0 M de concentración de ion sodio (u otras sales) a pH 7,0 a 8,3 y la temperatura es al menos de aproximadamente 30°C para sondas cortas (incluyendo, aunque no de forma limitativa, 10 a 50 nucleótidos) y al menos aproximadamente 60°C para sondas largas (incluyendo, aunque no de forma limitativa, más de 50 nucleótidos). También se pueden alcanzar condiciones rigurosas con la adición de agentes desestabilizantes, tales como formamida. Para la hibridación selectiva o específica, una señal positiva puede tener al menos un fondo de dos

veces, opcionalmente u a hibridación de fondo de 10 veces. Las condiciones rigurosas de hibridación ilustrativas pueden ser las siguientes: 50 % de formamida, 5X SSC y 1 % de SDS, incubando a 42°C o 5X SSC, 1% de SDS, incubando a 65°C, con lavado en 0,2X SSC, y 0,1% de SDS a 65°C. Dichos lavados se pueden llevar a cabo durante 5, 15, 30, 60, 120, o más minutos.

5 Tal como se usa en el presente documento, el término "eucariota" se refiere a organismos que pertenecen al dominio filogenético Eucarya tales como animales (incluyendo, aunque no de forma limitativa, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (incluyendo, aunque no de forma limitativa, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidia, protistas, etc.

10 Tal como se usa en el presente documento, la expresión "-no eucariota" se refiere a organismos no eucariotas. Por ejemplo, un organismo no eucariota puede pertenecer al dominio filogenético de las Eubacteria (incluyendo, aunque no de forma limitativa, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), o el dominio filogenético de las Archaea (incluyendo, aunque no de forma limitativa, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tales como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeuropyrum permix*, etc.).

15 El término "sujeto" tal como se usa en el presente documento, se refiere a un animal, en algunas realizaciones un mamífero, y en otras realizaciones un ser humano, que es el objeto del tratamiento, la observación o el experimento. Un animal puede ser un animal de compañía (por ejemplo, perros, gatos, y similares), animales de granja (por ejemplo, vacas, ovejas, cerdos, caballos, y similares) o animales de laboratorio (por ejemplo, ratas, ratones, cobayas y similares).

20 La expresión "cantidad eficaz", como se usa en el presente documento, se refiere a la cantidad de polipéptido aminoacídico no natural modificado que se está administrando que aliviará en cierta medida uno o más de los síntomas la enfermedad, dolencia o trastorno que se está tratando. Las composiciones que contienen el polipéptido aminoacídico no natural modificado descritas en el presente documento se pueden administrar para los tratamientos profilácticos, de mejora, y/o terapéuticos.

25 Los términos "potenciar" o "mejorar" significan aumentar o prolongar tanto la potencia como la duración de un efecto deseado. De esta manera, con respecto a la potenciación del efecto de los agentes terapéuticos, el término "potenciar" se refiere a la capacidad de aumentar o prolongar, tanto en potencia como en duración, el efecto de los diferentes agentes terapéuticos sobre un sistema. Una "cantidad potenciadora eficaz", tal como se usa en el presente documento, se refiere a una cantidad adecuada para potenciar el efecto de otro agente terapéutico en un sistema deseado. Cuando se usa en un paciente, las cantidades eficaces para este uso dependerán de la gravedad y el curso de la enfermedad, trastorno o dolencia, el tratamiento previo, el estado de salud del paciente y la respuesta a los fármacos, y del criterio del médico a cargo del tratamiento.

30 Tal como se usa en el presente documento, el término "selección positiva" o "marcador de selección" se refiere a un marcador que cuando está presente, por ejemplo, expresado, activado o similar, da como resultado la identificación de una célula con el marcador de selección positiva de aquellos sin el marcador de selección positiva.

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

40 Tal como se usa en el presente documento, el término "indicador" se refiere a un componente que se puede usar para seleccionar componentes diana de un sistema de interés. Por ejemplo, un indicador puede incluir una proteína, por ejemplo, una enzima, que confiere resistencia o sensibilidad a un antibiótico (incluyendo, aunque no de forma limitativa, β -lactamasa, cloranfenicol acetiltransferasa (CAT), y similares), un marcador de selección fluorescente (incluyendo, aunque no de forma limitativa, la proteína fluorescente verde (por ejemplo, GFP), YFP, EGFP, RFP, un marcador luminiscente (incluyendo, aunque no de forma limitativa, una proteína luciferasa de luciérnaga), un marcador de selección basado en la afinidad, o genes marcadores seleccionables positivos o negativos tales como LacZ, β -gal/lacZ (β -galactosidasa), ADH (alcohol deshidrogenasa), his3, ura3, leu2, lys2, o similares).

45 Tal como se usa en el presente documento, el término "eucariota" se refiere a organismos que pertenecen al dominio filogenético Eucarya tales como animales (incluyendo, aunque no de forma limitativa, mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas (incluyendo, aunque no de forma limitativa, monocotiledóneas, dicotiledóneas, algas, etc.), hongos, levaduras, flagelados, microsporidia, protistas, etc.

50 Tal como se usa en el presente documento, la expresión "-no eucariota" se refiere a organismos no eucariotas. Por ejemplo, un organismo no eucariota puede pertenecer al dominio filogenético de las Eubacteria (incluyendo, aunque no de forma limitativa, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.) o al dominio filogenético de las Archaea (por ejemplo, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tales como especies de *Haloferax volcanii* y *Halobacterium* NRC-1, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*,

Aeuropyrum pernix, etc.).

Se incluyen también los siguientes ejemplos no limitantes de proteínas de envolturas víricas: proteínas de envolturas de filovirus (tales como el virus del Ébola), ortomixovirus (tales como el virus de la gripe), VSV-G, alfavirus (tales como el virus del bosque Semliki y virus Sindbis), 5 arenavirus (tales como el virus de la coriomeningitis linfocítica), flavivirus (tales como el virus de la encefalitis transmitido por garrapatas y el virus del Dengue), rabdovirus (tales como el virus de la estomatitis vesicular y el virus de la rabia), el virus de la leucemia de Moloney, VHS, VZV, el virus de las paperas, Rinovirus, Sarampión, Rubéola, Arbovirus, Enterovirus (tales como Polio, Coxsackie, Ecovirus), Virus de la polio, Coxsackie B, A y Ecovirus, Rinovirus, Virus de la hepatitis, Virus Norwalk, Astrovirus, Togavirus, 10 Alfavirus, Pestivirus, Coronavirus, Virus paragripal, el virus de las paperas, Virus del sarampión, Virus sincitial respiratorio (VSR), Bunyaviridae, Reoviridae, Reovirus, Rotavirus, HTLV, Poliomaivirus, Papilomavirus, Adenovirus, Parvovirus, VEB, CMV, Virus de la varicela zoster, virus del herpes, y virus de la viruela.

Variante conservativa: La expresión "variante conservativa" se refiere a un componente de la traducción, por ejemplo, una variante conservativa O-ARNt o una variante conservativa O-RS, que funciona funcionalmente de la misma forma que el componente a partir del cual se basa la variante conservativa, por ejemplo, un O-ARNt u O-RS, 15 pero tiene variaciones en la secuencia. Por ejemplo, una O-RS aminoacilará un O-ARNt complementario o una variante conservativa de O-ARNt con un aminoácido seleccionado, por ejemplo, un aminoácido artificial, aunque el O-ARNt y la variante conservativa de O-ARNt no tienen la misma secuencia. De manera similar, un ARNt se aminoacilará con un aminoácido seleccionado, por ejemplo, un aminoácido artificial, mediante una O-RS complementaria o una variante conservativa de O-RS, aunque la O-RS y la variante conservativa de O-RS no 20 tendrán la misma secuencia. La variante conservativa puede tener, por ejemplo, una variación, dos variaciones, tres variaciones, cuatro variaciones, o cinco o más variaciones en la secuencia, siempre que la variante conservativa sea complementaria al O-ARNt o la O-RS correspondiente.

Selección de un agente de cribado: Tal como se usa en el presente documento, la expresión "selección de un agente de cribado" se refiere a un agente que, cuando está presente, permite una selección/cribado de 25 determinados componentes de una población. Por ejemplo, un agente de selección o cribado incluye, aunque no de forma limitativa, por ejemplo, un nutriente, un antibiótico, una longitud de onda de luz, un anticuerpo, un polinucleótido expresado o similares. Puede variarse el agente de selección, por ejemplo, mediante concentración, intensidad, etc.

La expresión "no reconocido eficazmente" se refiere a una eficacia, por ejemplo, menos de aproximadamente 10%, 30 menos de aproximadamente 5%, o menos de aproximadamente 1%, en la cual una RS de un organismo aminoacila O-ARNt.

Descripción detallada

Hasta ahora, se han limitado las vacunas a aquellas producidas con organismos muertos o atenuados. La muerte del microorganismos mediante tratamiento térmico, UV, formaldehído, da como resultado epítomos naturales reducidos, 35 y los virus atenuados son producidos normalmente a través de una deleción y truncamiento del gen que conduce a una replicación que no es de orden cero pero que es extremadamente baja, un riesgo para el paciente y particularmente para los pacientes más jóvenes, los pacientes mayores, aquellos con los sistemas inmunitarios comprometidos.

En una realización, la presente divulgación proporciona vacunas de organismos completos modificados. En otra 40 realización, la presente divulgación proporciona la replicación de vacunas de organismos completos genéticamente modificados dependiente de uno o más aminoácidos no naturales. La presente divulgación proporciona vacunas que incorporan microorganismos nativos, que se pueden usar con o sin adyuvantes. Estas pueden adicionalmente proporcionar una alta densidad de epítomos nativos que son muy inmunógenos y estimularán respuestas humorales y mediadas por linfocitos T, proporcionando por tanto una vacuna más eficaz. La presente divulgación proporciona 45 vacunas que incorporan uno o más aminoácidos no naturales, artificiales, o codificados de forma no natural en uno o más sitios en el virus o la bacteria. En algunas realizaciones de la presente divulgación, los aminoácidos codificados de forma no natural se incluyen en una parte del microorganismo requerida para la replicación, eliminando por tanto los riesgos asociados con los virus atenuados y eliminando la necesidad de matar el virus/bacteria. La vacuna con aminoácido(s) codificados no naturalmente proporcionará un microorganismo que estructuralmente está 50 extremadamente cercano al microorganismo nativo, incapaz sin embargo de replicación, o capaz de una replicación solo limitada en un entorno natural. Los controles absolutos de la replicación del microorganismo se realizan utilizando la incorporación de aminoácidos específica de sitios, detallada a continuación, para controlar la expresión del gen esencial (función) utilizando la supresión del codón de terminación. Para los virus, un sitio de línea celular de producción en el hospedador incorpora específicamente un aminoácido no natural requerido para el desarrollo. Para 55 las bacterias, el genoma se diseña mediante ingeniería genética para incluir uno o más aminoácido(s) no naturales incorporados específicamente en el sitio.

El control de la función del gen esencial puede realizarse a niveles genéticos y funcionales estructurales. Por ejemplo, Tabla 2 Al nivel genético, la longitud completa del gen esencial se expresa en presencia del aminoácido incorporado no naturalmente. Al nivel funcional estructural, el gen esencial se diseña mediante ingeniería genética

para ser funcional solo en presencia del aminoácido no natural en un sitio específico. La incorporación de cualquier aminoácido natural en este sitio dará como resultado la abolición de su función nativa. La libertad de elección de los aminoácido(s) no naturales permite a un experto en la materia modular la inmunogenicidad de la vacuna deseada y la libertad de una replicación excesiva o sobrecontrolada, donde tradicionalmente ha sido necesario equilibrar entre la atenuación de la replicación y la producción, la incorporación específica de sitio de un aminoácido no natural proporciona las herramientas de desarrollo una vacuna con cero capacidades de replicación en ausencia del aminoácido no natural.

Las bacterias y virus de las vacunas pueden desarrollarse utilizando esta tecnología incluyen virus conocidos. Para los ejemplos no limitantes, esto incluye, los virus que afectan el tracto respiratorio superior tal como los rinovirus humanos (RVH), adenovirus, coxsackievirus, gripe, virus paragripal, virus respiratorio sincitial (VRS), Virus de Epstein-Barr (VEB) y citomegalovirus (CMV); virus que afectan el tracto gastrointestinal (GI) tal como rotavirus, agente de Norwalk, hepatitis A (VHA), poliovirus y otros picornavirus; virus transmitidos sexualmente incluyendo, aunque no de forma limitativa, virus de la inmunodeficiencia humana (VIH), virus del papiloma humano (VPH), virus del herpes simple (VHS) 1, VHS-2, VZV, CMV, VEB, HHV-6, HHV-7 y HHV-8; CMV, virus de la hepatitis B (VHB), virus de la hepatitis C (VHC). Los ejemplos no limitantes de bacterias incluyen bacterias gram positivas y bacterias gram negativas, incluyendo *Staphylococcus*, *Streptococcus*, *Mycobacterium* (por ejemplo, *Mycobacterium avium*, y *Mycobacterium avium* subespecie *paratuberculosis*), *Enterococcus*, *Corynebacterium*, *Borrelia*, *Bacillus*, *Chlamydia*, *Mycoplasma*, y similares.

Sistemas de traducción que son adecuados para la preparación de proteínas que incluyen uno o más aminoácidos seleccionados, por ejemplo, un aminoácido artificial, se describen en las solicitudes de patentes de Estados Unidos 10/126.931, titulada "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGONAL tRNA-AMINOACYL tRNA SYNTHETASE PAIRS" y 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". Además, véase el documento USSN 10/825.867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". Dichos sistemas de traducción comprenden generalmente células que incluyen un ARNt ortogonal (O-ARNt), un aminoacil ARNt sintetasa ortogonal (O-RS), y un aminoácido seleccionado, por ejemplo, un aminoácido artificial, en el que la O-RS aminoacila el O-ARNt ortogonal con el aminoácido seleccionado. Una pareja ortogonal desvelada en el presente documento está compuesta de un O-ARNt, por ejemplo, un ARNt supresor, un ARNt de cambio de marco, o similar, y una O-RS. El O-ARNt reconoce un primer codón selector y tiene una actividad de supresión en presencia de una sintetasa análoga en respuesta a un codón selector. La célula utiliza los componentes para incorporar el aminoácido seleccionado en una cadena polipeptídica en crecimiento. Por ejemplo, puede estar también presente un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, donde el polinucleótido comprende un codón selector que es reconocido por el O-ARNt. El sistema de traducción puede ser también un sistema in vitro. Las moléculas de ARNt desveladas en el presente documento son útiles en cualquier sistema de traducción, incluyendo los sistemas que utilizan ribosomas en la traducción.

El sistema de traducción puede ser también un sistema de traducción exento de células (in vitro). En estos sistemas, que pueden incluir tanto ARNm como un molde (traducción in vitro) o ADN como molde (combinado en una transcripción y traducción in vitro), la síntesis in vitro es dirigida por los ribosomas. Se ha aplicado un esfuerzo considerable al desarrollo de los sistemas de expresión de proteínas exentos de células. Véanse, por ejemplo, Kim, D.M. y J.R. Swartz, *Biotechnology and Bioengineering*, 74 :309-316 (2001); Kim, D.M. y J.R. Swartz, *Biotechnology Letters*, 22, 1537-1542, (2000); Kim, D.M., y J.R. Swartz, *Biotechnology Progress*, 16, 385-390, (2000); Kim, D.M., y J.R. Swartz, *Biotechnology and Bioengineering*, 66, 180-188, (1999); y Patnaik, R. y J.R. Swartz, *Biotechniques* 24, 862-868, (1998); patente de Estados Unidos n.º 6.337.191; publicación de patente de Estados Unidos n.º 2002/0081660; documento WO 00/55353; documento WO 90/05785. Otra enfoque que se puede aplicar incluye la técnica de fusión de péptidos mediante ARNm. Véanse, por ejemplo, R. Roberts y J. Szostak, *Proc. Natl Acad. Sci. (USA)* 94:12297-12302 (1997); A. Frankel, y col., *Chemistry & Biology* 10:1043-1050 (2003). En este enfoque, un molde de ARNm unido a puromicina se traduce en el péptido en el ribosoma. Si se han modificado una o más moléculas de ARNt, se pueden incorporar aminoácidos no naturales también en el péptido. después que se ha leído el último codón de ARNm, la puromicina captura el extremo C del péptido. Si se encuentra que el conjugado de ARNm-péptido resultante tiene propiedades interesantes en un ensayo invitro, su identidad puede desvelarse fácilmente a partir de la secuencia de ARNm. De este modo, se pueden cribar las bibliotecas de polipéptidos que comprenden uno o más aminoácidos codificados no naturalmente para identificar polipéptidos que tengan las propiedades deseadas. Más recientemente, se han notificado traducciones de ribosomas in vitro con componentes purificados que permiten la síntesis de péptidos sustituidos con aminoácidos codificados no naturalmente. Véanse, por ejemplo, A. Forster y col., *Proc. Natl Acad. Sci. (USA)* 100:6353 (2003).

Se desvela una célula (por ejemplo, una bacteria) en la que se ha incorporado un aminoácido no natural en un producto génico esencial. En la presente invención, se ha modificado genéticamente una célula para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, una bacteria se ha modificado genéticamente para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. Se desvela también un virus en el que un aminoácido no natural se ha incorporado en un producto génico esencial. En otras realizaciones de la descripción, un virus que se ha modificado genéticamente para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, una célula de *mycobacterium avium* subespecie *paratuberculosis* se ha modificado genéticamente para incluir un aminoácido natural en un producto génico que se

requiere para la replicación. En otras realizaciones de la presente invención, una célula de meningitis se ha modificado genéticamente para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula de rabia para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula de alga para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula vírica para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula bacteriana para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula fúngica para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. Se desvelan también células de *mycobacterium avium* subespecie paratuberculosis, meningitis, rabia, algas, víricas, bacterianas y fúngicas en las que un aminoácido no natural se ha incorporado en un producto génico esencial.

En otras realizaciones de la presente invención, se ha modificado genéticamente una célula de polio para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula de *E. coli* para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. En otras realizaciones de la presente invención, se ha modificado genéticamente una célula de *mycobacterium* para incluir un aminoácido no natural en un producto génico que se requiere para la replicación. Se describen también células de polio, *E. coli* and *mycobacterium* en las que un aminoácido no natural se ha incorporado en un producto génico esencial.

En determinadas realizaciones, Las células dependientes de aminoácidos no naturales modificadas genéticamente de la presente invención pueden utilizarse para producir una vacuna. Se desvela también que las células dependientes de aminoácidos no naturales modificadas genéticamente de la presente invención pueden utilizarse para producir anticuerpos que se pueden administrar como una vacuna. En determinadas realizaciones, Las células dependientes de aminoácidos no naturales modificadas genéticamente de la presente invención pueden utilizarse en una inoculación.

En determinadas realizaciones, una célula de *E. coli* que comprende el ARNt de la presente divulgación incluye dicho sistema de traducción. Por ejemplo, la célula *E. coli* de la presente invención incluye un ARNt ortogonal (O-ARNt), en el que el O-ARNt comprende una actividad de supresión en presencia de una sintetasa análoga en respuesta a un codón selector; una aminoacil ARNt sintetasa ortogonal (O-RS); un aminoácido seleccionado; y, un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, donde el polinucleótido comprende un codón selector que es reconocido por el O-ARNt.

Se pueden usar también parejas de O-ARNt/O-RS múltiples en una célula, que permiten la incorporación de más de un aminoácido seleccionado. En determinadas realizaciones, la célula puede incluir además una pareja de O-ARNt/O-RS diferente y un segundo aminoácido seleccionado, donde el O-ARNt reconoce un segundo codón selector en la O-RS aminoacila preferentemente el O-ARNt con el segundo aminoácido seleccionado. Por ejemplo, una célula puede comprender además, por ejemplo, una pareja de un ARN supresor ámbar-aminoacil ARNt sintetasa derivada de la tirosil ARNt sintetasa de *Methanococcus jannaschii*.

El O-ARNt y/o la O-RS pueden producirse naturalmente o pueden derivarse mediante mutación de un ARNt y/o una RS que se producen naturalmente, por ejemplo, que genera bibliotecas de ARNt y/o bibliotecas de RS, procedentes de una variedad de organismos. Por ejemplo, una estrategia de producir una pareja de ARNt/aminoacil ARNt sintetasa ortogonal implica importar una pareja de ARNt/sintetasa heteróloga a partir de, por ejemplo, una fuente diferente que la célula hospedadora, o fuentes múltiples, en la célula hospedadora. Las propiedades de la sintetasa heteróloga candidata incluyen, por ejemplo, que no carga ningún ARNt de la célula hospedadora, y las propiedades del ARNt heterólogo candidato incluyen, por ejemplo, que no está aminoacilado por ninguna sintetasa de la célula hospedadora. Además, el ARNt heterólogo es ortogonal para todas las sintetasas de la célula hospedadora.

Una segunda estrategia para generar una pareja ortogonal implica generar bibliotecas mutantes a partir de las cuales cribar y/o seleccionar un O-ARNt o una O-RS. se pueden combinar también estas estrategias.

En diversas realizaciones, el O-ARNt y la O-RS se derivan de al menos un organismos. En otra realización, el O-ARNt se deriva de un ARNt que se produce naturalmente o que ha mutado naturalmente a partir de un primer organismo y la O-RS se deriva de una RS que se produce naturalmente o que ha mutado naturalmente a partir de un segundo organismo. En una realización, el primer y el segundo organismos son diferentes. Por ejemplo, una pareja ortogonal puede incluir una ARNt sintetasa derivada de *Methanobacterium, thermoautotrophicum*, y un ARNt derivado de un ARNt de archaea (por ejemplo, de *Halobacterium sp. NRC-1*). Como alternativa, el primer y el segundo organismos son iguales. Véase la sección titulada "fuentes y organismos hospedadores" en el presente documento para más información.

En determinadas realizaciones de la presente invención, un O-ARNt de la presente divulgación comprende, o está codificado por una secuencia de polinucleótidos en la SEQ ID NO.: 1, 2, o 3, o una secuencia de polinucleótido complementaria de la misma, o una variación conservativa de la misma. Véase también la sección titulada

"Secuencia de ácido nucleico y polipeptídica y variantes" en el presente documento.

ARNt ortogonal (O-ARNt)

Un ARNt ortogonal (O-ARNt) media en la incorporación de un aminoácido seleccionado es una proteína que está codificada por un polinucleótido que comprende un codón selector que es reconocido por el O-ARNt, por ejemplo, in vivo o in vitro. Un O-ARNt de la presente divulgación puede estar aminoacilado con un aminoácido deseado mediante cualquier procedimiento o técnica, incluyendo, aunque no de forma limitativa, aminoacilación química o enzimática. El O-ARNt aminoacilado de la presente divulgación puede añadirse directamente a un sistema de traducción. Un O-ARNt de la presente divulgación puede estar aminoacilado por una RS con un aminoácido seleccionado in vitro o in vivo. Además, La RS puede ser una O-RS. puede proporcionarse una O-ARNt de la presente divulgación al sistema de traducción (por ejemplo, componentes de la traducción in vitro, o una célula) directamente, o proporcionando un polinucleótido que codifica un O-ARNt o una parte del mismo. Por ejemplo, un O-ARNt, o una parte del mismo, está codificado por una secuencia de polinucleótido que se muestra en la SEQ ID NO.: 1, 2, 3, o una secuencia de polinucleótido complementaria del mismo, o una variación conservativa del mismo.

Un O-ARNt de la presente divulgación comprende una actividad de supresión en presencia de una sintetasa análoga en respuesta a un codón selector. La actividad de supresión puede determinarse mediante cualquier número de ensayos conocidos en la técnica. Por ejemplo, se puede usar un ensayo indicador de la β -galactosidasa. Se usa un derivado de un plásmido que expresa el gen *lacZ* bajo el control del promotor, por ejemplo, donde la Leu-25 del péptido VVLQRRDWEN of *lacZ* está sustituida por un codón selector, por ejemplo, los codones TAG, TGA, AGGA, etc., o codones de sentido directo (como un control) para la tirosina, serina, leucina, etc. El plásmido de *lacZ* derivatizado se introdujo en las células de un organismo adecuado (por ejemplo, un organismo donde se pueden usar los componentes ortogonales) junto con un plásmido que comprende un O-ARNt de la presente divulgación. Se puede introducir también una sintetasa análoga (tanto como un polipéptido o como un polinucleótido que codifica la sintetasa análoga cuando se expresa). Las células se hacen crecer en medios hasta una densidad deseada, por ejemplo, hasta una DO_{600} de aproximadamente 0,5, y se llevaron a cabo los ensayos de la β -galactosidasa, por ejemplo, utilizando el Kit de ensayo de la β -Galactosidasa BetaFluor™ (Novagen). Se calculó el porcentaje de supresión como el porcentaje de actividad de una muestra con respecto a un control comparable, por ejemplo, el valor observado a partir de la construcción de *lacZ* derivatizada, donde la construcción tiene un codón de sentido directo correspondiente en la posición deseada más bien que un codón selector.

En la molécula de ARNt, Timina (T) se sustituyó con uracilo (U). Además, pueden estar presentes más modificaciones en las bases. Se desvelan también variaciones conservativas del O-ARNt. Por ejemplo, las variaciones conservativas del O-ARNt incluyen aquellas moléculas que funcionan de forma análoga al O-ARNt y mantienen la estructura con forma de L del ARNt, pero que no tienen la misma secuencia (y son diferentes que las moléculas de ARNt naturales). Véase también la sección en el presente documento titulada "Secuencia de ácido nucleico y polipeptídica y variantes".

La composición que comprende un O-ARNt puede incluir además una aminoacil ARNt sintetasa ortogonal (O-RS), donde la O-RS aminoacila preferentemente el O-ARNt con un aminoácido seleccionado (por ejemplo, un aminoácido artificial). En determinadas realizaciones, una composición que incluye un O-ARNt puede incluir además un sistema de traducción (por ejemplo, un sistema de traducción in vitro o un sistema de traducción in vivo). Un ácido nucleico que comprende un polinucleótido que codifica un polipéptido de interés, en el que el polinucleótido comprende uno o más codones selectores reconocidos por el O-ARNt, o una combinación de uno o más de estos, puede también estar presente en la célula u otros sistemas de traducción. Véase también, la sección del presente documento titulada "Aminoacil ARNt sintetasa ortogonales (O-RS)".

Procedimientos para producir un ARNt ortogonal (O-ARNt), por ejemplo, un O-ARNt, se desvelan también en el presente documento. Un ARNt, por ejemplo, un O-ARNt, producido mediante el procedimiento es también una característica de la presente divulgación.

Los procedimientos para producir un ARNt ortogonal incluyen mutar el bucle anticodón de cada uno de un combinado de ARNt para permitir el reconocimiento de un codón selector (por ejemplo, un codón ámbar, un codón ópalo, un codón de cuatro bases, etc.), proporcionando por tanto una pluralidad de O-ARNt potenciales; y analizar la estructura secundaria de un miembro de la pluralidad de O-ARNt potenciales para identificar pares de bases no convencionales en la estructura secundaria, y, opcionalmente, mutar los pares de bases no convencionales (por ejemplo, los pares de bases no convencionales se hacen mutar a pares de bases convencionales). Los pares de bases no convencionales pueden localizarse en la región del tallo de la estructura secundaria. Un O-ARNt puede poseer una mejora de una o más características o actividades, tales como una mejora en la ortogonalidad para un organismo deseado en comparación con el material de partida, por ejemplo, la pluralidad de las secuencias de ARNt, preservando a la vez su afinidad hacia una RS deseada.

Como alternativa, Se pueden desarrollar los O-ARNt mutando un ARNt conocido para modular su interacción o la afinidad de unión con una o más moléculas que influyen la traducción o son componentes de la maquinaria de la traducción. Dichos componentes incluyen, aunque no de forma limitativa, factores de alargamiento. El factor EF-Tu de alargamiento bacteriano juega un papel clave en la etapa de alargamiento en la síntesis de proteínas. Tras la

aminoacilación de la ARNt por la ARNt sintetasa, EF-Tu se une al ARNt aminoacilado y lleva este al sitio A del ribosoma. El enlace éster entre el aminoácido cargado y el ARNt está protegido de la hidrólisis espontánea debida a la unión entre EF-Tu y el ARNt aminoacilado. Stortchevoi y col. investigaron los mutantes del par de bases inestables U50:G64 del tRNA^{Met} de inicio de *E. coli* en el tallo TΨC, debido a que se encontró que este par de bases era un determinante negativo secundario que bloqueaba la actividad del ARNt en el alargamiento, debido presumiblemente a una interacción debilitada entre el EF-Tu.GTP y el ARNt aminoacilado (JBC 2003 278(20):17672-17679). Asimismo, LaRiviere y col. describieron en Science del 5 de octubre de 2001;294(5540):165-8 las contribuciones termodinámicas del aminoácido y el cuerpo del ARNt para la afinidad de unión global con EF-Tu. Ellos indicaron que las contribuciones del cuerpo del ARNt y el aminoácido son independientes entre sí y que se compensan entre sí cuando los ARNt están correctamente acilados. Las alteraciones de la interacción entre EF-Tu.GTP y el ARNt aminoacilado con el aminoácido artificial puede afectar a la eficacia de la carga del ARNt en el sitio A del ribosoma. Se pueden encontrar los sitios de mutaciones potenciales pueden también encontrarse analizando las estructuras cristalinas de los complejos entre el ARNt y otros componentes de la maquinaria de traducción tales como EF-Tu. Por ejemplo, Nissen y col. han indicado que EF-Tu.GTP se une directamente a la estructura principal del fosfato del tallo TΨC del fenilalanil-ARN de transferencia de levadura (Phe-ARNt) (Science 1995 270(5241):1464-1472).

Los procedimientos incluyen opcionalmente analizar la homología de las secuencias de los ARNt y/o las aminoacil ARNt sintetasa para determinar los candidatos potenciales para un O-ARNt, O-RS y/o las parejas de los mismos, que parecen ser ortogonales para un organismo específico. Se pueden usar los programas informáticos conocidos en la técnica y descritos en el presente documento para el análisis. En un ejemplo, para seleccionar los componentes de la traducción ortogonales potenciales para uso en un organismo procarionta, se seleccionan una sintetasa y/o un ARNt que no presentan una homología inusual para los organismos procariontas.

Se puede producir también un combinado de ARNt mediante una estrategia de consenso. Por ejemplo, el combinado de ARNt se produce alineando una pluralidad de secuencias de ARNt; determinando una secuencia consenso; y generando una biblioteca de ARNt que utiliza al menos una parte, la mayoría, o la secuencia consenso completa. Por ejemplo, se puede compilar una secuencia consenso con un programa informático, por ejemplo, el programa GCG *pileup*. Opcionalmente, se cambiaron las posiciones degeneradas determinadas por el programa para la base más frecuente en aquellas posiciones. Se sintetizó una biblioteca mediante las técnicas conocidas en la materia utilizando la secuencia consenso. Por ejemplo, se puede usar la extensión del solapamiento de los oligonucleótidos en que cada sitio del gen del ARNt se puede sintetizar como una mezcla dopada del 90% de la secuencia consenso y un 10% de una mezcla de las otras 3 bases para proporcionar la biblioteca basada en la secuencia consenso. Se pueden usar también otras mezclas, por ejemplo, 75% de la secuencia consenso y 25% de una mezcla de las otras 3 bases, 80% de la secuencia consenso y 20% de una mezcla de las otras 3 bases, 95% de la secuencia consenso y 5% de una mezcla de las otras 3 bases, etc.

Se pueden generar bibliotecas de los ARNt mutantes utilizando diversas técnicas de mutagénesis conocidas en la materia. Por ejemplo, Se pueden generar ARNt mutantes mediante mutaciones específicas de sitios, mutaciones puntuales aleatorias, recombinación homóloga, transposición de secuencias de ADN u otros procedimientos recurrentes de mutagénesis, la construcción quimérica de cualquier combinación de los mismos.

Se pueden introducir mutaciones adicionales en posición(ones) específicas, por ejemplo, en una(s) posición(ones) no conservativa(s), o en una(s) posición(ones) conservativa(s), en una(s) posición(ones) aleatorizada(s), o una combinación de las mismas en un bucle o región deseado de un ARNt, por ejemplo, un bucle anticodón, el tallo aceptor, el brazo o bucle D, el bucle variable, el brazo o bucle TΨC, otras regiones de la molécula de ARNt, o una combinación de las mismas. Las mutaciones pueden incluir pares de bases emparejados en la región del tallo.

Normalmente, se obtiene un O-ARNt sometiendo a selección negativa una población de células de una primera especie, donde las células comprenden un miembro de la pluralidad de O-ARNt potenciales. la selección negativa elimina las células que comprenden un miembro de la pluralidad de O-ARNt potenciales que está aminoacilado por una aminoacil ARNt sintetasa (RS) que es endógena para las células. Esto proporciona un combinado de ARNt que son ortogonales para la célula de la primera especie.

En determinadas realizaciones de la selección negativa, un(os) codón(ones) selector(es) se introduce(n) en un polinucleótido que codifica un marcador de selección negativo, por ejemplo, una enzima que confiere resistencia a antibióticos, por ejemplo, β-lactamasa, una enzima que confiere un producto detectable, por ejemplo, β-galactosidasa, cloranfenicol acetiltransferasa (CAT), por ejemplo, un producto tóxico, tal como barnasa, en una posición no esencial, etc. Se puede llevar a cabo la criba/selección haciendo crecer la población de células en presencia de un agente de selección (por ejemplo, un antibiótico, tal como ampicilina). En una realización, se hace variar la concentración del agente de selección.

Por ejemplo, para medir la actividad de los ARNt supresores, se utiliza un sistema de selección que está basado en la supresión *in vivo* del codón selector, por ejemplo, mutaciones sin sentido o de cambio de marco introducidas en un polinucleótido que codifica un marcador de selección negativo, por ejemplo, un gen de la β-lactamasa (*bla*). Por ejemplo, variantes de polinucleótidos, por ejemplo, variantes *bla*, con, por ejemplo, se construyen codones TAG, AGGA, y TGA, en la posición de una determinada posición. Se transforman células, por ejemplo, bacterias, con

estos polinucleótidos. En el caso de un ARN ortogonal, que no puede cargarse eficazmente mediante *E. coli* sintetasas endógenas, resistencia a antibióticos, por ejemplo, resistencia a la ampicilina, deben ser aproximadamente o menores que eso para una bacteria transformada sin plásmidos. Si el ARNt no es ortogonal, o si una sintetasa heteróloga capaz de cargar el ARNt se expresa simultáneamente en el sistema, se va a observar un mayor nivel de resistencia al antibiótico, por ejemplo, ampicilina. Se seleccionan células, por ejemplo, bacterias, que son incapaces de crecer en placas de agar LB con concentraciones de antibiótico aproximadamente iguales a las de las células transformadas sin plásmidos.

en el caso de un producto tóxico (por ejemplo, la ribonucleasa barnasa), cuando un miembro de la pluralidad de ARNt potenciales se aminoacila por el hospedador endógeno, por ejemplo, sintetasas de *Escherichia coli* (es decir, no es ortogonal para el hospedador, por ejemplo, sintetasas de *Escherichia coli*), el codón selector se suprime y el producto del polinucleótido tóxico producido conduce a la muerte de la célula. Las células que albergan ARNt ortogonal O ARNt no funcional sobreviven.

En una realización, el combinado de ARNt que son ortogonales para un organismo deseado se somete a continuación a una selección positiva en la que un codón selector se coloca en un marcador de selección positiva, por ejemplo, codificado mediante un gen de resistencia a fármacos, tal como un gen de la β -lactamasa. se lleva a cabo la selección positiva en una célula que comprende un polinucleótidos que codifica o comprende un miembro del combinado de ARNt, un polinucleótido que codifica un marcador de selección positiva, y un polinucleótido que codifica una RS análoga. Estos polinucleótidos se expresan en la célula y la célula se hace crecer en presencia de un agente de selección, por ejemplo, ampicilina. A continuación se seleccionan los ARNt para su capacidad de aminoacilarse mediante la sintetasa análoga expresada simultáneamente y para insertar un aminoácido en respuesta a este codón selector. Normalmente, estas células muestran una potenciación en la eficacia de supresión en comparación con las células que albergan ARNt no funcionales, o ARNt que no pueden ser reconocidos eficazmente por la sintetasa de interés. La célula que alberga los ARNt no funcionales o los ARNt que no son reconocidos eficazmente por la sintetasa de interés son sensibles al antibiótico. Por lo tanto, los ARNt que: (i) no son sustratos para el hospedador endógeno, por ejemplo, las sintetasas de *Escherichia coli*; (ii) pueden aminoacilarse mediante la sintetasa de interés; y (iii) son en la traducción, sobreviven a ambas selecciones.

El rigor de la selección, por ejemplo, la selección positiva, la selección negativa o la selección positiva y negativa, en los procedimientos anteriormente descritos, pueden, opcionalmente, variarse. Por ejemplo, debido a que la barnasa es una proteína extremadamente tóxica, el rigor de la selección negativa puede controlarse introduciendo diferentes números de codones selectores en el gen de la barnasa y/o utilizando un promotor inducible. En otro ejemplo, se hace variar la concentración de la selección del agente de cribado (por ejemplo, ampicilina). En un aspecto, se hace variar el rigor debido a que la actividad deseada puede ser baja durante los ciclos iniciales. De esta manera, se aplican criterios de selección menos rigurosos en los ciclos iniciales y se aplican criterios más rigurosos en los ciclos finales de la selección. En determinadas realizaciones, la selección negativa, la selección positiva, o la selección negativa y la selección positiva pueden repetirse múltiples veces. se pueden usar múltiples marcadores de selección negativa diferentes, marcadores de selección positiva o marcadores de selección negativa y positiva. En determinadas realizaciones, el marcador de selección positiva y negativa puede ser el mismo.

Se pueden utilizar otros tipos de selecciones/cribados en la descripción para producir componentes de traducción ortogonales, por ejemplo, un O-ARNt, una O-RS y una pareja de O-ARNt/O-RS. Por ejemplo, el marcador de selección negativa, el marcador de selección positiva o los marcadores de selección positiva y negativa pueden incluir un marcador que fluoresce o cataliza una reacción luminiscente en presencia de un reactivo adecuado. En otra realización, se detecta un producto del marcador mediante la clasificación celular activada por fluorescencia (FACS) o mediante luminiscencia. Opcionalmente, el marcador incluye un marcador de cribado basado en la afinidad. Véase, Francisco, J. A., y col., (1993) Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A. 90:10444-8.

Se pueden encontrar procedimientos adicionales para producir un ARNt ortogonal recombinante, por ejemplo, en las solicitudes de patentes de Estados Unidos 10/126.931, titulada "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs" y 10/126.127, titulada "In vivo Incorporation of Unnatural Amino Acids", y en el documento USSN 10/825.867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". Véanse también, Forster y col., (2003) Programming peptidomimetic synthetases by translating genetic codes designed de novo. PNAS 100(11):6353-6357; y, Feng y col., (2003), Expanding tRNA recognition of a tRNA synthetase by a single amino acid change, PNAS 100(10): 5676-5681.

Un ARNt de la presente divulgación puede estar aminoacilado con un aminoácido deseado mediante cualquier procedimiento o técnica, incluyendo, aunque no de forma limitativa, aminoacilación química o enzimática.

La aminoacilación puede llevarse a cabo por aminoacil ARNt sintetasas o por otras moléculas enzimáticas, incluyendo, aunque no de forma limitativa, ribozimas. El término "ribozima" es intercambiable con "ARN catalítico" Cech y colaboradores (Cech, 1987, Science, 236:1532-1539; McCorkle y col., 1987, Concepts Biochem. 64:221-226) demostraron la presencia de ARN que se producen naturalmente que pueden actuar como catalizadores (ribozimas). Sin embargo, aunque estos catalizadores de ARN natural han mostrado actuar únicamente sobre sustratos de ácido ribonucleico para la escisión y el corte y empalme, el desarrollo reciente de la evolución artificial de ribozimas ha

expandido el repertorio de la catálisis a diversas reacciones químicas. Los estudios han identificado moléculas de ARN que pueden catalizar enlaces de aminoacil-ARN en sus propios términos (2')3' (Illangakekare y col., 1995 Science 267:643-647), y una molécula de ARN que puede transferir un aminoácido desde una molécula de ARN a otra (Lohse y col., 1996, Nature 381:442-444).

- 5 La publicación de solicitud de patente de Estados Unidos n.º 2003/0228593, describe procedimientos para construir ribozimas y su uso en la aminoacilación de los ARNt con aminoácidos codificados naturalmente y aminoácidos no codificados naturalmente. Las formas inmovilizadas de sustratos de las moléculas enzimáticas que pueden aminoacilar ARNt, incluyendo, aunque no de forma limitativa, ribozimas, pueden permitir una purificación por afinidad eficaz de los productos aminoacilados. Los ejemplos de sustratos adecuados incluyen agarosa, sefarosa, y perlas magnéticas. se describe la producción y uso de una forma inmovilizada de sustrato de ribozima para aminoacilación en Chemistry and Biology 2003, 10:1077-1084 y en la publicación de solicitud de patente de Estados Unidos 2003/0228593.

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

25 Los procedimientos biosintéticos que emplean aminoacil ARNt modificado químicamente se han utilizado para incorporar algunas sondas biofísicas en proteínas sintetizadas in vitro. Véanse las siguientes publicaciones y las referencias citadas en las mismas: Brunner, J. New Photolabeling and crosslinking methods, Annu. Rev Biochem, 62:483-514 (1993); y, Krieg, U.C., Walter, P., Hohnson, A.E. Photocrosslinking of the signal sequence of nascent preprolactin of the 54-kilodalton polypeptide of the signal recognition particle, Proc. Natl. Acad. Sci, 83(22):8604-8608 (1986).

30 Anteriormente, Se ha mostrado que los aminoácidos artificiales se pueden incorporar específicamente en el sitio en proteínas in vitro mediante la adición de ARNt supresores aminoacilados químicamente a las reacciones de síntesis de proteínas programadas con un gen que contiene una mutación ámbar sin sentido deseada. Utilizando estas soluciones, Se pueden sustituir numerosos de los veinte aminoácidos comunes con homólogos estructurales cercanos, por ejemplo, fluorofenilalanina para fenilalanina, utilizando cepas auxotróficas para un aminoácido concreto. Véanse, por ejemplo, Noren, C.J., Anthony-Cahill, Griffith, M.C., Schultz, P.G. A general method for site-specific incorporation of unnatural amino acids into proteins, Science, 244: 182-188 (1989); M.W. Nowak, y col., Science 268:439-42 (1995); Bain, J.D., Glabe, C.G., Dix, T.A., Chamberlin, A.R., Diala, E.S. Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, J. Am Chem Soc, 111:8013-8014 (1989); N. Budisa y col., FASEB J. 13:41-51 (1999); Ellman, J.A., Mendel, D., Anthony-Cahill, S., Noren, C.J., Schultz, P.G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins, Methods in Enz., vol. 202, 301-336 (1992); y, Mendel, D., Cornish, V.W. & Schultz, P.G. Site-Directed Mutagenesis with an Expanded Genetic Code, Annu Rev Biophys. Biomol Struct. 24, 435-62 (1995).

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

55 Los procedimientos para generar ARN catalítico pueden implicar combinados separados de secuencias de ribozimas aleatorizadas, llevando a cabo la evolución dirigida en los combinados, el cribado de los combinados para la actividad de aminoacilación deseable, y la selección de secuencias de aquellas ribozimas que presentan la actividad de aminoacilación deseada.

60 Las ribozimas pueden comprender motivos y/o regiones que facilitan la actividad de acilación, tales como un motivo CGU y una región rica en U. Por ejemplo, se ha notificado que las regiones ricas en U pueden facilitar el

reconocimiento de un sustrato de aminoácido, y un motivo CDU puede formar pares de bases con los términos 3' de un ARNt. En combinación, el CGU y el motivo y la región rica en U facilitan el reconocimiento simultáneo del aminoácido y el ARNt simultáneamente, y facilitan por tanto la aminoacilación del término 3' del ARNt.

5 Se pueden generar ribozimas mediante selección *in vitro* utilizando un r24mini parcialmente aleatorizado conjugado con ARN^{Asn}_{CCCG}, seguido por un diseño mediante ingeniería genética sistemático de una secuencia consenso que se encuentra en los clones activos. Una ribozima ilustrativa obtenida mediante este procedimiento se denomina "ribozima Fx3" y se describe en la solicitud de publicación de Estados Unidos n.º 2003/0228593, actúa como un catalizador versátil para la síntesis de los diversos aminoacil-ARNt cargados con aminoácidos análogos no naturales.

10 Se puede utilizar la inmovilización sobre un sustrato para permitir una purificación por afinidad eficaz de los ARNt aminoacilados. Los ejemplos de sustratos adecuados incluyen, aunque no de forma limitativa, agarosa, sefarosa, y perlas magnéticas. Se pueden inmovilizar ribozimas sobre resinas aprovechando la estructura química del ARN, de tal manera que el 3'-cis-diol en la ribosa del ARN se puede oxidar con peryodato para dar como resultado el correspondiente dialdehído, para facilitar la inmovilización del ARN sobre la resina. Se pueden usar diversos tipos de
15 resinas incluyendo las resinas de hidrazida baratas en las que la aminación reductora hace de la interacción entre la resina y la ribozima un enlace irreversible. La síntesis de los aminoacil ARNt puede estar significativamente facilitada por esta técnica de aminoacilación en columna. Kourouklis y col. Methods 2005; 36:239-4 describen un sistema de aminoacilación basado en una columna.

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

Los ARNt aminoacilados de la presente divulgación pueden añadirse a las reacciones de traducción a fin de incorporar el aminoácido con el cual el ARNt se aminoaciló en una posición de elección en un polipéptido preparado mediante la reacción de traducción. Los ejemplos de sistemas de traducción en los que se pueden usar los ARNt aminoacilados de la presente divulgación incluyen, aunque no de forma limitativa, lisados celulares. Los lisados celulares proporcionan los componentes de reacción necesarios para la traducción *in vitro* de un polipéptido a partir de un ARNm de entrada. Los ejemplos de dichos componentes de reacción incluyen, aunque no de forma limitativa, proteínas ribosómicas, ARNr, aminoácidos, ARNt, GTP, ATP, el inicio de la traducción y los factores de alargamiento y factores adicionales asociados con la traducción. Además, los sistemas de traducción pueden ser traducciones discontinuas o traducciones compartimentalizadas. Los sistemas de traducción discontinuos combinan componentes de reacción en un único compartimento mientras que los sistemas de traducción compartimentalizados separan los componentes de la reacción de traducción procedentes de los productos de reacción que pueden inhibir la eficacia de la traducción. Dichos sistemas de traducción están disponibles comercialmente.

35 Además, se puede usar un sistema de transcripción/traducción acoplado. Los sistemas de transcripción/traducción acoplados permiten para ambos la transcripción de un ADN de entrada en un ARNm correspondiente, que a la vez se traduce por los componentes de reacción. Un ejemplo de un sistema de transcripción/traducción acoplado comercialmente disponible es el Sistema de Traducción Rapid (RTS, Roche Inc.). El sistema incluye una mezcla que contiene lisados de *E. coli* para proporcionar componentes de la traducción tales como ribosomas y factores de traducción. Además, se incluye una ARN polimerasa para la transcripción del ADN de entrada en un molde de ARNm para uso en la traducción. El RTS puede utilizar la compartimentalización de los componentes de reacción por medio de una membrana intercalada entre los compartimentos de reacción, incluyendo un compartimento de residuos de suministros y un compartimento de transcripción/traducción.

45 La aminoacilación del ARNt puede llevarse a cabo por otros agentes, incluyendo, aunque no de forma limitativa, transferasas, polimerasas, anticuerpos catalíticos, proteínas multifuncionales, y similares.

Aminoacil ARNt sintetisas ortogonales (O-RS)

Una O-RS aminoacila preferentemente un O-ARNt de la presente divulgación con un aminoácido seleccionado *in vitro* o *in vivo*. Una O-RS puede proporcionarse al sistema de traducción (por ejemplo, componentes de la traducción *in vitro*, o una célula) mediante un polinucleótido que incluye una O-RS y/o mediante un polinucleótido que codifica una O-RS o una parte de la misma. Una O-RS, o una parte de la misma, está codificada por una secuencia de polinucleótidos o una secuencia de polinucleótidos complementaria de la misma, o una variación conservativa de la misma. Un O-ARNt de la presente divulgación puede estar aminoacilado por numerosas moléculas de O-RS diferentes, incluyendo, aunque no de forma limitativa, las desveladas en el presente documento.

55 Se desvelan también los procedimientos para identificar una aminoacil ARNt sintetasa ortogonal (O-RS); por ejemplo, una O-RS, para uso con un O-ARNt, por ejemplo, un O-ARNt. Se obtiene un O-ARNt. Por ejemplo, un procedimiento incluye someter a selección positiva una población de células de una primera especie, donde las células pueden comprender: 1) un miembro de una pluralidad de aminoacil ARNt sintetisas (RS), donde la pluralidad de RS comprenden los RS mutantes, Las RS derivadas de especies diferentes de la primera especie o ambos

mutantes de las RS y las RS derivadas de una especie diferente de la primera especie; 2) el ARNt ortogonal (O-ARNt) de una segunda especie; y 3) un polinucleótido que codifica un marcador de selección positiva y comprende al menos un codón selector. Las células se seleccionan o criban para aquellas que muestran una potenciación en la eficacia de la supresión en comparación con las células que carecen o tienen una cantidad reducida del miembro de la pluralidad de las RS. Las células que tienen una potenciación en la eficacia de la supresión comprenden una RS activa que aminoacila el O-ARNt. Un nivel de aminoacilación (in vitro o in vivo) por la RS activa de un primer conjunto de ARNt procedente de la primera especie se compara con el nivel de aminoacilación (in vitro o in vivo) por la RS activa de un segundo conjunto de ARNt procedente de la segunda especie. El nivel de aminoacilación puede determinarse por una sustancia detectable (por ejemplo, un aminoácido marcado o un aminoácido artificial). Se selecciona la RS activa que aminoacila más eficazmente el segundo conjunto de ARNt en comparación con el primer conjunto de ARNt, proporcionando por tanto la aminoacil ARNt sintetasa ortogonal para uso con el O-ARNt.

Se puede usar cualquiera de numerosos ensayos para determinar la aminoacilación. Estos ensayos se pueden llevar a cabo in vitro o in vivo. Por ejemplo, los ensayos de aminoacilación in vitro se describen en, por ejemplo, Hoben, P., y Soll, D. (1985) *Methods Enzymol.* 113:55-59 y en la solicitud de patente de Estados Unidos n.º 2003/0228593. Se puede determinar también la aminoacilación utilizando un indicador junto con componentes de traducción ortogonales y detectando el indicador en una célula que expresa un polinucleótido que comprende al menos un codón selector que codifica una proteína. Véanse también, la solicitud de patente de Estados Unidos 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". y, el documento USSN 10/825.867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE".

Una O-RS identificada se puede manipular adicionalmente para alterar la especificidad del sustrato de la sintetasa de tal manera que únicamente un aminoácido artificial deseado, pero no ninguno de los 20 aminoácidos comunes, se carga en el O-ARNt. Los procedimientos para generar unas aminoacil ARNt sintetasas ortogonales con una especificidad por el sustrato por un aminoácido artificial incluyen mutar la sintetasa, por ejemplo, en el sitio activo en la sintetasa, y el sitio del mecanismo de edición en la sintetasa, en sitios diferentes combinando diferentes dominios de sintetasas, o de forma similar, y aplicando un proceso de selección. Se usa una estrategia que está basada en la combinación de una selección positiva seguida por una selección negativa. En la selección positiva, la supresión del codón selector introducido es una(s) posición(ones) no esencial(es) de un marcador positivo permite a las células sobrevivir bajo presión de selección positiva. en presencia de aminoácidos naturales y artificiales, los sobrevivientes codifican por tanto las sintetasas activas cargando el ARNt supresor ortogonal tanto con un aminoácido natural como con un aminoácido artificial. En la selección negativa, la supresión de un codón selector introducido en posición(ones) no esenciales de un marcador negativo elimina las sintetasas con las especificidades de aminoácidos naturales. Los sobrevivientes de la selección negativa y la selección positiva codifican las sintetasas que aminoacilan (cargan) el ARNt supresor ortogonal únicamente con aminoácidos artificiales. Estas sintetasas pueden a continuación someterse a mutagénesis adicional, por ejemplo, transposición de secuencias de ADN u otros procedimientos recurrentes de mutagénesis.

Se pueden generar bibliotecas de las O-RS utilizando diversas técnicas de mutagénesis conocidas en la materia. Por ejemplo, Se pueden generar RS mutantes mediante mutaciones específicas de sitios, mutaciones puntuales aleatorias, recombinación homóloga, transposición de secuencias de ADN u otros procedimientos recurrentes de mutagénesis, la construcción química de cualquier combinación de los mismos. Por ejemplo, se puede producir una biblioteca de RS mutantes de dos o más diferentes, por ejemplo, "sub-bibliotecas" más pequeñas, menos diversas. Debe señalarse que las bibliotecas de ARNt sintetasas procedentes de diversos organismos (por ejemplo, microorganismos tales como eubacteria o archaeobacteria) tales como bibliotecas que comprenden una diversidad natural (véanse, por ejemplo, la patente de Estados Unidos n.º 6.238.884 de Short y col; la patente de Estados Unidos n.º 5.756.316 de Schallenberger y col; la patente de Estados Unidos n.º 5.783.431 de Petersen y col; la patente de Estados Unidos n.º 5.824.485 de Thompson y col; la patente de Estados Unidos n.º 5.958.672 de Short y col), se construyen y criban opcionalmente para parejas ortogonales.

Una vez que las sintetasas se someten a la estrategia de selección/cribado positiva y negativa, estas sintetasas pueden a continuación someterse a mutagénesis adicional. Por ejemplo, se puede aislar un ácido nucleico que codifica la O-RS; se puede generar un conjunto de polinucleótidos que codifican las O-RS mutadas (por ejemplo, mediante mutagénesis aleatoria, mutagénesis específica de 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 que se obtiene una O-RS mutada que aminoacila preferentemente el O-ARNt que el aminoácido artificial. En un aspecto de la presente divulgación, las etapas se llevan a cabo múltiples veces, por ejemplo, al menos dos veces.

Se pueden usar también niveles adicionales de rigor en la selección/cribado en los procedimientos de la presente divulgación, para producir O-ARNt, O-RS o las parejas de los mismos. Se puede variar el rigor de la selección o el cribado en una o ambas etapas del procedimiento para producir una O-RS. Esto podría incluir, por ejemplo, variar la cantidad de agente de selección/cribado que se usa, etc. Se pueden llevar a cabo también ciclos adicionales de selecciones positivas y/o negativas. La selección o el cribado pueden comprender también una o más selecciones o cribados positivos o negativos que incluyen, por ejemplo, un cambio en la permeabilidad de los aminoácidos, un cambio en la eficacia de la traducción, un cambio en la fidelidad de la traducción, etc. Normalmente, el uno o más cambios se basa en una mutación en uno o más genes en un organismo en el que una pareja de ARNt-ARNt sintetasa ortogonal se usa para producir una proteína.

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

Un polinucleótido que codifica el marcador de selección positiva puede estar unido operativamente a un elemento de respuesta. Puede estar presente también 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. La incorporación del aminoácido artificial en la proteína moduladora de la transcripción por el O-ARNt aminoacilado con el aminoácido artificial da como resultado la transcripción del polinucleótido (por ejemplo, gen indicador) que codifica el marcador de selección positiva. Opcionalmente, el codón selector se localiza en o sustancialmente próximo a una parte 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 puede estar también unido a un elemento de respuesta a partir del cual la transcripción está mediada por la proteína moduladora de la transcripción. Véanse, por ejemplo, A. J. DeMaggio, y col., (2000), The yeast split-hybrid system, *Method Enzymol.* 328:128-137; H. M. Shih, y col., (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, y col., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system, *Proc. Natl. Acad. Sci. U. S. A.* 93:10321-10326; y, M. Vidal, y col., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions (*Proc. Natl. Acad. Sci. U. S. A.* 93:10315-10320). La incorporación de un aminoácido natural en la proteína moduladora de la transcripción por el O-ARNt aminoacilado con un aminoácido natural da como resultado la transcripción del marcador de selección negativa. Opcionalmente, el marcador de selección negativa comprende un codón selector. El marcador de selección positiva y/o el marcador de selección negativa de la descripción puede comprender al menos dos codones selectores, de los cuales, cada uno o ambos pueden comprender 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 (de forma directa o indirecta) a una secuencia de ácido nucleico (por ejemplo, un elemento de respuesta) y modula la transcripción de una secuencia que está unida operativamente al elemento de respuesta. Una proteína moduladora de la transcripción puede ser una proteína activadora de la transcripción (por ejemplo, GAL4, receptores de hormonas nucleares, AP1, CREB, miembros de la familia LEF/tcf, SMAD, VP16, SP1, etc.), una proteína represora de la transcripción (por ejemplo, receptores de hormonas nucleares, la familia Groucho/tle, la 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 que está reconocida por la proteína moduladora de la transcripción o un agente adicional que actúa de forma concertada con la proteína moduladora de la transcripción.

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

El medio utilizado para la selección negativa puede comprender un agente de selección que se convierte en una sustancia detectable por el marcador de selección negativa. La sustancia detectable, como se desvela en el presente documento, puede ser 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 los sitios de unión del ADN a GAL4. Cuando se produce el marcador de selección negativa, por ejemplo, mediante la traducción de un polinucleótido que codifica el GAL4 con codones selectores, GAL4 activa la transcripción de URA3. La selección negativa se lleva a cabo en un medio que comprende ácido 5-fluoroorótico (5-FOA), que se convierte en una sustancia detectable (por ejemplo, una sustancia tóxica que destruye la célula) por el producto génico del gen *ura3*. Véanse, por ejemplo, J. D. Boeke, y col., (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, y col., (1996), Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system., *Proc. Natl. Acad. Sci. U. S. A.* 93:10321-10326; y, M. Vidal, y col., (1996), Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions., *Proc. Natl. Acad. Sci. U. S. A.* 93:10315-10320.

Como con el marcador de selección positiva, el marcador de selección negativa puede ser también cualquiera de una variedad de moléculas. El marcador de selección positiva y/o el marcador de selección negativa puede ser un polipéptido que fluoresce o cataliza una reacción luminiscente en presencia de un reactivo adecuado. Por ejemplo, los marcadores de selección negativa incluyen, aunque no de forma limitativa, por ejemplo, luciferasa, proteína fluorescente verde (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. El marcador de selección positiva y/o el marcador de selección negativa puede detectarse mediante clasificación celular activada por fluorescencia (FACS) o mediante luminiscencia. El marcador de selección positiva y/o el marcador de selección negativa puede comprender un marcador de cribado basado en la afinidad. El mismo polinucleótido puede codificar el marcador de selección positiva y el marcador de selección negativa. Por ejemplo, la etapa de selección positiva, la etapa de selección negativa o las etapas de selección positiva y selección negativa y puede incluir utilizar un indicador, en el que el indicador se detecta mediante clasificación celular activada por fluorescencia (FACS).se detecta un producto del marcador mediante la clasificación celular activada por fluorescencia (FACS) o mediante luminiscencia. Por ejemplo, se puede llevar a cabo una selección positiva en primer lugar con un marcador de selección positiva, por ejemplo, el gen de la cloranfenicol acetiltransferasa (CAT), donde el gen CAT comprende un codón selector, por ejemplo, un codón de terminación ámbar, en el gen CAT, que está seguido por un cribado de selección negativa, que se basa en la incapacidad de suprimir un(os) codón(ones) selector(es), por ejemplo, dos o más, en las posiciones en un marcador negativo, por ejemplo, el gen de la T7 ARN polimerasa. El marcador de selección positiva y el marcador de selección negativa pueden encontrarse en el mismo vector, por ejemplo, un plásmido. La expresión del marcador negativo impulsa la expresión del indicador, por ejemplo, la proteína fluorescente verde (GFP).proteína fluorescente verde (GFP). Pueden variarse el rigor de la selección y el cribado, por ejemplo, se puede variar la intensidad de la luz necesaria para la fluorescencia del indicador. Se puede llevar a cabo una selección positiva con un indicador como un marcador de selección positiva, que se criba mediante FACS, seguido por un cribado de selección negativa, que se basa en la incapacidad de suprimir un(os) codón(ones) selector(es), por ejemplo, dos o más, en las posiciones en un marcador negativo, por ejemplo, el gen de la barnasa.

Opcionalmente, el indicador se presenta en una superficie celular, por ejemplo, en una expresión de fagos o similar. La presentación en la superficie celular, por ejemplo, el sistema de presentación en la superficie celular basado en OmpA, se basa en la expresión de un epítipo concreto, por ejemplo, un péptido de un poliovirus C3 fusionado con una membrana externa de porina OmpA, sobre la superficie de la célula de *Escherichia coli*. El epítipo se presenta sobre la superficie celular únicamente cuando el codón selector en el mensaje de la proteína se suprime durante la traducción. El péptido presentado contiene a continuación el aminoácido reconocido por una de las aminoacil ARNt sintetetas mutantes en la biblioteca, y se puede aislar la célula que contiene el gen de la sintetasa correspondiente con anticuerpos sensibilizados contra los péptidos que contienen aminoácidos artificiales específicos. El sistema de presentación en la superficie celular basado en OmpA se desarrolló y optimizó por Georgiou y col. como una alternativa a la presentación en fagos. Véanse, Francisco, J. A., Campbell, R., Iverson, B. L. & Georgiou, G. Production and fluorescence-activated cell sorting of *Escherichia coli* expressing a functional antibody fragment on the external surface. *Proc. Natl. Acad. Sci. USA* 90:10444-8 (1993).

Otras realizaciones de la presente divulgación incluyen llevar a cabo una o más de las etapas de selección in vitro. El componente seleccionado, por ejemplo, sintetasa y/o ARNt, puede a continuación introducirse en una célula para el uso en la incorporación in vivo de un aminoácido artificial.

Pueden encontrarse detalles adicionales para producir O-RS, y alterar la especificidad por el sustrato de la sintetasa en la solicitud de patente de Estados Unidos 10/126.931 titulada "Methods and Compositions for the Production of Orthogonal tRNA-Aminoacyl tRNA Synthetase Pairs;" y, el documento USSN 10/825,867 titulado "EXPANDING THE EUKARYOTIC GENETIC CODE". se pueden encontrar detalles adicionales para producir O-RS en Hamano-Takaku y col., (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 y col. (2002), Una tirosil ARNt sintetasa de *Escherichia coli* diseñada mediante ingeniería genética para la incorporación específica en el sitio de un aminoácido artificial en proteína en la traducción eucariota y su aplicación en un sistema exento de células de

germen de trigo, PNAS 99(15): 9715-9723; y, Francklyn y col., (2002), Aminoacyl-tRNA synthetases: Versatile players in the changing theater of translation; RNA, 8:1363-1372.

ORGANISMOS FUENTE Y HOSPEDADORES

5 Los componentes de la traducción de la presente invención se derivan normalmente de organismos no eucarióticos. Por ejemplo, el O-ARNt ortogonal puede derivarse de un organismo no eucariótico, por ejemplo, un archaeobacterium, tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, o similares, o un eubacterium, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares, aunque la O-RS ortogonal puede derivarse de un organismo
10 no eucariota, por ejemplo, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, o similares, o un eubacterium, tal como *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, o similares. En una realización, se pueden usar también fuentes eucariotas, incluyendo, aunque no de forma limitativa, plantas, algas, protistas, hongos, levaduras, animales (por ejemplo, mamíferos, insectos,
15 artrópodos, etc.), o similares.

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

El O-ARNt, la O-RS o la pareja O-ARNt/O-RS se pueden seleccionar o cribarse in vivo o in vitro y/o utilizarse en una célula, por ejemplo, células no eucariotas (tales como una célula de *E. coli*), o una célula eucariota, para producir un polipéptido con un aminoácido seleccionado (por ejemplo, un aminoácido artificial). una célula no eucariota puede ser de una variedad de fuentes, tales como del dominio filogenético de las Archaea, incluyendo, aunque no de forma limitativa, *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tal como *Haloferax volcanii* y especies de *Halobacterium NRC-1*, *Archaeoglobus fulgidus*, *Pyrococcus furiosus*, *Pyrococcus horikoshii*, *Aeopyrum pernix*, o similares, o pueden pertenecer al dominio filogenético de las Eubacteria (incluyendo, aunque no de forma limitativa, *Escherichia coli*, *Thermus thermophilus*, *Bacillus stearothermophilus*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, etc.), o similares. Una célula eucariota puede ser de una variedad de fuentes, incluyendo, aunque no de forma limitativa, una planta (por ejemplo, una planta compleja, tal como monocotiledóneas, o dicotiledóneas), un alga, un protista, un hongo, una levadura (incluyendo, aunque no de forma limitativa, *Saccharomyces cerevisiae*), un animal (incluyendo, aunque no de forma limitativa, un mamífero, un insecto, un artrópodo, etc.), o similares. Las composiciones de células con componentes de la traducción de la presente divulgación son también una característica de la presente invención. Véase también el documento USSN 10/825.867 titulado "Expanding the Eukaryotic Genetic Code" para cribar O-ARNt y/u O-RS en una especie para el uso en otras especies.
25
30
35

Para expresar un polipéptido de interés con un aminoácido seleccionado en una célula hospedadora, se pueden subclonar los polinucleótidos que codifican un polipéptido de interés en un vector de expresión que contiene un promotor para dirigir la transcripción, un terminador de la transcripción/traducción, y para un ácido nucleico que codifica una proteína, un sitio de unión a ribosoma para el inicio de la traducción. Los promotores bacterianos adecuados son bien conocidos en la materia y se describen, por ejemplo, en Sambrook y col. y Ausubel y col.
40

Los sistemas de expresión bacterianos para expresar un polipéptido de interés están disponibles en, incluyendo, aunque no de forma limitativa, *E. coli*, *Bacillus sp.*, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, *Pseudomonas putida*, y *Salmonella* (Palva y col., Gene 22:229-235 (1983); Mosbach y col., Nature 302:543-545 (1983)). Están comercialmente disponibles los kits para dichos sistemas de expresión. Los sistemas de expresión eucariotas para células de mamíferos, levaduras, y células de insectos son bien conocidos en la materia y están también comercialmente disponibles.
45

Un ARNt y/o una RS de la presente divulgación y/o un polipéptido de interés pueden utilizarse y/o expresarse en cualquier número de sistemas de expresión adecuados incluyendo, por ejemplo, levaduras, células de insectos, células de mamíferos, y bacterias. Se proporciona a continuación una descripción de los sistemas de expresión ilustrativos.
50

Levadura Tal como se usa en el presente documento, el término "levadura" incluye cualquiera de las diversas levaduras capaces de expresar un polipéptido de interés. Dichas levaduras incluyen, aunque no de forma limitativa, levaduras ascoesporógenas (*Endomycetales*), levaduras basidioesporógenas y levaduras que pertenecen al grupo de los hongos imperfectos (*Blastomycetes*). Las levaduras ascoesporógenas se dividen en dos familias, *Spermophthoraceae* y *Saccharomycetaceae*. La última está comprendida por cuatro subfamilias, *Schizasaccharomycoidae* (por ejemplo, género *Schizosaccharomyces*), *Nadsonioideae*, *Lipomycoideae* y *Saccharomycoidae* (por ejemplo, géneros *Pichia*, *Kluyveromyces* y *Saccharomyces*). Las levaduras basidioesporógenas incluyen los géneros *Leucosporidium*, *Rhodospodium*, *Sporidiobolus*, *Filobasidium*, y
55

Filobasidiella. Las levaduras que pertenecen al grupo de los hongos imperfectos (*Blastomycetes*) se dividen en dos familias, *Sporobolomycetaceae* (por ejemplo, los géneros *Sporobolomyces* y *Bullera*) y *Cryptococcaceae* (por ejemplo, el género *Candida*).

De particular interés para el uso con la presente invención son las especies comprendidas en los géneros *Pichia*, *Kluyveromyces*, *Saccharomyces*, *Schizosaccharomyces*, *Hansenula*, *Torulopsis*, y *Candida*, incluyendo, aunque no de forma limitativa, *P. pastoris*, *P. guillerimondii*, *S. cerevisiae*, *S. carlsbergensis*, *S. diastaticus*, *S. douglasii*, *S. kluyveri*, *S. norbensis*, *S. oviformis*, *K. lactis*, *K. fragilis*, *C. albicans*, *C. maltosa*, y *H. polymorpha*. Las levaduras están generalmente disponibles a partir de una variedad de fuentes, incluyendo, aunque no de forma limitativa, the Yeast Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA), y la American Type Culture Collection ("ATCC") (Manassas, VA).

La expresión "hospedador de levadura" o "célula hospedadora de levadura" incluye levaduras que pueden ser, o han sido, utilizadas como un receptor de vectores recombinantes u otro ADN de transferencia. La expresión incluye la progenie de la de la célula hospedadora de levadura original que ha recibido los vectores recombinantes u otra ADN de transferencia. Se entiende que la progenie de una única célula precursora puede no ser de forma necesaria completamente idéntica en la morfología o en la genómica al complemento de ADN total en la célula precursora, debido a mutación accidental o deliberada. La progenie de la célula precursora que es suficientemente similar a precursora para caracterizarse mediante la propiedad relevante, tal como la presencia de una secuencia de nucleótidos que codifica un polipéptido de interés, está incluida en la progenie prevista por esta definición.

Se han desarrollado vectores de expresión y transformación, incluyendo replicones extracromosómicos o vectores de integración, para la transformación en muchos hospedadores de levaduras. Por ejemplo, se han desarrollado vectores de expresión para *S. cerevisiae* (Sikorski y col., GENETICS (1989) 122:19; Ito y col., J. BACTERIOL. (1983) 153:163; Hinnen y col., PROC. NATAL. ACAD. SCI. USA (1978) 75:1929); *C. albicans* (Kurtz y col., MOL. CELL. BIOL. (1986) 6:142); *C. maltosa* (Kunze y col., J. BASIC MICROBIOL. (1985) 25:141); *H. polymorpha* (Gleeson y col., J. GEN. MICROBIOL. (1986) 132:3459; Roggenkamp y col., MOL. GENETICS AND GENOMICS (1986) 202:302); *K. fragilis* (Das y col., J. BACTERIOL. (1984) 158:1165); *K. lactis* (De Louvencourt y col., J. BACTERIOL. (1983) 154:737; Van den Berg y col., BIOTECHNOLOGY (NY) (1990) 8:135); *P. guillerimondii* (Kunze y col., J. BASIC MICROBIOL. (1985) 25:141); *P. pastoris* (Patentes de Estados Unidos números 5.324.639; 4.929.555; y 4.837.148; Cregg y col., MOL. CELL. BIOL. (1985) 5:3376); *Schizosaccharomyces pombe* (Beach y col., NATURE (1982) 300:706); e *Y. lipolytica*; *A. nidulans* (Ballance y col., BIOCHEM. BIOPHYS. RES. COMMUN. (1983) 112:284-89; Tilburn y col., GENE (1983) 26:205-221; y Yelton y col., PROC. NATAL. ACAD. SCI. USA (1984) 81:1470-74); *A. niger* (Kelly y Hynes, EMBO J. (1985) 4:475-479); *T. reesia* (documento EP 0 244 234); y hongos filamentosos tales como, por ejemplo, *Neurospora*, *Penicillium*, *Tolypocladium* (documento WO 91/00357).

Las personas normalmente expertas en la materia conocen las secuencias control para los vectores de levaduras e incluye, aunque no de forma limitativa, regiones promotores procedentes de genes tales como la alcohol deshidrogenasa (ADH) (EP 0 284 044); enolasa; glucoquinasa; glucosa-6-fosfato isomerasa; gliceraldehído-3-fosfato-deshidrogenasa (GAP o GAPDH); hexoquinasa; fosfofructoquinasa; 3-fosfoglicerato mutasa; y piruvato quinasa (PyK) (EP 0 329 203). El gen PHO5 de levadura, que codifica la fosfatasa ácida, puede proporcionar también secuencias promotoras útiles (Miyanojara y col., PROC. NATAL. ACAD. SCI. USA (1983) 80:1). Otras secuencias promotoras adecuadas para uso con los hospedadores de levaduras pueden incluir los promotores de la 3-fosfoglicerato quinasa (Hitzeman y col., J. BIOL. CHEM. (1980) 255:12073); y otras enzimas glicolíticas, tales como piruvato decarboxilasa, triosafosfato isomerasa, y fosfoglucoasa isomerasa (Holland y col., BIOCHEMISTRY (1978) 17:4900; Hess y col., J. ADV. ENZYME REG. (1969) 7:149). Los promotores de levaduras inducibles que tienen la ventaja adicional de la transcripción controlada por las condiciones de crecimiento pueden incluir las regiones promotoras de la alcohol deshidrogenasa 2; el isocitocromo C; la fosfatasa ácida; metalotioneína; la gliceraldehído-3-fosfato deshidrogenasa; las enzimas degradativas asociadas con el metabolismo del nitrógeno; y las enzimas responsables de la utilización de la maltosa y la galactosa. Los vectores y promotores adecuados para uso en la expresión de levaduras se describen adicionalmente en el documento EP 0 073 657.

Los potenciadores de las levaduras pueden utilizarse también como promotores de las levaduras. Además, los promotores sintéticos pueden funcionar también como promotores de las levaduras. Por ejemplo, las secuencias de activación en la dirección 5' (UAS) de un promotor de levadura pueden unirse con la región de activación de la transcripción de otro promotor de levadura, creando un promotor híbrido sintético. Los ejemplos de dichos promotores híbridos incluyen la secuencia reguladora de ADH unida a la región de activación de la transcripción de GAP. Véanse las patentes de Estados Unidos números 4.880.734 y 4.876.197. Otros ejemplos de promotores híbridos incluyen promotores que consisten en las secuencias reguladoras de los genes ADH2, GAL4, GAL10, o PHO5, combinados con la región de activación de la transcripción de un gen de una enzima glicolítica tal como GAP o PyK. Véase el documento EP 0 164 556. Adicionalmente, un promotor de levadura puede incluir promotores que se producen naturalmente de origen no de levadura que tienen la capacidad de unirse a la ARN polimerasa de levadura e iniciar la transcripción.

Otros elementos control que pueden comprender parte de los vectores de expresión de levaduras incluyen terminadores, por ejemplo, de los genes de GAPDH o de la enolasa (Holland y col., J. BIOL. CHEM. (1981) 256:1385). Además, el origen de replicación procedente del origen del plásmido 2 μ es adecuada para las levaduras.

Un gen de selección adecuado para uso en levadura es el gen *trp1* presente en el plásmido de levadura. Véase Tschumper y col., GENE (1980) 10:157; Kingsman y col., GENE (1979) 7:141. El gen *trp1* proporciona un marcador de selección para una cepa mutante de una levadura que carece de la capacidad de crecer en triptófano. De manera similar, Las cepas de levadura deficientes en Leu2 (ATCC 20.622 o 38.626) están complementadas por plásmidos conocidos que soportan el gen *Leu2*.

Las personas normalmente expertas en la materia conocen los procedimientos para introducir ADN exógeno en hospedadores de levaduras, y normalmente incluyen, aunque no de forma limitativa, cualquiera de la transformación de esferoplastos o de células hospedadoras de levaduras intactas con cationes alcalinos. Por ejemplo, se puede llevar a cabo la transformación de levaduras de acuerdo con el procedimiento descrito en Hsiao y col., PROC. NATL. ACAD. SCI. USA (1979) 76:3829 y Van Solingen y col., J. BACT. (1977) 130:946. Sin embargo, se pueden usar también otros procedimientos para introducir ADN en las células tales como mediante inyección nuclear, electroporación, o fusión de protoplastos, como se describe generalmente en SAMBROOK Y COL., MOLECULAR CLONING: A LAB. MANUAL (2001). las células de hospedadores de levaduras pueden a continuación cultivarse utilizando técnicas normalizadas conocidas por las personas normalmente expertas en la materia.

Las personas normalmente expertas en la materia conocen otros procedimientos para expresar proteínas heterólogas en células hospedadoras de levaduras. Véanse generalmente la publicación de patente de Estados Unidos n.º 20020055169, las patentes de Estados Unidos números 6.361.969; 6.312.923; 6.183.985; 6.083.723; 6.017.731; 5.674.706; 5.629.203; 5.602.034; y 5.089.398; las patentes reexaminadas de Estados Unidos RE37.343 y RE35.749; las publicaciones de patentes PCT publicadas WO 99/07862; WO 98/37208; y WO 98/26080; las publicaciones de patentes europeas EP 0 946 736; EP 0 732 403; EP 0 480 480; WO 90/10277; EP 0 340 986; EP 0 329 203; EP 0 324 274; y EP 0 164 556. Véanse también Gellissen y col., ANTONIE VAN LEEUWENHOEK (1992) 62(1-2):79-93; Romanos y col., YEAST (1992) 8(6):423-488; Goedel, METHODS IN ENZYMOLOGY (1990) 185:3-7.

Las cepas hospedadoras de levaduras pueden hacerse crecer en fermentadores durante la etapa de amplificación utilizando procedimientos convencionales de fermentación alimentación discontinua conocidos por las personas normalmente expertas en la materia. Los procedimientos de fermentación pueden adaptarse para tener en cuenta las diferencias en la ruta de utilización del carbono por el hospedador de la levadura o el modo de control de la expresión. Por ejemplo, la fermentación de un hospedador de la levadura *Saccharomyces* puede requerir como único alimento la glucosa, una fuente de nitrógeno complejo (por ejemplo, hidrolizados de caseína), y una múltiple suplementación de vitaminas. Por el contrario, la levadura metilotrófica *P. pastoris* puede requerir glicerol, metanol, y alimentos minerales traza, pero sol sales de amonio simples (nitrógeno) para el crecimiento y la expresión óptima. Véanse, por ejemplo, patente de Estados Unidos n.º 5.324.639; Elliott y col., J. PROTEIN CHEM. (1990) 9:95; y Fieschko y col., BIOTECH. BIOENG. (1987) 29:1113.

Dichos procedimientos de fermentación, sin embargo, pueden tener determinadas características comunes independientemente de la cepa hospedadora de levaduras empleada. Por ejemplo, se puede añadir un nutriente limitante del crecimiento, normalmente carbono, al fermentador durante la fase de amplificación para permitir un crecimiento máximo. Además, los procedimientos de fermentación emplean generalmente un medio de fermentación diseñado para contener cantidades adecuadas de carbono, nitrógeno, sales basales, fósforo, y otros nutrientes menores (vitaminas, minerales traza y sales, etc.). Los ejemplos de medios de fermentación adecuados para el uso con *Pichia* se describen en las patentes de Estados Unidos. 5.324.639 y 5.231.178.

Células de insectos infectadas con baculovirus. La expresión "hospedador de insecto" o "célula hospedadora de insecto" se refiere a un insecto que puede o se ha utilizado, como un receptor de vectores recombinantes u otro ADN de transferencia. La expresión incluye la progenie de la célula hospedadora de insecto original que se ha transfectado. Se entiende que la progenie de una única célula precursora puede no ser de forma necesaria completamente idéntica en la morfología o en la genómica al complemento de ADN total en la célula precursora, debido a mutación accidental o deliberada. La progenie de la célula precursora que es suficientemente similar a precursora para caracterizarse mediante la propiedad relevante, tal como la presencia de una secuencia de nucleótidos que codifica un polipéptido de interés, está incluida en la progenie prevista por esta definición.

Las personas normalmente expertas en la materia conocen la selección de células de insecto adecuadas para la expresión de un polipéptido de interés. Algunas especies de insectos están bien descritas en la técnica y están comercialmente disponibles, incluyendo *Aedes aegypti*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, y *Trichoplusia ni*. en la selección de hospedadores de insectos para la expresión, los hospedadores adecuados pueden incluir aquellos que muestran tener, entre otras, buena capacidad de secreción, baja actividad proteolítica, y solidez global. los insectos están generalmente disponibles a partir de una variedad de fuentes, incluyendo, aunque no de forma limitativa, the Insect Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); y la American Type Culture Collection ("ATCC") (Manassas, VA).

En general, los componentes de un sistema de expresión de insecto infectado con baculovirus incluyen un vector de transferencia, normalmente un plásmido bacteriano, que contiene un fragmento del genoma de baculovirus, y un sitio de restricción conveniente para la inserción del gen heterólogo que se va a expresar; un baculovirus natural con secuencias homólogas para el fragmento específico de baculovirus en el vector de transferencia (esto permite la

recombinación homóloga del gen heterólogo en el genoma del baculovirus); y células hospedadoras de insectos adecuadas y medios de crecimiento. Los materiales, procedimientos y técnicas utilizados en los vectores de construcción, las células transfectantes, las placas para repicación, las células en crecimiento en el cultivo y similares, se conocen en la técnica y los manuales están disponibles describiendo estas técnicas.

5 Tras insertar el gen heterólogo en el vector de transferencia, el vector, y el genoma vírico natural se transfectan en una célula hospedadora de insecto donde el vector y el genoma vírico se recombinan. El virus recombinante empaquetado se expresa y las placas recombinantes se identifican y purifican. Los materiales y procedimientos para los sistemas de expresión baculovirus/célula de insecto están comercialmente disponibles en forma de kit de, por ejemplo, Invitrogen Corp. (Carlsbad, CA). Estas técnicas son generalmente conocidas por las personas normalmente expertas en la materia y se describen completamente en SUMMERS Y SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN NO. 1555 (1987). Véanse también, RICHARDSON, 39 METHODS IN MOLECULAR BIOLOGY: BACULOVIRUS EXPRESSION PROTOCOLS (1995); AUSUBEL Y COL., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY 16.9-16.11 (1994); KING Y POSSEE, THE BACULOVIRUS SYSTEM: A LABORATORY GUIDE (1992); y O'REILLY Y COL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

De hecho, las personas normalmente expertas en la materia conocen la producción de diversas proteínas heterólogas utilizando sistemas de expresión de baculovirus/células de insecto. Véanse, por ejemplo, las patentes de Estados Unidos números 6.368.825; 6.342.216; 6.338.846; 6.261.805; 6.245.528, 6.225.060; 6.183.987; 6.168.932; 6.126.944; 6.096.304; 6.013.433; 5.965.393; 5.939.285; 5.891.676; 5.871.986; 5.861.279; 5.858.368; 5.843.733; 5.762.939; 5.753.220; 5.605.827; 5.583.023; 5.571.709; 5.516.657; 5.290.686; WO 02/06305; WO 01/90390; WO 01/27301; WO 01/05956; WO 00/55345; WO 00/20032; WO 99/51721; WO 99/45130; WO 99/31257; WO 99/10515; WO 99/09193; WO 97/26332; WO 96/29400; WO 96/25496; WO 96/06161; WO 95/20672; WO 93/03173; WO 92/16619; WO 92/02628; WO 92/01801; WO 90/14428; WO 90/10078; WO 90/02566; WO 90/02186; WO 90/01556; WO 89/01038; WO 89/01037; documento WO 88/07082.

25 Se conocen en la técnica los vectores que son útiles en los sistemas de expresión de baculovirus/células de insecto e incluyen, por ejemplo, vectores de expresión y transferencia de insectos derivados del virus de la polihedrosis nuclear de *Autographa californica* infectada con baculovirus (AcNPV), que es un vector de expresión vírico auxiliar independiente. Los vectores de expresión víricos derivados de este sistema utilizan normalmente el promotor fuerte del gen de la polihedrina vírica para impulsar la expresión de genes heterólogos. Véase generalmente, O'Reilly Y COL., BACULOVIRUS EXPRESSION VECTORS: A LABORATORY MANUAL (1992).

Antes de insertar el gen extraño en el genoma del baculovirus, los componentes anteriormente descritos, que comprenden un promotor, una secuencia líder (si se desea), la secuencia de codificación de interés, y la secuencia de terminación de la transcripción, se ensamblan normalmente en una secuencia de transposición intermedia (vector de transferencia). Las construcciones de transposición intermedias se mantienen a menudo en un replicó, tal como un elemento extracromosómico (por ejemplo, plásmidos) capaz de mantenimiento estable en un hospedador, tal como bacterias. el replicón tendrá un sistema de replicación, que le permite de esta manera mantenerse en un hospedador adecuado para la clonación y la amplificación. Más específicamente, el plásmido puede contener la señal de poliadenilación de la polihedrina (Miller, ANN. REV. MICROBIOL. (1988) 42:177) y un gen procarionta de resistencia a la ampicilina (*amp*) y un origen de replicación para la selección y propagación en *E. coli*.

40 Un vector de transferencia comúnmente utilizado para introducir genes extraños en AcNPV es pAc373. Se han diseñado otros muchos vectores, conocidos por los expertos en la técnica. incluyendo, por ejemplo, pVL985, que altera el codón de inicio de la polihedrina de ATG a ATT y que introduce 32 pares de bases en un sitio de clonación BamHI en la dirección 3' desde el ATT. Véase Luckow y Summers, VIROLOGY 170:31 (1989). Otros vectores comercialmente disponibles incluyen, por ejemplo, PBlueBac4.5/V5-His; pBlueBacHis2; pMelBac; pBlueBac4.5 (Invitrogen Corp., Carlsbad, CA).

Tras la inserción del gen heterólogo, el vector de transferencia y el genoma de baculovirus natural se transfectan simultáneamente en un hospedador de célula de insecto. Se conocen en la técnica los procedimientos para introducir ADN heterólogo en el sitio deseado en el virus baculovirus. Véanse SUMMERS Y SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN NO. 1555 (1987); Smith y col., MOL. CELL. BIOL. (1983) 3:2156; Luckow y Summers, VIROLOGY (1989) 170:31. Por ejemplo, la inserción puede ser en un gen tal como el gen de la polihedrina, mediante recombinación homóloga con doble entrecruzamiento genético; la inserción puede ser también en un sitio de un enzima de restricción diseñado mediante ingeniería genética en el gen de baculovirus deseado. Véase Miller y col., BIOESSAYS (1989) 11(4):91.

La transfección puede llevarse a cabo mediante electroporación. Véanse TROTTER y WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Mann y King, J. GEN. VIROL. (1989) 70:3501. Como alternativa, se pueden usar liposomas para transfectar las células de insecto con el vector de expresión recombinante y el baculovirus. Véanse, por ejemplo, Liebman y col., BIOTECHNIQUES (1999) 26(1):36; Graves y col., BIOCHEMISTRY (1998) 37:6050; Nomura y col., J. BIOL. CHEM. (1998) 273(22):13570; Schmidt y col., PROTEIN EXPRESSION AND PURIFICATION (1998) 12:323; Siffert y col., NATURE GENETICS (1998) 18:45; TILKINS Y COL., CELL BIOLOGY: A LABORATORY HANDBOOK 145-154 (1998); Cai y col., PROTEIN EXPRESSION AND PURIFICATION (1997)

10:263; Dolphin y col., NATURE GENETICS (1997) 17:491; Kost y col., GENE (1997) 190:139; Jakobsson y col., J. BIOL. CHEM. (1996) 271:22203; Rowles y col., J. BIOL. CHEM. (1996) 271(37):22376; Reverey y col., J. BIOL. CHEM. (1996) 271(39):23607-10; Stanley y col., J. BIOL. CHEM. (1995) 270:4121; Sisk y col., J. VIROL. (1994) 68(2):766; y Peng y col., BIOTECHNIQUES (1993) 14(2):274. Los liposomas comercialmente disponibles incluyen, por ejemplo, Cellfectin® y Lipofectin® (Invitrogen, Corp., Carlsbad, CA). Además, se puede usar la transfección con fosfato de calcio. Véanse TROTTER y WOOD, 39 METHODS IN MOLECULAR BIOLOGY (1995); Kitts, NAR (1990) 18(19):5667; y Mann y King, J. GEN. VIROL. (1989) 70:3501.

Los vectores de expresión de baculovirus contienen normalmente un promotor de baculovirus. Un promotor de baculovirus es cualquier secuencia de ADN capaz de unirse a una ARN polimerasa de baculovirus e iniciar la transcripción en la dirección 3' (3') de una secuencia de codificación (por ejemplo, un gen estructural) en el ARNm. Un promotor tendrá una región de inicio de la transcripción que se coloca normalmente próxima al extremo 5' de la secuencia de codificación. Esta región de inicio de la transcripción incluye normalmente un sitio de unión de la ARN polimerasa y un sitio de inicio de la transcripción. Un promotor de baculovirus puede tener también un segundo dominio denominado un potenciador, que, si está presente, es usualmente distal al gen estructural. Además, la expresión puede ser tanto regulada como constitutiva.

Los genes estructurales, se transcriben abundantemente en los últimos momentos en el ciclo de infección, proporcionan secuencias promotoras particularmente útiles. Los ejemplos incluyen las secuencias derivadas del gen que codifica la proteína del polihedron vírico (FRIESEN Y COL., The Regulation of Baculovirus Gene Expression in THE MOLECULAR BIOLOGY OF BACULOVIRUSES (1986); documentos EP 0 127 839 y 0 155 476) y el gen que codifica la proteína p10 (Vlak y col., J. GEN. VIROL. (1988) 69:765).

El vector de expresión de baculovirus formado recientemente se empaqueta en un baculovirus recombinante infeccioso y posteriormente, las placas de crecimiento se pueden purificar mediante técnicas conocidas por las personas normalmente expertas en la materia. Véase Miller y col., BIOESSAYS (1989) 11(4):91; SUMMERS Y SMITH, TEXAS AGRICULTURAL EXPERIMENT STATION BULLETIN NO. 1555 (1987).

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

Las células y los medios de cultivo están comercialmente disponibles para la expresión directa y la expresión de la fusión de los polipéptidos heterólogos en un sistema de expresión de baculovirus, y las personas normalmente expertas en la materia conocen generalmente la tecnología del cultivo de células.

E. Coli, *Especies de Pseudomonas*, y otras *Procariontas* Las técnicas de expresión bacteriana son conocidas de los expertos en la materia. Están disponibles una amplia variedad de vectores para su uso en hospedadores bacterianos. Los vectores pueden ser de una única copia o vectores multicopia bajos o altos. Los vectores pueden servir para la clonación y/o la expresión. A la vista de la amplia bibliografía que se refiere a los vectores, la disponibilidad comercial de muchos vectores, e incluso a los manuales que describen los vectores y sus cartografías y características de restricción, no se requiere aquí una extensa descripción. Como se sabe bien, los vectores implican normalmente marcadores que permiten la selección, cuyos marcadores pueden proporcionar resistencia a agentes citotóxicos, prototrofia o inmunidad. Frecuentemente, está presente una pluralidad de marcadores, que proporcionan diferentes características.

Un promotor bacteriano es cualquier secuencia de ADN capaz de unirse a una ARN polimerasa bacteriana e iniciar la transcripción en la dirección 3' (3') de una secuencia de codificación (por ejemplo, un gen estructural) en el ARNm. Un promotor tendrá una región de inicio de la transcripción que se coloca normalmente próxima al extremo 5' de la secuencia de codificación. Esta región de inicio de la transcripción incluye normalmente un sitio de unión de la ARN polimerasa y un sitio de inicio de la transcripción. Un promotor bacteriano puede tener también un segundo dominio denominado un operador que puede solaparse con un sitio de unión de una ARN polimerasa adyacente en el cual comienza la síntesis de ARN. El operador permite la transcripción regulada negativa (inducible), ya que una proteína represora génica puede unirse al operador e inhibir por tanto la transcripción de un gen específico. Se puede producir la expresión constitutiva en ausencia de elementos reguladores negativos, tales como el operador. Además, se puede conseguir la regulación positiva mediante una secuencia de unión de la proteína activadora del gen, que, si está presente está normalmente próxima (5') a la secuencia de unión de la ARN polimerasa. Un ejemplo de una proteína activadora de un gen es la proteína activadora de catabolitos (CAP), que ayuda a iniciar la transcripción del operón lac en *Escherichia coli* (*E. coli*) [Raibaud y col., ANNU. REV. GENET. (1984) 18:173]. La expresión regulada puede por tanto ser tanto positiva como negativa, potenciando o reduciendo por tanto la transcripción.

Las secuencias que codifican las enzimas de la ruta metabólica proporcionan secuencias promotoras particularmente útiles. Los ejemplos incluyen secuencias promotoras derivadas de las enzimas que metabolizan el azúcar, tales como galactosa, lactosa (lac) [Chang y col., NATURE (1977) 198:1056], y maltosa. Los ejemplos adicionales incluyen las secuencias promotoras derivadas de enzimas biosintéticas tales como triptófano (trp) [Goeddel y col., Nuc. ACIDS RES. (1980) 8:4057; Yelverton y col., NUCL. ACIDS RES. (1981) 9:731; Patente de Estados Unidos n.º 4.738.921; Publicaciones EP. números 036 776 y 121 775]. The β -galactosidase (bla) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In Interferon 3 (Ed. I. Gresser)], Los sistemas promotores del bacteriófago lambda PL [Shimatake y col., NATURE (1981) 292:128] y T5 [Patente de Estados Unidos n.º 4.689.406] proporcionan también útiles secuencias promotoras. Los promotores fuertes, tales como el promotor T7 pueden utilizarse para inducir el polipéptido de interés a altos niveles. Las personas normalmente expertas en la materia conocen los ejemplos de dichos vectores e incluyen la serie pET29 de Novagen, y los vectores pPOP descritos en el documento WO99/05297. Dichos sistemas de expresión producen altos niveles de polipéptido en el hospedador sin comprometer la viabilidad de la célula hospedadora o los parámetros de crecimiento. pET19 (Novagen) es otro vector conocido en la técnica.

Además, los promotores sintéticos que no se producen en la naturaleza funcionan también como promotores bacterianos. Por ejemplo, las secuencias de activación de la transcripción de un promotor bacteriano o de bacteriófago pueden unirse con las secuencias del operón de otro promotor bacteriano o de bacteriófago, creando un promotor híbrido sintético [patente de Estados Unidos n.º 4,551.433]. Por ejemplo, el promotor tac es un promotor trp-lac híbrido comprendido por las secuencias del promotor trp y las secuencias del operón lac, que está regulado por el represor de lac [Amann y col., GENE (1983) 25:167; de Boer y col., PROC. NATAL. ACAD. SCI. (1983) 80:21]. Adicionalmente, un promotor bacteriano puede incluir promotores que se producen naturalmente de origen no bacteriano que tienen la capacidad de unirse a la ARN polimerasa bacteriana e iniciar la transcripción. Un promotor que se produce naturalmente de origen no bacteriano puede acoplarse también con una ARN polimerasa compatible para producir altos niveles de expresión de algunos genes en procariontes. El sistema ARN polimerasa/promotor del bacteriófago T7 es un ejemplo de un sistema promotor acoplado [Studier y col., J. MOL. BIOL. (1986) 189:113; Tabor y col., Proc Natl. Acad. Sci. (1985) 82:1074]. Además, un promotor híbrido puede estar también comprendido por un promotor de bacteriófago y una región operadora de *E. coli* (publicación EP. n.º 267:851).

Además de una secuencia promotora del funcionamiento, un sitio de unión a ribosoma eficaz es también útil para la expresión de los genes extraños en procariontes. En *E. coli*, el sitio de unión a ribosoma se denomina la secuencia de Shine-Dalgarno (SD) e incluye un codón de inicio (ATG) y una secuencia de 3-9 nucleótidos de longitud localizada 3-11 nucleótidos en la dirección 5' del codón de inicio [Shine y col., NATURE (1975) 254:34]. Se piensa que la secuencia SD promueve la unión del ARNm al ribosoma emparejando las bases entre la secuencia SD y el 3' y el ARNr de 16S de *E. coli* [Steitz y col. "Genetic signals and nucleotide sequences in messenger RNA", In Biological Regulation and Development: Gene Expression (Ed. R. F. Goldberger, 1979)]. Para expresar genes eucariotas y genes procariontes con un sitio de unión a ribosoma débil [Sambrook y col. "Expression of cloned genes in *Escherichia coli*", Molecular Cloning: A Laboratory Manual, 1989].

la expresión "hospedador bacteriano" o "célula hospedadora bacteriana" incluye bacterias que pueden, o han sido, utilizadas como un receptor de vectores recombinantes u otro ADN de transferencia. La expresión incluye la progenie de la célula hospedadora bacteriana original que se ha transfectado. Se entiende que la progenie de una única célula precursora puede no ser de forma necesaria completamente idéntica en la morfología o en la genómica al complemento de ADN total en la célula precursora, debido a mutación accidental o deliberada. La progenie de la célula precursora que es suficientemente similar a precursora para caracterizarse mediante la propiedad relevante, tal como la presencia de una secuencia de nucleótidos que codifica un polipéptido de interés, está incluida en la progenie prevista por esta definición.

Las personas normalmente expertas en la materia conocen la selección de bacterias hospedadoras adecuadas para la expresión de polipéptidos. En la selección de hospedadores bacterianos para la expresión, los hospedadores adecuados pueden incluir aquellos que muestran tener, *entre otros*, una buena capacidad de formación de cuerpos de inclusión, baja actividad proteolítica, y solidez global. Los hospedadores bacterianos están generalmente disponibles a partir de una variedad de fuentes, incluyendo, aunque no de forma limitativa, the Bacterial Genetic Stock Center, Department of Biophysics and Medical Physics, University of California (Berkeley, CA); y la American Type Culture Collection ("ATCC") (Manassas, VA). La fermentación industrial/farmacéutica utiliza generalmente bacterias derivadas de cepas K (por ejemplo, W3110) o de bacterias derivadas de cepas B (por ejemplo, BL21). Estas cepas son particularmente útiles debido a que sus parámetros de crecimiento son extremadamente bien conocidos y sólidos. Además, estas cepas no son patógenas, lo que es comercialmente importante por motivos de seguridad y ambientales. Otros ejemplos de hospedadores de *E. coli* adecuados incluyen, aunque no de forma limitativa, cepas BL21, DH10B, o derivados de las mismas. En otra realización de los procedimientos de la presente divulgación, el hospedador de *E. coli* host es una cepa proteasa menos que incluye, aunque no de forma limitativa, OMP- y LON-. la cepa de la célula hospedadora puede ser de una especie de *Pseudomonas*, incluyendo, aunque no de forma limitativa, *Pseudomonas fluorescens*, *Pseudomonas aeruginosa*, and *Pseudomonas putida*. *Pseudomonas fluorescens* biovar 1, designada cepa MB101, se conoce por ser útil para la producción recombinante y está disponible para los procesos de producción de proteínas terapéuticas. Los ejemplos de sistemas de expresión de *Pseudomonas* incluyen el sistema disponible de The Dow Chemical Company como una cepa hospedadora (Midland, MI disponible en el sitio web en dow.com). las patentes de Estados Unidos números 4.755.465 y

4.859.600, describen el uso de cepas de *Pseudomonas* como un hospedador celular para la producción de hGH.

Una vez que se ha establecido la cepa de la célula hospedadora recombinante (es decir, la construcción de expresión se ha introducido en la célula hospedadora y las células hospedadoras con la construcción de expresión adecuada se aíslan), la cepa de la célula hospedadora recombinante se cultiva en condiciones adecuadas para la producción del polipéptido de interés. Como será evidente para una persona experta en la materia, el procedimiento de cultivo de la cepa de célula hospedadora recombinante será dependiente de la naturaleza de la construcción de expresión utilizada y de la identidad de la célula hospedadora. Las cepas hospedadoras recombinantes se cultivan normalmente utilizando los procedimientos que son bien conocidos en la técnica. Las células hospedadoras recombinantes se cultivan normalmente en medio líquido que contiene fuentes asimilables de carbono, nitrógeno, y sales inorgánicas y, opcionalmente, que contiene vitaminas, aminoácidos, factores de crecimiento, y otros suplementos de cultivos proteínicos conocidos por las personas normalmente expertas en la materia. Los medios líquidos para el cultivo de células hospedadoras pueden contener opcionalmente antibióticos o antifúngicos para evitar el crecimiento de microorganismos indeseables y/o compuestos que incluyen, aunque no de forma limitativa, antibióticos para seleccionar las células hospedadoras que contienen el vector de expresión.

Están disponibles algunos procedimientos bien conocidos de introducir ácidos nucleicos diana en las células, cualquiera de los cuales se puede usar en la invención. Estos incluyen: La fusión de células receptoras con protoplastos bacterianos que contienen el ADN, electroporación, bombardeo de proyectiles, e infección con vectores víricos (descrita adicionalmente a continuación), etc. Se pueden usar las células bacterianas para amplificar el número de plásmidos que contienen las construcciones de ADN de la presente invención. Las bacterias se hacen crecer hasta la fase logarítmica y los plásmidos en las bacterias se pueden aislar mediante una variedad de procedimientos conocidos en la técnica (véase, por ejemplo, Sambrook). Además, están comercialmente disponibles kits para la purificación de plásmidos a partir de bacterias, (véase, por ejemplo, EasyPrep™, FlexiPrep™, ambos de Pharmacia Biotech; StrataClean™ de Stratagene; y, QIAprep™ de Qiagen). Los plásmidos aislados y purificados se manipulan adicionalmente a continuación para producir otros plásmidos, utilizados para transfectar células o incorporados en vectores relacionados para infectar organismos. Los vectores típicos contienen terminadores de la transcripción y la traducción, las secuencias de inicio de la transcripción y la traducción, y los promotores útiles para la regulación de la expresión del ácido nucleico diana concreto. Los vectores comprenden opcionalmente casetes de expresión genéricos que contiene al menos una secuencia terminadora independiente, las secuencias que permiten la replicación del casete en eucariotas, o procariotas, o ambos, (incluyendo, aunque no de forma limitativa, los vectores lanzadera) y los marcadores de selección para los sistemas procariotas y eucariotas. Los vectores son adecuados para la replicación y la integración en procariotas, eucariotas, o ambos. Véanse, Gillam & Smith, *Gene* 8:81 (1979); Roberts, y col., *Nature*, 328:731 (1987); Schneider, E., y col., *Protein Expr. Purif.* 6(1):10-14 (1995); Ausubel, Sambrook, Berger (todos anteriormente). Se proporciona un catálogo de bacterias y bacteriófagos útiles para la clonación, por ejemplo, por la ATCC, por ejemplo, *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna y col. (eds) publicado por la ATCC. Los procedimientos adicionales básicos para la secuenciación, la clonación y otros aspectos de la biología molecular y que subyacen bajo las consideraciones teóricas se encuentran también en el Watson y col. (1992) *Recombinant DNA Segunda Edición Scientific American Books*, NY. Además, esencialmente cualquier ácido nucleico /y, virtualmente, cualquier ácido nucleico marcado, tanto normalizado como no normalizado) puede ordenarse de forma personalizada o normalizada a partir de cualquiera de una variedad de fuentes comerciales, tales como la Midland Certified Reagent Company (Midland, TX disponible en el sitio web en mrcr.com), The Great American Gene Company (Ramona, CA disponible en el sitio web en genco.com), ExpressGen Inc. (Chicago, IL disponible en el sitio web en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchas otras.

Las células hospedadoras recombinantes pueden cultivarse en formatos discontinuos o continuos, tanto recogiendo las células (en el caso donde el polipéptido de interés se acumula intracelularmente) como recogiendo el sobrenadante del cultivo bien en formato discontinuo o bien en formato continuo. Para la producción en células hospedadoras procariotas, se prefieren el cultivo discontinuo y la cosecha de células.

CODONES SELECTORES

Los codones selectores de la presente divulgación expanden el marco del codón genético de la maquinaria biosintética de proteínas. Por ejemplo, un codón selector incluye, por ejemplo, un único codón de tres bases, un codón finalizador, tal como un codón de terminación, incluyendo, aunque no de forma limitativa, un codón ámbar (UAG), un codón ocre, o un codón ópalo (UGA), un codón artificial, un codón de cuatro bases (o más), un codón raro, o similares. Se pueden introducir numerosos codones selectores en un gen o polinucleótido deseado, por ejemplo, uno o más, dos o más, tres o más, etc.

Los codones selectores de la descripción expanden el marco del codón genético de la maquinaria biosintética de la proteína. Por ejemplo, un codón selector incluye, aunque no de forma limitativa, un único codón de tres bases, un codón finalizador, tal como un codón de terminación, incluyendo, aunque no de forma limitativa, un codón ámbar (UAG), un codón ocre, o un codón ópalo (UGA), un codón artificial, un codón de cuatro o más bases, un codón raro, o similares. Es fácilmente evidente para las personas normalmente expertas en la materia que existe un amplio intervalo en el número de codones selectores que se pueden introducir en un gen o polinucleótido deseado, incluyendo, aunque no de forma limitativa, uno o más, dos o más, tres o más, 4, 5, 6, 7, 8, 9, 10 o más en un único

polinucleótido que codifica el antígeno o el organismo completo, como ejemplos no limitantes de organismos completos modificados genéticamente utilizando la presente invención, MAP y *E. coli*.

En una realización, los procedimientos implican el uso de un codón selector que es un codón de terminación para la incorporación de un aminoácido seleccionado, por ejemplo, un aminoácido artificial, in vivo. Por ejemplo, se produce un O-ARNt que reconoce el codón de terminación y está aminoacilado por una O-RS con un aminoácido seleccionado. Este O-ARNt no está reconocido por las aminoacil ARNt sintetizadas del hospedador que se producen naturalmente. Se puede usar la mutagénesis dirigida al sitio convencional para introducir el codón de terminación en el sitio de interés en un polipéptido de interés. Véanse, por ejemplo, Sayers, J.R., y col. (1988), 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis. *Nucleic Acids Res*, 16:791-802. Cuando la O-RS, el O-ARNt y el ácido nucleico que codifica un polipéptido de interés se combinan, por ejemplo, in vivo, el aminoácido seleccionado se incorpora en respuesta al codón de terminación para proporcionar un polipéptido que contiene el aminoácido seleccionado, por ejemplo, un aminoácido artificial, en la posición especificada. En una realización de la presente divulgación, un codón de terminación utilizado como un codón selector es un codón ámbar, UAG, y/o un codón ópalo, UGA. Por ejemplo, Véase la SEQ ID NO: 6 para un ejemplo de un O-ARNt que reconoce un codón ámbar, y véase la SEQ ID NO: 7 para un ejemplo de un O-ARNt que reconoce un codón ópalo. Un código genético en el que UAG y UGA se utilizan como un codón selector puede codificar 22 aminoácidos preservando a la vez el codón finalizador ocre, UAA, que es la señal de terminación más abundante.

La incorporación de aminoácidos seleccionados, por ejemplo, aminoácidos artificiales, in vivo puede llevarse a cabo sin perturbación significativa de la célula hospedadora. Por ejemplo en células no eucariotas, tal como *Escherichia coli*, debido a que la eficacia de supresión del codón UAG depende de la competición entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y el factor 1 de liberación (RF1) (que se une al codón UAG e inicia la liberación del péptido en crecimiento desde el ribosoma), se puede modular la eficacia de la supresión, por ejemplo, tanto aumentando el nivel de expresión del O-ARNt, por ejemplo, el ARNt supresor, como utilizando una cepa deficiente en RF1. en células eucariotas, debido a que la eficacia de supresión del codón UAG depende de la competición entre el O-ARNt, por ejemplo, el ARNt supresor ámbar, y un fr de liberación eucariota (por ejemplo, eRF) (que se une a un codón de terminación e inicia la liberación del péptido en crecimiento desde el ribosoma), se puede modular la eficacia de la supresión, por ejemplo, aumentando el nivel de expresión del O-ARNt, por ejemplo, el ARNt supresor.

Se pueden codificar también los aminoácidos artificiales con codones raros. Por ejemplo, cuando se reduce la concentración de arginina en una reacción de síntesis de proteínas in vitro, el codón de arginina raro, AGG, ha demostrado ser eficaz para la inserción de Ala por un ARNt sintético acilado con alanina. Véanse, por ejemplo, Ma y col., *Biochemistry*, 32:7939 (1993). En este caso, el ARNt sintético compite con el ARNtArg que se produce naturalmente, que existe como una especie menor en *Escherichia coli*. Algunos organismos no utilizan todos los codones de tripletes. Se ha utilizado un codón AGA sin asignar en *Micrococcus luteus* para la inserción de aminoácidos en un extracto de transcripción/traducción in vitro. Véanse, por ejemplo, Kowal y Oliver, *Nucl. Acid. Res.*, 25:4685 (1997). Se pueden generar los componentes de la presente divulgación para utilizar estos codones raros in vivo.

Los codones selectores comprenden también codones extendidos, por ejemplo, codones de cuatro o más bases, tales como, codones de cuatro, cinco, seis o más bases. Los ejemplos de codones de cuatro bases incluyen, aunque no de forma limitativa, AGGA, CUAG, UAGA, CCCU, y similares. Los ejemplos de codones de cinco bases incluyen, aunque no de forma limitativa, AGGAC, CCCCU, CCCUC, CUAGA, CUACU, UAGGC y similares. Una característica puede incluir utilizar codones extendidos basados en la supresión del cambio de marco. Codones de cuatro o más bases pueden insertar, por ejemplo, uno o múltiples aminoácidos seleccionados, incluyendo, aunque no de forma limitativa, aminoácidos artificiales, en la misma proteína. Por ejemplo, en presencia de O-ARNt mutados, por ejemplo, un ARNt supresor del cambio de marco especial, con bucles anticodón, por ejemplo, con una secuencia CU(X)_nXXXAA (donde n=1), el codón de cuatro o más bases se lee como un único aminoácido. Por ejemplo, véanse las SEQ ID NO: 6, 12 del documento PCT/US04/22061 para los O-ARNt que reconocen un codón de cuatro bases. En otras realizaciones, los bucles anticodón 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. debido a que existen 256 posibles codones de cuatro bases, se pueden codificar múltiples aminoácidos artificiales en la misma célula utilizando un codón de cuatro o más bases. Véanse, Anderson y col., (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 procedimientos biosintéticos in vitro. Véanse, por ejemplo, Ma y col., (1993) *Biochemistry*, 32:7939; y Hohsaka y col., (1999) *J. Am. Chem. Soc.*, 121:34. Se utilizaron CGGG y AGGU para incorporar simultáneamente 2-naftilalanina y un derivado NBD de lisina en estreptavidina in vitro con dos ARNt supresores del cambio de marco químicamente acilados. Véase, por ejemplo, Hohsaka y col., (1999) *J. Am. Chem. Soc.*, 121:12194. En un estudio in vivo, Moore y col. examinaron la capacidad de los derivados de ARNtLeu con los anticodones NCUA de suprimir los codones UAGN (N puede ser U, A, G, o C), y encontraron que el cuadruplete UAGA puede ser decodificado por el ARNtLeu con el anticodón UCUA con una eficacia del 13 al 26% con una pequeña descodificación en el marco 0 o -1. Véase, Moore y col., (2000) *J. Mol. Biol.*, 298:195. En una realización, se pueden utilizar en la invención codones

extendidos basados en codones raros o codones finalizadores, que pueden reducir la lectura de sentido incorrecto y la supresión del cambio de marco en otros sitios no deseados.

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

Los codones selectores incluyen opcionalmente pares de bases artificiales. Estos pares de bases artificiales expanden además el alfabeto genético existente. Un par de bases extra aumenta el número de codones de tripletes de 64 a 125. Las propiedades del tercer par de bases incluyen el emparejamiento de bases estable y selectivo, una incorporación enzimática eficaz en el ADN con una alta fidelidad por una polimerasa, y continuó la extensión eficaz del cebador después de la síntesis del par de bases artificial nascente. Las descripciones de los pares de bases artificiales que se pueden adaptar para los procedimientos y composiciones incluye, por ejemplo, Hirao, y col., (2002) An unnatural base pair for incorporating amino acid analogues into protein, *Nature Biotechnology*, 20:177-182. Véase, también, Wu, Y., y col., (2002) *J. Am. Chem. Soc.* 124:14626-14630. Se relacionan a continuación otras publicaciones relevantes.

Para la utilización in vivo, el nucleótido artificial es permeable a la membrana y se fosforila 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 aprovechan los modelos de unión del hidrógeno que son diferentes de los pares de Watson-Crick convencionales, el ejemplo más notable del cual es el par iso-C:iso-G. Véase, por ejemplo, Switzer y col., (1989) *J. Am. Chem. Soc.*, 111:8322; y Piccirilli y col., (1990) *Nature*, 343:33; Kool, (2000) *Curr. Opin. Chem. Biol.*, 4:602. Estas bases se emparejan en general incorrectamente en algún grado con bases naturales y no pueden replicarse enzimáticamente. Kool y colaboradores demostraron que las interacciones de empaquetamiento hidrófobas entre bases pueden sustituir el enlace de hidrógeno para impulsar la formación del par de bases. Véase, Kool, (2000) *Curr. Opin. Chem. Biol.*, 4:602; y Guckian y Kool, (1998) *Angew. Chem. Int. Ed. Engl.*, 36, 2825. En un esfuerzo por desarrollar un par de bases artificial que satisfaga todos los requisitos anteriores, Schultz, Romesberg y colaboradores han sintetizado y estudiado sistemáticamente una serie de bases hidrófobas artificiales. Se encuentra que un autopar PICS:PICS es más estable que los pares de bases naturales, y puede incorporarse eficazmente en el ADN por el fragmento Klenow de la ADN polimerasa I de *Escherichia coli* (KF). Véase, por ejemplo, McMinn y col., (1999) *J. Am. Chem. Soc.*, 121: 11585-6; y Ogawa y col., (2000) *J. Am. Chem. Soc.*, 122:3274. Se puede sintetizar un autopar 3MN:3MN por el KF con una eficacia y selectividad suficientes para la función biológica. Véase, por ejemplo, Ogawa y col., (2000) *L Am. Chem. Soc.*, 122:8803. Sin embargo, ambas bases actúan como un terminador de cadena para la replicación adicional. Una ADN polimerasa mutante que se ha hecho evolucionar recientemente se puede usar para replicar el autopar de PICS. Además, se puede replicar un autopar 7AI. Véase, por ejemplo, Tae y col., (2001) *J. Am. Chem. Soc.*, 123:7439. Se ha desarrollado también un novedoso par de metalobases, Dipic:Py, que forma un par estable tras la unión de Cu(II). Véase, Meggers y col., (2000) *J. Am. Chem. Soc.*, 122:10714. Debido a que los codones extendidos y los codones artificiales son intrínsecamente ortogonales a los codones naturales, los procedimientos de la presente divulgación puede aprovecharse de esta propiedad para generar ARNt ortogonales para los mismos.

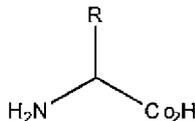
Se puede usar también un sistema de derivación de la traducción para incorporar un aminoácido seleccionado, por ejemplo, un aminoácido artificial, en un polipéptido deseado. En un sistema de derivación de la traducción, se inserta una secuencia grande en un gen, pero no se traduce en la proteína. La secuencia contiene una estructura que sirve como una señal para inducir al ribosoma a saltar sobre la secuencia y resumir la traducción en la dirección 3' de la inserción.

AMINOÁCIDOS SELECCIONADOS Y ARTIFICIALES

Tal como se usa en el presente documento, un aminoácido seleccionado se refiere a cualquier aminoácido que se produce naturalmente o aminoácido artificial. Un aminoácido que se produce naturalmente incluye uno cualquiera de los veinte alfa-aminoácidos codificados genéticamente: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. En una realización, El aminoácido seleccionado se incorpora en un polipéptido en crecimiento con una alta fidelidad, por ejemplo, de más de aproximadamente un 70% de eficacia para un codón selector dado, de más de un 75% de eficacia para un codón selector dado, de más de aproximadamente un 80% de eficacia para un codón selector dado, de más de aproximadamente un 85% de eficacia para un codón selector dado, de más de aproximadamente un 90% de eficacia para un codón selector dado, a más de aproximadamente un 95% de eficacia para un codón selector dado, o más de aproximadamente un 99% o más de eficacia para un codón selector dado.

Tal como se usa en el presente documento, un aminoácido artificial se refiere a cualquier aminoácido, aminoácido modificado, o análogo de aminoácido diferente que la selenocisteína y/o pirrolisina y los siguiente veinte alfa-aminoácidos clásicos codificados genéticamente: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. La estructura genérica de un alfa-aminoácido se ilustra por la Fórmula I:

I

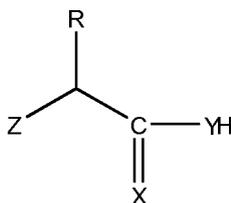


Un aminoácido artificial es normalmente cualquier estructura que tiene la Fórmula I en la que el grupo R es cualquier sustituyente diferente que uno utilizado en los veinte aminoácidos naturales. Véase, por ejemplo, Biochemistry by L. Stryer, 3ª ed. 1988, Freeman and Company, Nueva York, para las estructuras de los veinte aminoácidos naturales. Obsérvese que, los aminoácidos artificiales de la presente invención pueden ser compuestos que se producen naturalmente diferentes de los veinte aminoácidos anteriores.

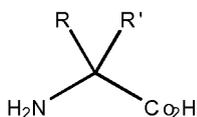
debido a que los aminoácidos artificiales de la presente invención difieren normalmente de los aminoácidos naturales únicamente en la estructura de la cadena secundaria, los aminoácidos artificiales forman enlaces amida con otros aminoácidos, incluyendo, aunque no de forma limitativa, naturales y artificiales, de la misma manera en la que se forman las proteínas que se producen naturalmente. Sin embargo, los aminoácidos artificiales tiene grupos en la cadena secundaria que los distinguen de los aminoácidos naturales. Por ejemplo, R en la Fórmula I puede comprender un alquilo, arilo, acilo, ceto, azido, hidroxilo, hidrazina, ciano, halo, hidrazida, alquenilo, alquinilo, éter, tiol, seleno, sulfonilo, borato, boronato, fosfo, fosfona, fosfina, heterocíclico, enona, imina, aldehído, éster, tioácido, hidroxilamina, amina, y similares, o cualquier combinación de los mismos. Otros aminoácidos que no se producen naturalmente incluyen, aunque no de forma limitativa, aminoácidos que comprenden un reticulador fotoactivable, aminoácidos marcados con espín, aminoácidos fluorescentes, aminoácidos de unión a metales, aminoácidos que contienen metales, aminoácidos radioactivos, aminoácidos con novedosos grupos funcionales, aminoácidos que actúan de forma covalente o no covalente con otras moléculas, aminoácidos fotoenjaulados y/o fotoisomerizables, aminoácidos que comprende biotina o un análogo de biotina, aminoácidos glicosilados tales como una serina sustituida con azúcar, otros aminoácidos modificados con hidratos de carbono, aminoácidos que contienen ceto, aminoácidos que comprenden polietilenglicol o poliéter, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles y/o fotoescindibles, aminoácidos con cadenas secundarias alargadas en comparación con los aminoácidos naturales, incluyendo, aunque no de forma limitativa, poliéteres o hidrocarburos de cadena larga, incluyendo, aunque no de forma limitativa, mayores de aproximadamente 5 o mayores de aproximadamente 10 carbonos, aminoácidos que contienen azúcares unidos a carbono, aminoácidos activos redox, aminoácidos que contienen aminoácidos, y aminoácidos que comprenden uno o más restos tóxicos. Véase, también, publicaciones de solicitudes de patente de Estados Unidos 2003/0082575 y 2003/0108885. Los aminoácidos artificiales pueden tener un reticulador fotoactivable que se usa, por ejemplo, para unir una proteína a un soporte sólido. Los aminoácidos artificiales pueden tener un resto sacárido unido a la cadena secundaria del aminoácido.

Además de los aminoácidos artificiales que contienen novedosas cadenas secundarias, los aminoácidos artificiales comprenden opcionalmente también estructuras principales modificadas, por ejemplo, como se ilustra por las estructuras de Fórmula II y III:

II



III



en las que Z comprende normalmente OH, NH₂, SH, NH-R', o S-R'; X e Y, que pueden ser iguales o diferentes, normalmente comprenden S u O, y R y R', que son opcionalmente iguales o diferentes, se seleccionan normalmente entre 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 pueden comprender sustituciones en el grupo amino o carboxilo, como se ilustra por las Fórmulas II y III. Los aminoácidos artificiales de este tipo

incluyen, aunque no de forma limitativa, α -hidroxiácidos, α -tioácidos, α -aminotiocarboxilatos, por ejemplo, con cadenas secundarias que corresponden a los veinte aminoácidos comunes naturales o cadenas secundarias artificiales. Además, las sustituciones en el carbono α incluyen opcionalmente aminoácidos L, D, o α - α -disustituidos tales como D-glutamato, D-alanina, D-metil-O-tirosina, ácido aminobutírico, y similares. Otras alternativas estructurales incluyen aminoácidos cíclicos, tales como análogos de prolina así como análogos de prolina con un anillo de 3, 4, 6, 7, 8 y 9 miembros, β y γ aminoácidos tales como β -alanina y ácido γ -amino butírico sustituidos.

Muchos aminoácidos artificiales se basan en aminoácidos naturales, tales como tirosina, glutamina, fenilalanina, y similares. Los análogos de tirosina incluyen tirosinas parasustituidas, tirosinas ortosustituidas, y tirosinas metasustituidas, en las que la tirosina sustituida comprende un grupo ceto (incluyendo, aunque no de forma limitativa, un grupo acetilo), un grupo benzoílo, un grupo amino, una hidrazina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo de cadena lineal $C_6 - C_{20}$ o un hidrocarburo de cadena ramificada, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, o similares. Además, se contemplan también múltiples anillos de arilo sustituidos. Los análogos de glutamina incluyen, aunque no de forma limitativa, α -hidroxiderivados, derivados γ -sustituidos, derivados cíclicos, y derivados de glutamina sustituidos con amida. Los análogos de fenilalanina ilustrativos incluyen, aunque no de forma limitativa, fenilalaninas parasustituidas, fenilalaninas ortosustituidas, y fenilalaninas metasustituidas, en las que el sustituyente comprende un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído, un azido, un yodo, un bromo, un grupo ceto (incluyendo, aunque no de forma limitativa, un grupos acetilo), o similares. Los ejemplos específicos de aminoácidos artificiales incluyen, aunque no de forma limitativa, una p-acetil-L-fenilalanina, una p-propargil fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metil-fenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una tri-O-acetil-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 fosfotirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina, una isopropil-L-fenilalanina, y una p-propargiloxi-fenilalanina, y similares. Se proporcionan ejemplos de estructuras de una variedad de aminoácidos artificiales en, por ejemplo, el documento WO 2002/085923 titulado "In vivo incorporation of unnatural amino acids". Véase también Kiick y col., (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, PNAS 99:19-24erein, para los análogos de metionina adicionales.

Un aminoácido no natural incorporado en un polipéptido en el extremo amino puede estar compuesto de un grupo R que es cualquier sustituyente diferente de uno utilizado en los veinte aminoácidos naturales y un 2º grupo reactivo diferente del grupo NH_2 normalmente presente en los α -aminoácidos (véase la Fórmula I). Puede incorporarse un aminoácido no natural en el extremo carboxilo con un 2º grupo reactivo diferente del grupo COOH normalmente presente en los α -aminoácidos (véase la Fórmula I).

Se pueden seleccionar los aminoácidos artificiales de la invención o diseñarse para proporcionar características adicionales no disponibles en los veinte aminoácidos naturales. Por ejemplo, el aminoácido artificial puede diseñarse o seleccionarse opcionalmente para modificar las propiedades biológicas de una proteína, por ejemplo, en la que se incorporan. Por ejemplo, las siguientes propiedades pueden modificarse opcionalmente mediante la inclusión de un aminoácido artificial en una proteína: toxicidad, biodistribución, solubilidad, estabilidad, por ejemplo, térmica, hidrolítica, oxidativa, resistencia a la degradación enzimática, y similares, facilidad de purificación y procesamiento, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, actividad catalítica, potencial redox, semivida, capacidad de reaccionar con otras moléculas, por ejemplo, de forma covalente o no covalente, y similares.

Se proporcionan las estructuras de una variedad de aminoácidos artificiales en, por ejemplo, Las Figuras 16, 17, 18, 19, 26, y 29 del documento WO 2002/085923 titulado "In vivo incorporation of unnatural amino acids". No se entiende que los ejemplos sean limitantes en ningún modo de los aminoácidos que se pueden unir al ARNt de la presente divulgación.

Una ventaja de un aminoácido artificial es que este presenta restos químicos adicionales que se pueden utilizar para añadir moléculas adicionales. Estas modificaciones se pueden realizar in vivo en células eucariotas o no eucariotas, o in vitro. De esta manera, en determinadas realizaciones, la modificación posterior a la traducción es a través del aminoácido artificial. Se puede usar un aminoácido artificial en un polipéptido para unir otras molécula al polipoéptido, incluyendo, aunque no de forma limitativa, una etiqueta, un colorante, un polímero, un polímero soluble en agua, un derivado de polietilenglicol, un fotorreticulador, un radionucleido, un compuesto citotóxico, un fármaco, una marca de afinidad, una marca de fotoafinidad, un compuesto reactivo, una resina, una segunda proteína o polipéptido o un análogo de polipéptido, un anticuerpo o fragmento de anticuerpo, un quelante metálico, un cofactor, un ácido graso, un hidrato de carbono, un polinucleótido, un ADN, un ARN, un polinucleótido de sentido contrario, un sacárido, un dendrímero soluble en agua, una ciclodextrina, un ácido ribonucleico inhibidor, un biomaterial, una nanopartícula, una marca de espín, un fluoróforo, un resto que contiene metal, un resto radioactivo, un grupo funcional novedoso, un grupo que interactúa de forma covalente o no covalente con otras moléculas, un resto fotoenjaulado, un resto excitable por radiación actínica, un resto fotoisomerizable, biotina, un derivado de biotina, un análogo de biotina, un resto que incorpora un átomo pesado, un grupo escindible químicamente, un grupo fotoescindible, una cadena secundaria alargada, un azúcar unido a carbono, un agente activo redox, un aminotioácido, un resto tóxico, un resto marcado isotópicamente, una sonda biofísica, un grupo fosforescente, un

grupo quimioluminiscente, un grupo denso en electrones, un grupo magnético, un grupo intercalante, un cromóforo, un agente de transferencia de energía, un agente biológicamente activo, una marca detectable, una molécula pequeña, un punto cuántico, un nanotransmisor, o cualquier combinación de los anteriores o cualquier otro compuesto o sustancia deseable, que comprende un segundo grupo reactivo para al menos un aminoácido artificial que comprende un primer grupo reactivo que utiliza una metodología química que es conocida por una persona normalmente experta en la técnica por ser adecuado para los grupos reactivos concretos.

Por ejemplo, la modificación posterior a la traducción puede ser a través de una reacción nucleófila-electrófila. La mayoría de reacciones utilizadas actualmente para la modificación selectiva de proteínas implica la formación de enlaces covalentes entre moléculas reactivas en reacciones nucleófilas y electrófilas, incluyendo, aunque no de forma limitativa la reacción de las α -halocetonas con cadenas secundarias de histidina o cisteína. La selectividad en estos casos se determina por el número y la accesibilidad de los restos nucleófilos en la proteína. En proteínas de la descripción, se pueden usar otras reacciones más selectivas tales como la reacción de un cetoaminoácido artificial con hidrazidas o aminooxicompuestos, *in vitro* e *in vivo*. Véase, *por ejemplo*, Cornish, y col., (1996) J. Am. Chem. Soc., 118:8150-8151; Mahal, y col., (1997) Science, 276:1125-1128; Wang, y col., (2001) Science 292:498-500; Chin, y col., (2002) J. Am. Chem. Soc. 124:9026-9027; Chin, y col., (2002) Proc. Natl. Acad. Sci., 99:11020-11024; Wang, y col., (2003) Proc. Natl. Acad. Sci., 100:56-61; Zhang, y col., (2003) Biochemistry, 42:6735-6746; y, Chin, y col., (2003) Science, 301:964-7. Esto permite el marcado selectivo de virtualmente cualquier proteína con un hospedador de reactivos incluyendo fluoróforos, agentes de reticulación, derivados de sacáridos y moléculas citotóxicas. Véase *también*, patente de Estados Unidos n.º 6.927.042 titulada "Glycoprotein synthesis". Pueden realizarse también modificaciones posteriores a la traducción, incluyendo, aunque no de forma limitativa, a través de un azido aminoácido, mediante la ligadura de Staudinger (incluyendo, aunque no de forma limitativa, con reactivos de triarilfosfina). Véase, *por ejemplo*, Kiick y col., (2002) Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation, PNAS 99:19-24.

Síntesis química de aminoácidos artificiales

Están comercialmente disponibles muchos aminoácidos artificiales, por ejemplo, de Sigma-Aldrich (St. Louis, MO, EE.UU.), Novabiochem (una división de EMD Biosciences, Darmstadt, Alemania), o Peptech (Burlington, MA, EE.UU.). Aquellos que no están comercialmente disponibles se sintetizan opcionalmente como se proporciona en el presente documento o utilizando los procedimientos normalizados conocidos por las personas normalmente expertas en la materia. Para las técnicas de síntesis orgánicas, véanse, por ejemplo, Organic Chemistry de Fessenden y Fessenden, (1982, Segunda Edición, Willard Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Third Edition, 1985, Wiley and Sons, Nueva York); y Advanced Organic Chemistry by Carey and Sundberg (Third Edition, Partes A y B, 1990, Plenum Press, Nueva York). Additional publications describing the synthesis of unnatural amino acids include, por ejemplo, documento WO 2002/085923 título "In vivo incorporation of Unnatural Amino Acids;" Matsoukas y col., (1995) J. Med. Chem., 38, 4660-4669; King, F.E. & Kidd, D.A.A. (1949) A New Synthesis of Glutamine and of γ Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319; Friedman, O.M. 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. y col. (1988) Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1-methylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. y Frappier, F. (1991) Glutamine analogues as Potential Antimalarials, Eur. J. Med. Chem. 26, 201-5; Koskinen, A.M.P. y Rapoport, H. (1989) Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859-1866; Christie, B.D. y Rapoport, H. (1985) Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization. J. Org. Chem. 50:1239-1246; Barton y col., (1987) Synthesis of Novel α -Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D- α -Amino-Adipic Acids, L- α -aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron 43:4297-4308; y, Subasinghe y col., (1992) Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35:4602-7. Véase también, publicación de patente de Estados Unidos n.º 2004/0198637 titulada "Protein Arrays".

Captación celular de aminoácidos artificiales

La captación de aminoácidos artificiales por una célula es un problema que se considera normalmente cuando se diseñan y se seleccionan aminoácidos artificiales, por ejemplo, para la incorporación a una proteína. Por ejemplo, la alta densidad de carga de los α -aminoácidos sugiere que estos compuestos es improbable que sean permeables para las células. Los aminoácidos naturales son capturados en la célula mediante una serie de sistemas de transporte basados en proteínas. Se puede llevar a cabo un rápido cribado que evalúe cuales de los aminoácidos artificiales, si los hay, son capturados por las células. Véase, *por ejemplo*, los ensayos de toxicidad en, por ejemplo, la publicación de patente de Estados Unidos n.º US 2004/0198637 titulada "Protein Arrays", y Liu, D.R. & Schultz, P.G. (1999) Progress toward the evolution of an organism with an expanded genetic code. PNAS United States 96:4780-4785. Aunque la captación se analiza fácilmente con diversos ensayos, una alternativa para diseñar los aminoácidos artificiales que son adecuados para las rutas de captación celular es proporcionar rutas biosintéticas para crear aminoácidos *in vivo*.

Biosíntesis de aminoácidos artificiales

Muchas rutas biosintéticas existen ya en las células para la producción de aminoácidos y otros compuestos. Aunque un procedimiento biosintético para un aminoácido artificial concreto puede no existir en la naturaleza, incluyendo, aunque no de forma limitativa, en una célula, la descripción proporciona dichos procedimientos. Por ejemplo, las rutas biosintéticas para aminoácidos artificiales se generan opcionalmente en una célula hospedadora añadiendo nuevas enzimas o modificando las rutas de las células hospedadoras existentes. Las nuevas enzimas adicionales son enzimas que se producen naturalmente opcionalmente o enzimas evolucionadas artificialmente. Por ejemplo, la biosíntesis de p-aminofenilalanina (como 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 de estas enzimas se pueden introducir en una célula transformando la célula con un plásmido que comprende los genes. Los genes, cuando se expresan en la célula, proporcionan una ruta enzimática para sintetizar el compuesto deseado. En los siguientes ejemplos se proporcionan ejemplos de los tipos de enzimas que se añaden opcionalmente. Se encuentran las secuencias de enzimas adicionales, por ejemplo, en el Genbank. Las enzimas evolucionadas artificialmente se añaden también opcionalmente en una célula de la misma manera. De esta manera, La maquinaria celular y los recursos de una célula se manipulan para producir aminoácidos naturales.

Están disponibles una variedad de procedimientos para producir novedosas enzimas para el uso en las rutas biosintéticas o para la evolución de las rutas existentes. Por ejemplo, una recombinación recurrente, por ejemplo, como desarrolla Maxygen, Inc (disponible en el sitio web en maxygen.com), se utiliza opcionalmente para desarrollar novedosas enzimas y rutas. Véase, por ejemplo, Stemmer (1994), Rapid evolution of a protein in vitro by DNA shuffling, Nature 370(4):389-391; y, Stemmer, (1994), DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution, Proc. Natl. Acad. Sci. USA., 91:10747-10751. Similarly DesignPath™, desarrollada por Genencor (disponible en el sitio web en genencor.com) se utiliza para diseñar mediante ingeniería genética rutas metabólicas, por ejemplo, para diseñar mediante ingeniería genética una ruta para crear O-metil-L-tirosina en una célula. Esta tecnología reconstruye las rutas existentes en organismos hospedadores utilizando una combinación de nuevos genes, incluyendo, aunque no de forma limitativa, aquellas identificadas a través de la genómica funcional, y la evolución y el diseño molecular. La Diversa Corporation (disponible en el sitio web en diversa.com) proporciona también tecnología para cribar rápidamente bibliotecas de genes y rutas de genes, incluyendo, aunque no de forma limitativa, para crear nuevas rutas.

Normalmente, el aminoácido artificial producido con una ruta biosintética diseñada mediante ingeniería genética de la presente divulgación se produce en una concentración suficiente para una biosíntesis de la proteína eficaz, por ejemplo, una cantidad celular natural, pero no hasta tal grado que afecte la concentración de los otros aminoácidos o agote los recursos celulares. Las concentraciones típicas producidas *in vivo* de esta manera son aproximadamente de 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 usan opcionalmente selecciones *in vivo* para optimizar adicionalmente el aminoácido artificial para la síntesis de las proteínas ribosómicas y el crecimiento celular.

SECUENCIA DE ÁCIDO NUCLEICO Y VARIANTES

Como se describe anteriormente y a continuación, se desvelan secuencias de polinucleótidos de ácidos nucleicos y secuencias de aminoácidos de polipéptidos, por ejemplo, ARNt y RS, y, por ejemplo, las composiciones y procedimientos que comprenden dichas secuencias. Se describen en el presente documento los ejemplos de dichas secuencias por ejemplo, ARNt y RS. Sin embargo, una persona experta en la materia apreciará que la descripción no está limitada a las secuencias desveladas en el presente documento, por ejemplo, los Ejemplos. Un experto apreciará que la descripción proporciona también muchas secuencias relacionadas y no relacionadas con las funciones descritas en el presente documento, por ejemplo, que codifican un O-ARNt o una O-RS.

La descripción proporciona polipéptidos (O-RS) y polinucleótidos, por ejemplo, un O-ARNt, los polinucleótidos que codifican la O-RS o las partes de los mismos, los oligonucleótidos utilizados para aislar los clones de aminoácil ARNt sintetasa, etc. Los polinucleótidos incluyen aquellos que codifican las proteínas o polipéptidos de interés con uno o más codones selectores. Además, los polinucleótidos incluyen, por ejemplo, un polinucleótido que comprende una secuencia de nucleótidos que se muestra en una cualquiera de la SEQ ID NO: 1, 2, 3; un polinucleótido que es complementario, o una variación conservativa del mismo. De manera similar, un ácido nucleico que se hibrida a un polinucleótido indicado anteriormente en condiciones muy rigurosas sobre sustancialmente la longitud completa del ácido nucleico es un polinucleótido de la presente divulgación.

En determinadas realizaciones, un vector (por ejemplo, un plásmido, un cósmido, un bacteriófago, una bacteria, un virus, un polinucleótido puro, un polinucleótido conjugado, 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 operativamente 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 presente divulgación.

Un experto apreciará también que se incluyen en la invención muchas variantes de las secuencias desveladas. Por

ejemplo, están incluidas en la invención variaciones conservativas de las secuencias desveladas que dan como resultado una secuencia funcionalmente idéntica. Las variantes de secuencias de polinucleótidos de ácidos nucleicos, en las que las variantes se hibridan con al menos una secuencia desvelada, se consideran que están incluidas en la invención. Las subsecuencias únicas de las secuencias desveladas en el presente documento, como se determinan mediante, por ejemplo, las técnicas de comparación de secuencias normalizadas, están también incluidas en la invención.

Variaciones conservativas

Debido a la degeneración inherente del código genético, las "sustituciones silenciosas" (*es decir*, las sustituciones en una secuencia de ácido nucleico que no dan como resultado una alteración en un polipéptido codificado) son una característica implícita de cada secuencia de ácido nucleico que codifica un aminoácido. De manera similar, las "sustituciones de aminoácidos conservativas", en uno o en unos pocos aminoácidos en una secuencia de aminoácidos están sustituidos con diferentes aminoácidos con propiedades muy similares, se identifican también fácilmente siendo muy similares a una construcción desvelada. Dichas variaciones conservativas de cada secuencia desvelada son una característica de la presente divulgación.

"Variaciones conservativa de una secuencia de ácido nucleico concreta se refieren a aquellos ácidos nucleicos que codifican secuencias de aminoácidos idénticas o esencialmente idénticas, o, donde el ácido nucleico no codifica una secuencia de aminoácido, en secuencias esencialmente idénticas. Una persona normalmente experta en la materia reconocerá que las sustituciones, deleciones o adiciones individuales que alteran, añaden o eliminan un único aminoácido o un pequeño porcentaje de aminoácidos en una secuencia codificada son "variaciones modificadas conservativamente" o "variantes modificadas conservativamente" donde las alteraciones ran como resultado la deleción de un aminoácido, la adición de un aminoácido, o la sustitución de un aminoácido con un aminoácido químicamente similar. De esta manera, las "variaciones conservativas" de una secuencia polipeptídica relacionada de la presente divulgación incluyen sustituciones de un pequeño porcentaje, normalmente menos del 5%, más normalmente menos del 4%, 2% o 1%, de los aminoácidos de la secuencia polipeptídica, con un aminoácido seleccionado conservativamente del mismo grupo de sustituciones conservativas. 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 son bien conocidas por las personas normalmente expertas en la materia. Los siguientes ocho grupos contienen aminoácidos que son sustituciones conservativas unos de otros:

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

(véase, por ejemplo, Creighton, Proteins: Structures and Molecular Properties (W H Freeman & Co.; 2ª edición (Diciembre 1993)

Hibridación del ácido nucleico

Se puede usar la hibridación comparativa para identificar ácidos nucleicos de la presente divulgación, tales como la SEQ ID NO: 1-3, que incluye variaciones conservativas de ácidos nucleicos de la presente divulgación, y este procedimiento de hibridación comparativa es un procedimiento preferido de distinguir ácidos nucleicos de la presente divulgación. Además, los ácidos nucleicos diana que se hibridan con los ácidos nucleicos representados por la SEQ ID NO: 1-3 en condiciones de rigor elevado, rigor ultraelevado, y/o rigor ultra-ultra elevado son una característica de la presente divulgación. Los ejemplos de dichos ácidos nucleicos incluyen aquellos con una o unas pocas sustituciones de ácidos nucleicos silenciosas o conservativas en comparación con una secuencia de ácido nucleico dada.

Se dice que un ácido nucleico de ensayo hibrida específicamente a una sonda de ácido nucleico cuando hibrida también al menos 1/2 de la sonda tal como la diana complementaria perfectamente, es decir, con una relación señal a ruido al menos 1/2 tal alta como la hibridación de la sonda a la diana en condiciones en las que la sonda perfectamente emparejada se une a la diana complementaria perfectamente emparejada con una relación señal a ruido que es al menos aproximadamente 5x-10x tan alta como la observada para la hibridación de cualquiera de los ácidos nucleicos diana no emparejados.

Los ácidos nucleicos se "hibridan" cuando se asocian, normalmente en solución. Los ácidos nucleicos se hibridan debido a una variedad de fuerzas fisicoquímicas bien caracterizadas, tales como el enlace de hidrógeno, la exclusión del disolvente, el apilado de bases y similares. La frase "condiciones de hibridación rigurosas" se refiere a

condiciones de baja fuerza iónica y alta temperatura como se conoce en la técnica. Normalmente, en condiciones rigurosas, una sonda se hibridará con su diana posterior en una mezcla compleja de ácido nucleico (que incluye, aunque no forma limitativa, ADN o ARN celular total o de biblioteca) pero no se hibrida con otras secuencias en la mezcla compleja. Se encuentra una guía extensa de la hibridación de ácidos nucleicos en Tijssen (1993) *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes* part I capítulo 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, Nueva York), así como en Ausubel y col., *Current Protocols in Molecular Biology* (1995). Hames y Higgins (1995) *Gene Probes 1* IRL Press at Oxford University Press, Oxford, Inglaterra, (Hames y Higgins 1) y Hames y Higgins (1995) *Gene Probes 2* IRL Press at Oxford University Press, Oxford, Inglaterra (Hames y Higgins 2) proporcionan detalles de la síntesis, marcado, detección y cuantificación del ADN y el ARN, incluyendo los oligonucleótidos. En general, se seleccionan las condiciones rigurosas para ser aproximadamente 5-10°C menores que el punto de fusión térmica (T_m) para la secuencia específica un pH con una fuerza iónica definida. La T_f es la temperatura (para una fuerza iónica, pH, y concentración de ácido nucleico definidas) a la cual el 50% de las sondas complementarias con la diana se hibridan a la secuencia diana en equilibrio) ya que las secuencias diana están presentes en exceso, a T_f , el 50% de las sondas están ocupadas en equilibrio). Las condiciones rigurosas pueden ser aquellas en las que la concentración de sal es menor de aproximadamente 1,0 M de ion sodio, normalmente aproximadamente 0,01 a 1,0 M de concentración de ion sodio (u otras sales) a pH 7,0 a 8,3 y la temperatura es al menos de aproximadamente 30°C para sondas cortas (incluyendo, aunque no de forma limitativa, 10 a 50 nucleótidos) y al menos aproximadamente 60°C para sondas largas (incluyendo, aunque no de forma limitativa, más de 50 nucleótidos). También se pueden alcanzar condiciones rigurosas con la adición de agentes desestabilizantes, tales como formamida. Para la hibridación selectiva o específica, una señal positiva puede tener al menos un fondo de dos veces, opcionalmente u a hibridación de fondo de 10 veces. Las condiciones rigurosas de hibridación ilustrativas pueden ser las siguientes: 50 % de formamida, 5X SSC y 1 % de SDS, incubando a 42°C o 5X SSC, 1% de SDS, incubando a 65°C, con lavado en 0,2X SSC, y 0,1% de SDS a 65°C. Dichos lavados se pueden llevar a cabo durante 5, 15, 30, 60, 120, o más minutos.

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

Las "condiciones rigurosas del lavado de la hibridación" en el contexto de los experimentos de hibridación de ácidos nucleicos tales como las hibridaciones Southern y Northern son dependientes de la secuencia y son diferentes bajo parámetros ambientales diferentes. Las secuencias más largas se hibridan específicamente a temperaturas más altas. Se encuentra una extensa guía de la hibridación de ácidos nucleicos en Tijssen (1993), *anteriormente*. y en Hames y Higgins, 1 y 2, anteriormente citados. La hibridación y las condiciones de lavado rigurosa pueden determinarse fácilmente de forma empírica para cualquier ácido nucleico de ensayo. Por ejemplo, en la determinación de las condiciones de hibridación y lavado muy rigurosas, la hibridación y las condiciones de lavado se aumentan gradualmente (por ejemplo, aumentando la temperatura, disminuyendo la concentración de sales, aumentando la concentración del 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, la hibridación y las condiciones de lavado se aumentan gradualmente hasta que la sonda se une a una diana complementaria perfectamente emparejada con una relación señal a ruido que es al menos 5c tan elevada como la observada para la hibridación de la sonda a una diana no emparejada.

Las condiciones "muy rigurosas" se seleccionan para ser iguales al punto de fusión térmica (T_m) para una sonda concreta. La T_f es la temperatura (para una fuerza iónica y pH definidos) a la cual el 50% de la secuencia de ensayo se hibrida con una sonda perfectamente emparejada. Para los fines de la presente invención, en general, las condiciones de hibridación y lavado "muy rigurosas" se seleccionan para ser aproximadamente 5°C menos que la T_f de la secuencia específica a una fuerza iónica y pH definidos.

Las condiciones de hibridación y lavado "ultra muy rigurosas" son aquellas en las que el rigor de las condiciones de hibridación y lavado se aumentan hasta que la relación señal a ruido para la unión de la sonda al ácido nucleico complementario perfectamente emparejado es al menos 10x tal alta como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no emparejados. Un ácido nucleico diana que se hibrida a una sonda bajo dichas condiciones, con una relación señal a ruido de al menos 1/2 de la del ácido nucleico diana complementario perfectamente emparejado se dice que se une a la sonda en condiciones de rigor ultraelevado.

De manera similar, se pueden determinar niveles incluso mayores de rigor aumentando gradualmente las condiciones de hibridación y/o lavado del ensayo de hibridación relevante. Por ejemplo, aquellos en los cuales las condiciones de rigor de la hibridación y el lavado se aumentan hasta que la relación señal a ruido para la unión de la sonda al ácido nucleico diana complementario perfectamente emparejado es al menos de 10X, 20X, 50X, 100X, o

500X o más tan elevada como la observada para la hibridación con cualquiera de los ácidos nucleicos diana no emparejados. Un ácido nucleico diana que se hibrida a una sonda bajo dichas condiciones, con una relación señal a ruido de al menos 1/2 de la del ácido nucleico diana complementario perfectamente emparejado se dice que se une a la sonda en condiciones de rigor ultra-ultra elevadas.

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

Subsecuencias únicas

- 10 Se proporciona también un ácido nucleico que comprende una única subsecuencia en un ácido nucleico seleccionado entre las secuencias de los O-ARNt y las O-RS desveladas en el presente documento. La única subsecuencia es única en comparación con un ácido nucleico que corresponde a cualquier secuencia de ácido nucleico del O-ARNt o la O-RS. Se puede llevar a cabo la alineación utilizando, por ejemplo, el conjunto BLAST para los parámetros por defecto. Cualquier subsecuencia única es útil, por ejemplo, como una sonda para identificar los ácidos nucleicos de la presente divulgación.

- 15 De manera similar, la descripción incluye un polipéptido que comprende una única subsecuencia en un polipéptido seleccionado entre las secuencias de las O-RS desveladas en el presente documento. En este caso, la única subsecuencia es única en comparación con un polipéptido correspondiente a cualquiera de las secuencias polipeptídicas conocidas.

- 20 Se desvelan también ácidos nucleicos que se hibridan en condiciones rigurosas a un único oligonucleótido de codificación que codifica una única subsecuencia en un polipéptido seleccionado entre las secuencias de las O-RS en las que la única subsecuencia es única en comparación con un polipéptido correspondiente a cualquiera de los polipéptidos control (por ejemplo, secuencias precursoras a partir de las cuales se derivan las sintetasas de la presente divulgación, por ejemplo, mediante mutación). Las secuencias únicas se determinan como se ha indicado anteriormente.

Comparación de secuencias, identidad y homología

- 25 Los términos "idéntico" o porcentaje de "identidad" en el contexto de dos o más secuencias de ácidos nucleicos o polipéptidos, se refieren a dos o más secuencias o subsecuencias que son iguales o que tienen un porcentaje especificado de restos de aminoácidos o nucleótidos que son iguales, cuando se comparan y alinean para la máxima correspondencia sobre una ventana de comparación, o una región designada, como se mide utilizando uno de los algoritmos de comparación de secuencias descritos a continuación (u otros algoritmos disponibles para las personas normalmente expertas en la materia) o mediante alineación manual e inspección visual.

- 30 la frase "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 tiene al menos aproximadamente 60%, aproximadamente 80%, aproximadamente 90-95%, aproximadamente 98%, aproximadamente 99% o más restos de nucleótidos o aminoácidos de identidad, cuando se comparan y alinean para la máxima correspondencia sobre una ventana de comparación, o una región designada, como se mide utilizando un algoritmo de comparación de secuencias (u otros algoritmos disponibles para personas normalmente expertas en la materia) o mediante alineación manual e inspección visual. Dichas secuencias "sustancialmente idénticas" se consideran normalmente por ser "homólogas", sin referencia al antecesor real. La "identidad sustancial" puede existir sobre una región de las secuencias que tiene al menos aproximadamente 50 restos de longitud, una región de al menos aproximadamente 100 restos, o una región de al menos aproximadamente 150 restos, o sobre la longitud completa de las dos secuencias que se van a comparar.

- 35 Para la comparación de secuencias y la determinación de la homología, normalmente, una secuencia actúa como una secuencia de referencia con la cual se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, se designan las coordenadas de la subsecuencia, en caso necesario, y se designan los parámetros del programa del algoritmo de secuencias. El algoritmo de comparación de secuencias calcula a continuación el porcentaje de identidad de secuencia para la una o varias secuencias de prueba con respecto a la secuencia de referencia, basándose en los parámetros del programa designados.

- 40 Los procedimientos de alineación de secuencias para comparación son bien conocidos por las personas normalmente expertas en la materia. La alineación óptima de las secuencias para comparación puede llevarse a cabo, *por ejemplo*, mediante el algoritmo de homología local de Smith y Waterman, Adv. Appl. Math. 2:482c (1981), mediante el algoritmo de alineación de homología de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), mediante la búsqueda para el procedimiento de similitud de Pearson y Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), mediante las 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 mediante alineación manual e inspección visual (véase *por ejemplo*, Ausubel y col., Current Protocols in Molecular Biology (suplemento de 1995)).

Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul y col. (1997) Nuc. Acids Res. 25:3389-3402, y Altschul y col., J. Mol. Biol. 215:403-410 (1990). El software para realizar los análisis BLAST está disponible para el público a través del National Center for Biotechnology Information en el sitio web en ncbi.nlm.nih.gov. Este algoritmo implica identificar en primer lugar pares de secuencias de puntuación elevada (HSP) identificando palabras cortas de longitud W en la secuencia de búsqueda, que se empareja con cualquiera de o satisface alguna puntuación T umbral de valor positivo cuando se alinea con una palabra de la misma longitud en una secuencia de una base de datos. T es el umbral de la puntuación de la palabra vecina (Altschul y col., citado anteriormente). Estas coincidencias de la palabra vecina inicial actúan como semillas para iniciar las búsquedas para encontrar los HSP que las contienen. Las coincidencias con la palabra se extienden a continuación en ambas direcciones a lo largo de cada secuencia todo lo que la puntuación de la alineación acumulada se pueda incrementar. Las puntuaciones acumuladas se calculan utilizando, para las secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de restos emparejados; siempre > 0) y N (puntuación de penalización para residuos emparejados incorrectamente; siempre < 0). Para las secuencias de aminoácidos, se usa una matriz de puntuación para calcular la puntuación acumulada. La extensión para las coincidencias de la palabra en cada dirección se detienen cuando: la puntuación de alineación acumulada está fuera de la cantidad X de su valor máximo conseguido; la puntuación acumulada baja a cero o por debajo, debido a la acumulación de una o más alineaciones de restos con puntuación negativa; o se alcanza el final de cualquier secuencia. Los parámetros W, T, y X del algoritmo BLAST determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para las secuencias de nucleótidos) utiliza por defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, un corte de 100, M=5, N=-4, y una comparación de ambas hebras. Para las secuencias de aminoácidos, El programa BLASTNP utiliza por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (véase Henikoff & Henikoff (1992) Proc. Natl. Acad. Sci. USA 89:10915) alineaciones (B) de 50, una expectativa (E) de 10, M=5, N=-4, y una comparación de ambas hebras. El algoritmo BLAST se lleva a cabo normalmente con el filtro de "baja complejidad" apagado.

Además de calcular el porcentaje de identidad de la secuencia, El algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias secuencias (véase, por ejemplo, Karlin y Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). Una medida de la similitud proporcionada por el algoritmo BLAST es la menor probabilidad de la suma (P(N)), que proporciona una indicación de la probabilidad por la cual se produciría un emparejamiento entre las dos secuencias de nucleótidos o aminoácidos por casualidad. Por ejemplo, un ácido nucleico se puede considerar similar a una secuencia de referencia si la menor probabilidad de la suma en una comparación del ácido nucleico de ensayo con el ácido nucleico de referencia es inferior a aproximadamente 0,2, menor de aproximadamente 0,01, o menor de aproximadamente 0,001.

Mutagénesis y otras técnicas de biología molecular

Se pueden manipular polinucleótidos y polipéptidos utilizando técnicas de biología molecular. Una secuencia de nucleótidos se puede modificar convenientemente mediante mutagénesis dirigida a sitio de acuerdo con los procedimientos convencionales. Como alternativa, la secuencia de nucleótidos puede prepararse mediante síntesis química, incluyendo, aunque no de forma limitativa, utilizando un sintetizador de oligonucleótidos, en el que los oligonucleótidos se diseñan basándose en la secuencia de aminoácidos del polipéptido deseado, y seleccionado preferentemente aquellos codones que están favorecidos en la célula hospedadora en que se producirá el polipéptido recombinante. Por ejemplo, algunos oligonucleótidos pequeños que codifican partes del polipéptido deseado pueden sintetizarse y ensamblarse mediante la PCR, ligadura o reacción en cadena de ligadura. Véase, por ejemplo, Barany, y col., Proc. Natl. Acad. Sci. 88: 189-193 (1991); Patente de Estados Unidos 6.521.427.

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

Los textos generales que describen las técnicas de biología molecular incluyen Berger y Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook y col., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, Nueva York, 1989 ("Sambrook") y Current Protocols in Molecular Biology, F.M. Ausubel y col., eds., Current Protocols, una unión temporal de empresas entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (suministrada a lo largo de 1999) ("Ausubel"). estos textos describen la mutagénesis, el uso de vectores, promotores y otros muchos temas relevantes relacionados con, por ejemplo, la generación de genes o polinucleótidos que incluyen codones selectores para la producción de proteínas que incluyen aminoácidos seleccionados (por ejemplo, aminoácidos artificiales), ARNt ortogonales, sintetisas ortogonales y parejas de las mismas.

Se utilizan diversos tipos de mutagénesis en la invención para una variedad de fines, incluyendo, aunque no de forma limitativa, para producir sintetisas novedosas o ARNt, para mutar moléculas de ARNt, para producir bibliotecas de ARNt, para mutar moléculas de RS, para producir bibliotecas de sintetisas, para producir codones selectores, para insertar codones selectores que codifican un aminoácido seleccionado en una proteína o polipéptido

de interés. Incluyen, aunque no de forma limitativa, mutagénesis dirigida a sitio, mutagénesis puntuales aleatorias, recombinación homóloga, transposición de secuencias de ADN u otros procedimientos recurrentes de mutagénesis, construcción quimérica, mutagénesis utilizando moldes que contienen uracilo, mutagénesis dirigida a oligonucleótidos, mutagénesis de ADN modificada por fosforotioato, mutagénesis utilizando ADN dúplex hueco o similares, o cualquier combinación de las mismas. Los procedimientos adecuados adicionales incluyen la reparación del emparejamiento incorrecto puntual, mutagénesis utilizando cepas hospedadoras deficientes en la reparación, restricción-selección y restricción-purificación, mutagénesis por delección, mutagénesis por síntesis total de genes, reparación de rotura bicatenaria, y similar. La mutagénesis, incluyendo, aunque no de forma limitativa, la implicación de construcciones quiméricas, se incluye también en la presente invención. En una realización, la mutagénesis puede guiarse mediante información conocida de la molécula que se produce naturalmente o la molécula que se produce naturalmente alterada o mutada, incluyendo, aunque no de forma limitativa, secuencias, comparaciones de secuencias, propiedades físicas, estructura secundaria, terciaria, o cuaternaria, estructura cristalina o similares.

Los textos y ejemplos que se encuentran en el presente documento describen estos procedimientos. Se encuentra información adicional en las siguientes publicaciones y referencias citadas en: Ling y col., Approaches to DNA mutagenesis: an overview, *Anal Biochem.* 254(2): 157-178 (1997); Dale y col., Oligonucleotide-directed random mutagenesis using the phosphorothioate method, *Methods Mol. Biol.* 57:369-374 (1996); Smith, In vitro mutagenesis, *Ann. Rev. Genet.* 19:423-462 (1985); Botstein y Shortle, Strategies and applications of in vitro mutagenesis, *Science* 229:1193-1201 (1985); Carter, Site-directed mutagenesis, *Biochem. J.* 237:1-7 (1986); Kunkel, The efficiency of oligonucleotide directed mutagenesis, in *Nucleic Acids & Molecular Biology* (Eckstein, F. y Lilley, D.M.J. eds., Springer Verlag, Berlin) (1987); Kunkel, Rapid and efficient site-specific mutagenesis without phenotypic selection, *Proc. Natl. Acad. Sci. USA* 82:488-492 (1985); Kunkel y col., Rapid and efficient site-specific mutagenesis without phenotypic selection, *Methods in Enzymol.* 154, 367-382 (1987); Bass y col., Mutant Trp repressors with new DNA-binding specificities, *Science* 242:240-245 (1988); Zoller y Smith, Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment, *Nucleic Acids Res.* 10:6487-6500 (1982); Zoller y Smith, Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors, *Methods in Enzymol.* 100:468-500 (1983); Zoller y Smith, Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template, *Methods in Enzymol.* 154:329-350 (1987); Taylor y col., The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor y col., The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA, *Nucl. Acids Res.* 13: 8765-8785 (1985); Nakamaye & 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 y col., 5'-3' Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis, *Nucl. Acids Res.* 16:791-802 (1988); Sayers y col., 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 y col., The gapped duplex DNA approach to oligonucleotide-directed mutation construction, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer y Fritz Oligonucleotide-directed construction of mutations via gapped duplex DNA, *Methods in Enzymol.* 154:350-367 (1987); Kramer y col., Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations, *Nucl. Acids Res.* 16: 7207 (1988); Fritz y col., Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro, *Nucl. Acids Res.* 16: 6987-6999 (1988); Kramer y col., Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*, *Cell* 38:879-887 (1984); Carter y col., 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 y col., Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986); Nambiar y col., Total synthesis and cloning of a gene coding for the ribonuclease S protein, *Science* 223: 1299-1301 (1984); Sakmar y Khorana, Total synthesis and expression of a gene for the alpha-subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin), *Nucl. Acids Res.* 14: 6361-6372 (1988); Wells y col., Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites, *Gene* 34:315-323 (1985); Grundström y col., Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis, *Nucl. Acids Res.* 13: 3305-3316 (1985); Mandecki, Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis, *Proc. Natl. Acad. Sci. USA*, 83:7177-7181 (1986); Arnold, Protein engineering for unusual environments, *Current Opinion in Biotechnology* 4:450-455 (1993); Sieber, y col., *Nature Biotechnology*, 19:456-460 (2001); W. P. C. Stemmer, *Nature* 370, 389-91 (1994); e, I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995). Se pueden encontrar detalles adicionales de muchos de los anteriores procedimientos en *Methods in Enzymology* Volumen 154, que describe también los controles útiles para resolver problemas con diversos procedimientos de mutagénesis.

Los oligonucleótidos, por ejemplo, para uso en la mutagénesis de la presente invención, por ejemplo, someter a mutación bibliotecas de sintetasas, o alterar los ARNt, se sintetizan de forma normal químicamente de acuerdo con el procedimiento de la fosforamidita triéster en fase sólida descrito por Beaucage y Caruthers, *Tetrahedron Letts.* 22(20):1859-1862, (1981) por ejemplo., utilizando un sintetizador automatizado, como se describe en Needham-VanDevanter y col., *Nucleic Acids Res.*, 12:6159-6168 (1984).

Además, esencialmente cualquier ácido nucleico puede personalizarse u ordenarse de forma normalizada a partir de cualquiera de una variedad de fuentes comerciales, tales como The Midland Certified Reagent Company (mcrcc@oligos.com), The Great American Gene Company (www.genco.com), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, Calif.) y otras muchas.

5 Se desvelan también células hospedadoras eucariotas, células hospedadoras no eucariotas, y organismos para la incorporación in vivo de un aminoácido artificial mediante parejas de ARNt/RS ortogonales. Las células hospedadoras se diseñan mediante ingeniería genética (incluyendo, aunque no de forma limitativa, transformadas, transducidas o transfectadas) con los polinucleótidos o construcciones de la presente divulgación que incluyen un polinucleótido de la presente divulgación, incluyendo, aunque no de forma limitativa, un vector de la presente
10 divulgación, que puede ser, por ejemplo, un vector de clonación o un vector de expresión. Por ejemplo, las regiones de codificación para el ARNt ortogonal, la ARNt sintetasa ortogonal, y la proteína que se va a derivatizar se unen de manera opertiva a los elementos de control de la expresión génica que son funcionales en la célula hospedadora deseada. El vector puede estar, por ejemplo, en la forma de un plásmido, un cósmido, un bacteriófago, una bacteria, un virus, un polinucleótido puro, o un polinucleótido conjugado. Los vectores se introducen en las células y/o
15 microorganismos mediante procedimientos normalizados incluyendo electroporación (Fromm y col., Proc. Natl. Acad. Sci. USA 82, 5824 (1985)), Infección por vectores víricos, penetración balística a alta velocidad por partículas pequeñas con el ácido nucleico tanto en la matriz de pequeñas perlas o partículas, como sobre la superficie (Klein y col., Nature 327, 70-73 (1987)), y/o similares.

Están disponibles algunos procedimientos bien conocidos de introducir ácidos nucleicos diana en las células,
20 cualquiera de los cuales se puede usar en la invención. Estos incluyen: La fusión de células receptoras con protoplastos bacterianos que contienen el ADN, electroporación, bombardeo de proyectiles, e infección con vectores víricos (descrita adicionalmente a continuación), etc. Se pueden usar las células bacterianas para amplificar el número de plásmidos que contienen las construcciones de ADN de la presente divulgación. Las bacterias se hacen crecer hasta la fase logarítmica y los plásmidos en las bacterias se pueden aislar mediante una variedad de
25 procedimientos conocidos en la técnica (véase, por ejemplo, Sambrook). Además, están comercialmente disponibles kits para la purificación de plásmidos a partir de bacterias, (véase, por ejemplo, EasyPrep™, FlexiPrep™, ambos de Pharmacia Biotech; StrataClean™ de Stratagene; e, QIAprep™ de Qiagen). Los plásmidos aislados y purificados se manipulan adicionalmente a continuación para producir otros plásmidos, utilizados para transfectar células o incorporados en vectores relacionados para infectar organismos. Los vectores típicos contienen terminadores de la
30 transcripción y la traducción, las secuencias de inicio de la transcripción y la traducción, y los promotores útiles para la regulación de la expresión del ácido nucleico diana concreto. Los vectores comprende opcionalmente casetes de expresión genéricos que contiene al menos una secuencia terminadora independiente, las secuencias que permiten la replicación del casete en eucariotas, o procariotas, o ambos, (incluyendo, aunque no de forma limitativa, los vectores lanzadera) y los marcadores de selección para los sistemas procariotas y eucariotas. Los vectores son
35 adecuados para la replicación y/o la integración en procariotas, eucariotas, o ambos. Véase, Gillam & Smith, Gene 8:81 (1979); Roberts, y col., Nature, 328:731 (1987); Schneider, E., y col., Protein Expr. Purif. 6(1)10-14 (1995); Ausubel, Sambrook, Berger (*todos anteriormente*). Se proporciona un catálogo de bacterias y bacteriófagos útiles para la clonación, por ejemplo, por la ATCC, por ejemplo, The ATCC Catalogue of Bacteria and Bacteriophage (1992) Gherna y col. (eds) publicado por la ATCC. Los procedimientos adicionales básicos para la secuenciación, la
40 clonación y otros aspectos de la biología molecular y que subyacen bajo las consideraciones teóricas se encuentran también en el Watson y col. (1992) Recombinant DNA Segunda Edición Scientific American Books, NY. Además, esencialmente cualquier ácido nucleico /y, virtualmente, cualquier ácido nucleico marcado, tanto normalizado como no normalizado) puede ordenarse de forma personalizada o normalizada a partir de cualquiera de una variedad de fuentes comerciales, tales como la Midland Certified Reagent Company (Midland, TX disponible en el sitio web en
45 mcrcc.com), The Great American Gene Company (Ramona, CA disponible en el sitio web en genco.com), ExpressGen Inc. (Chicago, IL disponible en el sitio web en expressgen.com), Operon Technologies Inc. (Alameda, CA) y muchas otras.

Las células hospedadoras diseñadas mediante ingeniería genética se pueden cultivar en medios nutrientes convencionales modificados según sea adecuado para dichas actividades como, por ejemplo, etapas de cribado,
50 promotores de la activación o transformantes de la selección. Estas células se pueden cultivar opcionalmente en organismos transgénicos. Otras referencias útiles, por ejemplo, para el aislamiento y el cultivo de células (por ejemplo, para el aislamiento posterior de ácidos nucleicos) incluyen Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, tercera edición, Wiley-Liss, Nueva York y las referencias citadas en el anterior; Payne y col. (1992) Plant Cell and Tissue Culture in Liquid, Systems John Wiley & Sons, Inc. Nueva York, NY; Gamburg y
55 Phillips (eds.) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds.) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

La capacidad de incorporar aminoácidos artificiales directamente en proteínas in vivo ofrece una amplia variedad de ventajas, incluyendo, aunque no de forma limitativa, altos rendimientos de proteínas mutantes, facilidad técnica, el
60 potencial de estudiar las proteínas mutantes en células, o posiblemente en organismos vivos y el uso de estas proteínas mutantes en tratamientos terapéuticos y usos diagnósticos. La capacidad de incluir aminoácidos no naturales con diversos tamaños, acideces, nucleofilicidades, hidrofobicidades, y otras propiedades en proteínas pueden expandir mucho la capacidad de manipular lógica y sistemáticamente las estructuras de las proteínas, para

sondar la función de la proteína y crear nuevas proteínas u organismos con novedosas propiedades.

PROTEÍNAS Y POLIPÉPTIDOS DE INTERÉS

se puede llevar a cabo la incorporación de un aminoácido artificial para una variedad de fines, incluyendo, aunque no de forma limitativa, cambios a medida en la estructura y/o la función de la proteína, tamaño del cambio, acidez, nucleofilicidad, enlaces de hidrógeno, hidrofobicidad, accesibilidad a los sitios diana de las proteasas, direccionamiento a un resto (incluyendo, aunque no de forma limitativa, para una matriz de proteínas), añadir una molécula biológicamente activa, unión a un polímero, unión a un radionucleido, modular la semivida en suero, modular la penetración en el tejido (por ejemplo, tumores), modular el transporte activo, modular la especificidad o distribución del tejido, célula u órgano, modular la inmunogenicidad, modular la resistencia a las proteasas, etc. Las proteínas que incluyen un aminoácido artificial pueden tener propiedades catalíticas o biofísicas mejoradas o incluso completamente nuevas. Por ejemplo, las siguientes propiedades están opcionalmente modificadas mediante la inclusión de un aminoácido artificial en una proteína: toxicidad, biodistribución, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, capacidad catalítica, semivida (incluyendo, aunque no de forma limitativa, semivida en suero), capacidad de reaccionar con otras moléculas, incluyendo, aunque no de forma limitativa, de forma covalente o no covalente, y similares. Las composiciones incluyendo proteínas que incluyen al menos un aminoácido artificial son útiles para, incluyendo, aunque no de forma limitativa, novedosas terapéuticas, diagnósticos, enzimas catalíticas, enzimas industriales, proteínas de unión (incluyendo, aunque no de forma limitativa, anticuerpos) e incluyendo, aunque no de forma limitativa, el estudio de la estructura y función de la proteína. Véase, por ejemplo, Dougherty, (2000) Unnatural Amino Acids as Probes of Protein Structure and Function, Current Opinion in Chemical Biology, 4:645-652.

Una proteína puede tener al menos uno, incluyendo, aunque no de forma limitativa, 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, incluyendo, aunque no de forma limitativa, puede haber 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más sitios diferentes en la proteína que comprenden 1, 2, 3, 4, 5, 6, 7, 8, 9, o 10 o más aminoácidos artificiales diferentes. una proteína puede tener al menos uno, pero menos que todos, de un aminoácido concreto presente en la proteína que está sustituido con el aminoácido artificial. Para una proteína dada con más de un aminoácido artificial, los aminoácidos artificiales pueden ser idénticos o diferentes (incluyendo, aunque no de forma limitativa, la proteína puede incluir dos o más tipos diferentes, o puede incluir dos de los mismo aminoácidos artificiales). 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.

Produciendo proteínas o polipéptidos de interés con al menos un aminoácido artificial en células eucariotas, las proteínas o polipéptidos incluirán normalmente modificaciones posteriores a la traducción eucariotas. En determinadas realizaciones, una proteína incluye al menos un aminoácido artificial y al menos una modificación posterior a la traducción que se lleva a cabo in vivo por una célula eucariota, donde la modificación posterior a la traducción no se realiza por una célula procariota. Por ejemplo, la modificación posterior a la traducción incluye, incluyendo, aunque no de forma limitativa, acetilación, acilación, modificación lipídica, palmitoilación, adición de palmitato, fosforilación, modificación del enlace glicolípido, glicosilación, y similares. En otro aspecto más, la modificación posterior a la traducción incluye el procesamiento proteolítico de los precursores (incluyendo, aunque no de forma limitativa, el precursor de calcitonina, precursor del péptido relacionado con el gen de la calcitonina, hormona preproparatiroidea, preproinsulina, proinsulina, prepro-opiomelanocortina, pro-opiomelanocortina y similares), ensamblaje en una proteína multisubunitaria o ensamblaje macromolecular, traducción en otro sitio en la célula (incluyendo, aunque no de forma limitativa, orgánulos, tales como el retículo endoplásmico, el aparato de Golgi, el núcleo, lisosomas, peroxisomas, mitocondria, cloroplastos, vacuolas, etc, o a través de la ruta secretora). En determinadas realizaciones, la proteína comprende una secuencia de secreción o de localización, una etiqueta de epítipo, una etiqueta FLAG, una etiqueta de polihistidina, una fusión GST, o similares.

Se desvelan también procedimientos para producir una proteína en una célula con un aminoácido seleccionado en una posición especificada. Por ejemplo, un procedimiento incluye crecer, en un medio adecuado, la célula, donde la célula comprende un ácido nucleico que comprende al menos un codón selector y codifica una proteína; e, proporcionar el aminoácido seleccionado; donde la célula comprende además: un ARNt ortogonal (O-ARNt) que funciona en la célula y reconoce el codón selector; y, una aminoacil-ARNt sintetasa ortogonal (O-RS) que aminoacila preferentemente el O-ARNt con el aminoácido seleccionado. Normalmente, el O-ARNt comprende una actividad de supresión en presencia de una sintetasa análoga en respuesta a un codón selector. Se desvela también una proteína producida mediante este procedimiento.

Las composiciones de la presente divulgación y las composiciones preparadas mediante los procedimientos de la presente divulgación están opcionalmente en una célula. Las parejas de O-ARNt/O-RS o los componentes individuales de la presente divulgación pueden usarse a continuación en la maquinaria de traducción de un sistema hospedador, que da como resultado un aminoácido seleccionado, por ejemplo, un aminoácido artificial, que se incorpora en una proteína. Las solicitudes de patente USSN 10/825.867, titulada "Expanding the Eukaryotic Genetic Code;" y 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS", describen este proceso. Por ejemplo, cuando una pareja O-ARNt/O-RS se introduce en un hospedador, por ejemplo, *Escherichia*

coli, la pareja conduce a la incorporación in vivo del aminoácido seleccionado, tal como un aminoácido artificial, por ejemplo, un aminoácido sintético, tal como un derivado de un aminoácido leucina, que se puede añadir exógenamente al medio de crecimiento, en una proteína, en respuesta a un codón selector. Opcionalmente, las composiciones de la presente divulgación pueden estar en un sistema de traducción invitro, o en un sistema(s) in vivo.

Cualquier proteína (o parte de la misma) que incluye un aminoácido seleccionado, por ejemplo, se puede producir un aminoácido artificial, (y cualquier ácido nucleico codificante correspondiente, por ejemplo, que incluye uno o más codones selectores) utilizando las composiciones y procedimientos en el presente documento. Cualquier polipéptido es adecuado para la incorporación de uno o más aminoácidos seleccionados. no se han intentado identificar los cientos de miles de proteínas conocidas, cualquiera de las cuales puede modificarse para incluir uno o más aminoácidos artificiales, por ejemplo, para adaptar cualesquiera procedimientos de mutación disponibles para incluir uno o más codones selectores adecuados en un sistema de traducción relevante. Los repositorios de secuencias comunes de las proteínas conocidas incluyen EMBL, DDBJ y NCBI del GenBank. Otros repositorios pueden identificarse fácilmente buscando 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 diagnóstica, una enzima industrial, o una parte de la misma, y similares), y comprenden uno o más aminoácidos seleccionados. Los ejemplos de proteínas terapéuticas, diagnósticas, y otras proteínas que se pueden modificar para comprender uno o más aminoácidos seleccionados, por ejemplo, un aminoácido artificial, se pueden encontrar, aunque no de forma limitativa, en el documento USSN 10/825.867 titulado "Expanding the Eukaryotic Genetic Code", y, en la solicitud de patente de Estados Unidos con número de serie 10/126.927, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS".

En determinadas realizaciones, la proteína o el polipéptido de interés (o parte de la misma) en los procedimientos y/o composiciones de la presente divulgación está codificada por un ácido nucleico. Normalmente, el ácido nucleico comprende al menos un codón selector, al menos dos codones selectores, al menos tres codones selectores, al menos cuatro codones selectores, al menos cinco codones selectores, al menos seis codones selectores, al menos siete codones selectores, al menos ocho codones selectores, al menos nueve codones selectores, diez o más codones selectores.

Los genes que codifican las proteínas o polipéptidos de interés se pueden mutagenizar utilizando los procedimientos bien conocidos por una persona normalmente experta en la técnica y descritos en el presente documento como "Mutagénesis y otras técnicas de biología molecular". para incluir por ejemplo, uno o más codones selectores para la incorporación de un aminoácido seleccionado, por ejemplo, un aminoácido artificial. Por ejemplo, un ácido nucleico de una proteína de interés se mutageniza para incluir uno o más codones selectores, proporcionando la inserción del uno o más aminoácidos seleccionados, por ejemplo, el aminoácido artificial. la invención incluye cualquiera de dichas variantes, por ejemplo, los mutantes, las versiones de cualquier proteína, por ejemplo, incluyendo al menos un aminoácido seleccionado. De manera similar, la invención incluye también los ácidos nucleicos correspondientes, es decir, cualquier ácido nucleico con uno o más codones selectores que codifica uno o más aminoácidos seleccionados.

Para preparar una proteína que incluye un aminoácido seleccionado, se pueden usar células hospedadoras y organismos que se adaptan para la incorporación in vivo del aminoácido seleccionado mediante parejas de ARNt/RS ortogonales. Las células hospedadoras están diseñadas mediante ingeniería genética (por ejemplo, transformadas, transducidas o transfectadas) con uno o más vectores que expresan el ARNt ortogonal, la ARNt sintetasa ortogonal, y un vector que codifica la proteína que se va a derivatizar. Cada uno de estos componentes puede estar en el mismo vector, o cada uno puede estar en un vector separado, dos componentes pueden estar en un vector y el tercer componente en un segundo vector. El vector puede estar, por ejemplo, en la forma de un plásmido, un cósmido, un bacteriófago, una bacteria, un virus, un polinucleótido puro, o un polinucleótido conjugado.

SISTEMAS ALTERNATIVOS

Se han empleado algunas estrategias para introducir aminoácidos artificiales en proteínas en células hospedadoras no recombinantes, células hospedadoras mutagenizadas, o en sistemas exentos de células. La derivatización de aminoácidos con cadenas secundarias reactivas tales como Lys, Cys y Tyr da como resultado la conversión de la lisina en N²-acetil-lisina. La síntesis química proporciona también un procedimiento directo para incorporar aminoácidos artificiales. Con el desarrollo reciente de la ligadura enzimática y la ligadura química nativa de los fragmentos peptídicos, es posible preparar proteínas más grandes. Véase, por ejemplo, P. E. Dawson y S. B. H. Kent, Annu. Rev. Biochem, 69:923 (2000). La ligadura peptídica química y la ligadura química nativa se describen en la patente de Estados Unidos 6.184.3.44, publicación de patente de Estados Unidos n.º 2004/0138412, publicación de patente de Estados Unidos n.º 2003/0208046, documentos WO 02/098902 y WO 03/042235. Un procedimiento biosintético in vitro general en el que un ARNt supresor se acila químicamente con un aminoácido artificial deseado y se añade a un extracto in vitro capaz de soportar la biosíntesis de la proteína, se ha utilizado para incorporar específicamente del sitio aproximadamente 100 aminoácidos artificiales en una variedad de proteínas de virtualmente cualquier tamaño. Véase, por ejemplo, V. W. Cornish, D. Mendel y P. G. Schultz, Angew. Chem. Int. Ed.

Engl., 1995, 34:621 (1995); C.J. Noren, S.J. Anthony-Cahill, M.C. Griffith, P.G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins, *Science* 244:182-188 (1989); y, J.D. Bain, C.G. Glabe, T.A. Dix, A.R. Chamberlin, E.S. Diala, Biosynthetic site-specific incorporation of a non-natural amino acid into a polypeptide, *J. Am. Chem. Soc.* 111:8013-8014 (1989). Se ha introducido un amplio intervalo de grupos funcionales en proteínas para los estudios de estabilidad de las proteínas, plegado de las proteínas, mecanismo enzimático, y transducción de la señal.

Se ha desarrollado un procedimiento in vivo, denominado incorporación mediante presión selectiva, para explicar la promiscuidad de las sintetasas naturales. Véase, por ejemplo, N. Budisa, C. Minks, S. Alefelder, W. Wenger, F. M. Dong, L. Moroder y R. Huber, *FASEB J.*, 13:41 (1999). Una cepa auxotrofa, en la que se interrumpe el suministro de la ruta metabólica relevante a la célula un aminoácido concreto, se hace crecer en medios mínimos que contienen concentraciones limitadas del aminoácido natural, a la vez que se reprime la transcripción del gen diana. En el inicio de la fase de crecimiento estacionaria, se agota el aminoácido natural y se sustituye con el análogo del aminoácido artificial. La inducción de la expresión de la proteína recombinante da como resultado la acumulación de una proteína que contiene el análogo artificial. Por ejemplo, utilizando esta estrategia, o, m y p se han incorporado p-fluorofenilalaninas en proteínas, y presentan dos hombros característico en el espectro del UV que se pueden identificar fácilmente, véanse, por ejemplo, C. Minks, R. Huber, L. Moroder y N. Budisa, *Anal. Biochem.*, 284:29 (2000); Se ha utilizado la trifluorometionina para sustituir metionina en la lisozima del bacteriófago T4 para estudiar su interacción con los ligandos del quitooligosacárido mediante ¹⁹F NMR, véanse, por ejemplo, H. Duewel, E. Daub, V. Robinson y J. F. Honek, *Biochemistry*, 36:3404 (1997); y se ha incorporado trifluoroleucina en lugar de leucina, dando como resultado una estabilidad térmica y química de una proteína de la cremallera de leucina. Véase, por ejemplo, Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado y D. A. Tirrell, *Angew. Chem. Int. Ed. Engl.*, 40:1494 (2001). Además, se incorporan selenometionina y telurometionina en diversas proteínas recombinantes para facilitar la solución de las fases en la cristalografía de rayos X. Véase, por ejemplo, W. A. Hendrickson, J. R. Horton y D. M. Lemaster, *EMBO J.*, 9:1665 (1990); J. O. Boles, K. Lewinski, M. Kunkle, J. D. Odom, B. Dunlap, L. Lebioda y M. Hatada, *Nat. Struct. Biol.*, 1:283 (1994); N. Budisa, B. Steipe, P. Demange, C. Eckerskorn, J. Kellermann y R. Huber, *Eur. J. Biochem.*, 230:788 (1995); y, N. Budisa, W. Karnbrock, S. Steinbacher, A. Humm, L. Prade, T. Neufeind, L. Moroder y R. Huber, *J. Mol. Biol.*, 270:616 (1997). Los análogos de metionina con funcionalidades alqueno o alquino se han incorporado también eficazmente, permitiendo la modificación adicional de proteínas por medios químicos. Véase, por ejemplo, J. C. van Hest y D. A. Tirrell, *FEBS Lett.*, 428:68 (1998); J. C. van Hest, K. L. Kiick y D. A. Tirrell, *J. Am. Chem. Soc.*, 122:1282 (2000); y, K. L. Kiick y D. A. Tirrell, *Tetrahedron*, 56:9487 (2000); patente de Estados Unidos n.º 6.586.207; publicación de patente de Estados Unidos 2002/0042097.

El éxito de este procedimiento depende del reconocimiento de los análogos de aminoácido artificiales por las aminoacil ARNt sintetasas, que, en general, requieren una alta selectividad para asegurar la fidelidad de la traducción de proteínas. Una vía para expandir el alcance de este procedimiento es relajar la especificidad del sustrato de las aminoacil ARNt sintetasas, lo que se puede conseguir en un número limitado de casos. Por ejemplo, la sustitución de Ala²⁹⁴ por Gly en la fenilalanil-ARNt sintetasa (PheRS) de *Escherichia coli* aumenta el tamaño de la bolsa de unión del sustrato, y da como resultado la acilación del ARNtPhe por la p-Cl-fenilalanina (p-Cl-Phe). Véase, M. Ibba, P. Kast y H. Hennecke, *Biochemistry*, 33:7107 (1994). Una cepa de *Escherichia coli* que incluye este PheRS mutante permite la incorporación de p-Cl-fenilalanina o p-Br-fenilalanina en lugar de la fenilalanina. Véase, por ejemplo, M. Ibba y H. Hennecke, *FEBS Lett.*, 364:272 (1995); y, N. Sharma, R. Furter, P. Kast and D. A. Tirrell, *FEBS Lett.*, 467:37 (2000). De manera similar, se mostró que la mutación puntual Phe130Ser próxima al sitio de unión del aminoácido de la tirosil-ARNt sintetasa de *Escherichia coli* permite incorporar más eficazmente la azatirosina que la tirosina. Véase, F. Hamano-Takaku, T. Iwama, S. Saito-Yano, K. Takaku, Y. Monden, M. Kitabatake, D. Soll and S. Nishimura, *J. Biol. Chem.*, 275:40324 (2000).

Otra estrategia para incorporar aminoácidos artificiales en proteínas in vivo es modificar las sintetasas que tienen mecanismos de liberación. No se pueden discriminar estas sintetasas y por tanto activan los aminoácidos que son estructuralmente similares para los aminoácidos naturales análogos. Este error se corrige en un sitio separado, que desactiva el aminoácido incorrectamente cargado de la ARNt para mantener la fidelidad de la traducción de proteínas. Si se desactiva la actividad de liberación de la sintetasa, los análogos estructurales que se desactivaron incorrectamente pueden escapar a la función de edición e incorporarse. Esta solución se ha demostrado recientemente con la valil-ARNt sintetasa (ValRS). Véase, V. Doring, H. D. Mootz, L. A. Nangle, T. L. Hendrickson, V. de Crecy-Lagard, P. Schimmel y P. Marliere, *Science*, 292:501 (2001). ValRS puede aminocilar incorrectamente ARNtVal con Cys, Thr, o aminobutirato (Abu); estos aminoácidos no análogos se hidrolizan posteriormente por el dominio de edición. Tras la mutagénesis aleatoria del cromosoma de *Escherichia coli*, se seleccionó una cepa mutante de *Escherichia coli* que tiene una mutación en el sitio de edición de ValRS. Este ValRS defectivo en la edición carga incorrectamente ARNtVal con Cys. Debido a que Abu se asemeja estéricamente a Cys (el grupo -SH de Cys está sustituido con -CH₃ en Abu), el ValRS mutante incorpora también Abu en las proteínas cuando esta cepa mutante de *Escherichia coli* se hace crecer en presencia de Abu. El análisis espectrométrico de masas muestra que aproximadamente el 24% de las valinas se sustituyen por Abu en cada posición de valina en la proteína nativa.

La síntesis en fase sólida y los procedimientos semisintéticos han permitido también la síntesis de numerosas proteínas que contienen novedosos aminoácidos. Por ejemplo, véanse las siguientes publicaciones y las referencias citadas en las mismas, que son las siguientes: Crick, F.H.C., Barrett, L. Brenner, S. Watts-Tobin, R. General nature

of the genetic code for proteins. *Nature*, 192:1227-1232 (1961); Hofmann, K., Bohn, H. Studies on polypeptides. XXXVI. The effect of pyrazole-imidazole replacements on the S-protein activating potency of an S-peptide fragment, *J. Am Chem*, 88(24):5914-5919 (1966); Kaiser, E.T. Synthetic approaches to biologically active peptides and proteins including enzymes, *Acc Chem Res*, 22:47-54 (1989); Nakatsuka, T., Sasaki, T., Kaiser, E.T. Peptide segment coupling catalyzed by the semisynthetic enzyme thiosubtilisin, *J Am Chem Soc*, 109:3808-3810 (1987); Schnolzer, M., Kent, S B H. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease, *Science*, 256(5054):221-225 (1992); Chaiken, I.M. Semisynthetic peptides and proteins, *CRC Crit Rev Biochem*, 11(3):255-301 (1981); Offord, R.E. Protein engineering by chemical means? *Protein Eng.*, 1(3):151-157 (1987); y, Jackson, D.Y., Burnier, J., Quan, C., Stanley, M., Tom, J., Wells, J.A. A Designed Peptide Ligase for Total Synthesis of Ribonuclease A with Unnatural Catalytic Residues, *Science*, 266(5183):243 (1994).

Se han usado modificaciones químicas para introducir una variedad de cadenas secundarias artificiales, incluyendo cofactores, marcas de espín y oligonucleótidos en proteínas in vitro. Véase, por ejemplo, Corey, D.R., Schultz, P.G. Generation of a hybrid sequence-specific single-stranded deoxyribonuclease, *Science*, 238(4832):1401-1403 (1987); Kaiser, E.T., Lawrence D.S., Rokita, S.E. The chemical modification of enzymatic specificity, *Annu Rev Biochem*, 54:565-595 (1985); Kaiser, E.T., Lawrence, D.S. Chemical mutation of enzyme active sites, *Science*, 226(4674):505-511 (1984); Neet, K.E., Nanci A, Koshland, D.E. Properties of thiol-subtilisin, *J Biol. Chem*, 243(24):6392-6401 (1968); Polgar, L. et M.L. Bender. A new enzyme containing a synthetically formed active site. Thiol-subtilisin. *J. Am Chem Soc*, 88:3153-3154 (1966); y, Pollack, S.J., Nakayama, G. Schultz, P.G. Introduction of nucleophiles and spectroscopic probes into antibody combining sites, *Science*, 242(4881):1038-1040 (1988).

Definición de polipéptidos según la inmunoreactividad

Puesto que los polipéptidos desvelados en el presente documento proporcionan una variedad de nuevas secuencias de polipéptidos (por ejemplo, que comprenden aminoácidos seleccionados) (por ejemplo, aminoácidos artificiales) en el caso de proteínas sintetizadas en los sistemas de traducción del presente documento, o, por ejemplo, en el caso de sintetetas novedosas, novedosas secuencias de aminoácidos convencionales), los polipéptidos también proporcionan nuevos rasgos estructurales que se pueden reconocer, por ejemplo, en ensayos inmunológicos. La generación de antisuero, que se une específicamente a los polipéptidos de la presente divulgación, así como los polipéptidos que se unen a dicho antisuero, se desvela. El término "anticuerpo", tal como se usa en el presente documento, incluye, aunque no de forma limitativa, un polipéptido codificado sustancialmente mediante un gen de inmunoglobulina o genes de inmunoglobulinas, o fragmentos de los mismos que se unen y reconocen específicamente en un analito (antígeno). Los ejemplos incluyen anticuerpos policlonales, monoclonales, quiméricos, y monocatenerios, y similares. Los fragmentos de inmunoglobulinas, incluidos los fragmentos Fab y los fragmentos producidos por una biblioteca de expresión, que incluye la expresión in vivo, también están incluidos en el término "anticuerpo" que se utiliza en el presente documento. Véase, por ejemplo, Paul, *Fundamental Immunology*, 4ª Ed., 1999, Raven Press, Nueva York, para conocer la estructura del anticuerpo y su terminología. Se proporcionan ejemplos del uso de aminoácidos no naturales y técnicas inmunológicas en el documento WO/2009/099672 de The Scripps Research Institute titulado *Breaking Immunological Tolerance With a Genetically Encoded Unnatural Amino Acid*.

Por ejemplo, también se desvelan proteínas fabricadas usando los ARNt y/o RS de la presente divulgación, que se unen específicamente o que son específicamente inmunorreactivos con un anticuerpo o antisuero generado contra un inmunógeno que comprende una secuencia de aminoácidos. Para eliminar la reactividad cruzada con otros homólogos, se resta del anticuerpo o antisuero la proteína disponible, tal como el polipéptido natural, por ejemplo, los polipéptidos de "control". Donde la proteína natural corresponde a un ácido nucleico, se genera un polipéptido codificado por el ácido nucleico, que se usa con fines de resta del anticuerpo/antisuero.

En un formato típico, el inmunoensayo utiliza un antisuero policlonal sensibilizado contra uno o más polipéptidos o una subsecuencia sustancial de los mismos (es decir, al menos un 30% de la secuencia de longitud completa proporcionada). El conjunto de potenciales polipéptidos inmunógenos derivados de la proteína se denominan colectivamente en los sucesivos como "los polipéptidos inmunogénicos". El antisuero resultante se selecciona opcionalmente para que tenga una baja reactividad cruzada contra los homólogos de la sinteteta de control, y se eliminan todas estas reactividades cruzadas, por ejemplo, mediante inmunoabsorción, con uno o más de los homólogos de control, antes de usar el antisuero policlonal en el inmunoensayo.

Para producir un antisuero para su uso en un inmunoensayo, se produce uno o más de los polipéptidos inmunogénicos que se purifica como se describe en el presente documento. Por ejemplo, la proteína recombinante se puede producir en una célula recombinante. Una cepa de ratones endogámicos (usados en este ensayo porque los resultados son más reproducibles debido a la virtual identidad genética de los ratones) se inmuniza con uno o más de las proteínas inmunógenas junto con un adyuvante normalizado, tal como adyuvante de Freund, y un protocolo de inmunización de ratones convencional (véase, por ejemplo, Harlow y Lane (1988) *Antibodies*, A Laboratory Manual, Cold Spring Harbor Publications, Nueva York, para una descripción convencional de la generación de anticuerpos, formatos de inmunoensayo y condiciones que se pueden usar para determinar la inmunorreactividad específica).

También se pueden encontrar referencias adicionales y descripción de anticuerpos, y se puede aplicar aquí para

definir los polipéptidos según la inmunoreactividad). Como alternativa, uno o más polipéptidos sintéticos o recombinantes derivados de las secuencias divulgadas en el presente documento se conjugan a una proteína transportadora y se utilizan como inmunógeno.

5 El suero policlonal se recoge y se titula contra el polipéptido inmunógeno en un inmunoensayo, por ejemplo, un inmunoensayo en fase sólida con una o más proteínas inmunogénicas inmovilizadas en un soporte sólido. El antisuero policlonal con un título de 10^6 o superior se selecciona, se combina y se resta de los polipéptidos de la sintetasa de control para producir antisuero policlonal combinado con el título restado.

10 El antisuero policlonal combinado con el título restado se analiza para determinar la reactividad cruzada contra los homólogos de control en un inmunoensayo comparativo. En este ensayo comparativo, se determinan las condiciones de discriminación de la unión para el antisuero policlonal con el título restado que da como resultado una señal aproximadamente 5-10 veces mayor respecto al ruido por la unión del antisuero policlonal titulado con la proteína inmunógena en comparación con la unión a los homólogos de la sintetasa de control. Es decir, la rigurosidad de la reacción de unión se ajusta mediante la adición de competidores no específicos tales como albúmina o leche desnatada y/o ajustando las condiciones de sal, temperatura, y/o similares. Estas condiciones de
15 unión se utilizan en ensayos posteriores para determinar si un polipéptido de ensayo (un polipéptido a compara con los polipéptidos inmunógenos y/o los polipéptidos de control) se une específicamente mediante el antisuero policlonal combinado restado.

20 En otro ejemplo, los inmunoensayos en el formato de unión competitiva se usan para la detección de un polipéptido de ensayo. Por ejemplo, como se ha señalado, los anticuerpos con reactividad cruzada se eliminan de la mezcla de antisuero mezclado mediante inmunoabsorción con los polipéptidos de control. A continuación, el uno o más polipéptidos se inmovilizan en un soporte sólido que se expone al antisuero combinado restado. Se añaden proteínas de prueba al ensayo para competir por la unión al antisuero combinado restado. La capacidad de la una o varias proteínas de prueba para competir por la unión al antisuero combinado restado en comparación con la una o varias proteínas se compara con la capacidad del uno o más polipéptidos inmunogénicos añadidos al ensayo para
25 competir por la unión (los polipéptidos inmunogénicos compiten eficazmente con los polipéptidos inmunogénicos inmovilizados) para su unión al antisuero combinado). La reactividad cruzada porcentual de las proteínas de prueba se calcula, mediante cálculos normalizados.

30 En un ensayo paralelo, la capacidad de las proteínas de control para combinar el antisuero combinado restado se determina opcionalmente por comparación con la capacidad del uno o más polipéptidos inmunogénicos para competir por la unión con el antisuero. Igualmente, la reactividad cruzada porcentual de los polipéptidos de control se calcula, mediante cálculos normalizados. Cuando la reactividad cruzada porcentual es al menos 5-10x tan alta para los polipéptidos de ensayo en comparación con los polipéptidos de control y/o donde la unión de los polipéptidos de ensayo está aproximadamente en el intervalo de la unión de los polipéptidos inmunógeno, se dice que los polipéptidos de ensayo se unen específicamente al antisuero combinado restado.

35 En general, el antisuero inmunoabsorbido y combinado se puede usar en un inmunoensayo de unión competitiva tal como se describe en el presente documento, para comparar cualquier polipéptido de ensayo con el uno o más polipéptidos inmunógenos y/o de control. Para realizar esta comparación, los polipéptidos inmunógenos, de ensayo y de control se analizan, cada uno de ellos, para una amplia gama de concentraciones, y la cantidad de cada polipéptido necesaria para inhibir el 50% de la unión al antisuero restado respecto de, por ejemplo, una proteína de control inmovilizada, de prueba o inmunógena se determina por técnicas convencionales. Si la cantidad de polipéptido de ensayo necesaria para su unión en el ensayo competitivo es menos de dos veces la cantidad de polipéptido inmunógeno necesaria, entonces, se dice que el polipéptido de ensayo se une específicamente a un anticuerpos generado por la proteína inmunógena, siempre que la cantidad sea al menos 5-10x tan alta como para e polipéptido de control.

45 Como determinación adicional de la especificidad, el antisuero combinado se inmunosorbe completamente de manera opcional con el uno o varios polipéptidos inmunógenos (en lugar de los polipéptidos de control) hasta que se detecte poco o nada de la unión de antisuero combinado restado de polipéptido inmunógeno resultante al uno o varios polipéptidos inmunógenos usados en la inmunosorción. Este antisuero completamente inmunosorbido se analiza a continuación para determinar su reactividad con el polipéptido de ensayo. Si se observa poca o ninguna reactividad (es decir, no superior a 2x la relación de señal a ruido observada por la unión del antisuero totalmente
50 inmunosorbido con el polipéptido inmunogénico), entonces, el polipéptido de ensayo se une específicamente al antisuero sensibilizado con la proteína inmunógena.

55 Detalles adicionales sobre proteínas, anticuerpos, antisueros, etc. se puede encontrar en el documento USSN 10/825,867 titulado "Expanding the Eukaryotic Genetic Code;" WO 2002/085923, titulada "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS". patente de EE.UU n.º 6.927.042 titulada "Glycoprotein synthesis"; y publicación de patente de EE.UU n.º US 2004/0198637 titulada "Protein Arrays".

KITS

También se desvelan kits. Por ejemplo, un kit para producir una proteína que comprende al menos un aminoácido

5 seleccionado, por ejemplo, un aminoácido artificial, en una célula, se proporciona, donde el kit incluye un recipiente que contiene una secuencia de polinucleótidos que codifica un O-ARnt, y/o un O-ARnt, y/o una secuencia de polinucleótidos que codifica un O-RS, y/o un O-RS. En una realización, el kit incluye además al menos un aminoácido seleccionado. En otra realización, el kit incluye un ARnt aminoacilado de la descripción. En otra realización, el kit comprende además materiales instructivos para producir la proteína.

10 Un ejemplo adicional es un kit para producir una proteína que comprende al menos un aminoácido seleccionado, por ejemplo, un aminoácido artificial, en un sistema de traducción exento de células, donde el kit incluye un recipiente que contiene una secuencia de polinucleótidos que codifica un O-ARnt, y/o un O-ARnt, y/o una secuencia de polinucleótidos que codifica un O-RS, y/o un O-RS. En una realización, el kit incluye además un aminoácido seleccionado. En otra realización, el kit incluye un ARnt aminoacilado de la descripción. En otra realización, el kit comprende además materiales instructivos para producir la proteína.

Ejemplos

Ejemplo 1:

Selección de aminoacil-ARnt sintetasa contra Para-Acetil Fenilalanina

15 Se seleccionaron dos genotecas para buscar aminoacil-ARnt sintetasa dirigidas contra para-acetil fenilalanina, un aminoácido codificado de forma no natural. Estas bibliotecas consisten en seis mutaciones en el gen de la tirosil ARnt sintetasa procedente de *Methanococcus janneschii* en el plásmido pBK.

20 Se llevó a cabo el procedimiento de selección, que consta de cinco rondas de selección alternantes, tres positivas, dos negativas. Las genotecas se combinaron en una relación 1:1 y se electroporaron en la línea celular de selección positiva (GeneHog con plásmido de selección positivo, pREP) y se sembraron en placas de medio mínimo (GMML) con los antibióticos apropiados y el aminoácido codificados de forma no natural para-acetil fenilalanina (pAF). Las placas se incubaron a 37°C durante aproximadamente 40 horas, en cuyo momento, las células se recogieron mediante rascado. El ADN se extrajo mediante un procedimiento Qiagen Mini-Prep y a continuación se purificó en gel de agarosa para aislar la genoteca de ADN plásmido.

25 Este ADN se electroporó a continuación en la línea celular de selección negativa line (GeneHog con el plásmido de selección negativa derivado de pBAD). Estos transformantes se sembraron en placas LB con el antibiótico adecuado sin el aminoácido codificado de forma no natural (pAF). Después de aproximadamente 17 horas, estas células se recogieron mediante rascado y el ADN plásmido se purificó usando el procedimiento Qiagen Mini-Prep y purificación en gel de agarosa.

30 Las rondas de selección posteriores se realizaron con el mismo procedimiento de electroporación, siembra, recogida, y purificación el ADN. En la última (quinta) ronda de selección, se realizaron diluciones en serie de las células transformadas con selección positiva que se sembraron en placas de medio mínimo. A continuación, las colonias individuales se repicaron y se hicieron crecer en un bloque de 96 pocillos durante la noche. Después, se sembró una réplica de este bloque en placas de medio mínimo con concentraciones variables de cloranfenicol (el antibiótico de selección positiva) con y sin el aminoácido no natural pAF. Después de aproximadamente 40 horas de crecimiento a 37°C, las placas se compararon visualmente para determinar qué colonias crecían para la mayor concentración de cloranfenicol, pero no crecían o crecían muy poco en ausencia del aminoácido codificado de forma no natural. Las colonias que cumplen estos criterios se dejaron crecer durante la noche. El ADN se aisló de los cultivos mediante Mini-Prep y purificación con gel de agarosa y se secuenció.

40 A partir de esta selección de pAF, se descubrió que 13 clones tenían secuencia de aminoácidos únicas, y se sometieron a caracterización adicional para determinar la fidelidad y la capacidad de procesamiento de la pAF-ARnt-sintetasa.

45 Para caracterizar estas sintetasa, se llevaron a cabo supresiones ámbar a pequeña escala para mostrar que el aminoácido codificado de forma no natural pAF se había incorporado a un polipéptido, y los resultados se visualizaron mediante SDS-PAGE. Se repicó una sola colonia que se dejó crecer durante la noche en caldo LB, que a continuación se utilizó para inocular 50 ml de LB. Las células se hicieron crecer hasta una DO de 0,3-0,4, en dicho momento, se tomaron alícuotas de 1,5 ml como puntos de preinducción, y el cultivo se dividió entre dos matraces. Se añadió pAF 1 mM a una mitad, y los dos se hicieron crecer durante 30 minutos. Después de 30 minutos de crecimiento, ambos cultivos (+/- pAF) se indujeron con L-arabinosa al 0,2% y se hicieron crecer durante 4,5 horas, y se registró al DO₆₀₀. A continuación, se tomaron alícuotas de 1,5 ml de los matraces +/- pAF para análisis SDS-PAGE.

55 Las alícuotas de 1,5 ml (Preinducción, + pAF, -pAF) se centrifugaron a 10.000 xg durante 10 minutos para aglomerar las células. A continuación, las células se suspendieron en reactivo de extracción de proteínas bacterianas proporcional (BPER, Pierce) en cantidades relativas respecto a su DO₆₀₀ en el momento de la recogida. Se añadió DNasa I a las células lisadas, y se incubaron a 4°C durante 20 minutos. A continuación, las muestras se combinaron con un agente reductor y colorante de carga y se analizaron en un gel Bis-TRIS 4-12% en tampón MES durante 30 minutos. El gel se lavó en H₂O DI dos veces durante diez minutos y se tiñó con colorante azul de coomassie. Las

bandas +/- pAF se compararon para fidelidad con el pAF-ARNt RS para dar como resultado la incorporación de pAF, y la banda + pAF se comparó con la pAF-ARNt RS anteriormente seleccionada.

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

10 De los pAF-ARNt RS identificados a partir de esta selección, se descubrió que una sintetasa (E9) incorpora pAF eficazmente, con una eficacia de incorporación superior al 95% de pAF en S4am-Myo. La incorporación se determinó mediante secuenciación de aminoácidos, mientras que la capacidad de procesamiento se demostró comparando las bandas de proteína en geles SDS-PAGE. La secuencia nucleótidos de E9 se muestra en la SEQ ID NO: 4, y la secuencia de aminoácidos de E9 se muestra en la SEQ ID NO: 5.

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

Ejemplo 2:

Mutagénesis de ARNt

15 Se generaron tres mutantes a partir del ARNt J17. La secuencia DNA de J17 natural se muestra como SEQ ID NO: 8 y en la publicación de patente de Estados Unidos 2003/0108885 como SEQ ID NO: 1 y en el documento US 2003/0082575 como SEQ ID NO: 1 (solicitudes de patente de Estados Unidos con números de serie 10/126.931 y 10/126.927, respectivamente). El ARNt de J17 tiene un par inestable U51:G63 en el tallo TΨC como se muestra en la Figura 1.

20 Tres mutantes J17 (F12, F13, yd F14) se generaron para producir una pareja de bases de Watson-Crick en las posiciones 51 y 63 del tallo TΨC. Se llevó a cabo la mutagénesis mediante PCR solapante, y las construcciones finales se clonaron en sitios EcoRI y NdeI de un plásmido pET19 que comprende la secuencia de polinucleótidos que codifica la aminoacil ARNt sintetasa E9 (SEQ ID NO: 4) y la secuencia de polinucleótidos que codifica la hormona del crecimiento humana (hGH) con una sustitución en el con un codón ámbar (SEQ ID NO: 16). La expresión de hGH está bajo el control del promotor T7.

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

GTAACGCTGAATTCCCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAAT
CCGCATGGCGC (FTam11; SEQ ID NO: 9).

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

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

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

GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACCGAGCGCCATGCGGATTTAGAGT
CCGCCGTTCTGC (FTam13; SEQ ID NO: 11).

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

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

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

40 CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTTTAAGC (cebador directo; FTam15; SEQ ID

NO: 13) y CAAATTCGTCCATATGGGATTCC (FTam 16; SEQ ID NO: 14). El cebador directo se usó para extender la secuencia desde el extremo 3' del ARNt hasta el sitio Nde I del plásmido. El producto resultante se purificó en gel.

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

10 **Expresión de proteínas**

Cada uno de los plásmidos que codifica ARNt (J17, F12, F13 o F14) se transformaron en células hospedadoras bacterianas de *E. coli* cepa 1 y cepa 2 por medios químicos y se sembraron en placas de agar LB con 50 ug/ml de carbenicilina. Se incubaron las placas a 37°C durante la noche. Para cada ARNt, se repicó una colonia individual para iniciar un cultivo nocturno a 37°C en 1 ml 2xYT con 50 ug/ml de carbenicilina. Este cultivo de 1 ml se usó para 15 inocular dos cultivos 2xYT de 10 ml con 50 ug/ml de carbenicilina a 37°C. Un cultivo de 10 ml se suplementó con para-acetilfenilalanina 4 mM. A DO₆₀₀=0,7, la expresión de hGH se indujo con IPTG 0,4 mM. Tras cultivar las células a 37°C durante 4 hrs con 250 rpm, las células se recogieron mediante centrifugación a 5000 x g durante 5 minutos. Las células se lisaron con B-PER Reagent (Pierce, Rockford, IL) suplementado con 5 ug/ml de DNase I. El lisado de células total se analizó con SDS PAGE 4-12%.

20 La Figura 2 muestra un análisis de los lisados de células totales de la cepa 1 de *E. coli* mediante SDS PAGE. La supresión de un codón de selección en la hormona del crecimiento humana se realizó usando ARNt de J17 o el mutante J17 (F12, F13, F14) de la aminoacil ARNt sintetasa E9. Las células que contienen los mutantes J17 crecieron algo menos que las células que contienen J17. No se observó producto de hGH de longitud completa mediante SDS-PAGE para los ARNt mutantes en ausencia de para-acetilfenilalanina 4 mM. En presencia de para-acetilfenilalanina 4 mM, el producto de longitud completa se produjo con cada uno de los ARNt mutantes, 25 demostrando que estos pares de ARNt mutante-RS de E9 son ortogonales a la maquinaria de *E. coli*. Según el SDS-PAGE, el rendimiento de la hGH suprimida en los mutantes J17 fue aproximadamente 1.5~2 veces superior que el de J17 de la cepa 1 de *E. coli*.

30 Un mutante J17, F13, se analizó adicionalmente en la línea celular bacteriana de la cepa 2 de *E. coli* para determinar la supresión ámbar, como se muestra en la Figura 3. En la cepa 2 de *E. coli*, los rendimientos tanto de la expresión como de la supresión ámbar se redujeron con respecto a la cepa 1 de *E. coli*. En ausencia de para-acetilfenilalanina, no se observó producto hGH de longitud completa mediante SDS-PAGE. En presencia de para-acetilfenilalanina 4 mM, se observó producto hGH de longitud completa ara ambos ARNt. Según el SDS-PAGE, el rendimiento de hGH suprimido de F13 fue aproximadamente tres veces superior al de J17.

35 Se llevó a cabo un ciclo de fermentación comparando J17 y F13 con un volumen final de aproximadamente 1,5 k. El plásmido que codifica el ARNt de J17 y el plásmido que codifica el ARNt de F13 se transformaron en la cepa 1 de *E. coli*. La densidad celular final de cada uno fue aproximadamente 190 g células húmedas/l. El título de hGH fue de 347 mg/l para el clon J17 y 542 mg/l para el clon F13.

TABLA 1

SEQ ID NO:	Marca	SECUENCIA
1	ADN F12	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCG CATGGCGCCGGTTCAAATCCGGCCCCGCCGGACCA
2	ADN F13	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCG CATGGCGCTGGTTCAAATCCAGCCCCGCCGGACCA
3	ADN F14	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCG CATGGCGCAGGTTCAAATCCTGCCCGCCGGACCA

40

(continuación)

SEQ ID NO:	Marca	SECUENCIA
4	ácido nucleico E9 RS	ATGGACGAATTGTGAAATGATAAAGAGAAAACACATCTGAAATTAT CAGCGAGGAAGAGTTAAGAGAGGTTTTAAAAAAGATGAAAAA TCTGCTGTTATAGGTTTTGAACCAAGTGGTAAAATACATTTAGG GCATTATCTCCAAATAAAAAAGATGATTGATTTACAAAATGCTG GATTTGATATAATTATATATTTGGCTGATTTACACGCCTATTTAA ACCAGAAAGGAGAGTTGGATGAGATTAGAAAAATAGGAGATTA TAACAAAAAGTTTTTGAAGCAATGGGGTTAAAGGCAAAATAT GTTTATGGAAGTGAACATGGTCTTGATAAGGATTATACACTGAA TGTCTATAGATTGGCTTTAAAAACTACCTTAAAAAGAGCAAGAA GGAGTATGGAACCTATAGCAAGAGAGGATGAAAATCCAAAGGT TGCTGAAGTTATCTATCCAATAATGCAGGTTAATGGGATTCATT ATGAGGGCGTTGATGTTGCAGTTGGAGGGATGGAGCAGAGAAA AATACACATGTTAGCAAGGGAGCTTTTACCAAAAAAGGTTGTTT GTATTCACAACCCTGTCTTAACGGGTTTGGATGGAGAAGGAAAG ATGAGTTCCTCAAAGGGAATTTTATAGCTGTTGATGACTCTCCA GAAGAGATTAGGGCTAAGATAAAGAAAGCATACTGCCAGCTG GAGTTGTTGAAGGAAATCCAATAATGGAGATAGCTAAATACTTC CTTGAATATCCTTTAACCATAAAAAGGCCAGAAAAATTTGGTGG AGATTTGACAGTTAATAGCTATGAGGAGTTAGAGAGTTTATTTA AAAATAAGGAATTGCATCCAATGGATTTAAAAAATGCTGTAGCT GAAGAACTTATAAAGATTTTAGAGCCAATTAGAAAGAGATTATA A
5	aminoácido E9 RS	MDEFEMIKRNTSEIIEEELREVLKKDEKSAVIGFEPGKIHLGHYLQ IKKMIDLQNAAGFDIIYLADLHAYLNQKGELDEIRKIGDYNKKVFEA MGLKAKYVYVYSEHGLDKDYTLNVYRLALKTTILKRARRSMELIAR EDENPKVAEVIYPIQVNGIHYEGVDVAVGGMEQRKIHMLARELL PKKVVCIHNPVLTGLDGEKMSSSKGNFIAVDDSPPEIRAKIKKAY CPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESF KNKELHPMDLKNVAEELIKILEPIRKRL
6	ARNt HL(TAG)3 ADN	CCCAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTAAATCC GTTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTGGGACCA
7	ARNt HL(TGA)1 ADN	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACTTCAAATCC GGTCCCGTAGGGGTTCGGGGTTCAAATCCCCGCCCCCGCACCA
8	J17 M <i>jannaschii</i>	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCG CATGGCGCTGGTTCAAATCCGGCCCCGCCGGACCA
	mtRNA ^{Tyr} _{CUA}	
	ADN	
9	cebador FTam11	GTAACGCTGAATCCCGGCGGTAGTTCAGCAGGGCAGAACGGC GGACTCTAAATCCGCATGGCGC
10	cebador FTam12	GATCTGCAGTGGTCCGGCGGGCCGGATTTGAACCGGCGCCATGC GGATTTAGAGTCCGCCGTTCTGC
11	cebador FTam13	GATCTGCAGTGGTCCGGCGGGCTGGATTTGAACAGCGCCATGC GGATTTAGAGTCCGCCGTTCTGC
12	cebador FTam14	GATCTGCAGTGGTCCGGCGGGCAGGATTTGAACCTGCGCCATGC GGATTTAGAGTCCGCCGTTCTGC

ES 2 660 000 T3

(continuación)

SEQ ID NO:	Marca	SECUENCIA
13	cebador FTam15	CGCCGGACCACTGCAGATCCTTAGCGAAAGCTAAGGATTTTTTT TAAGC
14	cebador FTam16	CAAATTCGTCCATATGGGATTCC
15	cebador FTam17	GTAACGCTGAATTCCCGGCG
16	hGH (ADN)	ATGGGCCACCACCACCACCACCCTTCCCAACCATTCCCTTATCC AGGCTTTTTGACAACGCTATGCTCCGCGCCCATCGTCTGCACCA GCTGGCCTTTGACACCTACCAGGAGTTTGAAGAAGCCTAGATCC CAAAGGAACAGAAGTATTCATTCCTGCAGAACCCCCAGACCTCC CTCTGTTTCTCAGAGTCTATTCCGACACCCTCCAACAGGGAGGA AACACAACAGAAATCCAACCTAGAGCTGCTCCGCATCTCCCTGC TGCTCATCCAGTCGTGGCTGGAGCCCGTGCAGTTCCTCAGGAGT GTCTTCGCCAACAGCCTGGTGTACGGCGCCTCTGACAGCAACGT CTATGACCTCCTAAAGGACCTAGAGGAAGGCATCCAAACGCTGA TGGGGAGGCTGGAAGATGGCAGCCCCGGACTGGGCAGATCTT CAAGCAGACCTACAGCAAGTTCGACACAACTCACACAACGAT GACGCACTACTCAAGAACTACGGGCTGCTCTACTGCTCAGGAA GGACATGGACAAGGTTCGAGACATTCCTGCGCATCGTGCAGTGCC GCTCTGTGGAGGGCAGCTGTGGCTTCTAA
17	D286R mutante de E9	MDEFEMIKRN TSEIISEBEL REVLKKDEKS AVIGFEPGSK IHLGHYLQIK KMIDLQNAGF DIIYLAADLH AYLNQGELD EIRKIGDYNK KVFEAMGLKA KYVYGSEHGL DKDYTLNVYR LALKTTLKRA RRSMELIARE DENPKVAEVI YPIMQVNGIH YEGVDVAVGG MEQRKIHMLA RELLPKKVVC IHNPLVLTGLD GEGKMSSSKG NFLAVDDSP EIRAKIKKAY CPAGVVEGNP IMEIAKYFLE YPLTIKRPEK FGGDLTVNSY EELESLFKNK ELHPMRLKNA VAEELIKILE PIRKRL
18	Cebador 1 Y31	gtattaccctgttatccctatggcgaaattcagaatgattctc
19	Cebador 2 Y31	cagaaaccgctgtgttcttaacaatgtaacattac
20	Cebador 3 Y31	gtaatgattacattgttaggaacaacacgcggttctg
21	Cebador 4 Y31	gagagcgggtttgcgtactg
22	Cebador 5 Y31	attaccctgttatccctagacgctcagtggaacgaaaactcag
23	Cebador 6 Y31	tagggataacagggtataacaatttcagggtg
24	Cebador 7 Y31	ggtcagctgtcgaaagtaccg
25	Cebador 8 Y31	ctacggttcgatcatctccagctaggataacagggtaat
26	Cebador 1 N51	gtattaccctgttatccctatggcgaaattcagaatgattctc
27	Cebador 2 N51	gattacattgttaggaacaacacgcggttctg

(continuación)

SEQ ID NO:	Marca	SECUENCIA
28	Cebador 3 N51	cagaaaccgctgtgttctctaaacaatgtaac
29	Cebador 4 N51	gagagcgggtttgcgtactg
30	Cebador 5 N51	attaccctgttatcccta gacgctcagtggaacgaaaactcacg
31	Cebador 6 N51	tagggataacagggtaatacaatttcagggtg
32	Cebador 7 N51	ggtcagcttgcgaaagtaccg
33	Cebador 8 N51	ctaaacaatgtaatctagggataacagggtaat
34	MAP Y31 Mutante ámbar A12	atggctatctcaatcaagaccccagaagatatcgaaaaaatgctgctgctggccgactggctgcccgaagtgctggagatgatcgaaccgttaggttaaaccggggcgtcagcaccggcgagctggatcgcatctgtaatgattacattgtttaatgaacaacacgcggttctgacctgacctcggctatcacggctatccgaaatccgtttgcatctctattaatgaagtgggtgtgccacgggtatcccgacgatgctaagctgctgaaagatggcgatctggttaacattgatgtcacccgtaatcaaagatgggtttccacggcgatacctcgaaaatgtttatcgtcggtaagccgaccatcatgggcgaacgtctgtgcccgcacacgcaagaaagcctgtacacctggcgctacgcgatggtaaaaccaggcattaatctgcgcgaaatcgggtgcccgcgattcagaaatttgtcgaagcagaaggcttctccgtcgttcgtgaatatggcggacacgggtattgggtcgcggcttccatgaagaaccgcaggtgctgcaactatgactcccgtaaaaccaacgtcgtactgaaacctgggatgacggtccacatcgagccaatgggtcaacgcgggtaaaaaagagatccgcaccatgaaagatggctggacggtaaaaaaccaaagatcgcagcttgtctgcacaatatgagcatactattgtgggtgactgataaacggctgcaaaattctgacgctacgcaaggatgacaccatccggcgataatctcgcacgacgaataa
35	MAP N51 Mutante ámbar D2	atggctatctcaatcaagaccccagaagatatcgaaaaaatgctgctgctggccgactggctgcccgaagtgctggagatgatcgaaccgtatgttaaaccggggcgtcagcaccggcgagctggatcgcatctgtaatgattacattgttttaggaacaacacgcggttctgacctgacctcggctatcacggctatccgaaatccgtttgcatctctattaatgaagtgggtgtgccacgggtatcccgacgatgctaagctgctgaaagatggcgatctggttaacattgatgtcacccgtaatcaaagatgggtttccacggcgatacctcgaaaatgtttatcgtcggtaagccgaccatcatgggcgaacgtctgtgcccgcacacgcaagaaagcctgtacacctggcgctacgcgatggtaaaaccaggcattaatctgcgcgaaatcgggtgcccgcgattcagaaatttgtcgaagcagaaggcttctccgtcgttcgtgaatatggcggacacgggtattgggtcgcggcttccatgaagaaccgcaggtgctgcaactatgactcccgtaaaaccaacgtcgtactgaaacctgggatgacggtccacatcgagccaatgggtcaacgcgggtaaaaaagagatccgcaccatgaaagatggctggacggtaaaaaaccaaagatcgcagcttgtctgcacaatatgagcatactattgtgggtgactgataaacggctgcaaaattctgacgctacgcaaggatgacaccatccggcgataatctcgcacgacgaataa

Ejemplo 3: Vacuna Ambrx viva para la paratuberculosis de subespecies de *Mycobacterium avium*

La paratuberculosis de subespecies de *Mycobacterium avium* (MAP) es un patógeno bacteriano olvidado que produce la enfermedad de Johne en el ganado vacuno y otros rumiantes. Un reciente estudio estimó que el 21% de las ganaderías estadounidenses estaban infectadas, dando como resultado cuantiosas pérdidas económicas en la industria láctea, totalizando más de 200 millones de dólares al año. La enfermedad de Johne solamente se puede tratar con una combinación de antibióticos tales como rifabutina y un macrólido tal como claritomicina. Los regímenes de tratamiento pueden durar años. Una vacuna eficaz contra MAP es fundamental porque las vacunas para la paratuberculosis han estado comercialmente disponibles pero, desafortunadamente, no han sido totalmente eficaces para prevenir la enfermedad.

5

Relación entre MAP y la enfermedad de Crohn humana

La enfermedad de Crohn afecta entre 400.000 y 600.000 personas en Norteamérica. La prevalencia estima que el Norte de Europa ha alcanzado una tasa de 27-48 por 100.000. Durante mucho tiempo se ha sospechado que es el causante de la enfermedad de Crohn en seres humanos; este vínculo está bajo controversia. En la hipótesis de que MAP es el agente causante de la enfermedad de Crohn, Gioconda busca comercializar una combinación de rifabutina, claritromicina y clofazimina (Myoconda) como fármaco potencial para Crohn. Con una respuesta clínica de casi el 95% en un ensayo clínico en Fase II, un ensayo clínico en Fase IIIa demostró una mejora estadísticamente significativa en la remisión conseguida a las 16 semanas en comparación con la terapia convencional. Sin embargo, Myoconda® no consiguió resultados estadísticamente significativos en el mantenimiento de la remisión tras completar la terapia en ese ensayo clínico.

Actualmente, las vacunas disponibles ofrecidas son vacunas basadas en células completas atenuadas (cepa 316 F), células completas destruidas (cepa 18), y también hay en curso un ensayo clínico con una vacuna de subunidades, sin embargo, las vacunas actuales solamente retrasan el desprendimiento fecal y la evolución a enfermedad clínica, pero no protegen contra la infección (Expert Rev. Vaccines 7(6) 2008). Este campo está buscando idealmente una vacuna atenuada lógica.

Como la secuencia del genoma completo de MAP se ha publicado (PNAS 2005 12344-12349) y se ha conseguido la expresión de genes extraños en MAP, en la actualidad están disponibles herramientas genéticas específicas para delección y mutación específica del sitio (Applied and Environmental Microbiology 2008 1687-1695). La vacuna Ambrx será un MAP de organismo completo genéticamente modificado dependiente de aminoácidos no naturales que se administrará en forma de vacuna para prevenir la infección por MAP en animales. Los genes esenciales se seleccionan y posteriormente se transforman para incluir una o más mutaciones ámbar específicas de sitio. El MAP se caracteriza mediante cultivo celular con y sin el aminoácido no natural presente en el medio de cultivo, se caracterizan los datos de reversión, y el organismo completo.

Ejemplo 4: Creación de glóbulos rojos competentes para supresión/lambda

Se preparó una cepa de *E. coli*. La cepa de *E. coli* es capaz tanto de supresión ámbar como de recombinación Lambda Red. Se prepararon *E. coli* W3110 que contenían el pKD46n mostrado en la Figura 5 que a continuación se transformaron mediante electroporación con el plásmido pVK10/camR mostrado en la Figura 4. Los transformantes positivos se seleccionaron sobre placas de agar LB que contenía 50 µg/ml de ampicilina y 34 µg/ml de cloranfenicol con crecimiento durante 48 horas a 30 °C.

Se prepararon células electrocompetentes que contenían ambos plásmidos haciendo crecer un cultivo de 500 ml que contenía ambos plásmidos en medio LB que contenía 50 µg/ml de ampicilina, 34 µg/ml de cloranfenicol y arabinosa al 0,2% hasta una OD600 final de 0,5. A continuación se lavaron las células tres veces en glicerol al 10% y se resuspendieron en 500 µl de glicerol al 10%. Alícuotas de 60 µl se almacenaron a -80 °C para uso en el futuro.

Ejemplo 5: Creación de una metionina aminopeptidasa condicional letal

Se generaron cepas de *E. coli* con metionina aminopeptidasa condicional letal. Se introdujo una mutación ámbar en uno de dos sitios del gen de la metionina aminopeptidasa, Y31 o N51. Para integrar estos mutantes en el cromosoma de *E. coli*, se generó un trozo lineal de ADN recombinante mediante la PCR solapante, como se muestra en el diagrama de la Figura 6. En primer lugar, usando los cebadores 1 y 2 (SEQ ID NO: 18 o 26), se clonó el extremo 5' del gen de la metionina aminopeptidasa (producto de la PCR A), finalizando en el aminoácido seleccionado (Y31 o N51) que se convirtió del codón natural en el codón ámbar. Se clonó el extremo 3' del gen de la metionina aminopeptidasa, partiendo de la mutación ámbar, usando los cebadores 3 (SEQ ID NO: 20 o 28) y 4 (SEQ ID NO: 21 o 29) (producto de la PCR B). Los dos productos de la reacción se solapan en la secuencia que flanquea la mutación ámbar, y estos productos se amplificaron con los cebadores 1 (SEQ ID NO: 18 o 26) y 4 (SEQ ID NO: 21 o 29) para dar el producto C, el gen completo de la metionina aminopeptidasa con una sustitución de codón ámbar en la posición 31 o 51.

Para crear la plantilla de mutagénesis completa, un marcador de resistencia a la kanamicina flanqueado por los sitios de reconocimiento de la endonucleasa de direccionamiento I-sceI se amplificó usando los cebadores 5 (SEQ ID NO: 22 o 30) y 6 (SEQ ID NO: 23 o 31) en el ADN de una cepa de *E. coli* anteriormente generada (productos de la PCR D). Los cebadores 7 (SEQ ID NO: 24 o 32) y 8 (SEQ ID NO: 25 o 33) se usaron para amplificar la secuencia genómica en dirección 5' y el extremo 5' de la metionina aminopeptidasa, para permitir la eliminación del marcador de resistencia a la kanamicina y la recombinación homóloga (producto de la PCR E). La plantilla de ADN final para la transformación se creó amplificando los fragmentos E, D, y C usando los cebadores 7 (SEQ ID NO: 24 o 32) y 4 (SEQ ID NO: 21 o 29).

Ejemplo 7: Transformación y selección de clones con metionina aminopeptidasa condicional letal

La plantilla de transformación generada mediante la PCR solapante se usó para transformar células electrocompetentes pKD46, pVK10-camR positivas. Los transformantes positivos se seleccionaron mediante crecimiento de colonias sobre placas de agar LB que contenía 50 µg/ml de ampicilina, 50 µg/ml de kanamicina y pAF

2 mM durante 48-72 horas a 30°C. Para analizar el crecimiento dependiente de pAF, se hicieron crecer 96 colonias de cada sitio en 1 ml de LB que contenía 50 µg/ml de ampicilina, 50 µg/ml de kanamicina, y pAF 1 mM durante 72 horas a 30°C, también mostrado en la Figura 7, Figura 8, y Figura 9. A continuación, réplicas de estos clones se sembraron en placa de agar que contenía 50 µg/ml de ampicilina, 50 µg/ml de kanamicina, y pAF 2 mM, y agar LB que contenía 50 µg/ml de ampicilina y 50 µg/ml de kanamicina sin pAF. Estas placas de réplicas se hicieron crecer durante 72 horas a 30 °C y a continuación se puntuó el crecimiento. Los clones que mostraron crecimiento solamente en presencia de pAF 2 mM se seleccionaron para análisis posterior.

Ejemplo 8: Selección y maduración de clones

Los clones con dependencia positiva para pAF se seleccionaron para estudio posterior. Estos clones fueron A12 de Y31, C5 de Y31, E3 de Y31, D2 de N51 y H9 de N51. Los clones se volvieron a cribar individualmente y todos crecieron solamente en medios de cultivo que contenían pAF. Todos los clones se secuenciaron para determinar la presencia de un codón ámbar (tag) en el sitio adecuado: Y31 o N51. Los cinco tenían un codón ambar en la situación adecuada. Los clones A12 de Y31 y D2 de N51 se aislaron a continuación, y el plásmido pKD46 se eliminó mediante un tratamiento térmico repetido a 37 grados centígrados.

Una vez perdido pKD46 y el gen de resistencia a la ampicilina, las células se transformaron con el plásmido que contenía la enzima I-SceI. Después del tratamiento con esta enzima, se seleccionaron algunos clones que se analizaron para determinar su resistencia a la kanamicina. Los clones sin resistencia a la kanamicina se volvieron a secuenciar y se congelaron en disoluciones madre de glicerol al 10%. Estos clones serán los clones finales de la metionina aminopeptidasa dependiente de pAF.

En algunas realizaciones de la presente invención, las tasas de reversión deseadas, consideradas sistemáticamente en la Figura 11, serán inferiores a 10^{-9} de frecuencia de reversión con un solo cambio. Se analizaron los clones Y31 y N51, y las tasas de reversión se muestran en la Tabla 2.

TABLA 2:

Clon	Ensayo 1	Ensayo 2
Y31	$2,78 \times 10^{-8}$	$<3 \times 10^{-8}$ (ninguno detectado)
N51	$6,82 \times 10^{-8}$	$<2 \times 10^{-8}$ (ninguno detectado)

TABLA 3:

Genes esenciales en <i>E. coli</i>		
Alt. Símbolo	Símbolo	Producto génico
b0026	ileS	Isoleucil-ARNt sintetasa (EC 6.1.1.5)
b0027	IspA	Lipoproteína de la peptidasa señal (EC 3.4.23.36)
b0029	ispH*	proteína IspH (IytB)
b0083	ftsL	Proteína de división celular ftsL
b0084	ftsI	Precursor ftsI de la peptidoglicano sintetasa
b0085	murE	UDP-N-acetilmuramoilalanil-D-glutamato--2,6-diaminopimelato ligasa (EC 6.3.2.13)
b0088	murD	UDP-N-acetilmuramoilalanina--D-glutamato ligasa (EC 6.3.2.9)
b0089	ftsW	Proteína de división celular ftsW
b0090	murG	UDP-N-acetilglucosamina--N-acetilmuramil-(Pentapéptido) pirofosforil-undecaprenol N-acetilglucosamina transferasa (EC 2.4.1.-)
b0091	murC	UDP-N-acetilmuramato--alanina ligasa (EC 6.3.2.8)
b0093	ftsQ	Proteína de división celular ftsQ
b0094	ftsA	Proteína de división celular ftsA
b0095	ftsZ	Proteína de división celular ftsZ
b0098	secA	subunidad secA de la preproteína translocasa
b0168	map	Metionina aminopeptidasa (EC 3.4.11.18)
b0169	rpsB	proteína ribosómica S2 de 30S
b0171	pyrH	Uridilato quinasa (EC 2.7.4.-)

ES 2 660 000 T3

(continuación)

Genes esenciales en <i>E. coli</i>		
Alt. Símbolo	Símbolo	Producto génico
b0172	frr	Factor de recirculación del ribosoma
b0174	uppS	Undecaprenil pirofosfato sintetasa (EC 2.5.1.31)
b0175	cdsA	Fosfatidato citidiltransferasa (EC 2.7.7.41)
b0180	fabZ	(3R)-hidroximiristoil-[proteína transportadora de acilo] deshidratasa (EC 4.2.1.-)
b0181	lpxA	Acil-[proteína transportadora de acilo]-UDP-N-acetilglucosamina O-aciltransferasa (EC 2.3.1.129)
b0182	lpxB	Lípido-A-disacárido sintasa (EC 2.4.1.182)
b0184	dnaE	subunidad alfa de la ADN polimerasa III (EC 2.7.7.7)
b0185	accA	Acetil-coenzima A carboxilasa carboxil transferasa, subunidad alfa (EC 6.4.1.2)
b0194	proS	Prolil-ARNt sintetasa (EC 6.1.1.15)
b0416	nusB	N proteína B utilización de sustancia
b0420	dxs	1-desoxi-D-xilulosa 5-fosfato sintasa (EC 4.1.3.37)
b0421	ispA	Geranil transferasa (EC 2.5.1.10)
b0470	dnaX	subunidad tau de la ADN polimerasa III (EC 2.7.7.7)
b0474	adk	Adenilato quinasa (EC 2.7.4.3)
b0526	cysS	Cisteinil-ARNt sintetasa (EC 6.1.1.16)
b0634	mrdB	Proteína rodA determinante de la forma de bacilo
b0635	mrdA	Proteína de unión a penicilina 2
b0640	holA	ADN polimerasa III, subunidad delta (EC 2.7.7.7)
b0642	leuS	Leucil-ARNt sintetasa (EC 6.1.1.4)
b0680	glnS	Glutaminil-ARNt sintetasa (EC 6.1.1.18)
b0893	serS	Seril-ARNt sintetasa (EC 6.1.1.11)
b0914	msbA	probable proteína msbA de unión a ATP de transporte
b0918	kdsB	3-desoxi-mano-octulosonato citidililtransferasa (EC 2.7.7.38)
b0930	asnS	Asparaginil-ARNt sintetasa (EC 6.1.1.22)
b0954	fabA	3-hidroxidecanoil-[proteína transportadora de acilo] deshidratasa (EC 4.2.1.60)
b1084	rne	Ribonucleasa E (EC 3.1.4.-)
b1092	fabD	proteína transportadora de malonil CoA trans acilasa (EC 2.3.1.39)
b1093	fabG	3-oxoacil-[proteína transportadora de acilo] reductasa (EC 1.1.1.100)
b1098	tmk	Timidilato quinasa (EC 2.7.4.9)
b1099	holB	ADN polimerasa III, subunidad delta' (EC 2.7.7.7)
b1116	lolC	proteína transmembrana del sistema liberador de lipoproteína lolC
b1117	lold	proteína de unión a ATP del sistema liberador de lipoproteína lolD
b1118	lolE	proteína transmembrana del sistema liberador de lipoproteína lolE
b1133	trmU	ARNt (5-metilaminometil-2-tiouridilato)-metiltransferasa (EC 2.1.1.61)
b1207	prsA	Ribosa-fosfato pirofosfoquinasa (EC 2.7.6.1)
b1274	topA	ADN topoisomerasa I (EC 5.99.1.2)
b1288	fabI	Enoil-[proteína transportadora de acilo] reductasa [NADH] (EC 1.3.1.9)
b1714	pheS	Fenilalanil-ARNt sintetasa cadena alfa (EC 6.1.1.20)
b1716	rpIT	proteína ribosómica L20 de 50S
b 1718	infC	Factor de inicio de la traducción IF-3
b1740	nadE	NAD(+) sintetasa dependiente de NH(3) (EC 6.3.5.1)
b1866	aspS	Aspartil-ARNt sintetasa (EC 6.1.1.12)

ES 2 660 000 T3

(continuación)

Genes esenciales en <i>E. coli</i>		
Alt. Símbolo	Símbolo	Producto génico
b1876	argS	Arginil-ARNt sintetasa (EC 6.1.1.19)
b2185	rplY	proteína ribosómica L25 de 50S
b2234	nrdA	Ribonucleósido-difosfato reductasa 1, cadena alfa (EC 1.17.4.1)
b2235	nrdB	Ribonucleósido-difosfato reductasa 1, cadena beta (EC 1.17.4.1)
b2323	fabB	3-oxoacil-[proteína transportadora de acilo] sintasa I (EC 2.3.1.41)
b2400	gltX	Glutamil-ARNt sintetasa (EC 6.1.1.17)
b2411	ligA	ADN ligasa (EC 6.5.1.2)
b2412	zipA	Proteína de división celular zipA
b2496	yfgE	Proteína teórica yfgE
b2514	hisS	Histidil-ARNt sintetasa (EC 6.1.1.21)
b2563	acpS	Holo-[proteína transportadora de acilo] sintasa (EC 2.7.8.7)
b2606	rplS	proteína ribosómica L19 de 50S
b2607	trmD	ARNt (Guanina-N(1)-)metiltransferasa (EC 2.1.1.31)
b2609	rpsP	proteína ribosómica S16 de 30S
b2610	ffh	Partícula de proteína de reconocimiento de señal
b2614	grpE	proteína GrpE
b3018	plsC	1-acil-sn-glicerol-3-fosfato aciltransferasa (EC 2.3.1.51)
b3019	parC	Subunidad A de la topoisomerasa IV (EC 5.99.1.-)
b3030	parE	Subunidad B de la topoisomerasa IV (EC 5.99.1.-)
b3041	ribB	3,4-dihidroxi-2-butanona 4-fosfato sintasa
b3067	rpoD	Factor sigma de la ARN polimerasa rpoD
b3165	rpsO	proteína ribosómica S15 de 30S
b3183	yhbZ	proteína teórica de unión a GTP yhbZ
b3230	rpsI	proteína ribosómica S9 de 30S
b3231	rplM	proteína ribosómica L13 de 50S
b3287	def	Péptido deformilasa (EC 3.5.1.88)
b3298	rpsM	proteína ribosómica S13 de 30S
b3300	secY	subunidad secY de la preproteína translocasa
b3301	rplO	proteína ribosómica L15 de 50S
b3303	rpsE	proteína ribosómica S5 de 30S
b3304	rplR	proteína ribosómica L18 de 50S
b3305	rplF	proteína ribosómica L6 de 50S
b3307	rpsN	proteína ribosómica S14 de 30S
b3309	rplX	proteína ribosómica L24 de 50S
b3310	rplN	proteína ribosómica L14 de 50S
b3311	rpsQ	proteína ribosómica S17 de 30S
b3316	rpsS	proteína ribosómica S19 de 30S
b3317	rplB	proteína ribosómica L2 de 50S
b3318	rplW	proteína ribosómica L23 de 50S
b3319	rplD	proteína ribosómica L4 de 50S
b3320	rplC	proteína ribosómica L3 de 50S
b3321	rpsJ	proteína ribosómica S10 de 30S
b3340	fusA	Elongation factor G

(continuación)

Genes esenciales en <i>E. coli</i>		
Alt. Símbolo	Símbolo	Producto génico
b3342	rpsL	proteína ribosómica S12 de 30S
b3384	trpS	Triptofanil-ARNt sintetasa (EC 6.1.1.2)
b3464	ftsY	Proteína de división celular ftsY
b3559	glyS	Glicil-ARNt sintetasa, cadena beta (EC 6.1.1.14)
b3633	kdtA	ácido 3-desoxi-D-mano-octulosónico transferasa (EC 2.-.-.-)
b3634	coaD	Fosfopanteína adeniltransferasa (EC 2.7.7.3)
b3640	dut	Desoxiutidina 5'-trifosfato nucleotidohidrolasa (EC 3.6.1.23)
b3648	gmk	Guanilato quinasa (EC 2.7.4.8)
b3702	dnaA	Proteína dnaA de inicio de la replicación cromosómica
b3730	glmU	proteína glmU bifuncional
b3865	engB	proteína de unión a GTP engB
b3967	murl	Glutamato racemasa (EC 5.1.1.3)
b3985	rplJ	proteína ribosómica L10 de 50S
b3986	rplL	proteína ribosómica de 50S L7/L12
b3987	rpoB	Cadena beta de la ARN polimerasa dirigida a ADN (EC 2.7.7.6)
b3988	rpoC	Cadena beta de la ARN polimerasa dirigida a ADN (EC 2.7.7.6)
b4052	dnaB	Helicasa replicativa del ADN (EC 3.6.1.-)
b4142	groS	chaperonina de 10 kDa
b4143	groL	chaperonina de 60 kDa
b4147	efp	Factor de elongación P
b4200	rpsF	proteína ribosómica S6 de 30S
b4226	ppa	Pirofosfatasa inorgánica (EC 3.6.1.1)
b4258	valS	Valil-ARNt sintetasa (EC 6.1.1.9)
b4361	dnaC	Proteína de replicación del ADN dnaC
b4362	dnaT	Proteína primosomal I

Ejemplo 9:

Se siembra una cepa del mismo tipo y paso utilizado para la vacuna de la poliomielitis (oral), transformada para incluir mutaciones ámbar en uno o más genes esenciales, y cultivar.

Cultivo celular:

- 5 i) Los diferentes sustratos celulares para la propagación del virus incluyen una línea celular diploide humana, células de riñón de mono, la línea de células Vero, o cualquier otro tipo de cultivo celular adecuado. ii) Producción de células a gran escala: 1. Preparación del banco de células de trabajo del fabricante (MWCB): MWCB en el caso de cultivos celulares primarios (células derivadas de tejido normal y almacenadas congeladas a menos 70° C) o de cualquier otra línea celular autorizada continua se preparó a partir de las células, y las células se combinaron después de subcultivos en serie dentro del número especificado de cultivos celulares. 2. Se puede usar cualquiera de los siguientes sistemas, siempre que sea adecuado, para el incremento gradual del volumen de células. Frascos sobre rodillos: Crecimiento en frascos sobre rodillos, a partir de cultivos estacionarios tras distribución de células sobre la superficie de vidrio y hasta la máxima densidad celular que se pueda conseguir. Los cultivos en frascos sobre rodillos no se deterioran menos rápidamente ya que pueden tolerar una densidad casi dos veces superior a la de los cultivos estacionarios. Las células de mamífero crecidas en cultivo de monocapa se adaptarán mejor a la técnica de frascos sobre rodillos. Fábricas celulares: Se trata de pilas de placas de cultivo que comparten una puerto de entrada y de salida común. Proporcionan a las células un superficie de crecimiento tan grande como el interior de los frascos sobre rodillos, pero ocupan menos espacio en la incubadora. Son ideales para cultivos celulares adherentes, tienen menor contaminación, riesgo, y son compactos. Sistema Roux flasks Micro-carrier 3. Las muestras de los cultivos celulares de control se incubaron por separado durante al menos dos semanas y se analizaron para determinar evidencias de cambios citopáticos.
- 10
- 15
- 20

Propagación del virus de la polio en cultivos celulares:

5 i) Sistema de siembra por lotes: a) Nivel de paso según la vacuna de la poliomielitis oral O b) Su siguiente nivel de paso según se deriva de la suspensión a granel monovalente que pueda usar para fabricar una vacuna de la poliomielitis oral. 1. La vacuna se fabrica sobre la base de la siembra de virus por lotes. Un lote de siembra de virus es una cantidad de virus que se procesa conjuntamente y de composición uniforme que se prepara a partir de un lote de siembra. 2. El subcultivo de la siembra de virus no debe realizarse más de 10 veces a partir de un lote de siembra usado en la producción de la vacuna sobre la que se realizaron los ensayos originales de laboratorio y de campo.

Cosecha única y combinaciones monovalentes:

10 El virus se propaga en cultivos celulares y se recoge de cultivos celulares derivados de un único lote de células, y se procesan conjuntamente. Esto se conoce como cosecha única. La única cosecha de un solo tipo de suspensión de virus se procesa a la vez para obtener la forma bruta de la combinación de lente monova.

Filtración o clarificación de la combinación monovalente:

15 La suspensión de virus en bruto de cada combinación monovalente se purifica gradualmente mediante filtros de porosidad decreciente. El objetivo de la etapa de filtración es eliminar el material en forma de partículas y otras sustancias que puedan afectar el proceso de inactivación puesto que dichos agregados tienden a aumentar en reposo.

Concentración de la combinación monovalente:

Cada combinación monovalente filtrada se concentra por ultra-filtración.

Purificación de la combinación monovalente:

20 Se utiliza cromatografía con una columna de DEAE Spharose/ADN-asa inmovilizada o una columna de inmunoadsorción para la purificación de la combinación monovalente. El proceso de purificación reduce de forma coherente el nivel de ADN celular del que tenía la cosecha inicial de virus en un factor de al menos 108. La concentración de antígeno D se determinó mediante ELISA y mostró una inmunogenicidad comparable en ratas. La potencia se ajusta de acuerdo con ello. Las combinaciones monovalentes de virus de la poliomielitis se combinan para formar el producto final trivalente a granel.

Formulación y Fabricación de la vacuna:

30 El volumen purificado se filtra, preferentemente a través de un filtro de membrana de 0,22 micrómetros o 0,45 micrómetros. Los virus monovalentes transformados siguen en presencia del aminoácido no natural, y se mezcla opcionalmente con estabilizantes y/o conservantes para preparar la vacuna monovalente, bivalente, trivalente o multivalente.

Ensayo de la eficacia de la vacuna:

La eficacia de la vacuna tal como se ha fabricado anteriormente, en comparación con la vacuna de referencia establecida, resulta similar en los procedimientos in vitro y/o in vivo.

35 Se llevaron a cabo otros ensayos basándose en los procedimientos aceptados convencionalmente sobre la vacuna anteriormente fabricada tales como determinación de pH, ensayo de esterilidad, ensayo de transformación eficaz (por ejemplo, división y comprobación de destrucción en ausencia del aminoácido no natural), estudio de toxicidad anómala, medición de volumen, determinación de la potencia in vitro, ensayo de endotoxina, ensayo de identidad, estimación en el contenido de nitrógeno, y determinación del contenido de estabilizante y del contenido de formaldehído para garantizar su eficacia, seguridad y potencia.

40 La potencia in vitro de la SIPV fabricada anteriormente se determinó utilizando ELISA en sándwich. Dos anticuerpos primarios sensibilizados en diferentes especies contra el virus de la poliomielitis se utilizaron para lo mismo, y un segundo anticuerpo conjugado con HRP se usó de nuevo como el segundo anticuerpo primario. Los resultados se analizaron por comparación con el patrón de referencia.

45 El ensayo de identidad de la vacuna se llevó a cabo usando, ELISA directo. El antígeno se revisó en placas ELISA de 96 pocillos y se utilizó suero anti-poliomielitis específico de tipo para determinar la presencia de los tres tipos de virus de la poliomielitis en la vacuna.

LISTADO DE SECUENCIAS

<110> Tian, Feng Hehli, Brad

<120> Vacunas y microorganismos dependientes de la replicación de aminoácidos no naturales

50 <130> AMBX-0157.00PCT

ES 2 660 000 T3

<150> 61/100.688
 <151> 26/09/2008

<160> 35

<170> PatentIn versión 3.4

5 <210> 1
 <211> 77
 <212> ADN
 <213> Secuencia artificial

<220>
 10 <223> ARNt mutante derivado del ARNt de Methanococcus jannaschii

<400> 1

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ccggttcaaa 60

tccggcccgc cggacca 77

<210> 2
 <211> 77
 15 <212> ADN
 <213> Secuencia artificial

<220>
 <223> ARNt mutante derivado del ARNt de Methanococcus jannaschii

<400> 2

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60

tccagcccgc cggacca 77

<210> 3
 <211> 77
 <212> ADN
 <213> Secuencia artificial

25 <220>
 <223> ARNt mutante derivado del ARNt de Methanococcus jannaschii

<400> 3

ccggcggtag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg caggttcaaa 60

tcctgcccgc cggacca 77

<210> 4
 <211> 921
 <212> ADN
 <213> Secuencia artificial

<220>
 <223> Sintetasa mutante derivada de la sintetasa de Methanococcus jannaschii

35 <400> 4

ES 2 660 000 T3

atggacgaat ttgaaatgat aaagagaaac acatctgaaa ttatcagcga ggaagagtta 60
 agagaggttt taaaaaaga tgaaaaatct gctgttatag gttttgaacc aagtggtaaa 120
 atacatttag ggcattatct ccaaataaaa aagatgattg atttacaaaa tgctggattt 180
 gatataatta tatatttggc tgatttacac gcctatttaa accagaaaagg agagttagat 240
 gagattagaa aaataggaga ttataacaaa aaagtttttg aagcaatggg gttaaaggca 300
 aaatatgttt atggaagtga acatggctct gataaggatt atacactgaa tgtctataga 360
 ttggctttta aactacctt aaaaagagca agaaggagta tggaacttat agcaagagag 420
 gatgaaaatc caaaggttgc tgaagttatc tatccaataa tgcaggttaa tgggattcat 480
 tatgagggcg ttgatgttgc agttggaggg atggagcaga gaaaaataca catgttagca 540
 agggagcttt taccaaaaaa gggtgtttgt attcacaacc ctgtcttaac gggtttggat 600
 ggagaaggaa agatgagttc ttcaaaaggg aattttatag ctgttgatga ctctccagaa 660
 gagattaggg ctaagataaa gaaagcatac tgcccagctg gagttgttga aggaaatcca 720
 ataatggaga tagctaaata cttccttgaa tacccttaa ccataaaaag gccagaaaaa 780
 tttggtggag atttgacagt taatagctat gaggagttag agagtttatt taaaaataag 840
 gaattgcatc caatggattt aaaaaatgct gtagctgaag aacttataaa gattttagag 900
 ccaattagaa agagattata a 921

<210> 5

<211> 306

<212> PRT

5 <213> Secuencia artificial

<220>

<223> Sintetasa mutante derivada de la sintetasa de Methanococcus jannaschii

<400> 5

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

Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val
 20 25 30

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln

ES 2 660 000 T3

35	40	45
Ile Lys Lys Met Ile Asp 50	Leu Gln Asn Ala Gly Phe 55 60	Asp Ile Ile Ile
Tyr Leu Ala Asp Leu His 65 70	Ala Tyr Leu Asn Gln Lys 75	Gly Glu Leu Asp 80
Glu Ile Arg Lys Ile Gly 85	Asp Tyr Asn Lys Lys Val 90	Phe Glu Ala Met 95
Gly Leu Lys Ala Lys Tyr 100	Val Tyr Gly Ser Glu His 105	Gly Leu Asp Lys 110
Asp Tyr Thr Leu Asn Val 115	Tyr Arg Leu Ala Leu Lys 120	Thr Thr Leu Lys 125
Arg Ala Arg Arg Ser Met 130	Glu Leu Ile Ala Arg 135	Glu Asp Glu Asn Pro 140
Lys Val Ala Glu Val Ile 145	Tyr Pro Ile Met Gln Val 150 155	Asn Gly Ile His 160
Tyr Glu Gly Val Asp Val 165	Ala Val Gly Gly Met Glu 170	Gln Arg Lys Ile 175
His Met Leu Ala Arg Glu 180	Leu Leu Pro Lys Lys Val 185	Val Val Cys Ile His 190
Asn Pro Val Leu Thr Gly 195	Leu Asp Gly Glu Gly Lys 200	Met Ser Ser Ser 205
Lys Gly Asn Phe Ile Ala 210	Val Asp Asp Ser Pro Glu 215 220	Glu Glu Ile Arg Ala
Lys Ile Lys Lys Ala Tyr 225	Cys Pro Ala Gly Val Val 230 235	Glu Gly Asn Pro 240
Ile Met Glu Ile Ala Lys 245	Tyr Phe Leu Glu Tyr Pro 250	Leu Thr Ile Lys 255
Arg Pro Glu Lys Phe Gly 260	Gly Asp Leu Thr Val Asn 265	Ser Tyr Glu Glu 270
Leu Glu Ser Leu Phe Lys 275	Asn Lys Glu Leu His Pro 280	Met Asp Leu Lys 285

ES 2 660 000 T3

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
 290 295 300

Arg Leu
 305

5 <210> 6
 <211> 88
 <212> ADN
 <213> Halobacterium sp. NRC-1
 <400> 6
 cccagggtag ccaagctcgg ccaacggcga cggactctaa atccgttctc gtaggagttc 60
 gagggttcga atcccttccc tgggacca 88

10 <210> 7
 <211> 88
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> ARNt ortogonal que reconoce un codón ópalo
 <400> 7
 gcggggggttg ccgagcctgg ccaaaggcgc cggacttcaa atccggtccc gtaggggttc 60
 cggggttcaa atccccgccc ccgcacca 88

15 <210> 8
 <211> 77
 <212> ADN
 <213> Methanococcus jannaschii
 <400> 8
 ccggcggttag ttcagcaggg cagaacggcg gactctaaat ccgcatggcg ctggttcaaa 60
 tccggcccgc cggacca 77

20 <210> 9
 <211> 65
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido sintético
 <400> 9
 gtaacgctga attcccggcg gtagttcagc agggcagaac ggcggactct aaatccgcat 60
 ggcgc 65

25 <210> 10
 <211> 67
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido sintético
 <400> 10

30
 35

ES 2 660 000 T3

gatctgcagt ggtccggcgg gccggatttg aaccggcgcc atgcggattt agagtccgcc 60
 gttctgc 67
 <210> 11
 <211> 67
 <212> ADN
 5 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido sintético
 <400> 11
 gatctgcagt ggtccggcgg gctggatttg aaccagcgcc atgcggattt agagtccgcc 60
 gttctgc 67
 10 <210> 12
 <211> 67
 <212> ADN
 <213> Secuencia artificial
 <220>
 15 <223> Cebador oligonucleótido sintético
 <400> 12
 gatctgcagt ggtccggcgg gcaggatttg aacctgcgcc atgcggattt agagtccgcc 60
 gttctgc 67
 <210> 13
 <211> 49
 20 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido sintético
 <400> 13
 25 cgccggacca ctgcagatcc ttagcgaaag ctaaggattt ttttaagc 49
 <210> 14
 <211> 23
 <212> ADN
 <213> Secuencia artificial
 30 <220>
 <223> Cebador oligonucleótido sintético
 <400> 14
 caaattcgtc catatggat tcc 23
 <210> 15
 <211> 20
 <212> ADN
 <213> Secuencia artificial
 <220>
 <223> Cebador oligonucleótido sintético
 40 <400> 15
 gtaacgctga attccggcg 20
 <210> 16

ES 2 660 000 T3

<211> 600
 <212> ADN
 <213> Homo sapiens

<400> 16

atggggccacc accaccacca ccacttccca accattccct tatccaggct ttttgacaac 60
 gctatgctcc gcgcccacgc tctgcaccag ctggcctttg acacctacca ggagtttgaa 120
 gaagcctaga tcccaaagga acagaagtat tcattcctgc agaaccacca gacctccctc 180
 tgtttctcag agtctattcc gacaccctcc aacagggagg aaacacaaca gaaatccaac 240
 ctagagctgc tccgcatctc cctgctgctc atccagtcgt ggctggagcc cgtgcagttc 300
 ctcaggagtg tcttcgcaa cagcctggtg tacggcgcct ctgacagcaa cgtctatgac 360
 ctccataaagg acctagagga aggcattcaa acgctgatgg ggaggctgga agatggcagc 420
 ccccgactg ggcagatctt caagcagacc tacagcaagt tcgacacaaa ctcacacaac 480
 gatgacgcac tactcaagaa ctacgggctg ctctactgct tcaggaagga catggacaag 540
 5 gtcgagacat tctgcgcat cgtgcagtgc cgctctgtgg agggcagctg tggcttctaa 600

<210> 17
 <211> 306
 <212> PRT
 <213> Secuencia artificial

10

<220>
 <223> Sintetasa mutante derivada de la sintetasa de Methanococcus jannaschii

<400> 17

Met Asp Glu Phe Glu Met Ile Lys Arg Asn Thr Ser Glu Ile Ile Ser
 1 5 10 15
 Glu Glu Glu Leu Arg Glu Val Leu Lys Lys Asp Glu Lys Ser Ala Val
 20 25 30

ES 2 660 000 T3

Ile Gly Phe Glu Pro Ser Gly Lys Ile His Leu Gly His Tyr Leu Gln
35 40 45

Ile Lys Lys Met Ile Asp Leu Gln Asn Ala Gly Phe Asp Ile Ile Ile
50 55 60

Tyr Leu Ala Asp Leu His Ala Tyr Leu Asn Gln Lys Gly Glu Leu Asp
65 70 75 80

Glu Ile Arg Lys Ile Gly Asp Tyr Asn Lys Lys Val Phe Glu Ala Met
85 90 95

Gly Leu Lys Ala Lys Tyr Val Tyr Gly Ser Glu His Gly Leu Asp Lys
100 105 110

Asp Tyr Thr Leu Asn Val Tyr Arg Leu Ala Leu Lys Thr Thr Leu Lys
115 120 125

Arg Ala Arg Arg Ser Met Glu Leu Ile Ala Arg Glu Asp Glu Asn Pro
130 135 140

Lys Val Ala Glu Val Ile Tyr Pro Ile Met Gln Val Asn Gly Ile His
145 150 155 160

Tyr Glu Gly Val Asp Val Ala Val Gly Gly Met Glu Gln Arg Lys Ile
165 170 175

His Met Leu Ala Arg Glu Leu Leu Pro Lys Lys Val Val Cys Ile His
180 185 190

Asn Pro Val Leu Thr Gly Leu Asp Gly Glu Gly Lys Met Ser Ser Ser
195 200 205

Lys Gly Asn Phe Ile Ala Val Asp Asp Ser Pro Glu Glu Ile Arg Ala
210 215 220

Lys Ile Lys Lys Ala Tyr Cys Pro Ala Gly Val Val Glu Gly Asn Pro
225 230 235 240

Ile Met Glu Ile Ala Lys Tyr Phe Leu Glu Tyr Pro Leu Thr Ile Lys
245 250 255

Arg Pro Glu Lys Phe Gly Gly Asp Leu Thr Val Asn Ser Tyr Glu Glu
260 265 270

ES 2 660 000 T3

Leu Glu Ser Leu Phe Lys Asn Lys Glu Leu His Pro Met Arg Leu Lys
 275 280 285

Asn Ala Val Ala Glu Glu Leu Ile Lys Ile Leu Glu Pro Ile Arg Lys
 290 295 300

Arg Leu
 305

5 <210> 18
 <211> 43
 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 1 de Y31
 <400> 18
 gtattaccct gttatcccta tggcgaaatt cagaatgatt ctc 43

10 <210> 19
 <211> 39
 <212> ADN
 <213> Artificial
 <220>
 15 <223> Cebador 2 de Y31
 <400> 19
 cagaaaccgc ggtgtgtcc taaacaatgt aatcattac 39

20 <210> 20
 <211> 39
 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 3 de Y31
 <400> 20

25 gtaatgatta cattgttag gaacaacacg cggtttctg 39
 <210> 21
 <211> 21
 <212> ADN
 <213> Artificial

30 <220>
 <223> Cebador 4 de Y31
 <400> 21
 gagagcggtg ttgcgtact g 21

35 <210> 22
 <211> 44
 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 5 de Y31

40 <400> 22

5 attaccctgt tatccctaga cgctcagtg aacgaaaact cacg 44
 <210> 23
 <211> 31
 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 6 de Y31
 <400> 23

10 tagggataac agggaatac aatttcaggt g 31
 <210> 24
 <211> 22
 <212> ADN
 <213> Artificial
 <220>
 15 <223> Cebador 7 de Y31
 <400> 24

 ggtcagcttg tcgaaagtac cg 22
 <210> 25
 <211> 40
 20 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 8 de Y31
 <400> 25

25 ctacgggtcg atcatctcca gctagggata acagggtaat 40
 <210> 26
 <211> 43
 <212> ADN
 <213> Artificial
 30 <220>
 <223> Cebador 1 de N51
 <400> 26

 gtattacct gttatcccta tggcgaaatt cagaatgatt ctc 43
 <210> 27
 <211> 34
 35 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 2 de N51
 40 <400> 27

 gattacattg ttaggaaca acacgcggtt tctg 34
 <210> 28
 <211> 34
 45 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 3 de N51

<400> 28
 cagaaaccgc ggtgttcc taaacaatgt aatc 34
 <210> 29
 <211> 21
 5 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 4 de N51
 <400> 29
 10 gagagcggtg tttgcgtact g 21
 <210> 30
 <211> 44
 <212> ADN
 <213> Artificial
 15 <220>
 <223> Cebador 5 de N51
 <400> 30
 attacctgt tatccctaga cgctcagtgg aacgaaaact cacg 44
 <210> 31
 20 <211> 31
 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 6 de N51
 25 <400> 31
 tagggataac agggaatac aattcaggt g 31
 <210> 32
 <211> 22
 30 <212> ADN
 <213> Artificial
 <220>
 <223> Cebador 7 de N51
 <400> 32
 ggtcagcttg tcgaaagtac cg 22
 35 <210> 33
 <211> 33
 <212> ADN
 <213> Artificial
 <220>
 40 <223> Cebador 8 de N51
 <400> 33
 ctaaacaatg taatctaggg ataacaggt aat 33
 <210> 34
 45 <211> 795
 <212> ADN
 <213> Artificial
 <220>

ES 2 660 000 T3

<223> MAP Y31 Mutante ámbar A12

<400> 34

```
atggctatct caatcaagac cccagaagat atcgaaaaaa tgcgcgctcg tggccgactg      60
gctgccgaag tgctggagat gatcgaaccg taggttaaac cgggcgctcag caccggcgag      120
ctggatcgca tctgtaatga ttacattggt aatgaacaac acgcggtttc tgctgcctc      180
ggctatcacg gctatccgaa atccgtttgc atctctatta atgaagtggg gtgccacggt      240
atcccggacg atgctaagct gctgaaagat ggcgatatcg ttaacattga tgtcacgta      300
atcaaagatg gtttccacgg cgatacctcg aaaatgttta tcgtcggtaa gccgaccatc      360
atgggcgaac gtctgtgccg catcacgcaa gaaagcctgt acctggcgct acgcatggta      420
aaaccaggca ttaatctgcg cgaaatcggg gcggcgattc agaaatttgt cgaagcagaa      480
ggcttctccg tcgttcgtga atattgcgga cacggtattg gtcgcggtt ccatgaagaa      540
ccgcaggtgc tgcactatga ctcccgtgaa accaacgctc tactgaaacc tgggatgacg      600
ttcaccatcg agccaatggt caacgcgggt aaaaaagaga tccgcacat gaaagatggc      660
tggacggtaa aaaccaaaga tcgcagcttg tctgcacaat atgagcatac tattgtggtg      720
actgataacg gctgcgaaat tctgacgcta cgcaaggatg acaccatccc ggcgataatc      780
tcgcacgacg aataa                                                                795
```

5

<210> 35

<211> 795

<212> ADN

<213> Artificial

<220>

<223> MAP N51 Mutante ámbar D2

10

<400> 35

ES 2 660 000 T3

atggctatct caatcaagac cccagaagat atcgaaaaaa tgcgcgtcgc tggccgactg	60
gctgccgaag tgctggagat gatcgaaccg tatgttaaac cgggcgtcag caccggcgag	120
ctggatcgca tctgtaatga ttacattggt taggaacaac acgcggtttc tgccctgcctc	180
ggctatcacg gctatccgaa atccgtttgc atctctatta atgaagtggg gtgccacggg	240
atcccggacg atgctaagct gctgaaagat ggcgatatcg ttaacattga tgtcaccgta	300
atcaaagatg gtttccacgg cgatacctcg aaaatgttta tcgtcggtaa gccgaccatc	360
atgggcgaac gtctgtgccg catcacgcaa gaaagcctgt acctggcgct acgcatggta	420
aaaccaggca ttaatctgcg cgaaatcggg gcggcgattc agaaatttgt cgaagcagaa	480
ggcttctccg tcgttcgtga atattgcgga cacggtattg gtcgcggctt ccatgaagaa	540
ccgcaggtgc tgcactatga ctcccgtgaa accaacgtcg tactgaaacc tgggatgacg	600
ttcaccatcg agccaatggg caacgcgggt aaaaaagaga tccgcacat gaaagatggc	660
tggacggtaa aaaccaaaga tcgcagcttg tctgcacaat atgagcatac tattgtggtg	720
actgataacg gctgcgaaat tctgacgcta cgcaaggatg acaccatccc ggcgataatc	780
tcgcacgacg aataa	795

REIVINDICACIONES

- 5 1. Una célula en la que un aminoácido no natural se ha incorporado de forma específica de sitio en un producto génico necesario para la replicación, en la que dicha célula es capaz de replicación en presencia de dicho aminoácido no natural, y tiene una capacidad limitada, o nula, de replicación en ausencia de dicho aminoácido no natural.
2. La célula de la reivindicación 1, en la que la célula es una célula de la subespecie paratuberculosis de *Mycobacterium avium* (MAP).
3. La MAP de la reivindicación 2, en la que más de un aminoácido no natural se incorpora en más de un producto génico necesario para la replicación.
- 10 4. La célula de la reivindicación 1, en la que la célula es una célula de *E. coli*.
5. La célula de la reivindicación 1, en la que la célula es una célula bacteriana.
6. La célula de acuerdo con cualquiera de las reivindicaciones 1 a 5, en la ue el aminoácido no natural se selecciona del grupo que consiste en p-carboxi fenilalanina, p-nitrofenilalanina, p-sulfotirosina, una p-carboxifenilalanina, una o-nitrofenilalanina, una m-nitrofenilalanina, una p-boronil fenilalanina, una o-boronilfenilalanina, una m-boronilfenilalanina, una p-aminofenilalanina, una o-aminofenilalanina, una m-aminofenilalanina, una p-acilfenilalanina, una o-acilfenilalanina, una m-acilfenilalanina, una p-OMe fenilalanina, una o-OMe fenilalanina, una m-OMe fenilalanina, una p-sulfofenilalanina, una o-sulfofenilalanina, una m-sulfofenilalanina, una 5-nitro His, una 3-nitro Tyr, una 2-nitro Tyr, una Leu sustituida con nitro, una His sustituida con nitro, una De sustituida con nitro, una Trp sustituida con nitro, una 2-nitro Trp, una 4-nitro Trp, una 5-nitro Trp, una 6-nitro Trp, una 7-nitro Trp, 3-aminotirosina, 2-aminotirosina, O-sulfotirosina, 2-sulfoxifenilalanina, 3-sulfoxioxifenilalanina, o-carboxifenilalanina, m-carboxifenilalanina, una p-acetil-L-fenilalanina, una p-propargil-fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metil-fenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una tri-O-acetil-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 fosfotirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina, una isopropil-L-fenilalanina, y una p-propargiloxi-fenilalanina.
7. La célula de la reivindicación 6, en la que el aminoácido no natural es p-acetil fenilalanina.
8. La célula de acuerdo con cualquiera de las reivindicaciones 5 a 7, en la que la célula es una célula de *E. coli* y el producto génico necesario para la replicación se selecciona del grupo que consiste en ileS, lspA, ispH, ftsL, ftsl, murE, murD, ftsW, murG, murC, ftsQ, ftsA, ftsZ, secA, map, rpsB, pyrH, frr, uppS, cdsA, fabZ, lpxA, lpxB, dnaE, accA, proS, nusB, dxs, ispA, dnaX, adk, cysS, mrdB, mrdA, holA, leuS, glnS, serS, msbA, kdsB, asnS, fabA, rne, fabD, fabG, tmk, holB, lolC, lolD, lolE, trmU, prsA, topA, fabI, pheS, rplT, infC, nadE, aspS, argS, rplY, nrdA, nrdB, fabB, gltX, ligA, zipA, yfgE, hisS, acpS, rplS, trmD, rpsP, ffh, grpE, plsC, parC, parE, ribB, rpoD, rpsO, yhbZ, rpsI, rplM, def, rpsM, secY, rplO, rpsE, rplR, rplF, rpsN, rplX, rplN, rpsQ, rpsS, rplB, rplW, rplD, rplC, rpsJ, fusA, rpsL, trpS, ftsY, glyS, kdtA, coaD, dut, gmk, dnaA, glmU, engB, murl, rplJ, rplL, rpoB, rpoC, dnaB, groS, groL, efp, rpsF, ppa, valS, dnaC, y dnaT.
9. La célula de la reivindicación 8, en la que el producto génico necesario para la replicación es metionina aminopeptidasa (map).
10. Una vacuna que comprende la célula de una cualquiera de las reivindicaciones 1 a 9.

40

Figura 1

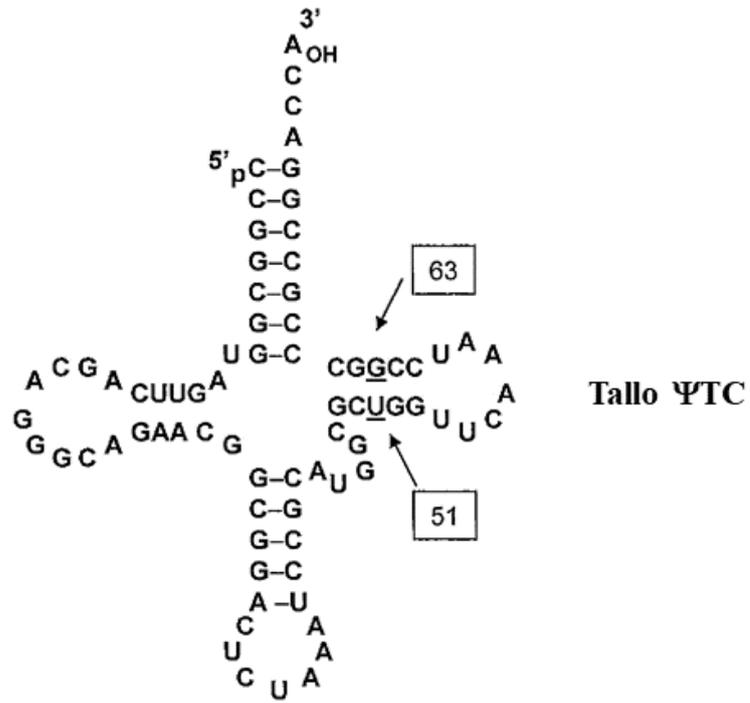


Figura 2

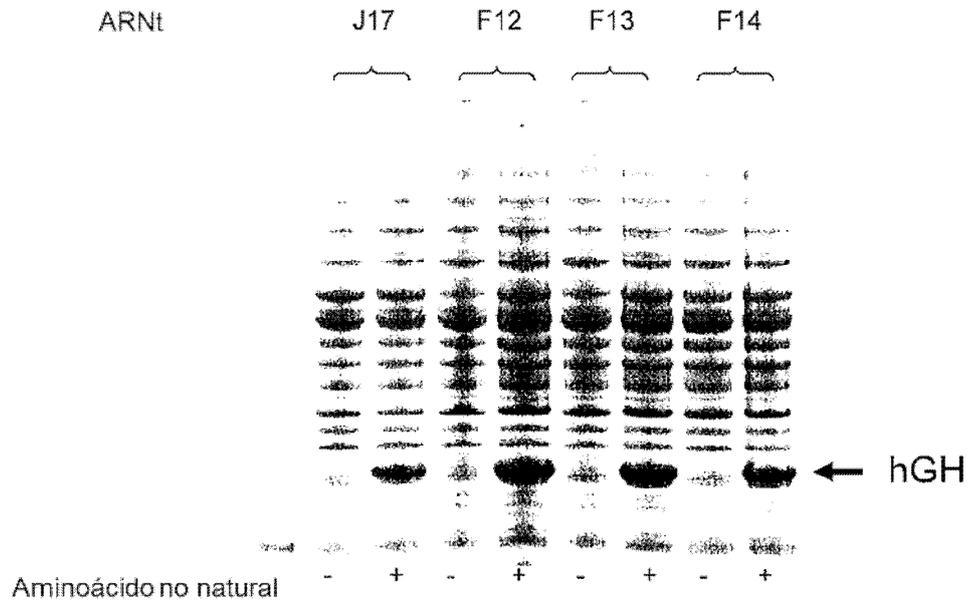


Figura 3

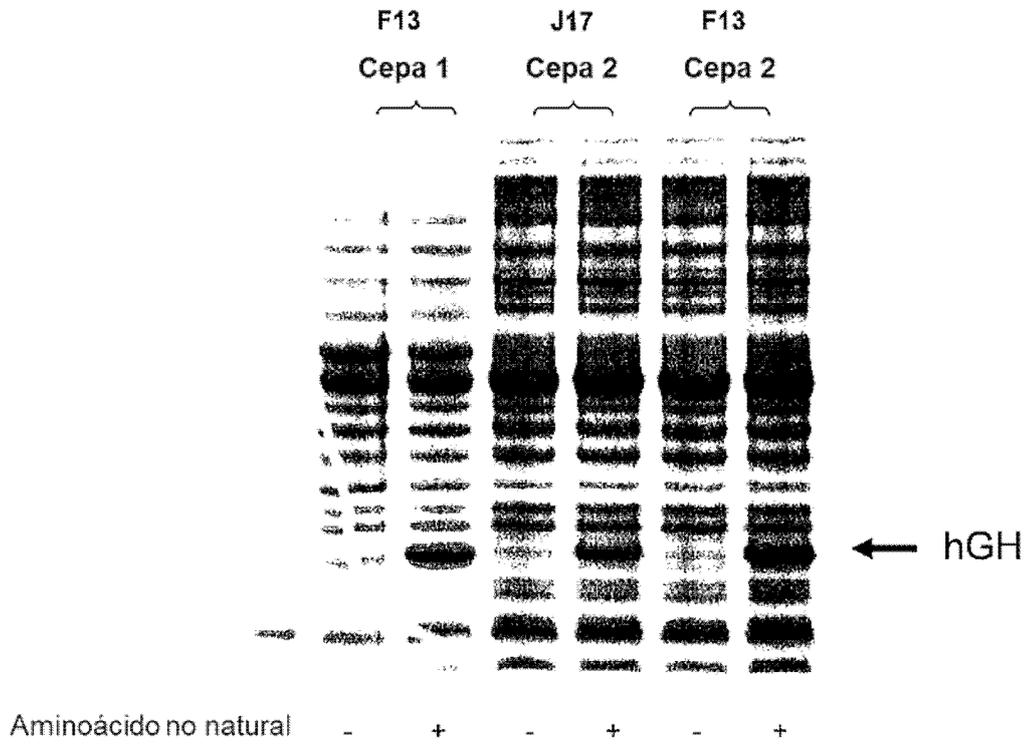


Figura 4

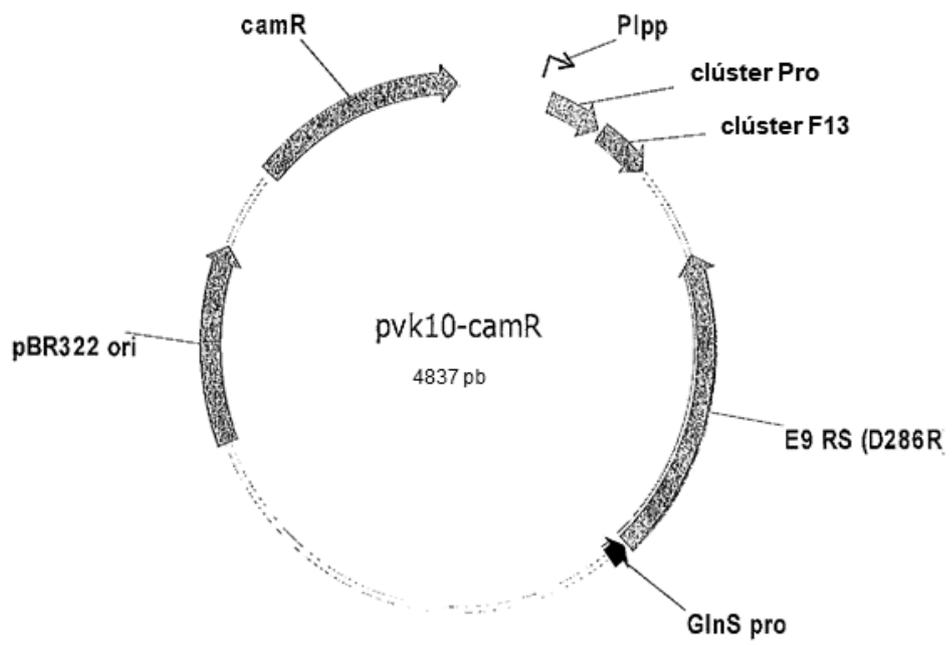


Figura 5

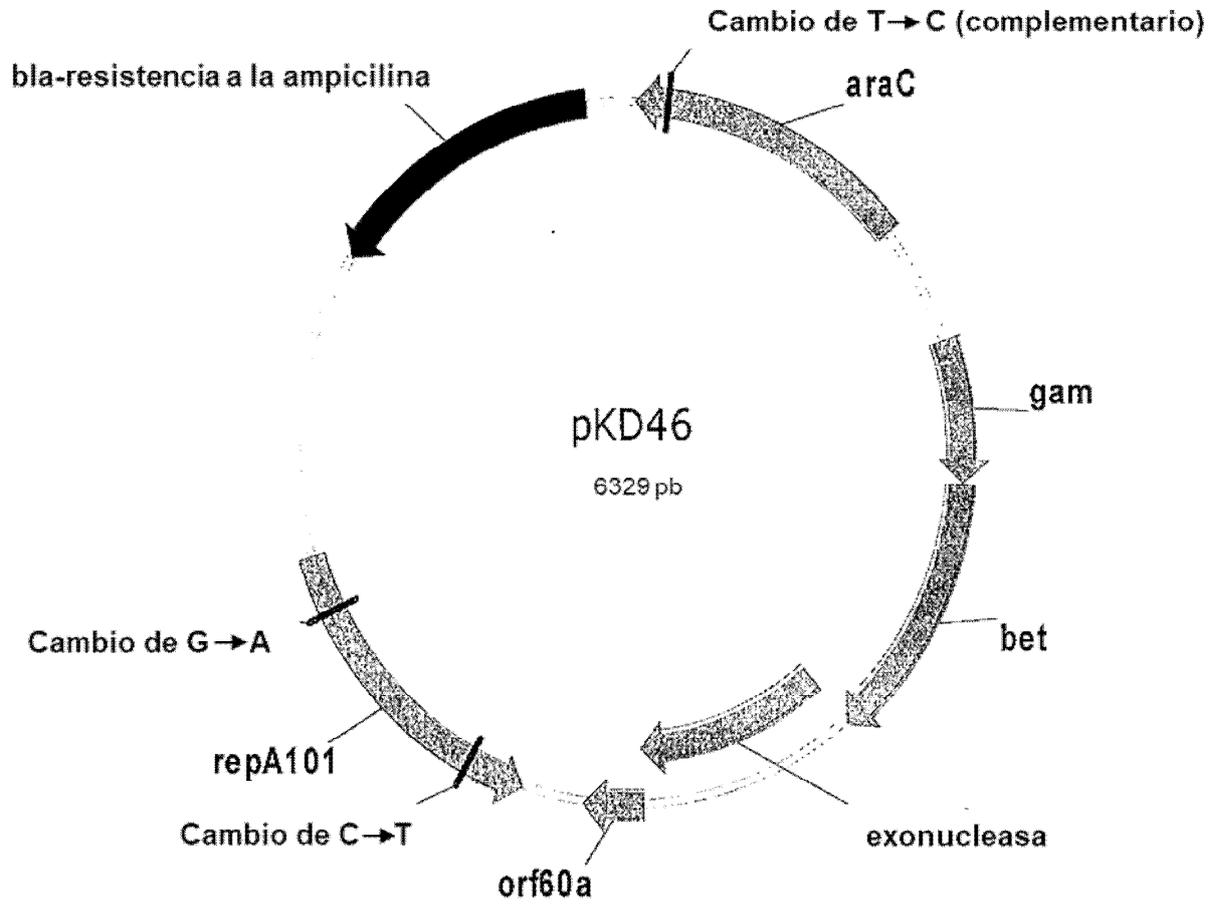


Figura 6

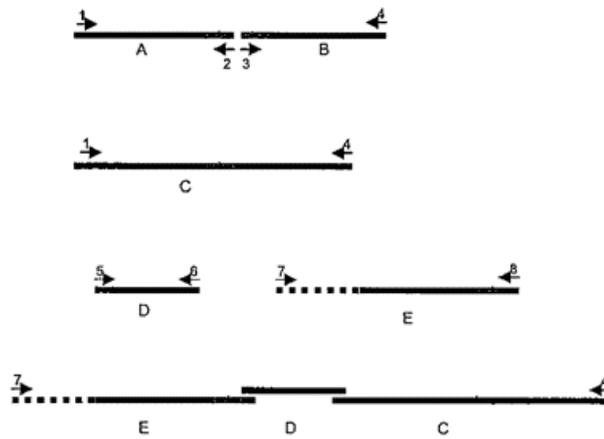


Figura 7

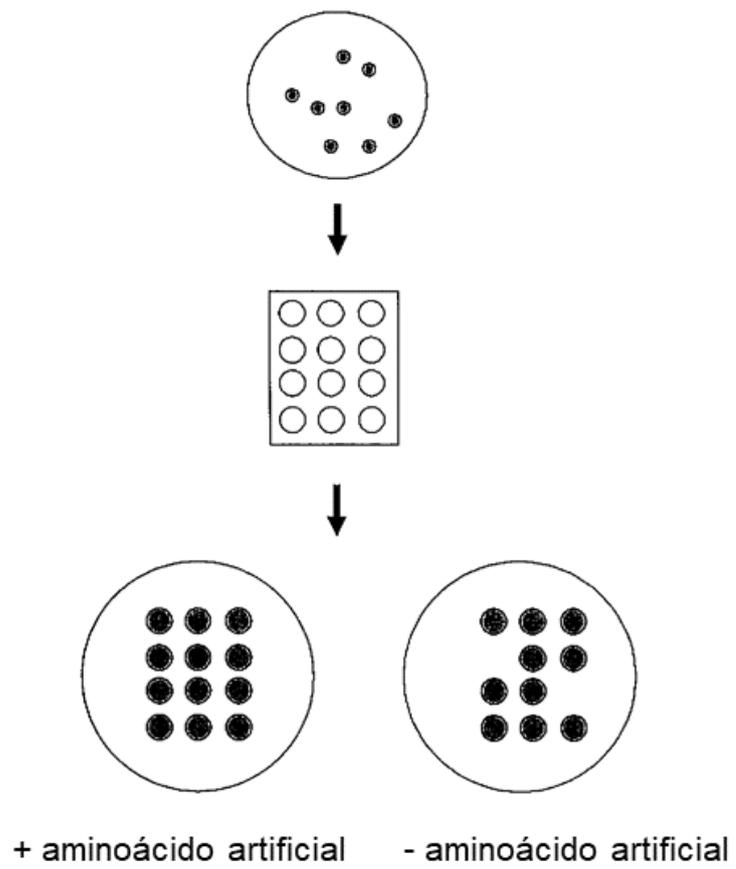


Figura 8

Y31 – 1^a Criba

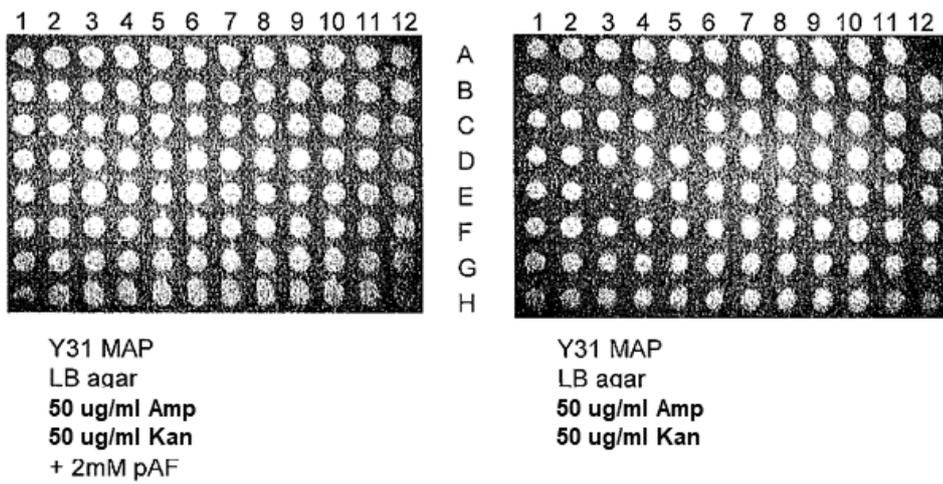


Figura 9

YN51 – 1ª Criba |

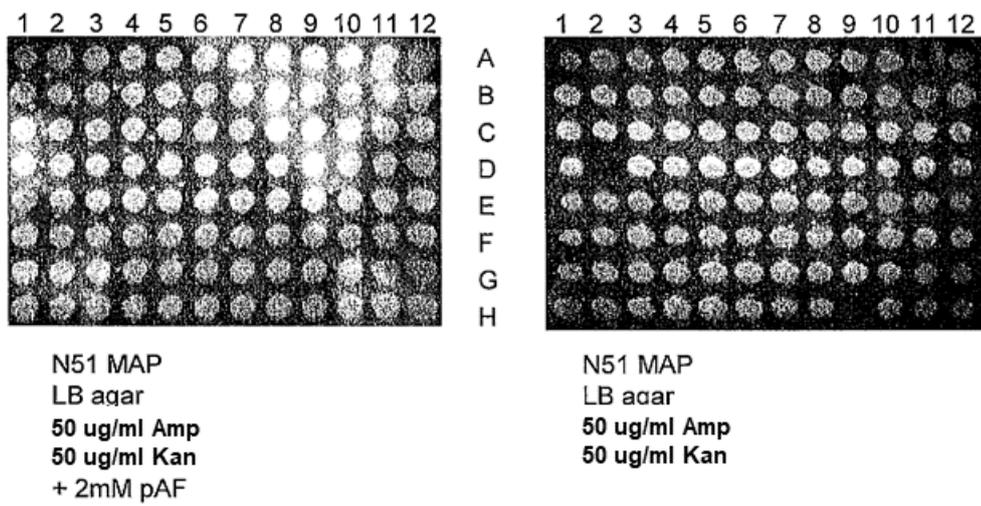


Figura 10A

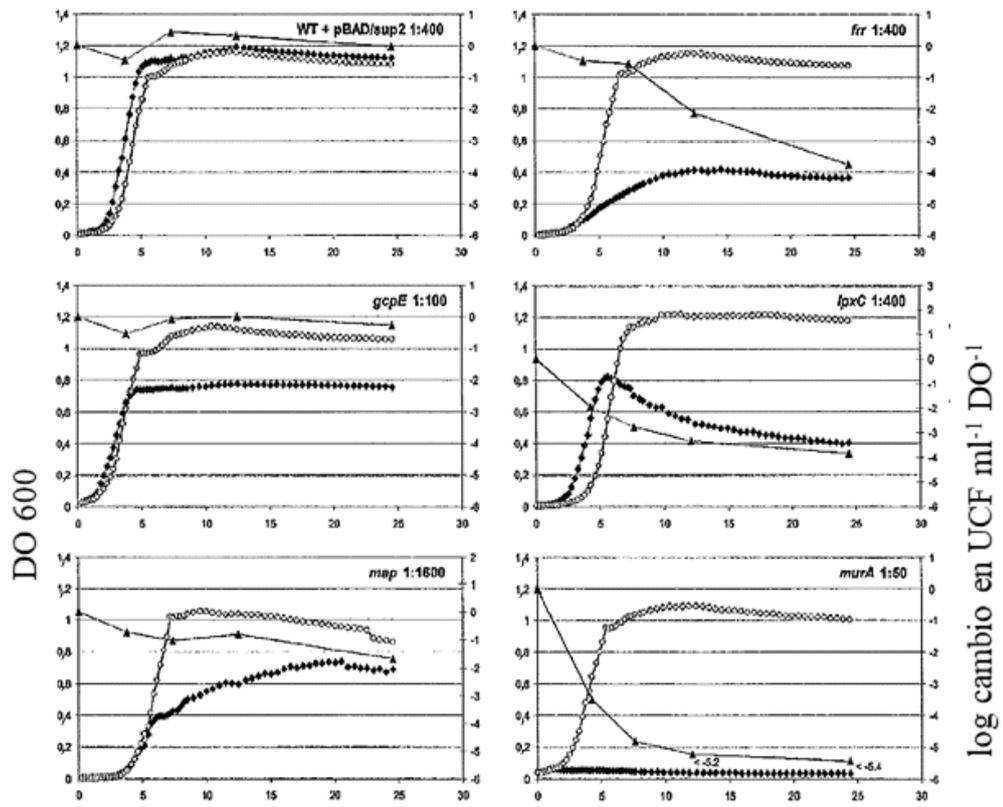


Figura 10B

Horas

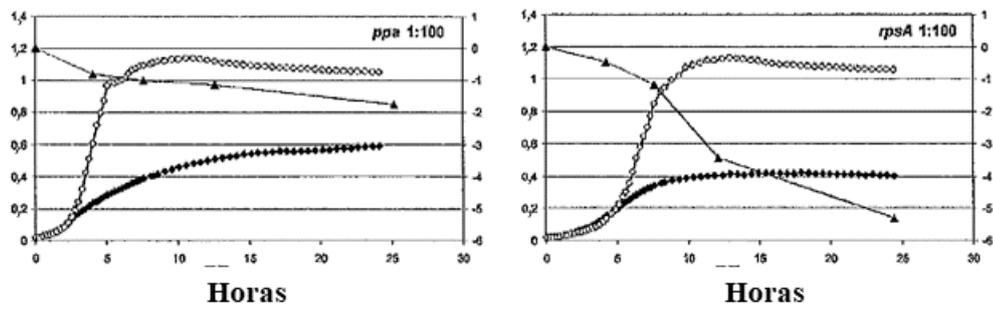


Figura 11

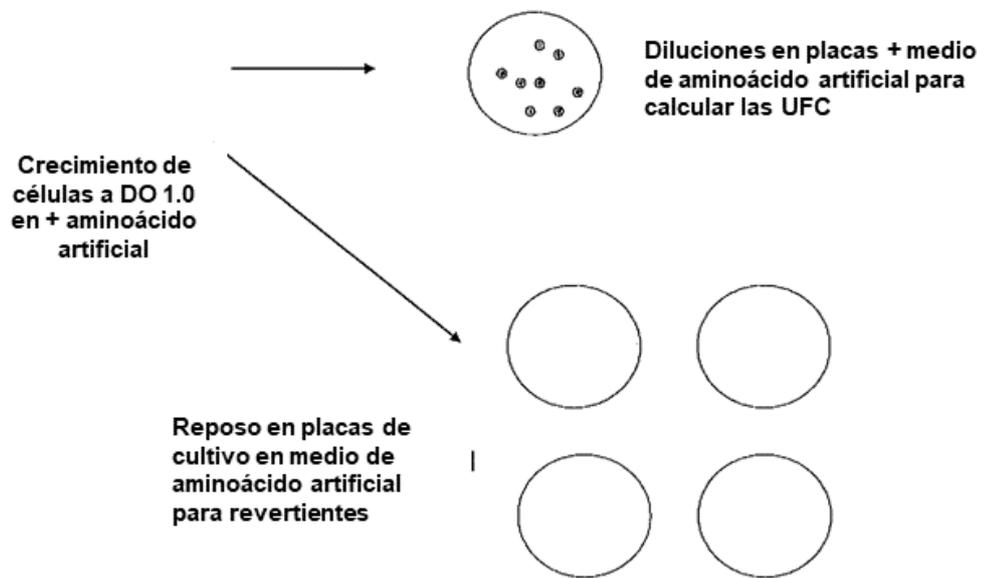


Figura 12

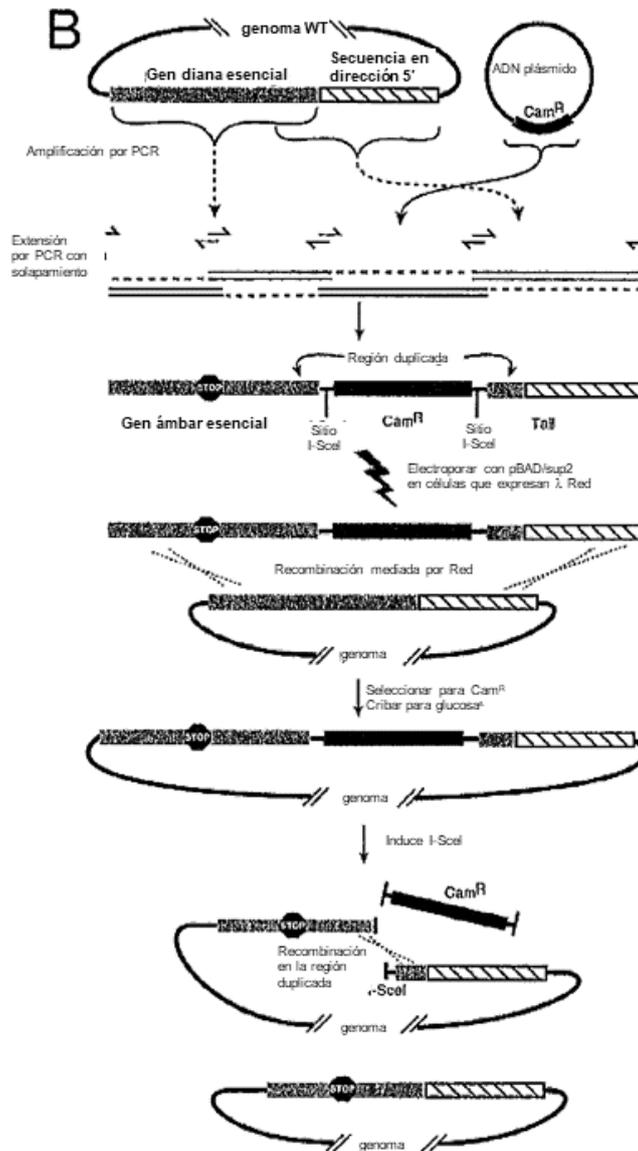


Figura 13

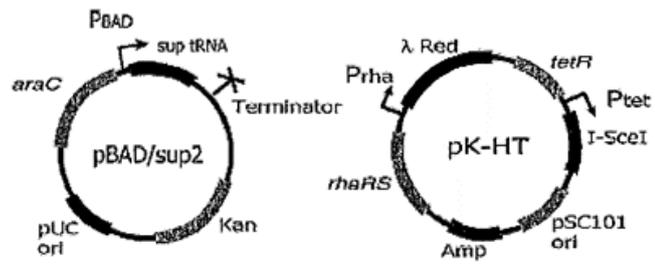


Figura 14

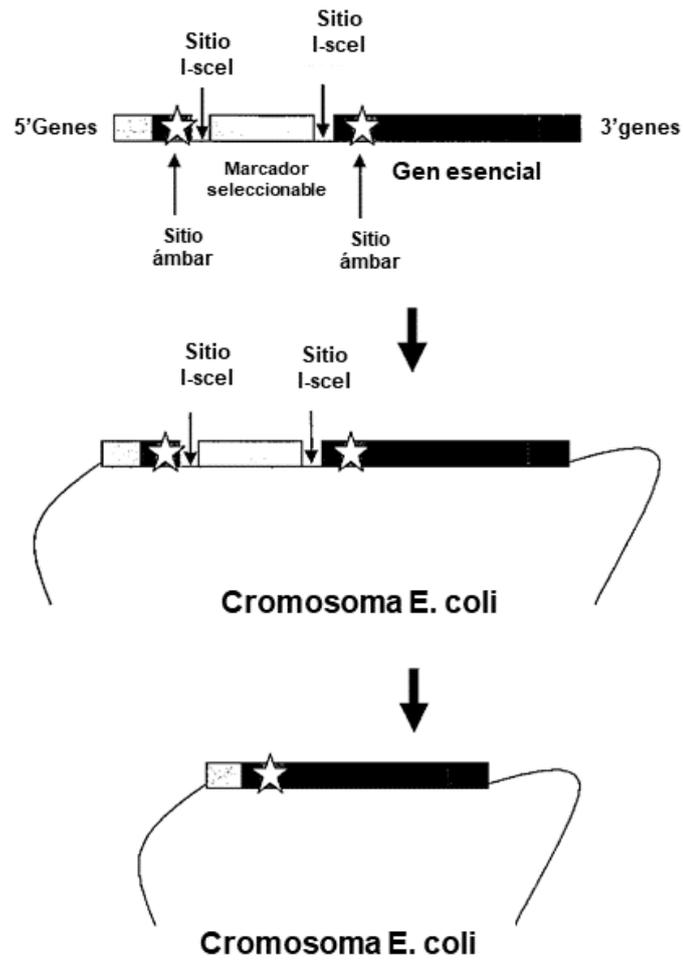


Figura 15

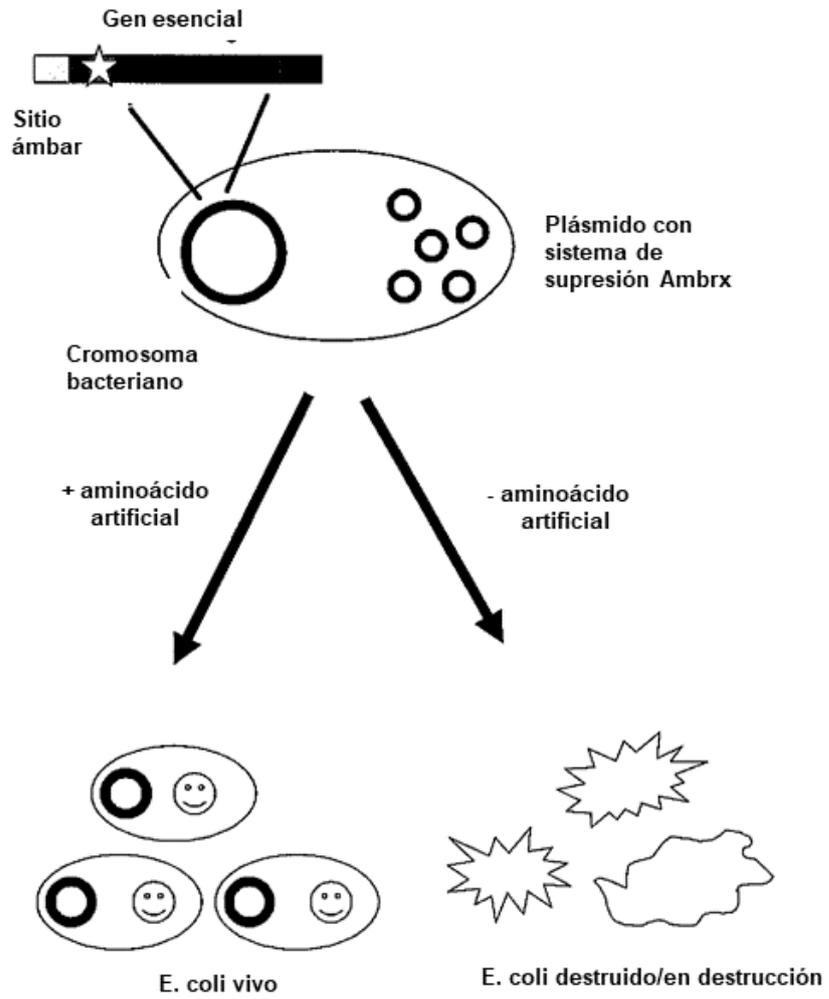


Figura 16

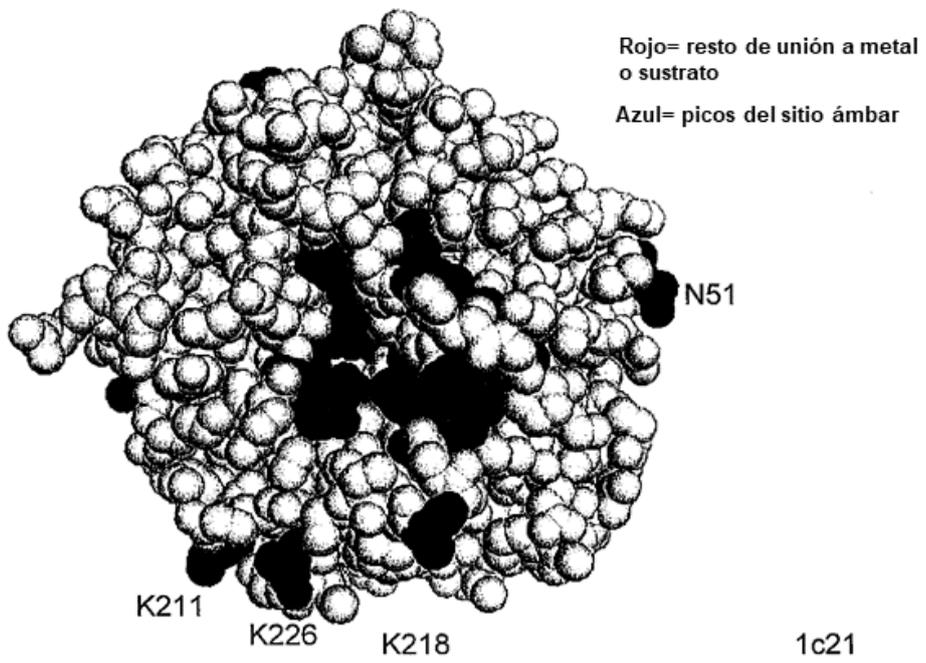
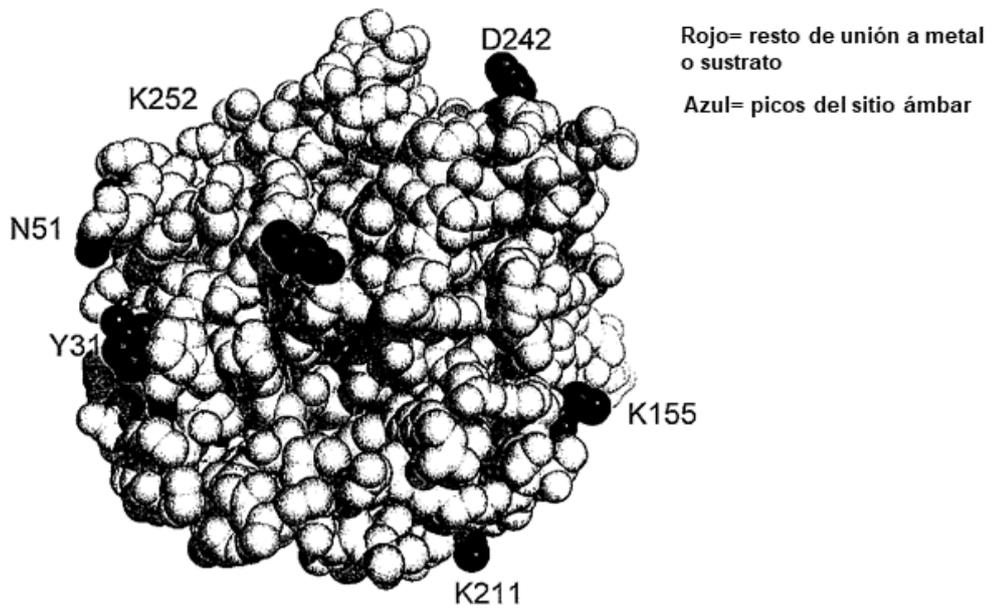


Figura 17



1c21