

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 504 521**

51 Int. Cl.:

C12N 15/09

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **05.03.2007** **E 07752295 (1)**

97 Fecha y número de publicación de la concesión europea: **25.06.2014** **EP 1999259**

54 Título: **Incorporación específica del sitio de aminoácidos en moléculas**

30 Prioridad:

03.03.2006 US 779375 P
03.03.2006 US 779376 P

45 Fecha de publicación y mención en BOPI de la
traducción de la patente:
08.10.2014

73 Titular/es:

CALIFORNIA INSTITUTE OF TECHNOLOGY
(100.0%)
1200 EAST CALIFORNIA BLVD. MS 201-85
PASADENA CA 91125, US

72 Inventor/es:

WANG, PIN;
KWON, INCHAN;
SON, SOOJIN;
TANG, YI y
TIRRELL, DAVID A.

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 504 521 T3

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

DESCRIPCIÓN

Incorporación específica del sitio de aminoácidos en moléculas

5 Declaración de interés del gobierno

Esta invención se realizó con el apoyo del gobierno federal bajo el número de concesión GM62523, otorgado por el NIH. El gobierno de Estados Unidos tiene ciertos derechos sobre la invención.

10 Antecedentes de la invención

La ingeniería de proteínas es una potente herramienta para la modificación de las propiedades catalíticas estructurales y de unión de proteínas naturales y para el diseño *de novo* de proteínas artificiales. La ingeniería de proteínas se basa en un mecanismo de reconocimiento eficiente para la incorporación de aminoácidos mutantes en las secuencias de proteínas deseadas. Aunque este proceso ha sido muy útil para el diseño de nuevas macromoléculas con un control preciso de la composición y la arquitectura, una limitación importante es que la mutagénesis está restringida a los 20 aminoácidos que se producen de forma natural. Sin embargo, cada vez es más evidente que la incorporación de aminoácidos no naturales puede ampliar el alcance y el impacto de los métodos de ingeniería de proteínas.

Los aminoácidos no naturales que portan una amplia diversidad de nuevos grupos funcionales han sido reemplazados globalmente para el reemplazo o la incorporación específica para el residuo en proteínas recombinantes. La asimilación biosintética de aminoácidos no canónicos en proteínas se ha logrado en gran medida explotando la capacidad del aparato de síntesis de tipo salvaje de utilizar análogos de aminoácidos de origen natural (Budisa 1995, *Eur. J. Biochem* 230: 788-796; Deming 1997, *J. Macromol. Sci. Pure Appl. Chem* A34: 2143-2150; Duewel 1997, *Biochemistry* 36: 3404-3416; van Hest y Tirrell 1998, *FEBS Lett* 428(1-2): 68-70; Sharma *et al.*, 2000, *FEBS Lett* 467(1): 37-40.). Sin embargo, hay situaciones en las que se requiere la sustitución o incorporación en un solo sitio de aminoácidos no naturales. Una metodología de este tipo permitiría la adaptación en una proteína (el tamaño, la acidez, nucleofilicidad, enlaces hidrógeno o propiedades hidrófobas, etc. de aminoácidos) para cumplir una propiedad estructural o funcional específica de interés. La capacidad de incorporar de modo específico para el sitio tales análogos de aminoácidos en proteínas ampliaría en gran medida nuestra capacidad de manipular de forma racional y sistemática las estructuras de proteínas, tanto para investigar la función de proteínas como para crear proteínas con nuevas propiedades. Por ejemplo, la capacidad de sintetizar grandes cantidades de proteínas que contienen átomos pesados facilitaría la determinación de la estructura de las proteínas, y la capacidad de sustituir de modo específico para el sitio fluoróforos o grupos foto-escindibles en proteínas en células vivas proporcionaría potentes herramientas para el estudio de las funciones de proteínas *in vivo*.

En los últimos años, varios laboratorios han llevado a cabo una expansión en el número de aminoácidos codificados genéticamente, utilizando un supresor sin sentido o un ARNt supresor del desplazamiento del marco para incorporar aminoácidos no canónicos en proteínas en respuesta a codones ámbar o de cuatro bases, respectivamente (Bain *et al.*, *J. Am. Chem. Soc.* 111: 8013, 1989; Noren *et al.*, *Science* 244: 182, 1989; Furter, *Protein Sci.* 7: 419, 1998; Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 100: 56, 2003; Hohsaka *et al.*, *FEBS Lett.* 344: 171: 1994; Kowal y Oliver, *Nucleic Acids Res.* 25: 4685, 1997). Tales métodos insertan aminoácidos no canónicos en posiciones del codón que normalmente terminarían la síntesis de péptidos de tipo salvaje (p. ej., un codón de terminación o una mutación de desplazamiento del marco). Estos métodos han funcionado bien para la inserción en un solo sitio de nuevos aminoácidos. Sin embargo, su utilidad en la sustitución o incorporación específica para la posición en múltiples sitios (frente a específica para residuos) está limitada por modestas eficacias de supresión (20-60%) (Anderson *et al.*, *J. Am. Chem. Soc.* 124: 9674, 2002; Bain *et al.*, *Nature* 356: 537, 1992; Hohsaka *et al.*, *Nucleic Acids Res.* 29: 3646, 2001). Esto es así, en parte debido a que una eficacia de supresión demasiado alta del codón de terminación interferirá con la terminación normal de la traducción de algunas proteínas no fijadas como objetivo en el organismo. Por otra parte, una baja eficacia de supresión será probablemente insuficiente para suprimir más de un sitio sin sentido o de mutación de desplazamiento del marco en la proteína diana, de tal manera que se hace más y más difícil o poco práctico sintetizar una proteína diana de longitud completa incorporando más y más aminoácidos no canónicos.

Una incorporación eficaz en múltiples sitios se ha logrado mediante la sustitución de aminoácidos naturales en cepas auxotróficas de *Escherichia coli*, por ejemplo utilizando aminoacil-ARNt sintetasas con especificidad relajada para el sustrato o actividad de edición alterada (Wilson y Hatfield, *Biochim. Biophys. Acta* 781: 205, 1984; Kast y Hennecke, *J. Mol. Biol.* 222: 99, 1991; Ibba *et al.*, *Biochemistry* 33: 7107, 1994; Sharma *et al.*, *FEBS Lett.* 467: 37, 2000; Tang y Tirrell, *Biochemistry* 41: 10635, 2002; Datta *et al.*, *J. Am. Chem. Soc.* 124: 5652, 2002; Doring *et al.*, *Science* 292: 501, 2001). Aunque este método proporciona una incorporación eficaz de análogos en múltiples

sitios, adolece de la limitación de que los nuevos aminoácidos deben "compartir" codones con uno de los aminoácidos naturales. Así, para cualquier posición del codón dada en la que se puedan insertar aminoácidos tanto naturales como nuevos, con excepción de una probabilidad de incorporación, hay relativamente poco control sobre qué aminoácido terminará siendo insertado. Esto puede ser indeseable, ya que para una enzima o proteína modificada, la incorporación no canónica de aminoácidos en un sitio no pretendido puede comprometer de forma inesperada la función de la proteína, mientras que omitir la incorporación del aminoácido no canónico en el sitio diseñado fracasará en lograr el objetivo de diseño.

En general, métodos de sustitución en múltiples sitios son relativamente simples de llevar a cabo, pero están reemplazados todos los sitios correspondientes a un aminoácido natural particular a lo largo de la proteína. El grado de incorporación del aminoácido natural y no natural también puede variar. Además de ello, la incorporación de múltiples sitios de análogos resulta, a menudo, en toxicidad cuando se utilizan células, lo que hace difícil estudiar la proteína mutante en células vivas. La presente invención supera estos obstáculos al permitir una mutación específica para el sitio de aminoácidos en proteínas.

Determinadas realizaciones descritas en esta memoria proporcionan una nueva técnica para la incorporación de aminoácidos de sustitución, incluidos aminoácidos de origen natural o aminoácidos no estándares o no canónicos, en proteínas que se basa en romper la degeneración del código genético. Específicamente, determinadas realizaciones en esta memoria permiten una sustitución específica para la posición de alta fidelidad o la incorporación de aminoácidos no naturales en proteínas.

El documento US2004/0053390 Datta et al.) describe métodos, reactivos y herramientas computacionales para diseñar análogos de sustrato no naturales para enzimas, especialmente para diseñar análogos de aminoácidos no naturales para aminoacil-ARNt sintetasas tal como la Phe-ARNt sintetasa.

Breve resumen de la invención

Determinadas realizaciones descritas en esta memoria proporcionan composiciones de componentes utilizados en la maquinaria biosintética de proteínas, que incluyen moléculas de aminoacil-ARNt sintetasa (AARS – siglas en inglés) mutantes externas, opcionalmente emparejadas con moléculas de aminoacil-ARNt mutantes externas.

También se proporcionan métodos para la generación y selección de ARNts mutantes externos, aminoacil-ARNt sintetasas mutantes externas, y pares de los mismos, que son capaces de incorporar aminoácidos, incluidos aminoácidos no naturales, en polipéptidos o proteínas. Determinadas composiciones de formas de realización específicas incluyen nuevos pares de ARNt mutante externo o aminoacil-ARNt mutante externo. Las nuevas moléculas de ARNt, moléculas de AARS mutantes externas, o pares de AARS-ARNt mutantes externos se pueden utilizar para incorporar un aminoácido no natural en un polipéptido *in vitro* e *in vivo*. Otras realizaciones de la invención incluyen la selección de pares de mutantes externos.

Composiciones de la presente invención incluyen una aminoacil-ARNt sintetasa mutante externa, en que la ARNt sintetasa mutante externa aminoacila preferiblemente un ARNt mutante externo con un aminoácido no natural, opcionalmente *in vivo*. En una forma de realización, se proporciona un ácido nucleico o polinucleótido que codifica una sintetasa mutante externa, o una secuencia de ácidos nucleicos complementaria de la misma.

Por lo tanto, determinadas formas de realización incluyen una composición que comprende un primer vector que contiene un polinucleótido que codifica una aminoacil fenilalanil ARNt sintetasa (PheRS) de levadura modificada, en donde dicho polinucleótido está mutado en codones que codifican el aminoácido en las posiciones de secuencia seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3, y en donde dicha sintetasa modificada es capaz de cargar una molécula de ARNt con un aminoácido no natural.

Al menos una forma de realización comprende, además, un segundo vector que contiene un polinucleótido que codifica una molécula de ARNt. En al menos una forma de realización, dichos primer y segundo vectores son el mismo vector. En otras formas de realización, dichos primer y segundo vectores son vectores diferentes.

En al menos una forma de realización, el ARNt es endógeno, y en al menos una forma de realización, el ARNt está modificado. En al menos una forma de realización, el ARNt está modificado de modo que contiene un anticodón mutado que se empareja en bases con un correspondiente codón degenerado de bamboleo con una afinidad mayor que la afinidad del ARNt natural. En algunas formas de realización, los AARS y el ARNt son del mismo o de diferentes organismos. En al menos una forma de realización, el aminoácido no natural se selecciona del grupo que consiste en: azidonorleucina, 3-(1-naftil)alanina, 3-(2-naftil)alanina, *p*-etnil-fenilalanina, *p*-propargil-oxi-

fenilalanina, *m*-etnil-fenilalanina, 6-etnil-triptófano, 5-etnil-triptófano, ácido (R)-2-amino-3-(4-etnil-1H-pirol-3-il)propanoico, *p*-bromofenilalanina, *p*-yodofenilalanina, *p*-azidofenilalanina, 3-(6 cloroindolil)alanina, 3-(6-bromoindolil)alanina, 3-(5-bromoindolil)alanina, azidohomoalanina y *p*-clorofenilalanina.

Otra forma de realización comprende un polipéptido que comprende una fenilalanil ARNt sintetasa (PheRS) de levadura modificada, en donde dicha sintetasa modificada está mutada en las posiciones de la secuencia de aminoácidos en codones que codifican el aminoácido en las posiciones de secuencia seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3.

Determinadas realizaciones incluyen un sistema de traducción que comprende un polinucleótido que codifica una fenilalanil ARNt sintetasa de levadura modificada, en donde dicho polinucleótido de sintetasa modificada está mutado en uno o más codones que codifican el aminoácido en las posiciones de secuencia seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3, y en donde dicha sintetasa modificada es capaz de cargar una molécula de ARNt con un aminoácido no natural. En al menos una forma de realización, el sistema comprende una célula huésped. En al menos una forma de realización, la aminoacil ARNt sintetasa modificada se deriva de un organismo diferente de la célula huésped. En otra forma de realización, el sistema de traducción comprende, además, un polinucleótido que codifica una molécula de ARNt modificado.

En determinadas formas de realización, la molécula de ARNt modificado se deriva de un organismo diferente de la célula huésped. En determinadas formas de realización, la molécula de ARNt modificado se deriva de una célula eucariota y la célula huésped es una célula procariota. En aún otras formas de realización, la célula es una auxótrofo.

En algunas formas de realización, el sistema de traducción comprende, además, un medio de cultivo que contiene uno o más aminoácidos no naturales. En todavía otras formas de realización, dichos uno o más aminoácidos no naturales se seleccionan del grupo que consiste en: azidonorleucina, 3-(1-naftil)alanina, 3-(2-naftil)alanina, *p*-etnil-fenilalanina, *p*-propargil-oxi-fenilalanina, *m*-etnil-fenilalanina, 6-etnil-triptófano, 5-etnil-triptófano, ácido (R)-2-amino-3-(4-etnil-1H-pirol-3-il)propanoico, *p*-bromofenilalanina, *p*-yodofenilalanina, *p*-azidofenilalanina, 3-(6 cloroindolil)alanina, 3-(6-bromoindolil)alanina, 3-(5-bromoindolil)alanina, azidohomoalanina y *p*-clorofenilalanina.

Otras formas de realización se refieren a un método para incorporar un aminoácido no natural en un polipéptido diana en una o más posiciones especificadas, comprendiendo el método las etapas de: proporcionar un sistema de traducción que contiene al menos un aminoácido no natural; proporcionar al sistema de traducción una o más fenilalanil ARNt sintetastas de levadura modificadas según se define arriba, proporcionar al sistema de traducción un polinucleótido que codifica un polipéptido diana de interés, y permitir la traducción del polipéptido diana de interés, incorporando con ello al menos un aminoácido no natural en el polipéptido diana, en donde las etapas se efectúan en cualquier orden.

En determinadas formas de realización, dicho sistema de traducción comprende una célula. En algunas formas de realización, se proporciona el aminoácido no natural poniendo en contacto dicho sistema de traducción con una disolución que contiene el aminoácido no natural. En al menos una forma de realización, la constante de especificidad (k_{cat} / K_M) para la activación de dicho aminoácido no natural por parte de dicha AARS modificada es al menos 5 veces mayor que la de dicho aminoácido natural. En determinadas formas de realización, la AARS modificada carga erróneamente un ARNt a una velocidad de no más de 1%, 2%, 3%, 4%, 5%, 6%, 7% u 8%. En aún otras formas de realización, el ARNt es un ARNt modificado. En determinadas formas de realización, dicho primer polinucleótido o dicho segundo polinucleótido comprende una secuencia de promotor constitutivamente activa o inducible que controla la expresión del ARNt o de la PheRS. En al menos una forma de realización, el método comprende, además, la etapa de rastreo de células que contienen una PheRS modificada. En otra forma de realización, el método comprende, además, la etapa de verificar la incorporación del aminoácido no natural. Aún otras formas de realización comprenden un polipéptido hecho por el método descrito.

Determinadas formas de realización descritas en esta memoria incluyen un método para la incorporación de al menos un aminoácido no natural en un polipéptido diana en una o más ubicaciones especificadas, comprendiendo el método proporcionar un sistema de traducción que contiene al menos un aminoácido no natural; proporcionar al sistema de traducción una o más PheRS modificadas; proporcionar al sistema de traducción un polinucleótido que codifica un polipéptido diana de interés; y permitir la traducción de interés, incorporando con ello al menos un aminoácido no natural en el polipéptido diana. Determinadas formas de realización incluyen un polipéptido hecho por este método.

Breve descripción de las diversas vistas de los dibujos

La Figura 1 muestra una alineación de secuencia de variantes PheRS a partir de secuencias conservadas de *Thermus thermophilus*, *Escherichia coli* y *Saccharomyces cerevisiae*.

La Figura 2 muestra varios aminoácidos ilustrativos (que se producen de forma natural o no natural) utilizados para algunas realizaciones descritas en esta memoria.

La Figura 3 muestra la secuencia de aminoácidos de un polipéptido ilustrativo utilizado para algunas formas de realización descritas en esta memoria, dihidrofolato reductasa (DHFR). Se utilizaron cuatro fragmentos de péptidos proteolíticos (marcados Péptido A, Péptido B, Péptido C y Péptido D) para MALDI y análisis de cromatografía líquida-Espectro de Masas/Espectro de masas (LC-MS/MS) tal como se destaca.

La Figura 4 muestra una MALDI-MS de fragmentos de péptidos proteolíticos derivados de mDHFR expresada en medios suplementados con (a) aminoácido 1 (3 mM); (B) aminoácidos 7 (3 mM) y 1 (0,03 mM); (c) aminoácidos 2 (3 mM) y 1 (0,03); (d) aminoácidos 2 (3 mM) y 1 (0,03 mm). No se suplementa nada de triptófano durante la inducción, excepto que 1 mM de triptófano se suplementa en medios en (c). Péptido B, que contiene un codón Phe, es el control. Péptido A contiene un codón ámbar (Z), el aminoácido para el que se asigna basado en las unidades de masa para el péptido A a diferentes condiciones de expresión. Debido a la incorporación de lisina con Péptido A (c), la lisina C-terminal se escinde para producir un péptido A más corto (NGDLPWPPLRNEK) (SEQ ID NO: 4).

Figura 5 Espectro de masas tándem de Péptido A (NGDLPWPPLRNEZK) (SEQ ID NO: 5) derivado de DHFR expresada en medios suplementados con aminoácido 7 (3 mM) y aminoácido 1 (0,03 mM). La secuencia parcial de PWPPLRNE (SEQ ID NO: 6) y del residuo Z (correspondiente al aminoácido 7) del Péptido A puede ser asignada a la serie de iones y y b anotadas, respectivamente.

La Figura 6 muestra una MALDI-MS de fragmentos de péptidos proteolíticos derivados de mDHFR expresada en medios suplementados con (a) aminoácido 2 (3 mM) y aminoácido 6 (3 mM) y aminoácido 1 (0,03 mM) y aminoácido 13 (0,2 mM); (b) aminoácido 2 (3 mM) y aminoácido 6 (0,01 mM) y aminoácido 1 (0,03 mM) y aminoácido 13 (0,2 mM); (c) aminoácido 3 (3 mM) y aminoácido 6 (0,01 mM) y aminoácido 1 (0,03 mM) y aminoácido 13 (0,01 mM). Péptido C, con un codón Phe, es el control.

La Figura 7 muestra los resultados de MALDI-MS de fragmentos de péptidos proteolíticos derivados de mDHFR expresada en medios suplementados con (a) aminoácido 9 (3 mM) y aminoácido 6 (0,03 mM) y aminoácido 1 (0,01 mM); (b) aminoácido 10 (3 mM) y aminoácido 6 (0,03 mM) y aminoácido 1 (0,03 mM); (c) aminoácido 11 (3 mM) y aminoácido 6 (0,01 mM) y aminoácido 1 (0,03 mM). Péptido D era el control.

La Figura 8 muestra la aminoacilación de ARNt^{Phe}_{CUA} de levadura (cuadrados) y ARNt^{Phe}_{CUA_UG} (círculo) con lisina mediante eLysS.

La Figura 9 muestra cromatogramas de LC-MS de digestiones tripticas de los polipéptidos mDHFR.

La figura 10 muestra las tasas de intercambio ATP-IPP para fenilalanina, triptófano y p-bromofenilalanina por parte de PheRS de levadura de tipo salvaje y PheRS de levadura T415G mutante externa o PheRS de levadura T415A mutante externa.

La figura 11 muestra la aminoacilación de fenilalanina y triptófano por parte de aminoacil ARNt sintetasa de tipo salvaje o ARNt sintetasa mutante externa (T415G o T415A).

La figura 12 muestra la incorporación de lisina (cuadrado en blanco), triptófano (rayado) o p-bromofenilalanina (tablero de ajedrez).

La Figura 13 muestra los espectros de masas de p-etilfenilalanina incorporada en un polipéptido utilizando una ARNt sintetasa modificada (T415G) y ARNt^{Phe} con un supresor ámbar en una célula huésped.

La Figura 14 muestra un FACS de proteína fluorescente verde (GFP) la incorporación de aminoácidos utilizando el codón de supresión ámbar en una proteína de ensayo.

La Figura 15 ilustra una representación en mapa ilustrativa de un plásmido de una aminoacil ARNt sintetasa mutante (T415G).

La Figura 16 ilustra mutaciones ilustrativas realizadas en una fenilalanina ARNt sintetasa de levadura.

Descripción detallada de la invención

Las proteínas se encuentran en la encrucijada de prácticamente cada uno de los procesos biológicos, desde la fotosíntesis y la visión a la transducción de señales y la respuesta inmune. La modificación de proteínas o polipéptidos para incluir aminoácidos no naturales tiene un gran potencial para su uso en productos terapéuticos humanos, la agricultura, los biocombustibles y otras áreas.

Aminoacil-ARNt sintetasas catalizan la reacción de aminoacilación para la incorporación de aminoácidos en proteínas a través de las moléculas de ARN de transferencia correspondientes. La manipulación precisa de la actividad sintetasa puede alterar la especificidad de aminoacilación para fijar de manera estable aminoácidos no canónicos en el ARNt pretendido. Entonces, a través de la interacción codón-anticodón entre el ARN mensajero (ARNm) y el ARNt, los análogos de aminoácidos pueden ser suministrados a una cadena polipeptídica en crecimiento. Por lo tanto, la incorporación de aminoácidos no naturales en proteínas se basa en la manipulación de especificidad de aminoácidos de aminoacil ARNt sintetasas (AARS).

Las aminoacil-ARNt sintetetas actúan para transformar las secuencias del código genético en proteínas biológicamente funcionales a través de una reacción de aminoacilación de dos etapas. Como una etapa inicial, el aminoácido cognato es activado por parte de AARS en presencia de ATP para formar el aminoácido-adenilato; posteriormente, la AARS cataliza la reacción de esterificación para unir el aminoácido a 2'- o 3'-OH del ribonucleótido terminal de su ARNt cognato. Una vez que se produce la reacción de aminoacilación, el aminoácido se dirige a la cadena polipeptídica en crecimiento mediante el ARNt cargado.

Determinadas formas de realización descritas en esta memoria se refieren a moléculas de PheRS mutantes o modificadas que han sido mutadas o modificadas de tal manera que las enzimas son capaces de cargar una molécula de ARNt con un aminoácido de sustitución, preferiblemente un aminoácido no natural debido a una interrupción entre la sintetasa y el aminoácido natural correspondiente.

Por ejemplo, la interrupción puede ser debida a interferir con el apareamiento de bases de Watson-Crick, interfiriendo con el apareamiento de bases de bamboleo, o a la creación de un nuevo bamboleo u otro apareamiento de bases.

Algunas formas de realización se refieren a un polinucleótido que codifica un ARNt mutante o modificado de un ARNt para un aminoácido natural, en donde el aminoácido natural es codificado por uno o más codones de bamboleo degenerados, el ARNt modificado comprende una secuencia de anticodon modificada que forma un apareamiento de bases de Watson-Crick con uno de los codones de bamboleo degenerados. Preferiblemente, el ARNt modificado no se carga o sólo lo hace ineficazmente mediante una aminoacil-ARNt sintetasa (AARS) endógena para el aminoácido natural. En una forma de realización, pueden utilizarse múltiples moléculas de PheRS modificadas con una o más moléculas de ARNt. En una forma de realización, se pueden utilizar una o más moléculas de PheRS modificadas o mutadas con uno o más moléculas de ARNt nativas, mientras que en otra forma de realización se puede utilizar una PheRS modificada o mutada con una molécula de ARNt modificada o mutada. En determinadas formas de realización se pueden utilizar uno o más pares de moléculas de PheRS/ARNt modificadas. En determinadas formas de realización se pueden utilizar pares heterólogos. En determinadas formas de realización, uno o más PheRS y/o ARNt modificado o mutado puede derivarse del mismo o de un organismo diferente.

En algunas formas de realización ilustrativas, una PheRS particular puede utilizar varios métodos para la incorporación de un aminoácido de sustitución (incluido un aminoácido no natural) en un polipéptido o proteína. Por ejemplo, una sola PheRS puede utilizar un codón sin sentido (tal como un codón de terminación ámbar) para la incorporación de un aminoácido de sustitución (tal como un aminoácido no natural) en una localización particular en el polipéptido. Además o en lugar de esto, se podría utilizar un codón de bamboleo (tal como UUU) para la incorporación del aminoácido de sustitución en el sitio del codón de bamboleo.

En otras formas de realización ilustrativas, múltiples aminoácidos de sustitución (tales como dos aminoácidos no naturales diferentes) se pueden incorporar en un polipéptido o proteína a través de la utilización de diversos métodos. Por ejemplo, un aminoácido no natural se puede incorporar en un sitio de bamboleo, mientras que un aminoácido no natural diferente puede ser incorporado en un codón de terminación ámbar. En algunas formas de realización ilustrativas, la incorporación de múltiples aminoácidos de sustitución (incluidos aminoácidos no naturales) incluye la utilización de una PheRS modificada para incorporar múltiples análogos de fenilalanina diferentes tales como bromofenilalanina y/o p-yodofenilalanina, en el mismo polipéptido o proteína.

Una estrategia similar implica el uso de una sintetasa heteróloga y un ARNt iniciador mutante del mismo organismo o de un organismo relacionado como una molécula de ARNt. (Véase, por ejemplo, Kowal, *et al.*, *PNAS*, 98, 2268 (2001)).

En determinadas formas de realización, la RS modificada o mutada interactúa con la sustitución de aminoácidos deseada (ya sea aminoácido que se produce de forma natural o no natural) con una especificidad alterada de unión y/o un evento catalítico alterado de la enzima hacia la sustitución de aminoácidos cuando se compara con la enzima RS de tipo salvaje o del correspondiente aminoácido de tipo salvaje.

En la cinética enzimática, k_{cat} es una constante de velocidad de primer orden que corresponde a la etapa o etapas más lentas en la vía catalítica global. La k_{cat} representa el número máximo de moléculas de sustrato que se puede convertir en producto por cada una de las moléculas de enzima por unidad de tiempo (lo cual se produce si la enzima está "saturada" con el sustrato) y, por lo tanto, se refiere a menudo como el número de cambio. La K_m es una constante de disociación aparente y se refiere a la afinidad de la enzima por el sustrato; es el producto de todas las constantes de disociación y de equilibrio antes de la primera etapa irreversible de la vía. A menudo, es una medida cercana a la constante de disociación enzima-sustrato. La k_{cat}/K_m es una constante de velocidad de

segundo orden que se refiere a la enzima libre (no al complejo enzima-sustrato) y es también una medida de la eficacia global de la catálisis de la enzima y a la que también se alude como la constante de especificidad.

- En determinadas formas de realización, la sintetasa mutante externa tiene propiedades enzimáticas mejoradas o potenciadas, p. ej., la K_m es mayor o menor, la k_{cat} es mayor o menor, el valor de k_{cat}/K_m es mayor o menor, o similar, para el aminoácido no natural en comparación con un aminoácido que se produce de forma natural, p. ej., uno de los 20 aminoácidos conocidos. La K_m de la AARS mutante o modificada es preferiblemente igual o menor para el aminoácido no natural que para el aminoácido natural de tipo salvaje correspondiente.
- En determinadas formas de realización, los valores k_{cat}/K_m de la variante RS puede variar de 3 veces, 5 veces, 10 veces, 25 veces, 50 veces, 100 veces, 150 veces, 200 veces, 250 veces, 300 veces, 350 veces, 385 veces, 400 veces más que para el aminoácido que se produce de forma natural.
- En determinadas formas de realización, el ARNt modificado interactúa con el codón de bamboleo degenerado con una afinidad a 37°C de al menos aproximadamente 1,0 kcal/mol, 1,5 kcal/mol, 2,0 kcal/mol, 2,5 kcal/mol, 3,0 kcal/mol, 3,5 kcal/mol, 4,0 kcal/mol, 4,5 kcal/mol, 5,0 kcal/mol o mayor (o cualquier valor intermedio) de manera más favorable que la interacción entre su versión sin modificar y el codón de bamboleo degenerado.
- Por ejemplo, fenilalanina (Phe) es codificada por dos codones, UUC y UUU. Ambos codones son leídos por un único ARNt, que está equipado con la secuencia de anticodón GAA. Por lo tanto, el codón UUC es reconocido a través de un apareamiento de bases de Watson-Crick estándar entre codón y anticodón; UUU es leído a través de un par de bases de bamboleo G-U en la primera posición del anticodón (Crick, *J. Mol. Biol.* 19: 548, 1966; Soll y Rajbhandary, *J. Mol. Biol.* 29: 113, 1967). La desnaturalización térmica de los dúplex de ARN ha proporcionado estimaciones de las energías libres de Gibbs de fusión de pares de bases G-U, G-C, A-U y A-C como 4,1, 6,5, 6,3 y 2,6 kcal/mol, respectivamente, a 37°C. Así, el par de bases de bamboleo, G-U, es menos estable que el par de bases de Watson-Crick, A-U. Un ARNt^{Phe} modificado equipado con el anticodón (tRNA^{Phe}_{AAA}) fue diseñado para leer el codón UUU, y fue predicho para leer tales codones de una forma más rápida que el ARNt^{Phe}_{GAA} de tipo salvaje.
- En algunas formas de realización, el bolsillo de unión de la RS se modifica de tal manera que la RS modificada exhibe una preferencia por el aminoácido no natural frente al aminoácido que se produce de forma natural correspondiente. En formas de realización preferidas, la RS se modifica en uno o más codones necesarios para el contacto estructural entre la RS y el aminoácido que está siendo cargado al ARNt. En determinadas formas de realización, el uno o más codones seleccionados para la mutación o modificación se seleccionan por medio de modelado por ordenador. Mientras que cualquier RS puede ser modificada, la presente invención se refiere a fenilalanil-ARNt sintetasa (PheRS). La RS modificada es de *Saccharomyces cerevisiae*, u otra célula de levadura. En determinadas formas de realización, en las que la RS es una PheRS, la enzima tiene una o más mutaciones puntuales según se define antes, en particular una mutación puntual (N412G), (T415G), (T415A), (S418C) o (S437F) en la subunidad alfa de la enzima. Las mutaciones puntuales (por ejemplo, la mutación T a G o A en la posición 415, o la mutación S a C en la posición 418, o la mutación N a G en la posición 412, o la mutación S a F en la posición 437) se encuentran en la región del bolsillo de unión de la aminoacil-ARNt sintetasa (RS).
- En algunas formas de realización ilustrativas, los valores típicos K_m para diferentes análogos con AARS pueden oscilar desde aproximadamente 15 microM, 20 microM, 30 microM, 50 microM, 75 microM, 100 microM, 150 microM, 200 microM, 300 microM, 400 microM, 440 microM, 500 microM, 1000 microM, 1500 microM, 2000 microM, 3000 microM, 4000 microM, 5000 microM, 6000 microM, o mayor, o cualquier valor intermedio.
- Asimismo, los valores de k_{cat} de la PheRS mutante son preferiblemente iguales o más altos para el análogo de aminoácido que para el aminoácido natural. Por ejemplo, los valores k_{cat} para diferentes análogos con el PheRS correspondiente puede oscilar entre aproximadamente 0,002 s⁻¹, 0,0018 s⁻¹, 0,0015 s⁻¹, 0,014 s⁻¹, 0,1 s⁻¹, 0,3 s⁻¹, 1 s⁻¹, 3 s⁻¹, 5 s⁻¹, 8 s⁻¹, 10 s⁻¹, 13,3 s⁻¹, 15 s⁻¹, o mayor.
- Por lo tanto, la k_{cat}/K_m del mutante PheRS es óptimamente igual a o mayor que para el análogo de aminoácido que para el aminoácido de tipo salvaje natural. Valores de k_{cat}/K_m típicos pueden variar de aproximadamente 0,0001 M⁻¹ s⁻¹, 0,0003 M⁻¹ s⁻¹, 0,005 M⁻¹ s⁻¹, 0,05 M⁻¹ s⁻¹, 0,5 M⁻¹ s⁻¹, 0,547 M⁻¹ s⁻¹, 1 M⁻¹ s⁻¹, 5 M⁻¹ s⁻¹, 10 M⁻¹ s⁻¹, 20 M⁻¹ s⁻¹, 30 M⁻¹ s⁻¹, 32 M⁻¹ s⁻¹, 500 M⁻¹ s⁻¹, 600 M⁻¹ s⁻¹, 1000 M⁻¹ s⁻¹, 5000 M⁻¹ s⁻¹, 11000 M⁻¹ s⁻¹.
- En determinadas formas de realización, las mutaciones puntuales PheRS se pueden alterar para cualquier aminoácido, dependiendo de las características del aminoácido no natural deseado para su incorporación en la proteína/el polipéptido de ensayo. En determinadas formas de realización, un aminoácido en el bolsillo de unión de la PheRS puede ser mutado a un codón para un aminoácido con una cadena lateral pequeña, un aminoácido con una cadena lateral alifática, un aminoácido cíclico, un aminoácido con cadenas laterales que contienen hidroxilo o azufre, un aminoácido aromático, un aminoácido de carácter básico, un aminoácido (o amida) de carácter ácido.

La selección del aminoácido para mutar la PheRS en un punto particular es de rutina, dependiendo del resultado deseado y del aminoácido no natural deseado a ser incorporado en el polipéptido/proteína diana o de ensayo. Por ejemplo, si el objetivo es agrandar el bolsillo de unión de la molécula de PheRS, entonces podría elegirse un aminoácido con una cadena lateral alifática, o una pequeña cadena lateral para mutar la PheRS. En otros casos, si se desea un bolsillo de unión que albergue un bolsillo cargado, entonces puede seleccionarse un aminoácido de carácter básico o ácido para la mutación puntual de la PheRS.

En algunas realizaciones, la RS modificada puede utilizarse en un sistema de traducción, incluida una célula huésped auxótrofa o célula huésped prototrófica junto con un ARNt supresor con el fin de permitir la asignación de un codón de terminación (tal como un codón ámbar, ocre u ópalo sin sentido, un codón de terminación que no está presente en un organismo particular, cualquier codón sin sentido, un codón de cuatro o cinco pares de bases, u otro aminoácido natural que no esté presente en niveles significativos en la proteína, tal como metionina) para incorporar otro aminoácido, incluido un análogo de aminoácido. Por lo tanto, las enzimas RS pueden ser "reprogramadas" para una especificidad promiscua para el sustrato con el fin de facilitar la incorporación de un aminoácido no natural en un polipéptido de una manera específica para el sitio. En formas de realización particulares, se puede utilizar cualquier aminoácido no natural aromático con la PheRS modificada. Esta reprogramación permite una incorporación de alta fidelidad de un aminoácido, incluidos aminoácidos no naturales, en polipéptidos o proteínas con o sin el uso de células huésped auxotróficas.

La reprogramación de una enzima AARS puede implicar el análisis estructural o bioquímico, incluido el modelado por ordenador o la alineación de secuencias. Dado que no hay información de la secuencia disponible para muchas moléculas de AARS, por ejemplo en GenBank, la comparación de alineaciones de secuencias es un proceso rutinario una vez que se determina la secuencia de la región particular de interés.

El uso de células huésped auxotróficas puede aumentar el nivel de incorporación del aminoácido no natural, o puede disminuir el nivel de incorporación errónea de otro aminoácido en lugar del aminoácido no natural deseado. Por ejemplo, después de potenciar la reactividad de aminoacilación celular mediante la expresión de PheRS de tipo salvaje en el huésped, los autores de la invención encontraron, sorprendentemente, que algunos de los lentos análogos de aminoácidos también podrían ser introducidos en las proteínas, incluso en ausencia de una célula huésped auxotrófica.

En determinadas formas de realización, la expresión de una o más moléculas de PheRS modificadas o mutantes, una o más moléculas de ARNt mutantes o modificadas, o ambos, puede ser regulada por un promotor constitutivo o inducible u otro sistema de expresión inducible.

Determinadas formas de realización descritas en esta memoria se refieren a permitir la inserción selectiva para el sitio de uno o más aminoácidos no naturales en cualquier posición deseada de cualquier proteína, (ii) es aplicable a células tanto procariotas como eucariotas, y permite estudios *in vivo* de proteínas mutantes, además de la generación de grandes cantidades de proteínas mutantes purificadas. Además, determinadas formas de realización se refieren a la adaptación para incorporar cualquiera de una gran diversidad de aminoácidos no naturales, en proteínas *in vivo*. Así, en una secuencia de polipéptido específica es posible un cierto número de diferentes inserciones selectivas del sitio de aminoácidos no naturales. Tales inserciones son opcionalmente todas del mismo tipo (p. ej., múltiples ejemplos de un tipo de aminoácido no natural insertado en múltiples puntos en un polipéptido) o son, opcionalmente, de diversos tipos (p. ej., diferentes tipos de aminoácidos no naturales se insertan en múltiples puntos en un polipéptido).

Un resultado sorprendente descrito en esta memoria muestra que el re-diseño del sitio sintético de una enzima PheRS puede ampliar la capacidad de introducir aminoácidos de sustitución (incluidos aminoácidos no naturales) en polipéptidos o proteínas. En algunas realizaciones, las composiciones y los métodos de modificar o mutar PheRS pueden incluir, opcionalmente, alterar la función de edición de la PheRS modificada o mutante. En algunas formas de realización, la capacidad de edición o corrección de la PheRS modificada o mutante es aproximadamente igual a la de la PheRS de tipo salvaje (inalterada). En otras formas de realización, se reduce la capacidad de editar o corregir la PheRS modificada o mutantes. En aún otras formas de realización, se elimina la capacidad de edición o corrección de la PheRS modificada o mutante. En algunas determinadas formas de realización, la alteración de la función de edición o de corrección de la PheRS es inherente a la modificación de la PheRS con el fin de alojar a un aminoácido de sustitución (por ejemplo, modificación del bolsillo de unión de la PheRS). En otras formas de realización, la alteración de la función de edición o corrección se puede realizar, además de la modificación de la PheRS, con el fin de acomodar a un aminoácido de sustitución. En aún otras formas de realización, la función de edición o corrección de la PheRS está inalterada. Así, además de la modificación o alteración del bolsillo de unión de una PheRS, también puede ser alterado el dominio de corrección o edición de la PheRS modificada con el fin de permitir una especificidad incrementada de aminoacilar el aminoácido de sustitución (incluidos aminoácidos no naturales) a un ARNt (ya sea ARNt endógeno o mutante externo), al tiempo de hidrolizar opcionalmente el o los aminoácidos-adenilato de tipo salvaje que se pueden

formar y que resultan en una mayor fidelidad o especificidad de incorporación del aminoácido de sustitución (incluido un aminoácido no natural) en el polipéptido o proteína.

5 En algunas formas de realización, las tasas de incorporación de un aminoácido no natural eran aproximadamente 65% o mayor, 70% o mayor, 75% o mayor, 80% o mayor, 85% o mayor, 90% o mayor, 91% o mayor, 92% o mayor, 93% o mayor, 94% o mayor, 95% o mayor, 96% o mayor, 97% o mayor, 98% o mayor o 99% o mayor utilizando una RS modificada.

10 Tal como se describe en esta memoria, entre otros puede utilizarse la estructura cristalina de la RS exacta a ser modificada, o la estructura cristalina de una RS homóloga para el modelado molecular de la interacción enzima - aminoácido con el fin de determinar los puntos de contacto entre la RS y el aminoácido de origen natural correspondiente, y/o los puntos de contacto entre la RS y el aminoácido no natural seleccionado deseado para ser incorporado en un polipéptido. Por ejemplo, en determinadas formas de realización de esta memoria, se utilizó la estructura cristalina de la PheRS de *Thermus thermophilus* complejada con fenilalanina para el diseño de
15 modelado molecular de un PheRS de *Saccharomyces cerevisiae*, debido a la identidad de la secuencia de aproximadamente un 40% en la región del sitio activo de las sintetetasas. La mutación de la treonina en la posición 415 a glicina o alanina (T415G o T415A, respectivamente) amplió el sitio activo y permitió el acomodamiento de análogos de fenilalanina mayores. Mutaciones como ésta que interrumpen el apareamiento de bases de Watson-Crick con el aminoácido que se produce de forma natural designadas para una RS particular permiten una
20 especificidad incrementada para la incorporación de un aminoácido no natural y una incorporación errónea disminuida de otro aminoácido. Tal como se expone en los Ejemplos y Figuras, la (T415A) fenilalanina aminoacil ARNt sintetasa de levadura (PheRS) reveló una preferencia quíntuple por bromofenilalanina que por fenilalanina que se produce de forma natural. Por lo tanto, es posible alterar una molécula de aminoacil ARNt sintetasa (RS) para incorporar preferentemente un aminoácido deseado en 2 veces, 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8
25 veces, 9 veces, 10 veces o más, dependiendo de las propiedades de la RS particular y del aminoácido deseado (incluido un aminoácido no natural).

En una forma de realización particular, se encontró que el espacio alrededor de la posición para del anillo de arilo de Phe unida podría reducirse ligeramente para excluir otros aminoácidos aromáticos, tales como Trp, al tiempo
30 que sigue acomodando a un aminoácido no natural, tal como pBrF. Por lo tanto, una forma de realización describe la incorporación de un grupo funcional bromuro de arilo en un polipéptido en una posición programada, proporcionando un ligamiento quimiosselectivo a través de un acoplamiento cruzado catalizado por paladio con participantes en la reacción de etino o acetileno. Aunque tales RSs reprogramadas o modificadas pueden utilizarse en cualquier número de células huésped, incluidas células huésped auxotróficas, el alto nivel de eficacia de
35 incorporación del aminoácido (o análogo) deseado de las RSs modificadas de determinadas formas de realización, así como los altos rendimientos de producción de proteínas, hacen innecesario el uso de células huésped auxotróficas.

40 Dado que se conoce o se deduce fácilmente la región del sitio activo para casi todas las aminoácidos sintetetasas, una técnica ilustrativa de este tipo se puede aplicar a otras AARSs en un esfuerzo para reprogramar la especificidad para el aminoácido de un aminoácido que se produce de forma natural a un aminoácido no natural con una expectativa de éxito y sin experimentación excesiva.

45 Como un ejemplo ilustrativo, la treonina en la posición 415 en PheRS de levadura es el equivalente a treonina 251 en PheRS de *E. coli*. Por lo tanto, la mutación de la PheRS de levadura (T415G) permitió la activación de una diversidad de análogos de Phe. (Véanse los Ejemplos en esta memoria). Mutaciones puntuales adicionales y/o el uso de una célula huésped auxotrófica permitió una incorporación errónea disminuida en la variante de PheRS de levadura T415G.

50 En otra forma de realización particular descrita en esta memoria, un ARN de transferencia de levadura mutante ($yARNt^{Phe}_{CUA}$) del cual se interrumpió un apareamiento de bases de Watson-Crick entre la posición de aminoácido 30 y la posición de aminoácido 40, se cargó con p-bromo-fenilalanina (pBrF) mediante una fenilalanina de levadura co-expresada. En determinadas formas de realización, el bolsillo de unión de aminoácidos de la AARS constituye aproximadamente 200, aproximadamente 100, aproximadamente 75, aproximadamente 50, aproximadamente 25,
55 aproximadamente 10, aproximadamente 5 o más o menos aminoácidos. En algunas formas de realización, los aminoácidos a mutar en el sitio activo o de unión son tramos contiguos de aminoácidos. En otras formas de realización, los aminoácidos a mutar están situados en estrecha proximidad entre sí, pero no son contiguos.

60 En determinadas formas de realización, el aminoácido natural es codificada por dos o más códigos genéticos (por lo tanto, es codificado por códigos genéticos degenerados). En la mayoría, si no todos los casos, esto incluye 18 de los 20 aminoácidos naturales, excepto Met y Trp. En estas circunstancias, para reconocer todos los códigos genéticos degenerados para el aminoácido natural, el bucle de anticodón del o de los ARNts de tipo salvaje se

basa tanto en un apareamiento de bases de bamboleo como de un apareamiento de bases de Watson-Crick puro. El presente ARNt modificado contiene al menos una modificación en su bucle de anticodón, de tal manera que el bucle de anticodón modificado forma ahora un apareamiento de bases de Watson-Crick a uno de los códigos genéticos degenerados, al que el ARNt se une previamente sólo a través de apareamiento de bases por bamboleo.

Ya que el apareamiento de bases de Watson-Crick es invariablemente más fuerte y más estable que el apareamiento de bases por bamboleo, el presente ARNt modificado se unirán preferiblemente a un código genético previo de apareamiento de bases por bamboleo (ahora a través de un apareamiento de bases de Watson-Crick), a lo largo de un apareamiento de bases de Watson-Crick previo (ahora a través de apareamiento de bases por bamboleo). Por lo tanto un análogo puede ser incorporado en el presente codón, si el ARNt modificado se carga con un análogo de un aminoácido natural, que puede o puede no ser el mismo que el aminoácido natural codificada por el codón en cuestión.

Así, en determinadas formas de realización, si es deseable incorporar determinados análogos de aminoácidos en los codones para Met o Trp, un ARNt para un aminoácido natural (p. ej., un ARNt Met, un ARNt Trp, o incluso un ARNt Phe, etc.) puede ser modificado para reconocer el codón Met o Trp. Bajo este tipo de situación única, tanto el ARNt modificado como el ARNt natural compiten por unirse al mismo código genético (único) a través de un apareamiento de bases de Watson-Crick. Algunos, pero no todos estos codones aceptarán sus aminoácidos naturales, mientras que otros pueden aceptar análogos de aminoácidos portados por el ARNt modificado. Otros factores, tales como la abundancia del aminoácido natural frente a la del análogo, pueden afectar al resultado final. (Véanse los Ejemplos descritos en esta memoria).

En determinadas formas de realización preferidas, el ARNt modificado no está cargado o sólo lo está ineficazmente por una aminoacil-ARNt sintetasa endógena (AARS) para cualquier aminoácido natural, de manera que el ARNt modificado en gran medida (si no exclusivamente) porta un análogo de aminoácido, pero no un aminoácido natural. Aunque un ARNt modificado objeto todavía puede ser útil si se puede cargar por parte de las AARS endógenas con un aminoácido natural.

En determinadas formas de realización, el ARNt modificado cargado con un análogo de aminoácido tiene tal forma y tamaño general que el ARNt análogo es un complejo ribosómicamente aceptable, es decir, el complejo ARNt-análogo puede ser aceptado por los ribosomas procariotas o eucariotas en un sistema de traducción *in vivo* o *in vitro*.

Preferiblemente, la PheRS modificada carga de forma específica o preferible el análogo al ARNt modificado en cualquier aminoácido natural. En una realización preferida, la constante de especificidad para la activación del análogo por parte de la AARS modificada (definida como k_{cat}/K_M) es igual o mayor que al menos aproximadamente 2 veces mayor que para el aminoácido natural, de preferencia aproximadamente 3 veces, 4 veces, 5 veces, 6 veces, 7 veces, 8 veces, 9 veces, 10 veces mayor o más que el aminoácido natural.

En determinadas formas de realización, el ARNt modificado comprende, además, una mutación en el cuarto sitio de anticodón extendido, para aumentar la eficacia de la traducción.

El uso de codones extendidos se basa en la supresión del marco de lectura de traducción. Cuatro codones base tienen el potencial para la inserción de múltiples aminoácidos no naturales en la misma proteína. Por ejemplo, el cuádruplete UAGA puede ser descodificado por un ARNt^{Leu} con un anticodón UCUA con una eficacia de 13 a 26%. (Véase, por ejemplo, Moore, *et al.*, *J. Mol. Biol.*, 298: 195 (2000)). El uso de codones extendidos solos tiene problemas potenciales, de modo que la lectura en marco de las tres primeras bases como un triplete en el codón extendido compite con la supresión global del marco de lectura. En algunos casos, los codones extendidos basados en codones raros o codones sin sentido pueden reducir la lectura sin sentido y la supresión del marco de lectura en otros sitios no deseados. Estos problemas pueden ser superados, sin embargo, con el uso de un codón/anticodón extendido y una AARS y/o ARNt modificado como se indica en algunas formas de realización descritas en esta memoria.

Por lo tanto, para resumir, los codones específicos están reservados para uso en los métodos descritos en esta memoria por la PheRS mutantes o modificada y/o por el ARNt modificado o mutante para la incorporación de un aminoácido de sustitución (incluido un aminoácido que se produce de forma natural o no natural). Tales métodos pueden incluir el uso de una decodificación ámbar (ocre, ámbar u otro ARNt supresor) que lee codones de terminación (TAG), sesgo de decodificación que explota ARNts no utilizados responsables del sesgo de codones, decodificación por bamboleo que crea nuevos ARNts que leen codones de bamboleo y codones extendidos (4-5 bases o más) que utilizan ARNts "supresores" mutantes que utilizan anticodones de 4 bases o 5 bases (o más).

Al menos una otra forma de realización proporciona un método para incorporar un análogo de aminoácido en una

- proteína diana en una o más posiciones especificadas, comprendiendo el método: (1) proporcionar a un entorno un primer polinucleótido objeto para un ARNt modificado, o un ARNt modificado objeto; (2) proporcionar al entorno un segundo polinucleótido objeto que codifica una PheRS modificada, en donde la PheRS modificada es capaz de cargar el ARNt modificado con el análogo; (3) proporcionar al entorno el análogo; (4) proporcionar un polinucleótido molde que codifica la proteína diana, en donde el codón en el polinucleótido molde para la posición especificada sólo forma un apareamiento de bases de Watson-Crick con el ARNt modificado; y (5) permitir la traducción del polinucleótido molde para proceder, incorporando así el análogo en la proteína diana en la posición especificada, en donde las etapas (1) - (4) se efectúan en cualquier orden.
- En determinadas formas de realización, el método comprende, además, verificar la incorporación del análogo, por ejemplo, mediante espectrometría de masas, secuenciación de proteínas, marcaje de aminoácidos tal como mediante fluorescencia, radiactividad, etc., ELISA, u otro rastreo de anticuerpos, ensayos o rastreos funcionales, u otros métodos.
- En determinadas formas de realización, el método incorpora el análogo en la posición con una eficacia de al menos aproximadamente 50%, o 60%, 70%, 80%, 90%, 95%, 99% o casi el 100%.

Definiciones

- Antes de describir determinadas formas de realización en detalle, debe entenderse que esta invención no se limita a composiciones o sistemas biológicos particulares, que pueden, por supuesto, variar. Ha de entenderse también que la terminología utilizada en esta memoria es con el fin de describir solamente realizaciones ilustrativas particulares, y no se pretende que sea limitante. Tal como se utiliza en esta memoria descriptiva y en las reivindicaciones adjuntas, las formas en singular "un", "una" y "el", "la" incluyen referentes en plural, a menos que el contenido dicte claramente lo contrario. Así, por ejemplo, la referencia a "una molécula" incluye opcionalmente una combinación de dos o más de tales moléculas, y similares.

- A menos que se defina específicamente a continuación, los términos y expresiones utilizados en esta memoria descriptiva tienen generalmente sus significados ordinarios en la técnica, dentro del contexto general de esta invención y en el contexto específico donde se utiliza cada uno de los términos y expresiones. Determinados términos y expresiones se discuten más adelante o en otra parte en la memoria descriptiva, para proporcionar una orientación adicional al practicante en la descripción de las composiciones y métodos de la invención y cómo hacer y utilizar los mismos. El alcance y el significado de cualquier uso de un término o expresión resultarán evidentes a partir del contexto específico en el que se utiliza el término o expresión.

- "Alrededor de" y "aproximadamente" deben dar a entender, en general, un grado aceptable de error para la cantidad medida dada la naturaleza o la precisión de las mediciones. Grados de error típicos ilustrativos están dentro del 20 por ciento (%), preferentemente dentro del 10%, y más preferiblemente dentro del 5% de un valor o intervalo de valores dado. Alternativamente, y en particular en los sistemas biológicos, las expresiones "alrededor de" y "aproximadamente" pueden significar valores que están dentro de un orden de magnitud, preferentemente dentro de 5 veces y más preferiblemente dentro de 2 veces de un valor dado. Cantidades numéricas dadas en esta memoria son aproximadas, a menos que se indique otra cosa, lo que significa que el término "alrededor de" o "aproximadamente" se puede inferir cuando no se indique expresamente.

- "Análogo de aminoácido", "aminoácido no canónico" o "aminoácido no estándar", "aminoácido no natural", "aminoácido antinatural" y similares se pueden utilizar de manera indistinta, y pretende incluir todos los compuestos de tipo aminoácido que son similares en estructura y/o forma general a uno o más de los veinte L-aminoácidos que se encuentran comúnmente en proteínas que se producen de forma natural (Ala o A, Cys o C, Asp o D, Glu o E, Phe o F, Gly o G, His o H, Ile o I, Lys o K, Leu o L, Met o M, Asn o N, Pro o P, Gln o Q, Arg o R, Ser o S, Thr o T, Val o V, Trp o W, Tyr o Y, según se definen y listan en la Norma ST.25 de la OMPI (1998), Apéndice 2, Tabla 3). Análogo de aminoácido también pueden ser aminoácidos naturales con cadenas laterales o cadenas principales modificadas. Los aminoácidos también pueden ser aminoácidos que se producen de forma natural en forma D más que en forma L. Preferiblemente, estos análogos normalmente no son "sustratos" para las aminoacil ARNt sintetasas (AARSs), debido a la normalmente alta especificidad de las AARSs. Aunque ocasionalmente, determinados análogos con estructuras o formas suficientemente cercanas a las de los aminoácidos naturales se pueden incorporar en las proteínas erróneamente mediante AARSs, especialmente AARSs modificadas con especificidad de sustrato relajado. En una forma de realización preferida, los análogos comparten estructuras de cadena principal, y/o incluso la mayoría de estructuras de cadena lateral de uno o más aminoácidos naturales, siendo la o las únicas diferencias el que contienen uno o más grupos modificados en la molécula. Una modificación de este tipo puede incluir, sin limitación, la sustitución de un átomo (tal como N) por un átomo relacionado (tal como S), la adición de un grupo (tal como metilo, o un grupo hidroxilo, etc.) o un átomo (tal como Cl o Br, etc.), la supresión de un grupo (supra), la sustitución de un enlace covalente (enlace sencillo por

un doble enlace, etc.), o combinaciones de los mismos. Análogos de aminoácidos pueden incluir α -hidroxiácidos y β - aminoácidos, a los que también se puede aludir como "aminoácidos modificados" o "sustratos de AARS antinaturales". Los análogos de aminoácidos se pueden producir de forma natural o no natural (por ejemplo, sintetizados). Como se apreciará por los expertos en la técnica, cualquier estructura para la cual se conoce o se puede generar un conjunto de rotámeros se puede utilizar como un análogo de aminoácido. Las cadenas laterales pueden estar en la configuración (R) o la configuración (S) (o la configuración D o L). En una forma de realización preferida, los aminoácidos están en la configuración el (S) o L.

Preferiblemente, la forma y el tamaño globales de los análogos de aminoácidos son tales que, al ser cargados a ARNts (naturales o modificados o re-diseñados) mediante AARS (naturales o re-diseñados), el análogo-ARNt es un complejo aceptado por los ribosomas, es decir, el complejo ARNt-análogo puede ser aceptado por los ribosomas procariotas o eucariotas en un sistema de traducción *in vivo* o *in vitro*.

"Residuos de anclaje" son las posiciones de residuos en AARS que mantienen interacciones críticas entre la AARS y la cadena principal del aminoácido natural.

"Cadena principal", o "molde" incluye los átomos de la cadena principal y cualesquiera cadenas laterales fijadas (tales como las cadenas laterales de los residuos de anclaje) de la proteína (p. ej., AARS). A efectos de cálculo, la cadena principal de un análogo se trata como parte de la cadena principal de la AARS.

"Estructura de la cadena principal de la proteína" o equivalentes gramaticales en esta memoria quieren dar a entender las coordenadas tridimensionales que definen la estructura tridimensional de una proteína particular. Las estructuras que comprenden una estructura de la cadena principal de la proteína (de una proteína que se produce de forma natural) son el nitrógeno, el carbono del carbonilo, el carbono α y el oxígeno del carbonilo, junto con la dirección del vector del carbono α al carbono β .

La estructura de la cadena principal de la proteína que se introduce en un ordenador para la predicción estructural o la interacción molecular por ordenador, puede incluir las coordenadas tanto para la cadena principal como para las cadenas laterales de aminoácidos, o simplemente la cadena principal, es decir, con las coordenadas de las cadenas laterales de los aminoácidos eliminadas. En el primer caso, los átomos de la cadena lateral de cada uno de los aminoácidos de la estructura de proteína puede ser "despojado" o retirado de la estructura de una proteína, tal como se conoce en la técnica, dejando sólo las coordenadas para los átomos de la "cadena principal" (el nitrógeno, el carbono y el oxígeno del carbonilo, y el carbono α , y los átomos de hidrógeno unidos al nitrógeno y el carbono α).

Opcionalmente, la estructura de la cadena principal de la proteína puede ser alterada antes del análisis que se esboza más adelante. En esta forma de realización, la representación de la estructura de la cadena principal de la proteína de partida se reduce a una descripción de la disposición espacial de sus elementos estructurales secundarios. Las posiciones relativas de los elementos estructurales secundarios están definidas por un conjunto de parámetros denominados parámetros de la estructura supersecundaria. A estos parámetros se asignan valores que se pueden variar de forma sistemática o aleatoria para alterar la disposición de los elementos de estructura secundaria para introducir una flexibilidad de la cadena principal explícita. Las coordenadas atómicas de la cadena principal son entonces cambiadas para reflejar los parámetros estructurales supersecundarios alterados, y estas nuevas coordenadas se introducen en el sistema para su uso en la subsiguiente automatización de diseño de proteínas. Para más detalles, véase la patente de EE.UU. Nº 6.269.312, cuyo contenido completo queda incorporado aquí como referencia.

"Energía conformacional" se refiere generalmente a la energía asociada con una "conformación" particular, o estructura tridimensional, de una macromolécula, tal como la energía asociada con la conformación de una proteína particular. Interacciones que tienden a estabilizar una proteína tienen energías que están representadas como valores de energía negativa, mientras que las interacciones que desestabilizan una proteína tienen valores de energía positiva. Así, la energía conformacional para cualquier proteína estable está cuantitativamente representada por un valor de energía conformacional negativo. Generalmente, la energía conformacional para una proteína particular se relacionará con la estabilidad de esa proteína. En particular, las moléculas que tienen una energía conformacional menor (es decir, más negativa) son típicamente más estables, p. ej., a temperaturas más altas (es decir, tienen una mayor "estabilidad térmica"). Por consiguiente, a la energía conformacional de una proteína también se puede aludir como la "energía de estabilización".

Típicamente, la energía conformacional se calcula utilizando una energía del "campo de fuerza" que calcula o estima la contribución de energía a partir de diversas interacciones que dependen de la conformación de una molécula. El campo de fuerza está constituido por términos que incluyen la energía conformacional de la cadena principal del carbono alfa, interacciones cadena lateral - cadena principal, e interacciones cadena lateral - cadena

lateral. Típicamente, las interacciones con la cadena principal o la cadena lateral incluyen términos para la rotación del enlace, torsión del enlace y longitud del enlace. Las interacciones cadena principal - cadena lateral y cadena lateral - cadena lateral incluyen interacciones de van der Waals, enlaces hidrógeno, electrostática y términos de solvatación. Interacciones electrostáticas pueden incluir interacciones de Coulomb, interacciones dipolo e interacciones cuadrupolo). También se pueden incluir otros términos similares. Campos de fuerza que pueden utilizarse para determinar la energía conformacional para un polímero son bien conocidos en la técnica e incluyen los campos de fuerza CHARMM (véase, Brooks *et al.*, *J. Comp.Chem.* 1983,4: 187-217; MacKerell *et al.*, en *The Encyclopedia of Computational Chemistry*, vol. 1: 271-277, John Wiley & Sons, Chichester, 1998), AMBER (véase, Cornell *et al.*, *J. Amer. Chem. Soc.* 1995, 117: 5179; Woods, *et al.*, *J. Phys. Chem.* 1995, 99: 3832-3846; Weiner *et al.*, *J. Comp. Chem.* 1986, 7: 230; y Weiner *et al.*, *J. Amer.Chem.Soc.* 1984, 106:765) y DREIDING (Mayo *et al.*, *J. Phys.Chem.* 1990, 94-: 8897), por nombrar sólo unos pocos. En una implementación preferida, las expresiones enlace de hidrógeno y electrostática son como se describe en Dahiyat y Mayo, (*Science* 1997 278: 82). El campo de fuerza también puede ser descrito para incluir términos atómicos conformacionales (ángulos de enlace, longitudes de enlace, torsiones), como en otras referencias. Véase, p. ej., Nielsen, *et al.*, *Prot. Eng.*, 12: 657-662 (1999); Stikoff, *et al.*, *Biophys. J.*, 67: 2251-2260 (1994); Hendsch, *et al.*, *Prot. Sci.*, 3: 211-226 (1994); Schneider, *et al.*, *J. Am. Chem. Soc.*, 119: 5742-5743 (1997); Sidelar, *et al.*, *Prot Sci.* 7: 1898-1914 (1998). También podrían incluirse términos y expresiones de solvatación. Véase, p. ej., Jackson, *et al.*, *Biochemistry*, 32: 11259 - 11269 (1993); Eisenberg, *et al.*, *Nature*, 319: 199-203 (1986); Street A G y Mayo S L, *Folding & Design*, 3: 253-258 (1998); Eisenberg y Wesson, *Prot Sci.*, 1: 227-235 (1992); Gordon & Mayo, *supra*. "Residuos acoplados" son los residuos en una molécula que interactúan a través de cualquier mecanismo. Por lo tanto, a la interacción entre los dos residuos se la alude como una "interacción de acoplamiento". Residuos acoplados generalmente contribuyen a la aptitud de polímero a través de la interacción de acoplamiento. Típicamente, la interacción de acoplamiento es una interacción física o química tal como una interacción electrostática, una interacción de van der Waals, una interacción de enlaces de hidrógeno, o una combinación de las mismas. Como resultado de la interacción de acoplamiento, cambiar la identidad de cualquier residuo afectará a la "aptitud" de la molécula, en particular si el cambio altera la interacción de acoplamiento entre los dos residuos. La interacción de acoplamiento también se puede describir por un parámetro de distancia entre los residuos en una molécula. Si los residuos se encuentran a una determinada distancia de corte, se consideran que interactúan.

La "aptitud" se utiliza para denotar el nivel o grado en el que se optimiza una propiedad particular o una combinación particular de propiedades de una molécula, p. ej. una proteína. En determinadas formas de realización de la invención, la aptitud de una proteína se determina preferiblemente por las propiedades que un usuario desea mejorar. Así, por ejemplo, la aptitud de una proteína puede referirse a la estabilidad térmica de la proteína, la actividad catalítica, la afinidad de unión, la solubilidad (p. ej., en un disolvente acuoso u orgánico), y similares. Otros ejemplos de propiedades de aptitud incluyen la enantioselectividad, la actividad hacia sustratos antinaturales, y mecanismos catalíticos alternativos. Las interacciones de acoplamiento se pueden modelar como una forma de evaluar o predecir la aptitud (estabilidad). La aptitud puede determinarse o evaluarse experimental o teóricamente, p. ej., por ordenador.

Preferiblemente, la aptitud se cuantifica de manera que cada una de las moléculas, p. ej., cada uno de los aminoácidos tendrá un "valor de aptitud" particular. Por ejemplo, la aptitud de una proteína puede ser la velocidad a la que la proteína cataliza una reacción química particular, o la afinidad de unión de la proteína por un ligando. En una forma de realización particularmente preferida, la aptitud de una proteína se refiere a la energía conformacional del polímero y se calcula, p. ej., utilizando cualquier método conocido en la técnica. Véase, p. ej., Brooks, *et al.*, *J. Comp. Chem.*, 4: 187-217 (1983); Mayo, *et al.*, *J. Phys. Chem.*, 94: 8897-8909 (1990); Pabo, *et al.*, *Biochemistry*, 25: 5987-5991 (1986), Lazar, *et al.*, *Prot. Sci.*, 6: 1167-1178 (1997); Lee, *et al.*, *Nature*, 352: 448-451 (1991); Colombo, *et al.*, *J. Am. Chem. Soc.*, 121: 6895-6903 (1999); Weiner, *et al.*, *J. Am. Chem. Soc.*, 106: 765-784 (1984). Generalmente, la aptitud de una proteína se cuantifica de modo que el valor de la aptitud aumente a medida que se optimiza la propiedad o combinación de propiedades. Por ejemplo, en realizaciones en las que la estabilidad térmica de una proteína se ha de optimizar (la energía conformacional disminuye preferiblemente), el valor de aptitud puede ser la energía conformacional negativa, es decir, $F = -E$.

La "contribución a la aptitud" de un residuo de proteína se refiere al nivel o grado $f(i_a)$ al que el residuo i_a , que tiene una identidad, contribuye a la aptitud total de la proteína. Así, por ejemplo, si se cambia o muta un residuo de aminoácido particular, se reducirá en gran medida la aptitud de la proteína, se dice que ese residuo tiene una alta contribución a la aptitud para el polímero. Por el contrario, típicamente algunos residuos i_a en una proteína pueden tener una diversidad de posibles identidades sin afectar a la aptitud de la proteína. Tales residuos, por lo tanto, tienen una baja contribución a la aptitud de la proteína.

La "Eliminación sin salida" (DEE) es un algoritmo de búsqueda determinista que busca eliminar sistemáticamente malos rotámeros y combinaciones de rotámeros hasta que quede una única solución. Por ejemplo, los residuos de aminoácidos se pueden modelar como rotámeros que interactúan con una cadena principal fija. La base teórica para DEE establece que, si la búsqueda de DEE converge, la solución es la conformación de energía mínima

global (GMEC) sin incertidumbre (Desmet *et al.*, 1992).

La eliminación sin salida se basa en el siguiente concepto. Consideremos dos rotámeros, i_r e i_t en el residuo i , el conjunto de todas las demás configuraciones de rotámeros $\{S\}$ en todos los residuos, excluido i (de los cuales el rotámero j_s es un miembro). Si la energía por parejas contribuida entre i_r y j_s es superior a la energía por parejas entre i_t y j_s para todos los $\{S\}$, entonces el rotámero i_r no puede existir en la conformación global de energía mínima, y puede ser eliminado. Esta noción se expresa matemáticamente por la desigualdad.

$$E(i_r) + \sum_{j \neq i}^N E(i_r, j_s) > E(i_t) + \sum_{j \neq i}^N E(i_t, j_s) \{ S \}$$

(Ecuación A)

Si esta expresión es verdadera, el único rotámero i_r puede ser eliminado (Desmet *et al.*, 1992)

- 10 En esta forma, la Ecuación A no es computacionalmente tratable porque, para realizar una eliminación, se requiere que se enumere todo el espacio de la secuencia (rotámero). Para simplificar el problema, se pueden utilizar límites implicados por la Ecuación A:

$$E(i_r) + \sum_{j \neq i}^N \min(s) E(i_r, j_s) > E(i_t) + \sum_{j \neq i}^N \max(s) E(i_t, j_s) \{ S \}$$

(Ecuación B)

- 15 Utilizando un argumento análogo, la Ecuación B se puede extender a la eliminación de pares de rotámeros incompatibles con la GMEC. Esto se realiza mediante la determinación de que un par de rotámeros i_r en el residuo i y j_s en el residuo j , siempre contribuyen a energías más altas que los rotámeros i_u y j_v con todas las combinaciones posibles de rotámeros $\{L\}$. De manera similar a la Ecuación B, el estricto límite de esta afirmación viene dado por:

$$\varepsilon(i_r, j_s) + \sum_{k \neq i, j}^N \min(t) \varepsilon(i_r, j_s, k_t) > \varepsilon(i_u, j_v) + \sum_{k \neq i, j}^N \max(t) \varepsilon(i_u, j_v, k_t)$$

(Ecuación C)

- 20 en que ε es las energías combinadas para pares de rotámeros

$$\varepsilon(i_r, j_s) = E(i_r) + E(j_s) + E(i_r, j_s)$$

(Ecuación D)

y

$$\varepsilon(i_r, j_s, k_t) = E(i_r, k_t) + E(j_s, k_t)$$

(Ecuación E)

- 25 Esto conduce a la eliminación de dobles del par de rotámeros i_r y j_s , pero no elimina los rotámeros individuales por completo, ya que cualquiera de ellos podría existir independientemente en la GMEC. La etapa de eliminación de dobles reduce el número de posibles pares (reduce S) que necesitan ser evaluados en el lado de la derecha de la Ecuación 6, permitiendo que sean eliminados individualmente más rotámeros.

- 30 Los criterios sencillos y dobles presentados por Desmet *et al.* dejan por descubrir condiciones especiales que conducen a la determinación de más rotámeros sin salida. Por ejemplo, es posible que la contribución energética de rotámero i_t sea siempre menor que i_r sin estar el máximo de i_r por debajo del mínimo de i_t . Para abordar este problema, Goldstein 1994 presentó una modificación de los criterios que determina si se cruzan los perfiles de energía de dos rotámeros. Si no lo hacen, se puede determinar que el rotámero de mayor energía es sin salida. El
- 35 cálculo de dobles utiliza significativamente más tiempo computacional que el cálculo de los sencillos. Para acelerar el proceso, se han desarrollado otros métodos computacionales para predecir los cálculos de dobles que serán los más productivos (Gordon y Mayo, 1998). Este tipo de modificaciones, denominados colectivamente dobles rápidos, mejoraron significativamente la velocidad y la eficacia de la DEE.

- 40 Varias otras modificaciones también potencian el DEE. Rotámeros de múltiples residuos se pueden combinar formando los llamados super-rotámeros para solicitar más eliminaciones (Desmet *et al.*, 1994; Goldstein, 1994). Esto tiene la ventaja de eliminar múltiples rotámeros en una sola etapa. Además, se ha demostrado que la "división" del espacio conformacional entre rotámeros mejora la eficacia del DEE (Pierce *et al.*, 2000). La división maneja el siguiente caso especial. Consideremos el rotámero i_r . Si un rotámero i_{t1} contribuye en una energía más
- 45 baja que i_r para una porción del espacio conformacional, y un rotámero i_{t2} tiene una energía más baja que i_r para la fracción restante, entonces se puede eliminar i_r . Este caso no sería detectado por los criterios de Desmet o Goldstein menos sensibles. En las implementaciones preferidas, como se describe en esta memoria, se utilizaron

todas las mejoras descritas para DEE.

Para más información sobre estos métodos, véase, Goldstein, *Biophysical Journal* 66, 1335-1340 (1994); Desmet, *et al.*, *Nature* 356, 539-542 (1992).; Desmet, *et al.*, *The Protein Folding Problem and Tertiary Structure Prediction* (Jr., K.M. y Grand, S.L., comps.), págs. 307-337 (Birkhauser, Boston, 1994); De Maeyer, *et al.*, *Folding & Design* 2, 53-66 (1997), Gordon y Mayo, *J. Comp. Chem.* 19, 1505-1514 (1998); Pierce, *et al.*, *J. Comp. Chem.* 21, 999- 1009 (2000).

Se puede utilizar otro cálculo, apodado SCREAM (siglas inglesas de Método de Análisis de la Energía de Rotámeros de la Cadena Lateral). El SCREAM permite el examen del mecanismo de discriminación frente a aminoácidos no afines, mediante el cálculo de las energías relativas de unión de los 20 aminoácidos naturales a un AARS particular. (Véase, por ejemplo, McClendon, *et al.*, *Prot. Eng. Design & Select.* 19: 195-203 (2006)).

Como una primera etapa, se calcula el espectro de energía del rotámero para un solo aminoácido en una cadena principal vacía, sin otras cadenas laterales movibles. A continuación, a partir de los rotámeros más bajos de la cadena principal vacía, llenar las cadenas laterales, pero eliminar los enfrentamientos. Por ejemplo, colocar las cadenas laterales en cada sitio, estimar las energías de las excitaciones bajas del espectro de la cadena principal vacía y calcular las interacciones por parejas, eliminando configuraciones que tengan enfrentamientos. Por lo tanto, $E_{tot}(A,B) = E_{self}(A) + E_{self}(B) + E_{int.}(A,B)$ y $E_{tot}(A,B,C) = E_{self}(A) + E_{self}(B) + E_{self}(C) + E_{int.}(A,B) + E_{int.}(A,C) + E_{int.}(B,C)$ establece la relación recursiva. A continuación, todas las excitaciones bajas del espectro de la cadena lateral deben ser analizadas hasta que deje de aumentar la distribución de energía de las energías de rotámeros en la cadena principal vacía. En síntesis, se evalúa la energía del estado fundamental para todos los residuos, seguido de un conjunto de rotámeros con la energía suma lineal más baja y, finalmente, la siguiente energía suma lineal más baja, y así sucesivamente. Además de ello, las interacciones electrostáticas deben ser abordadas a medida que las cargas polarizan el entorno para proteger las cargas y reducir la interacción de aminoácidos deseada. Dado que los métodos de modelado de dinámica molecular no representan, por lo general, la polarización, el sesgo es a favor de puentes de sal. Con el fin de superar esto, los residuos se neutralizan y los parámetros se evalúan de nuevo de acuerdo con la parametrización DREIDING.

Para la parametrización DREIDING, las interacciones cargado-cargado y cargado-dipolo se compensan mediante la introducción del término enlaces de hidrógeno. Esto puede hacerse en unión con otros programas, incluidos CHARMM, según se describe en esta memoria. Por lo tanto, utilizando una estructura cristalográfica a partir de una AARS particular, o una AARS homóloga de otro organismo, se puede utilizar un programa tal como SCREAM y HierDock, u otros, para predecir la conformación de unión y la energía de enlace de cada uno de los 20 aminoácidos naturales en el sitio de unión en el modo mejor de unión y el modo de activación, al ordenar los cálculos según los cuales los aminoácidos compiten por la unión a un AARS particular.

En particular, primero se realiza la unión selectiva, que proporciona el aminoácido y la molécula de ATP para unirse al sitio activo de la AARS. A veces, esto conduce a un cambio conformacional. Seguidamente, la activación selectiva de la AARS para catalizar la formación de un enlace covalente entre el aminoácido y el ATP, forma un aminoacil adenilato complejoado con el AARS y elimina pirofosfato inorgánico. En tercer lugar, si se produce una activación errónea de un aminoácido no afín, la AARS puede escindir hidrolíticamente el complejo aminoacil adenilato (como corrector de pre-transferencia). Finalmente, si ha sobrevivido un aminoacil adenilato no afín, la AARS puede escindir hidrolíticamente el complejo aminoacil-ARNt (corrección post-transferencia).

Por lo tanto, un programa computacional de este tipo permite una predicción fiable de la probabilidad de que los aminoácidos naturales compitan por unirse y la aminoacilación mediante enzimas AARS de tipo salvaje o mutantes. Utilizando programas múltiples (tales como SCREAM y HierDock juntos) se reduce la tasa de incorporación errónea y permite una mayor previsibilidad en la selección de localizaciones de aminoácidos que se unen específicamente al aminoácido.

Todavía, otros programas de modelado por ordenador incluyen SCAP (siglas inglesas de Programa de Predicción del Aminoácido de la Cadena lateral) (Xiang y Honig, *J. Mol. Biol.* 311, 421-430 (2001)), y SCWRL (siglas inglesas de Reemplazo de la Cadena Lateral con un Banco de Rotámeros), que es útil para la adición de cadenas laterales a una cadena principal de la proteína basado en el banco de rotámeros dependiente de la cadena principal. El banco SCWRL ofrece listas de valores χ_1 - χ_2 - χ_3 - χ_4 y sus probabilidades relativas de residuos en valores ϕ - ψ dados, y explora estas conformaciones para minimizar enfrentamientos cadena lateral-cadena principal y enfrentamientos cadena lateral-cadena lateral. (Véase, por ejemplo. Bower, *et al.*, *J. Mol. Biol.*, 267, 1268-1282 (1997)).

La previsibilidad computacional se debe, en gran parte, a utilizar secuencias de ácidos nucleicos y/o de

aminoácidos conocidas de las enzimas AARS. Por ejemplo, el dominio catalítico se conserva a lo largo de todos los miembros de una clase particular de enzima AARS. (O'Donoghue y Luthey-Schulten, *Microbiol. And Mol. Biol. Rev* : 550-573 (2003); Díaz-Lazcoz, *et al.*, *Mol. Biol. Evol.* 15 (11): 1548-1561 (1998); Wang, *et al.*, *Chem. Commun.* 1-11 (2002)).

"Sistema de expresión" significa una célula huésped, o componentes celulares y vector compatible bajo condiciones adecuadas, p. ej., para la expresión de una proteína codificada por ADN extraño transportado por el vector e introducido en la célula huésped. Los sistemas de expresión comunes incluyen células huésped de *E. coli* y vectores de plásmido, células huésped de insectos tales como células Sf9, Hi5 o S2 y vectores de Baculovirus, células de *Drosophila* (células de Schneider) y sistemas de expresión, y células huésped y vectores de mamíferos.

"Célula huésped" significa cualquier célula de cualquier organismo que ha sido seleccionada, modificada, transformada, cultivada o utilizada o manipulada de alguna forma para la producción de una sustancia por parte de la célula. Una célula huésped puede ser auxotrófica, que es incapaz de sintetizar al menos un compuesto orgánico particular requerido para su mantenimiento o crecimiento y debe obtener el compuesto de otra fuente tal como su entorno o medio de cultivo. Además, una célula huésped auxotrófica puede tener niveles de auxotrofia sencillo, doble, triple, cuádruple o mayor, de manera que es incapaz de sintetizar uno, dos, tres, cuatro o más compuestos orgánicos necesarios para su crecimiento o mantenimiento, respectivamente. Por ejemplo, una célula huésped puede ser una que es manipulada para expresar un gen particular, una secuencia de ADN o ARN, una proteína o una enzima. Las células huésped pueden cultivarse *in vitro* o una o más células en un animal no humano (p. ej., un animal transgénico o un animal transfectado transitoriamente).

Determinadas formas de realización descritas en esta memoria expresamente sólo utilizan un sistema de expresión o traducción exento de células y no una célula huésped. Determinadas otras formas de realización utilizan expresamente sólo una célula huésped auxotrófica. Todavía otras determinadas formas de realización utilizan expresamente sólo una célula huésped no auxotrófica, o una célula huésped prototrófica.

La similitud de secuencia puede ser relevante para determinadas formas de realización, ya que éstas pueden incluir etapas de comparación de secuencias entre sí, incluida la secuencia de tipo salvaje, con uno o más mutantes. Tales comparaciones comprenden típicamente alineaciones de secuencias de polímeros, p. ej., utilizando programas y/o algoritmos de alineación de secuencias que son bien conocidos en la técnica (por ejemplo, BLAST, FASTA y MEGALIGN, por nombrar unos pocos). El experto en la técnica puede apreciar fácilmente que, en tales alineaciones, en que una mutación contiene una inserción o delección de residuos, la alineación de la secuencia introducirá un "hueco" (por lo general representado por un guión, "-", o "Δ") en la secuencia de polímero que no contiene el residuo insertado o suprimido.

"Homóloga", en todas sus formas gramaticales y variaciones ortográficas, se refiere a la relación entre dos moléculas (p. ej., proteínas, ARNs, ácidos nucleicos) que poseen un "origen evolutivo común", incluidas proteínas de superfamilias en la misma especie de organismo, así como proteínas homólogas de diferentes especies de organismos. Tales proteínas (y sus ácidos nucleicos codificantes) tienen una homología de la secuencia y/o estructural, tal como se refleja por su similitud de secuencia, ya sea en términos de porcentaje de identidad o por la presencia de residuos o motivos específicos y posiciones conservadas. Moléculas homólogas también comparten, con frecuencia, funciones similares o incluso idénticas.

En algunos aspectos, homóloga puede incluir una secuencia que es al menos 50% homóloga, pero que presenta una estructura homóloga en tres dimensiones, es decir, incluye una carga de superficie sustancialmente similar o una presentación de grupos hidrófobos. Dado que los enlaces hidrógeno, fuerzas de van der Waals e interacciones hidrófobas pueden funcionar para unir un aminoácido al bolsillo de unión de una AARS, la manipulación de una estructura de la AARS también puede alterar una o más de estas fuerzas.

Por lo tanto, tal como se utiliza en esta memoria, proteínas y/o secuencias de proteínas son "homólogas" cuando se derivan, natural o artificialmente, de una proteína o secuencia de proteína ancestral común. De manera similar, ácidos nucleicos y/o secuencias de ácidos nucleicos son homólogos cuando se derivan, natural o artificialmente, de un ácido nucleico o secuencia de ácido nucleico ancestral común. Por ejemplo, cualquier ácido nucleico que se produzca de forma natural se puede modificar por cualquier método de mutagénesis disponible para incluir uno o más codones selectores. Cuando se expresa, este ácido nucleico mutagenizado codifica un polipéptido que comprende uno o más aminoácidos antinaturales. El proceso de mutación puede, por supuesto, alterar adicionalmente uno o más codones estándares, cambiando de este modo asimismo uno o más aminoácidos estándares en la proteína mutante resultante. La homología generalmente se infiere de la similitud de secuencias 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 un valor tan bajo como 25% de similitud de la secuencia se utiliza rutinariamente para establecer la

homología. Niveles más altos de similitud de la secuencia, por ejemplo 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95% o 99% o mayores también se pueden utilizar para establecer la homología. En esta memoria se describen y están generalmente disponibles métodos para determinar los porcentajes de similitud de secuencias (p. ej., BLASTP BLASTN y utilizando los parámetros por defecto).

La expresión "similitud de secuencias", en todas sus formas gramaticales, se refiere al grado de identidad o correspondencia entre secuencias de ácidos nucleicos o de aminoácidos que pueden o no compartir un origen evolutivo común (véase, Reeck *et al.*, supra). Sin embargo, en el uso común y en la presente solicitud, el término "homólogo", cuando se modifica con un adverbio tal como "altamente", puede referirse a la similitud de secuencias y puede o no puede estar relacionado con un origen evolutivo común.

Una molécula de ácido nucleico es "hibridable" a otra molécula de ácido nucleico, tal como un ADNc, ADN genómico, o ARN, cuando una forma de cadena sencilla de la molécula de ácido nucleico puede asociarse con la otra molécula de ácido nucleico en las condiciones apropiadas de temperatura y fuerza iónica en solución (véase Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Segunda Edición (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY). Las condiciones de temperatura y fuerza iónica determinan la "rigurosidad" de la hibridación. Para el rastreo preliminar de ácidos nucleicos homólogos, se pueden utilizar condiciones de hibridación de rigurosidad baja, correspondientes a una T_m (temperatura de fusión) de 55°C, por ejemplo, 5xSSC, SDS al 0,1%, leche al 0,25% y sin formamida; o formamida al 30%, 5xSSC, SDS al 0,5%). Condiciones de hibridación de rigurosidad moderada corresponden a una T_m superior, p. ej. formamida al 40%, con 5x o 6xSSC. Condiciones de hibridación de alta rigurosidad corresponden a la T_m más alta, por ejemplo, formamida al 50%, 5x o 6xSSC. SSC es NaCl 0,15 M, citrato de Na 0,015 M. La hibridación requiere que los dos ácidos nucleicos contengan tramos complementarios de secuencias genéticas o de aminoácidos, aunque dependiendo de la rigurosidad de la hibridación, son posibles uniones erróneas entre bases.

La rigurosidad apropiada para hibridar ácidos nucleicos depende de la longitud de los ácidos nucleicos y del grado de complementariedad, variables bien conocidas en la técnica. Cuanto mayor sea el grado de similitud u homología entre dos secuencias de nucleótidos, más alto será el valor de T_m para los híbridos de ácidos nucleicos que tienen esas secuencias. La estabilidad relativa (correspondiente a la T_m alta) de hibridaciones de ácidos nucleicos disminuye en el siguiente orden: ARN:ARN, ADN:ARN, ADN:ADN. Para híbridos de más de 100 nucleótidos de longitud, se han derivado ecuaciones para calcular la T_m (véase Sambrook *et al.*, supra, 9.50-9.51). Para la hibridación con ácidos nucleicos más cortos, es decir, oligonucleótidos, la posición de los errores de apareamiento se vuelve más importante, y la longitud del oligonucleótido determina su especificidad (véase Sambrook *et al.*, supra, 11.7- 11.8). Una longitud mínima para un ácido nucleico hibridable es de al menos aproximadamente 10 nucleótidos; preferiblemente de al menos aproximadamente 15 nucleótidos; y más preferiblemente la longitud es de al menos aproximadamente 20 nucleótidos.

Salvo que se especifique, la expresión "condiciones de hibridación estándares" se refiere a una T_m de aproximadamente 55°C, y utiliza las condiciones como se estableció anteriormente. En una forma de realización preferida, la T_m es 60°C; en una realización más preferida, la T_m es 65°C. En una realización específica, "alta rigurosidad" se refiere a las condiciones de hibridación y/o lavado a 68°C en 0,2xSSC, a 42°C en formamida al 50%, 4xSSC, o bajo condiciones que ofrezcan niveles de hibridación equivalentes a los observados bajo cualquiera de estas dos condiciones.

Condiciones de hibridación adecuadas para oligonucleótidos (p. ej. para sondas o cebadores de oligonucleótidos) son típicamente algo diferentes que para los ácidos nucleicos de longitud completa (p. ej., ADNc de longitud completa), debido a la menor temperatura de fusión de los oligonucleótidos. Debido a que la temperatura de fusión de los oligonucleótidos dependerá de la longitud de las secuencias de oligonucleótidos implicadas, temperaturas de hibridación adecuadas variarán dependiendo de las moléculas de oligonucleótidos utilizadas. Temperaturas ilustrativas pueden ser 37°C (para oligonucleótidos de 14 bases), 48°C (para oligonucleótidos de 17 bases), 55°C (para oligonucleótidos 20 bases) y 60°C (para oligonucleótidos de 23 bases). Condiciones de hibridación adecuadas ilustrativas para oligonucleótidos incluyen el lavado en 6xSSC/pirofosfato sódico al 0,05%, u otras condiciones que aportan niveles equivalentes de hibridación.

"Polipéptido", "péptido" o "proteína" se utilizan indistintamente para describir una cadena de aminoácidos que están unidos entre sí por enlaces químicos denominados "enlaces peptídicos". Una proteína o polipéptido, incluida una enzima, puede ser "nativo" o de "tipo salvaje", lo que significa que se produce en la naturaleza; o puede ser un "mutante", "variante" o "modificado", lo que significa que se ha hecho, alterado, derivado, o de alguna manera es diferente o se ha cambiado de una proteína nativa o de otro mutante.

"Rotámeros" se define como un conjunto de posibles conformeros para cada una de las cadenas laterales de aminoácidos o análogos. Véase Ponder, *et al.*, Acad. Press Inc. (London) Ltd. págs. 775-791 (1987); Dunbrack, *et al.*, *Struc. Biol.* 1(5): 334-340 (1994); Desmet, *et al.*, *Nature* 356: 539-542 (1992). Un "banco de rotámeros" es una

colección de un conjunto de conformaciones rotaméticas posibles/permitidas para un conjunto dado de aminoácidos o análogos. Existen dos tipos generales de bancos de rotámeros: "dependiente de la cadena principal" e "independiente de la cadena principal". Un banco de rotámeros dependiente de la cadena principal permite diferentes rotámeros dependiendo de la posición del residuo en la cadena principal; así, por ejemplo, se permiten determinados rotámeros de leucina si la posición está dentro de una hélice α , y se permiten diferentes rotámeros de leucina si la posición no está en una hélice α . Un banco de rotámeros independiente de la cadena principal utiliza todos los rotámeros de un aminoácido en cada una de las posiciones. En general, se prefiere un banco independiente de la cadena principal considerando residuos del núcleo, puesto que la flexibilidad en el núcleo es importante. Sin embargo, bancos independientes de la cadena principal son computacionalmente más caros y, por lo tanto, para posiciones en la superficie y los límites, se prefiere un banco dependiente de la cadena principal. Sin embargo, se puede utilizar cualquier tipo de banco en cualquier posición.

"Posición del residuo variable" en esta memoria quiera dar a entender una posición de aminoácido de la proteína a diseñar que no se fija en el método de diseño como un residuo o rotámero específico, generalmente el residuo o rotámero de tipo salvaje. Cabe señalar que incluso si una posición se elige como una posición variable, es posible que determinados métodos descritos en esta memoria optimicen la secuencia de tal manera como para seleccionar el residuo de tipo salvaje en la posición variable. En general, esto ocurre con mayor frecuencia para los residuos del núcleo, y con menos regularidad para residuos en la superficie. Además, asimismo es posible fijar residuos como aminoácidos de tipo no salvaje.

"Posición de residuo fija" significa que el residuo identificado en la estructura tridimensional se encuentra en una conformación establecida. En algunas formas de realización, una posición fija se deja en su conformación original (que puede o puede no correlacionarse con un rotámero específico del banco de rotámeros que se esté utilizando). Alternativamente, los residuos se pueden fijar como un residuo de tipo no salvaje en función de las necesidades de diseño; por ejemplo, cuando las técnicas de mutagénesis dirigida al sitio conocidas han demostrado que es deseable un residuo particular (por ejemplo, para eliminar un sitio proteolítico o alterar la especificidad del sustrato para una AARS), el residuo puede fijarse como un aminoácido particular. Los residuos que pueden fijarse incluyen, pero no se limitan a residuos estructural o biológicamente funcionales, por ejemplo los residuos de anclaje.

En determinadas formas de realización, una posición fija puede ser "flotando"; el aminoácido o análogo en esa posición es fijo, pero se someten a ensayo diferentes rotámeros de ese aminoácido o análogo. En esta forma de realización, los residuos variables pueden ser al menos uno, o en cualquier lugar de 0,1% a 99,9% del número total de residuos. Así, por ejemplo, puede ser posible cambiar sólo unos pocos residuos (o uno) de la mayoría de los residuos, con todas las posibilidades entremedias.

Tal como se utiliza en esta memoria, la expresión "mutante externa" se refiere a una molécula modificada (p. ej., un ARNt mutante externo y/o una aminoacil ARNt sintetasa mutante externa) que exhibe una eficacia reducida (en comparación con la de tipo salvaje o endógena) para la aminoacilación con el correspondiente aminoácido de tipo salvaje. "Mutante externa" se refiere a la incapacidad o eficacia reducida, p. ej., menos de 20% de eficacia, menos de 10% de eficacia, menos de 5% de eficacia o, por ejemplo, menos de 1% de eficacia, de un ARNt y/o una RS de actuar con el correspondiente aminoácido que se produce de forma natural en el sistema de traducción de interés. Por ejemplo, un mutante RS externos en un sistema de traducción de interés aminoacila cualquier ARNt endógeno de un sistema de traducción de interés con el aminoácido de tipo salvaje a una eficacia reducida o incluso cero, en comparación con la aminoacilación de un ARNt endógeno por parte de una RS endógena.

Cabe señalar, sin embargo, que una RS mutante externa aminoacila un ARNt endógeno con un aminoácido de sustitución (tanto si se produce de forma natural como no natural) con una eficacia incrementada en comparación con la capacidad de la RS endógena de aminoacilar un ARNt endógeno con un aminoácido de sustitución. De igual manera, un ARNt mutante externo funciona con una mayor eficacia hacia el aminoácido de sustitución (ya sea un aminoácido no natural u otro aminoácido que se produzca de forma natural) que hacia el correspondiente aminoácidos de tipo salvaje.

"Codón degenerado por bamboleo" se refiere a un codón que codifica un aminoácido natural, codón que, cuando está presente en ARNm, es reconocido por un anticodón de ARNt natural, a través de al menos un apareamiento de bases no de Watson-Crick, o por bamboleo (p. ej. apareamiento de bases A-C o G-U). El apareamiento de bases de Watson-Crick se refiere al apareamiento de bases G-C o A-U (ARN o híbrido de ADN/ARN) o A-T (ADN). Cuando se utiliza en el contexto de apareamiento de bases codón de ARNm – anticodón de ARNt, el apareamiento de bases de Watson-Crick significa que todos los apareamientos de bases codón-anticodón están mediados a través de cualquier apareamiento G-C o A-U.

La expresión "preferiblemente aminoacila" se refiere a una eficacia, p. ej., de aproximadamente el 20%, aproximadamente el 30%, aproximadamente el 40%, aproximadamente el 50%, aproximadamente el 60%,

aproximadamente el 70%, aproximadamente el 75%, aproximadamente el 85%, aproximadamente el 90%, aproximadamente el 95%, aproximadamente el 99% o eficacia mayor. La eficacia se puede medir mediante la cual una aminoacil ARNt sintetasa modificada o mutante externa aminoacila un ARNt con un aminoácido de sustitución, ya sea un aminoácido antinatural u otro aminoácido que se produce de forma natural cuando se compara con el correspondiente aminoácido natural asignado al ARNt particular, a la AARS, o a ambos. La expresión "preferiblemente aminoacila" puede referirse, además, a la eficacia de la aminoacil ARNt sintetasa modificada o mutante externa a aminoacilar o cargar un ARNt con cualquier aminoácido que no sea el correspondiente aminoácido natural asignado al ARNt particular, la AARS, o ambos. La expresión "preferiblemente aminoacila" puede referirse, además, a la eficacia de la aminoacil ARNt sintetasa modificada o mutante externa a aminoacilar un ARNt con un aminoácido no natural en comparación con la AARS no modificada o que se produce de forma natural.

Cabe señalar que la eficacia de aminoacilación del ARNt por parte de la AARS puede ser correlacionada con la eficacia de la especificidad, o la fidelidad de incorporación del aminoácido no natural en el polipéptido o proteína diana. Esto es debido a la función de la maquinaria de síntesis de proteínas en que, una vez que un ARNt se aminoacila con un aminoácido (ya sea el aminoácido de tipo salvaje, o un aminoácido no natural), el ARNt cargado se libera de la enzima AARS y el aminoácido se incorpora en el polipéptido diana. Cuando se altera la capacidad de corrección de la AARS, la enzima permitirá que el aminoácido de sustitución cargue el ARNt y sea liberado para su incorporación en la proteína diana. Por lo tanto, la eficacia de aminoacilación por parte del AARS se correlaciona directamente con la fidelidad o especificidad de incorporación del aminoácido no natural en el polipéptido diana.

El aminoácido de sustitución (ya sea no natural o que se produzca de forma natural) se incorpora a continuación en una cadena polipeptídica en crecimiento con alta fidelidad, p. ej. una eficacia mayor que aproximadamente 20%, 30%, 40%, 50%, 60%, 75%, 80 %, 90%, 95%, o mayor que aproximadamente 99% para un codón particular.

El término "complementario" se refiere a componentes de un par mutante externo, el ARNt mutante externo y la sintetasa mutante externa que pueden funcionar juntos, p. ej. la sintetasa mutante externa aminoacila el ARNt mutante externo.

La expresión "derivado de" se refiere a un componente que se aísla de un organismo, o se aísla y modifica, o se genera, p. ej., se sintetiza químicamente, utilizando información del componente del organismo.

La expresión "sistema de traducción" se refiere a los componentes necesarios para incorporar un aminoácido que se produce de forma natural o antinatural en una cadena polipeptídica en crecimiento (proteína). Por ejemplo, los componentes pueden incluir ribosomas, ARNt(s), sintetasa(s), ARNm y similares. Los componentes descritos en esta memoria pueden añadirse a un sistema de traducción, *in vivo* o *in vitro*. Un sistema de traducción *in vivo* puede ser una célula (célula eucariota o procariota). Un sistema de traducción *in vitro* puede ser un sistema exento de células tal como uno reconstituido con componentes de diferentes organismos (purificados o producidos de forma recombinante). En determinadas formas de realización, el sistema de traducción no comprende una célula. En determinadas formas de realización, el sistema de traducción no comprende una célula auxotrófica. Si el sistema de traducción no comprende una célula auxotrófica, puede comprender otra célula o componentes celulares.

La expresión "RS inactiva" se refiere a un sintetasa que ha sido mutada, de manera que ya no puede aminoacilar su ARNt afín con cualquier aminoácido, ya sea que se produce de forma natural o no natural. La expresión "RS modificada" se refiere a un sintetasa que ha sido mutada de manera que ya no puede aminoacilar su ARNt afín con el correspondiente aminoácido que se produce de forma natural, pero puede ser capaz de aminoacilar su ARNt afín con otro aminoácido, preferiblemente un aminoácido no natural.

La expresión "agente de selección" se refiere a un agente que, cuando está presente, permite una selección de determinados componentes de una población, p. ej., un antibiótico, longitud de onda de la luz, un anticuerpo, un nutriente o similares. El agente de selección se puede variar, p. ej., tal como concentración, intensidad, etc.

La expresión "marcador de selección positiva" se refiere a un marcador que cuando está presente, p. ej., expresado, activado o similares, da como resultado la identificación de un organismo con el marcador de selección positivo de los que no tienen el marcador de selección positivo.

La expresión "marcador de selección negativo" se refiere a un marcador que cuando está presente, p. ej., expresado, activado o similares, permite la identificación de un organismo que no posee la propiedad deseada (p. ej., en comparación con un organismo que sí posee la propiedad deseada).

El término "informador" se refiere a un componente que se puede utilizar para seleccionar los componentes descritos en la descripción. Por ejemplo, un informador puede incluir una proteína verde fluorescente, una proteína de luciferasa de luciérnaga, o genes tales como β -gal/lacZ (β -galactosidasa), Adh (alcohol deshidrogenasa) o similares.

La expresión "no reconocido de manera eficiente" se refiere a una eficacia, p. ej., menor que aproximadamente 10%, menor que aproximadamente 5% o menor que aproximadamente 1%, a la que una RS de un organismo aminoacila un ARNt mutante externo. En determinadas formas de realización, la RS puede ser del mismo o de un organismo diferente que el ARNt mutante externo. En algunas realizaciones, la RS ha sido modificada para aminoacilar un ARNt con un aminoácido particular, preferiblemente un aminoácido no natural.

El término "eucariota" se refiere a organismos que pertenecen al dominio filogenético Eucarya, tales como animales (p. ej., mamíferos, insectos, reptiles, aves, etc.), ciliados, plantas, hongos (p. ej., levaduras, etc.), flagelados, microsporidios, protistas, etc. Adicionalmente, el término "procariota" se refiere a organismos no eucariotas pertenecientes a los dominios filogenéticos Eubacteria (p. ej., *Escherichia coli*, *Thermus thermophilus*, etc.) y Archaea (p. ej., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium* tales como *Haloferax volcanii* y especies de *Halobacterium* NRC-1, *A. fulgidus*, *P. firiokus*, *P. horikoshii*, *A. pernix*, etc.).

El Código Genético, Células Huésped y los Codones Degenerados

El código genético estándar que utiliza la mayoría de las células se lista a continuación.

El Medio del Código Genético

Primero	U	C	A	G	Último
	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
U	Leu	Ser	Terminación		
(Ocre)	Terminación				
(Ámbar)	A				
	Leu	Ser	Terminación		
(Ámbar)	Trp	G			
	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
C	Leu	Pro	Gin	Arg	A
	Leu	Pro	Gin	Arg	G
	Ile	Thr	Asn	Ser	U
A	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
	Val	Ala	Asp	Gly	U
G	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

El código genético es degenerado, debido a que la maquinaria biosintética de proteínas utiliza 61 codones sentido ARNm para dirigir la polimerización con molde de los 20 monómeros de aminoácidos naturales. (Crick *et al.*, *Nature* 192: 1227, 1961). Sólo dos aminoácidos, es decir, metionina y triptófano, son codificados por tripletes de ARNm únicos.

El código genético estándar se aplica a la mayoría, pero no a todos los casos. Se han encontrado excepciones en el ADN mitocondrial de muchos organismos y en el ADN nuclear de unos pocos organismos inferiores. Algunos ejemplos se dan en la siguiente tabla.

Ejemplos de códigos genéticos no estándares

Mitocondrias	Vertebrados	UGC→Trp; AGA, AGG→TERMINACIÓN
	Invertebrados	UGA→Trp; AGA, AGG→Ser
	Levaduras	UGC→Trp; CUN→Thr
	Protistas	UGA→Trp;

Núcleo	Bacterias	GUG, UUG, AUU, CUG→Iniciación
	Levaduras	CUG→Ser
	Ciliados	UAA, UGA→Gln

* Las células vegetales utilizan el código genético estándar tanto en mitocondrias como en el núcleo.

El NCBI (Centro Nacional de Información Biotecnológica) mantiene una lista detallada del código genético estándar, y los códigos genéticos utilizados en diversos organismos, incluido el código mitocondrial de vertebrados; el código mitocondrial de levaduras; el código genético de hongos, protozoos y celentéreos y el código de micoplasma / espiroplasma; el código mitocondrial de invertebrados; el código nuclear de ciliados, dasycladaceae y hexamita; el código mitocondrial de equinodermos y platelmintos; el código nuclear de eulótida; el código bacteriano y vegetal de plastidio; el código nuclear de la levadura alternativa; el código mitocondrial de la ascidia; el código mitocondrial alternativo de platelmintos; el código nuclear de blepharisma; el código mitocondrial de clorofíceas; el código mitocondrial de trematodos; el código mitocondrial de scenedesmus obliquus; el código mitocondrial de thraustochytrium (todos ellos incorporados en esta memoria como referencia). Estas se basan principalmente en las revisiones de Osawa *et al.*, *Microbiol. Rev.* 56: 229-264, 1992, y Jukes y Osawa, *Comp. Biochem. Physiol.* 106B: 489-494, 1993.

Células Huésped

Algunos métodos descritos en esta memoria pueden ponerse en práctica dentro de una célula, lo que permite hacer niveles de producción de proteínas para fines prácticos. En realizaciones preferidas, las células utilizadas son células cultivables (es decir, células que pueden ser cultivadas en condiciones de laboratorio). Células adecuadas incluyen células de mamíferos (mamíferos humanos o no humanos), células bacterianas y células de insectos, etc.

Un ejemplo incluye la tecnología PFENEX™, que es una línea celular que utiliza una línea celular basada en *Pseudomonas fluorescens* que aumenta la expresión celular, al tiempo que mantiene determinadas características de solubilidad y actividad debido a su uso de diferentes vías en el metabolismo de determinados azúcares en comparación con *E. coli*.

Además, otras líneas de células huésped auxotróficas la cepa Phe auxotrófica basada en K10 (AF), las cepas auxotróficas dobles Phe/Trp (AFW), las cepas auxotróficas triples de Phe/Trp/Lys (AFWK), un auxótrofo Met (M 15MA sobre fondo M15), así como la cepa AF basada en DH10B.

Células que se pueden utilizar para practicar determinadas formas de realización descritas en esta memoria incluyen células huésped auxotróficas (ya sean procariotas o eucariotas). Células auxotróficas pueden exhibir niveles de auxotrofia sencillos, dobles, triples, cuádruples o mayores (indicando cada uno de los niveles de auxotrofia un compuesto orgánico particular de que el organismo es incapaz de sintetizar y deben ser suministrados a la célula). Determinadas formas de realización descritas en esta memoria expresamente no utilizan una célula huésped auxotrófica. En la medida en que no se utilice una célula huésped auxotrófica, se puede seguir utilizando otra célula o componentes celulares para poner en práctica determinadas formas de realización descritas en esta memoria. Otras formas de realización pueden utilizar uno, dos, tres o más células huésped auxotróficas diferentes que pueden ser de la misma o de diferentes cepas u organismos.

Las células huésped se modifican genéticamente (p. ej., se transforman, transducen o transfectan) con los vectores de esta descripción, que pueden ser, por ejemplo, un vector de clonación o un vector de expresión. El vector puede estar, por ejemplo, en forma de un plásmido, una bacteria, un virus, un polinucleótido desnudo o un polinucleótido conjugado. Los vectores se introducen en células y/o microorganismos por métodos estándares, incluidos electroporación (From *et al.*, *PNAS USA* 82, 5824 (1985)), Infección por vectores virales, penetración balística de alta velocidad por partículas pequeñas con el ácido nucleico ya sea dentro de la matriz de pequeñas perlas o partículas, o sobre la superficie (Klein *et al.*, *Nature* 327, 70-73 (1987)). Berger, Sambrook y Ausubel proporcionan una diversidad de métodos de transformación apropiados.

Las células huésped modificadas pueden cultivarse en medios nutrientes convencionales modificados según sea apropiado para actividades tales como, por ejemplo, etapas de rastreo, activación de promotores o selección de transformantes. Estas células pueden opcionalmente cultivarse en organismos transgénicos.

Determinadas formas de realización descritas en esta memoria incluyen, además, métodos de rastreo de AARSs modificadas y/o ARNts modificados. Por ejemplo, en una forma de realización, un banco de PheRS de levaduras se somete a rastreo por doble tamiz para detectar altos niveles de incorporación de un aminoácido no natural o la

incorporación errónea de aminoácidos naturales distintos de Phe dará lugar a un serio plegamiento erróneo o a un despliegue de GFP. Las células del banco de PheRS de levaduras son así sometidas a un rastreo de alto rendimiento basado en el análisis por citometría de flujo (FACS). En primer lugar, las células del banco de PheRS de levadura se expresan en presencia de células 2Nal y de baja fluorescencia, indicando una mayor incorporación de cualquiera de 2Nal u otros aminoácidos naturales son recogidos por FACS. A continuación, las células del banco de PheRS de levaduras se expresan sin 2Nal. Las células brillantes se recogen a fin de eliminar variantes de PheRS de levaduras que puedan incorporar erróneamente otros aminoácidos naturales. En una forma de realización ilustrativa, dos ciclos de cribado produjeron una PheRS de levadura mutante con mutaciones en N412G, S418C, T415G y S437F, que tenían una baja fluorescencia en presencia de 2Nal y una alta fluorescencia en ausencia de 2Nal. Esta técnica permite la incorporación de 2Nal en el codón UUU, aumentando hasta alrededor de 90%.

Otras referencias útiles, p. ej., para el aislamiento y cultivo de células (p. ej., para el subsiguiente aislamiento de ácidos nucleicos) incluyen Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, tercera edición, Wiley-Liss, Nueva York y las referencias citadas en la misma; Payne *et al.* (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. Nueva York, N.Y.; Gamborg y Phillips (comps.) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods* Springer Lab Manual, Springer-Verlag (Berlin Heidelberg Nueva York) y Atlas y Parks (comps.) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.

Están disponibles varios métodos bien conocidos para introducir ácidos nucleicos diana en células bacterianas, cualquiera de los cuales puede ser utilizado en determinadas formas de realización descritas en esta memoria. Éstos incluyen: la fusión de las células receptoras con protoplastos bacterianos que contienen el ADN, electroporación, bombardeo de proyectiles y la infección con vectores virales, etc. Las células bacterianas se pueden utilizar para amplificar el número de plásmidos que contienen construcciones de ADN de determinadas formas de realización descritas en esta memoria. Las bacterias se cultivan hasta la fase logarítmica y los plásmidos dentro de las bacterias se pueden aislar por una diversidad de métodos conocidos en la técnica (véase, por ejemplo, Sambrook). Además, está comercialmente disponible una plétora de kits para la purificación de plásmidos a partir de bacterias, (véase, p. ej., EASYPREP™, FLEXIPREP™, ambos de Pharmacia Biotech; STRATACLEAN™, de Stratagene; y, QIAPREP™ de Qiagen). Los plásmidos aislados y purificados se manipulan luego adicionalmente para producir otros plásmidos, utilizados para transfectar células o se incorporan en vectores relacionados para infectar organismos.

Vectores típicos contienen terminadores de la transcripción y traducción, secuencias de iniciación de la transcripción y traducción, y promotores útiles para la regulación de la expresión del ácido nucleico diana particular. Los vectores comprenden, opcionalmente, casetes de expresión genéricas que contienen al menos una secuencia de terminador independiente, secuencias que permiten la replicación de la casete en eucariotas o procariotas, o en ambos, (p. ej., vectores lanzadera) y marcadores de selección tanto para sistemas procariotas como eucariotas. Los vectores son adecuados para la replicación e integración en procariotas, eucariotas o preferiblemente en ambos. Véase Gilman y Smith, *Gene* 8:81 (1979); Roberts, *et al.*, *Nature*, 328: 731 (1987); Schneider, B., *et al.*, *Protein Expr. Purif.* 6435:10 (1995); Ausubel, Sambrook, Berger (todos supra). Se proporciona un catálogo de Bacterias y Bacteriófagos útiles para la clonación, p. ej., por la ATCC, p. ej., *The ATCC Catalogue of Bacteria and Bacteriophage* (1992) Gherna *et al.* (comps.) publicado por la ATCC. Procedimientos básicos adicionales para la secuenciación, la clonación y otros aspectos de la biología molecular y consideraciones teóricas subyacentes también se encuentran en Watson *et al.* (1992) *Recombinant DNA* 2ª Edición Scientific American Books, NY.

Selección de Codones Degenerados

Tal como se describió anteriormente, todos los aminoácidos, con la excepción de metionina y triptófano, son codificados por más de un codón. De acuerdo con algunos métodos descritos en esta memoria, un codón en el genoma que se utiliza normalmente para codificar un aminoácido natural se reprograma, en parte por la maquinaria transcripcional o de traducción en lugar de codificar un análogo de aminoácido. Un análogo de aminoácido puede ser un análogo de aminoácido que se produce de forma natural o canónico. En una forma de realización preferida, el análogo de aminoácido no es un aminoácido codificado canónicamente.

La estabilidad termodinámica de un par codón-anticodón se puede predecir o determinar experimentalmente. De acuerdo con algunas realizaciones, es preferible que el ARNt mutante externo interactúe con el codón degenerado con una afinidad (a 37°C) de al menos aproximadamente 1,0 kcal/mol más intensamente, incluso más preferiblemente 1,5 kcal/mol más intensamente, e incluso más preferiblemente más de 2,0 kcal/mol más intensamente que un ARNt natural en la célula que reconocería la misma secuencia. Estos valores son conocidos para un experto en la técnica y se pueden determinar mediante experimentos de desnaturalización térmica (véase, p. ej., Meroueh y Chow, *Nucleic Acids Res.* 27: 1118, 1999).

La siguiente Tabla lista algunas de las secuencias anti-codón conocidos para *E. coli*. En general, para cualquier organismo, la secuencia de anticodón de ARNt se puede determinar de forma rutinaria utilizando tecnologías reconocidas en la técnica. Por ejemplo, cualquier gen ARNt puede ser amplificado, por ejemplo, mediante PCR. La secuenciación se puede realizar para determinar las secuencias exactas del bucle anti-codón. Alternativamente, el ensayo de unión bioquímico se puede utilizar para determinar la afinidad de unión de un ARNt purificado a uno de los 2-6 posibles codones. El codón que se une al ARNt con la mayor especificidad / afinidad tiene, presumiblemente, un apareamiento de Watson-Crick puro en las tres posiciones del codón, determinando así la secuencia del bucle anti-codón.

En general, la base de bamboleo en el bucle de anti-codón tiende a ser G o U (en lugar de A o C).

Los Codones Degenerados para *E. coli*

Amino-ácido	Anti-codón	Apareamiento de bases en la 3ª base	Codón	Amino-ácido	Anti-codón	Apareamiento de bases	Codón
Ala	GGC	W/C ¹	GCC	His	GUG	W/C	CAC
		Bamboleo ²	GCU			Bamboleo	CAU
	UGC	W/C	GCA	Ile	GAU	W/C	AUC
		Bamboleo	GCG			Bamboleo	AUU, AUA
Asp	GUC	W/C	GAC	Leu	GAG	W/C	CUC, CUA, CUG, UUC, UUG
		Bamboleo	GAU			Bamboleo	CUU
Asn	GUU	W/C	AAC	Lys	UUU	W/C	AAA
		Bamboleo	AAU			Bamboleo	AAG
Cys	GCA	W/C	UGC	Phe	GAA	W/C	UUC
		Bamboleo	UGU			Bamboleo	UUU
Glu	UUC	W/C	GAA	Ser	GGA	W/C	UUC, AGU
		Bamboleo	GAG			Bamboleo	UCU, AGC, UCA, UCG
Gly	GCC	W/C	GGC, GGA, GGG	Tyr	GUA	W/C	UAC
		Bamboleo	GGU			Bamboleo	UAU
Met		W/C	AUG	Thr		W/C	ACC, ACA, ACG
Gin		W/C	CAA, CAG			Bamboleo	ACU
Arg		W/C	AGA, AGG, CGU, CGG	Pro		W/C	CCC, CCA, CCG
		Bamboleo	CGC, CGA	Trp		Bamboleo	CCU
						W/C	UGG
STOP		W/C	UGA, UAA	Val		W/C	GUC, GUA
		Bamboleo	UAG			Bamboleo	GUU, GUG

¹ Apareamiento de bases de Watson-Crick

² Apareamiento de bases por bamboleo

Cuando un ARNt sencillo reconoce un codón a través de una interacción complementaria perfecta entre el anticodón del ARNt y un codón, se denomina apareamiento de bases de Watson-Crick. Cuando un ARNt sencillo reconoce un segundo codón degenerado, se denomina un apareamiento de base por bamboleo u otro apareamiento de bases no estándar. En determinadas formas de realización descritas en esta memoria, un nuevo ARNt puede ser construido con una secuencia de anticodón que sea perfectamente complementaria a un codón degenerado o a un codón para un aminoácido no natural, utilizando así el apareamiento de bases por bamboleo o de Watson-Crick. Del mismo modo, se puede construir una nueva AARS que utiliza un aminoácido de sustitución (que no sea de tipo salvaje - puede ser otro aminoácido que se produzca de forma natural o un aminoácido no natural) para aminoacilar el ARNt correspondiente. Esto puede ser además de o en lugar de modificar una molécula de ARNt para la incorporación de un aminoácido de sustitución.

La AARS modificada se puede alterar de manera que la eficacia de unión al aminoácido no natural, u otro aminoácido que se produce de forma natural seleccionado, es mayor que la eficacia de unión de la AARS modificada para el aminoácido que se produce de forma natural correspondiente. De esta manera, una AARS modificada puede unirse preferiblemente a un aminoácido no natural con el fin de cargar un ARNt, incluso en presencia del aminoácido que se produce de forma natural que corresponde a la AARS en su estado no modificado. Esta "reprogramación" de una aminoacil ARNt sintetasa permite la incorporación de un aminoácido no natural en un polipéptido con niveles más bajos de incorporación errónea de otros aminoácidos en el sitio deseado.

La "reprogramación" adicional puede permitir el uso de la sintetasa modificada o mutante externa con altos niveles de incorporación en células huésped estándares, sin la necesidad de células huésped auxotróficas, y con o sin agotar los medios del aminoácido que se produce de forma natural correspondiente. Por lo tanto, mientras que determinadas formas de realización descritas en esta memoria pueden ponerse en práctica utilizando una célula huésped auxotrófica, determinadas otras formas de realización pueden ponerse en práctica sin utilizar una célula huésped auxotrófica. En el caso de no utilizar una célula huésped auxotrófica para poner en práctica determinadas formas de realización, se puede utilizar otra célula huésped, se pueden utilizar componentes, o se puede utilizar un sistema totalmente exento de células.

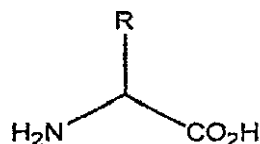
Cuando la célula tiene múltiples moléculas de ARNt para un aminoácido particular, y un ARNt tiene una secuencia de anticodón que es perfectamente complementaria a la del codón degenerado seleccionado, el gen que codifica el ARNt puede ser desactivado a través de cualquier medio disponible para un experto en la técnica, que incluye, por ejemplo, mutagénesis dirigida al sitio o delección del gen o de la secuencia promotora del gen. La expresión del gen también se puede desactivar a través de cualesquiera técnicas de interferencia antisentido o de ARN.

Aminoácidos Antinaturales o No naturales

La primera etapa en el proceso de ingeniería de proteínas es habitualmente seleccionar un conjunto de aminoácidos antinaturales o no naturales que tienen las propiedades químicas deseadas. La selección de los aminoácidos no naturales depende de las propiedades químicas predeterminadas que a uno le gustaría tener, y las modificaciones que a uno le gustaría hacer en la proteína diana. Los aminoácidos antinaturales, una vez seleccionados, se pueden adquirir de vendedores, o se pueden sintetizar químicamente.

En los métodos descritos en esta memoria se puede utilizar una amplia diversidad de aminoácidos antinaturales o no naturales. El aminoácido antinatural se puede seleccionar en base a las características deseadas del aminoácido antinatural, p. ej., la función del aminoácido antinatural tal como modificar las propiedades biológicas de la proteína tales como toxicidad, biodistribución, inmunogenicidad, o semi-vida, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, propiedades catalíticas, capacidad para reaccionar con otras moléculas (ya sea de forma covalente o no covalente), o similares.

Tal como se utiliza en esta memoria, un "aminoácido antinatural" se refiere a cualquier aminoácido, aminoácido modificado o análogo de aminoácido diferente de selenocisteína y los siguientes veinte alfa-aminoácidos genéticamente codificados: alanina, arginina, asparagina, ácido aspártico, cisteína, glutamina, ácido glutámico, glicina, histidina, isoleucina, leucina, lisina, metionina, fenilalanina, prolina, serina, treonina, triptófano, tirosina, valina. La estructura genérica de un alfa-aminoácido se ilustra por la Fórmula I:

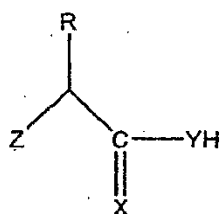


Fórmula I

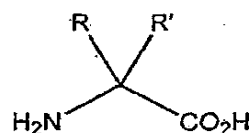
Un aminoácido antinatural es típicamente cualquier estructura que tenga la Fórmula I, en donde el grupo R es cualquier sustituyente distinto del utilizado en los veinte aminoácidos naturales. Véase, p. ej., cualquier texto de bioquímica tal como Biochemistry por L. Stryer, 3ª ed. 1988, Freeman and Company, Nueva York, para las estructuras de los veinte aminoácidos naturales. Obsérvese que los aminoácidos antinaturales descritos en esta memoria pueden ser compuestos que se producen de forma natural distintos de los veinte alfa-aminoácidos anteriores. Debido a que los aminoácidos antinaturales descritos en esta memoria difieren típicamente de los aminoácidos naturales solamente en la cadena lateral, los aminoácidos antinaturales forman enlaces amida con otros aminoácidos, p. ej., naturales o antinaturales, de la misma manera en que se forman en proteínas que se producen de forma natural. Sin embargo, los aminoácidos antinaturales tienen grupos de cadena lateral que los distinguen de los aminoácidos naturales. Por ejemplo, R en la Fórmula I comprende opcionalmente un, grupo alquilo, arilo, haluro de arilo, haluro de vinilo, haluro de alquilo, acetilo, cetona, aziridina, nitrilo, nitro, haluro, acilo, ceto, azido, hidroxilo, hidrazina, ciano, halo, hidrazida, alquenilo, alquinilo, éter, tioéter, epóxido, sulfona, ácido borónico, éster boronato, borano, ácido fenilborónico, tiol, seleno, sulfonilo, borato, boronato, fosfo, fosfono, fosfina, heterocíclico, piridilo, naftilo, benzofenona, un anillo restringido tal como una ciclo-octina, tioéster, enona, imina, aldehído, éster, tioácido, hidroxilamina, amino, ácido carboxílico, ácido alfa-cetocarboxílico, ácidos alfa o beta-insaturados y amidas, glioxil-amida, o grupo organosilano, o similares, o cualquier combinación de los mismos.

Sustituciones arilo pueden aparecer en diversas posiciones, p. ej., orto, meta, para, y con uno o más grupos funcionales colocados en el anillo de arilo. Otros aminoácidos antinaturales de interés incluyen, pero no se limitan a aminoácidos que comprenden un reticulante fotoactivable, aminoácidos spin-marcados, aminoácidos marcados con colorante, aminoácidos fluorescentes, aminoácidos de unión a metales, aminoácidos que contienen metales, aminoácidos radiactivos, aminoácidos con nuevos grupos funcionales, aminoácidos con hidrofili-
 5 hidrofobicidad, polaridad o capacidad de enlace de hidrógeno alterada, aminoácidos que interactúan de forma covalente o no covalente con otras moléculas, aminoácidos foto-enjaulados o fotoisomerizables, aminoácidos que comprenden 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
 10 que comprenden polietilenglicol o un poliéter, un polialcohol o un polisacárido, aminoácidos que puede sufrir metátesis, aminoácidos que pueden sufrir cicloadiciones, aminoácidos sustituidos con átomos pesados, aminoácidos químicamente escindibles y/o fotoescindibles, aminoácidos con cadenas laterales alargadas en comparación con aminoácidos naturales, p. ej., poliéteres o hidrocarburos de cadena larga, p. ej., más de aproximadamente 5 o más de aproximadamente 10 átomos de carbono, aminoácidos que contienen azúcar
 15 enlazado a carbono, aminoácidos con actividad redox, aminoácidos que contienen aminotioácido, aminoácidos que contienen un resto de fármaco y aminoácidos que comprenden uno o más restos tóxicos.

Además de aminoácidos antinaturales que contienen nuevas cadenas laterales, los aminoácidos antinaturales también comprenden opcionalmente estructuras de cadena principal modificadas, p. ej., como se ilustra por las
 20 estructuras de las Fórmulas II y III:



Fórmula II



Fórmula III

en donde Z comprende típicamente OH, NH₂, SH, NH₂O⁻, NH-R', R'NH-, R'S- o S-R'-; X e Y, que pueden ser
 25 iguales o diferentes, comprenden típicamente S, N u O, y R y R', que son opcionalmente iguales o diferentes, se seleccionan típicamente de la misma lista de constituyentes para el grupo R descrito anteriormente para los aminoácidos antinaturales que tienen la Fórmula I así como hidrógeno o (CH₂)_x o las cadenas laterales de aminoácidos naturales. Por ejemplo, los aminoácidos antinaturales descritos en esta memoria comprenden
 30 opcionalmente sustituciones en el grupo amino o carboxilo tal como se ilustra por las Fórmulas II y III. Aminoácidos antinaturales de este tipo incluyen, pero no se limitan a α-hidroxiácidos, α-aminotiocarboxilatos de α-tioácidos, o aminoácidos α-α-disustituidos, con cadenas laterales correspondientes, p. ej., a los veinte aminoácidos naturales o a cadenas laterales antinaturales. También incluyen pero no se limitan a β-aminoácidos o γ-
 35 aminoácidos, tales como β-alanina sustituida y ácido γ-aminobutírico. Además, sustituciones o modificaciones en el carbono α incluyen, opcionalmente, L o D isómeros tales como D-glutamato, D-alanina, D-metil-O- tirosina, ácido aminobutírico, y similares. Otras alternativas estructurales incluyen aminoácidos cíclicos tales como análogos de prolina, así como análogos de prolina de anillos de 3, 4, 6, 7, 8 y 9 miembros. Algunos aminoácidos
 40 no naturales tales como haluros de arilo (p-bromo-fenilalanina, p-yodofenilalanina) proporcionan reacciones de acoplamiento cruzado catalizadas por paladio versátil con reacciones de etino o acetileno que permiten la formación de enlaces carbono-carbono, carbono-nitrógeno y carbono-oxígeno entre haluros de arilo y una amplia diversidad de participantes en el acoplamiento.

Por ejemplo, muchos aminoácidos antinaturales se basan en aminoácidos naturales, tales como tirosina, glutamina, fenilalanina y similares. Análogos de tirosina incluyen tirosinas para-sustituidas, tirosinas orto-
 45 sustituidas y tirosinas meta-sustituidas, en donde la tirosina sustituida comprende 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 C6-C20 de cadena lineal o ramificado o, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, o similares. Además, también se contemplan anillos de arilo poli-
 50 sustituidos. Análogos de glutamina incluyen, pero no se limitan a, derivados α-hidroxi, derivados β-sustituidos, derivados cíclicos y derivados de glutamina sustituidos con amida. Análogos de fenilalanina ilustrativos incluyen, pero no se limitan a fenilalaninas meta-sustituidas, en donde el sustituyente comprende un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un grupo acetilo, o similares.

Ejemplos específicos de aminoácidos antinaturales incluyen, pero no se limitan a formas *o*, *m* y/o *p* de aminoácidos o análogos de aminoácidos (aminoácidos no naturales), incluidos homoalilglicina, cis- o trans-
 55 crotilglicina, ácido 6,6,6-trifluoro-2-aminoheptanoico, ácido 2-aminoheptanoico, norvalina, norleucina, O-metil-L-

tirosina, o-, m- o p-metil-fenilalanina, 4-O-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-azidofenilalanina, una p-acil-L-fenilalanina, una p-benzoil-L-fenilalanina, una L-fosfoserina, una fosfonoserina, una fosfonotirosina, una p-yodo-fenilalanina, o-, m- o p-bromofenilalanina, 2-, 3- ó 4-piridilalanina, p-yodofenilalanina, ácido diaminobutírico, ácido aminobutírico, benzofuranilalanina, 3-bromo-tirosina, 3-(6-cloroindolil)alanina, 3-(6-bromoindolil)alanina, 3-(5-bromonindolil)alanina, p-clorofenilalanina, p-etinil-fenilalanina, p-propargil-oxi-fenilalanina, m-etinil-fenilalanina, 6-etinil-triptófano, 5-etinil-triptófano, ácido (R)-2-amino-3-(4-etinil-1H-pirol-3-il)propanoico, azidonorleucina, azidohomoalanina, p-acetilfenilalanina, p-amino-L-fenilalanina, homopropargilglicina, p-etil-fenilalanina, p-etinil-fenilalanina, p-propargil-oxi-fenilalanina, isopropil-L-fenilalanina, una 3-(2-naftil)alanina, 3-(1-naftil)alanina, 3-idio-tirosina, O-propargil-tirosina, homoglutamina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una 3-nitro-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-acetil-L-fenilalanina, una m-acetil-L-fenilalanina, selenometionina, telurometionina, selenocisteína, una alquino-fenilalanina, una O-alil-L-tirosina, una O-(2-propinil)-L-tirosina, una p-etiltiocarbonil-L-fenilalanina, una p-(3-oxobutanoil)-L-fenilalanina, una p-benzoil-L-fenilalanina, una L-fosfoserina, una fosfonoserina, una fosfonotirosina, homopropargilglicina, azidohomoalanina, una p-yodo-fenilalanina, una p-bromo-L-fenilalanina, dihidroxi-fenilalanina, dihidroxil-L-fenilalanina, una p-nitro-L-fenilalanina, una m-metoxi-L-fenilalanina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina y una isopropil-L-fenilalanina, trifluoroleucina, norleucina, 4-, 5- ó 6-fluoro triptófano, 4-aminotriptófano, 5-hidroxitriptófano, biocitina, ácido amino-oxiacético, m-hidroxifenilalanina, m-alilfenilalanina, grupo m-metoxifenilalanina, β -GlcNAc-serina, α -GalNAc-treonina, p-acetoacetilfenilalanina, para-halo- fenilalanina, seleno-metionina, etionina, S-nitroso-homocisteína, tia-prolina, 3-tienil-alanina, homo-alil-glicina, trifluoroisoleucina, ácido trans- y cis-2-amino-4-hexenoico, 2-butilil-glicina, alil-glicina, para-azido-fenilalanina, para-ciano-fenilalanina, para-etinil-fenilalanina, hexafluoroleucina, 1,2,4-triazol-3-alanina, 2-fluoro-histidina, L-metil-histidina, 3-metil-L-histidina, β -2-tienil-L-alanina, β -(2-tiazolil)-DL-alanina, homopropargilglicina (HPG) y azidohomoalanina (AHA) y similares. Las estructuras de una diversidad de aminoácidos antinaturales no limitantes se proporcionan en las figuras, p. ej., las FIGs. 29, 30 y 31 del documento US 2003/0108885 A1, cuyo contenido completo se incorpora aquí como referencia.

Análogos de tirosina incluyen tirosinas para-sustituidas, tirosinas orto-sustituidas y tirosinas meta-sustituidas, en donde la tirosina sustituida comprende un grupo acetilo, un grupo benzoilo, un grupo amino, una hidrazina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo C6-C20 de cadena lineal o ramificado, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, o similares. Además, también se contemplan anillos arilo polisustituidos. Análogos de glutamina de la invención incluyen, pero no se limitan a derivados α -hidroxi, derivados β -sustituidos, derivados cíclicos y derivados de glutamina sustituidos con amida. Análogos de fenilalanina ilustrativos incluyen, pero no se limitan a fenilalaninas meta-sustituidas, en donde el sustituyente comprende un grupo hidroxil, un grupo metoxi, un grupo metilo, un grupo alilo, un grupo acetilo, o similares.

Adicionalmente, otros ejemplos incluyen opcionalmente (pero no se limitan a) un análogo antinatural de un aminoácido tirosina; un análogo antinatural de un aminoácido glutamina; un análogo antinatural de un aminoácido fenilalanina; un análogo antinatural de un aminoácido serina; un análogo antinatural de un aminoácido treonina; un aminoácido sustituido con alquilo, arilo, acilo, azido, ciano, halo, hidrazina, hidrazida, hidroxilo, alquenilo, alquinilo, éter, tiol, sulfonilo, seleno, éster, tioácido, borato, boronato, fosfo, fosfono, fosfina, heterociclo, enona, imina, aldehído, hidroxilamina, ceto o amino, o cualquier combinación de los mismos; un aminoácido con un reticulante fotoactivable; un aminoácido spin-marcado; un aminoácido fluorescente; un aminoácido con un nuevo grupo funcional; un aminoácido que interactúa de forma covalente o no covalente con otra molécula; un aminoácido de unión a metales; un aminoácido que contiene metales; un aminoácido radiactivo; un aminoácido foto-enjaulado; un aminoácido fotoisomerizable; un aminoácido que contiene biotina o un análogo de biotina; un aminoácido glicosilado o modificado con hidratos de carbono; un aminoácido que contiene ceto; un aminoácido que comprende polietilenglicol; un aminoácido que comprende poliéter; un aminoácido sustituido con átomos pesados; un aminoácido químicamente escindible o fotoescindible; un aminoácido con una cadena lateral alargada; un aminoácido que contiene un grupo tóxico; un aminoácido sustituido con azúcar, p. ej., una serina sustituida con azúcar o similares; un aminoácido que contiene azúcar enlazado a carbono; un aminoácido redox-activo; un ácido que contiene α -hidroxi; un aminoácido que contiene un amino-tio-ácido; un aminoácido α,α -disustituido; un β -aminoácido; y un aminoácido cíclico.

Típicamente, los aminoácidos antinaturales utilizados en esta memoria para determinadas formas de realización se pueden seleccionar o diseñar para proporcionar características adicionales no disponibles en los veinte aminoácidos naturales. Por ejemplo, los aminoácidos antinaturales están opcionalmente diseñados o se seleccionan para modificar las propiedades biológicas de una proteína, p. ej., en la que están incorporados. Por ejemplo, las siguientes propiedades se modifican opcionalmente mediante la inclusión de un aminoácido antinatural en una proteína: toxicidad, biodistribución, solubilidad, estabilidad, p. ej., térmica, hidrolítica, oxidativa,

resistencia a la degradación enzimática, y similares, facilidad de purificación y tratamiento, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, actividad catalítica, potencial redox, semi-vida, capacidad de reaccionar con otras moléculas, p. ej., de forma covalente o no covalente, y similares.

Otros ejemplos de análogos de aminoácidos incluyen opcionalmente (pero no se limitan a) un análogo antinatural de un aminoácido tirosina; un análogo antinatural de un aminoácido glutamina; un análogo antinatural de un aminoácido fenilalanina; un análogo antinatural de un aminoácido serina; un análogo antinatural de un aminoácido treonina; un grupo alquilo, arilo, acilo, azido, ciano, halo, hidrazina, hidrazida, hidroxilo, alquenilo, alquinilo, éter, tiol, sulfonilo, seleno, éster, tioácido, borato, boronato, fosfo, fosfono, fosfina, heterocíclico, enona, imina, aldehído, hidroxilamina, ceto o aminoácido amino-sustituido, o cualquier combinación de los mismos; un aminoácido con un reticulante fotoactivable; un aminoácido spin-marcado; un aminoácido fluorescente; un aminoácido con un nuevo grupo funcional; un aminoácido que interactúa de forma covalente o no covalente con otra molécula; un aminoácido de unión a metales; un aminoácido que contiene metales; un aminoácido radiactivo; un aminoácido foto-enjaulado; un aminoácido fotoisomerizable; un aminoácido que contiene biotina o un análogo de biotina; un aminoácido glicosilado o modificado con hidratos de carbono; un aminoácido que contiene ceto; un aminoácido que comprende polietilenglicol; un aminoácido que comprende poliéter; un aminoácido sustituido con átomos pesados; un aminoácido químicamente escindible o fotoescindible; un aminoácido con una cadena lateral alargada; un aminoácido que contiene un grupo tóxico; un aminoácido sustituido con azúcar, p. ej., una serina sustituida con azúcar o similares; un aminoácido que contiene azúcar enlazado a carbono; un aminoácido redox-activo; un ácido que contiene α -hidroxi; un aminoácido que contiene un amino-tio-ácido; un aminoácido α,α -disustituido; un β -aminoácido; y un aminoácido cíclico.

Aminoacil-ARNt sintetasa

La aminoacil-ARNt sintetasa (se utiliza de forma indistinta en este documento con AARS, RS o "sintetasa") utilizada en determinados métodos descritos en esta memoria puede ser una sintetasa que se produce de forma natural derivada de un organismo, ya sea la misma (homóloga) o diferente (heteróloga), una sintetasa mutada o modificada, o una sintetasa diseñada.

La sintetasa utilizada puede reconocer el análogo de aminoácido (antinatural) deseado selectivamente de aminoácidos disponibles relacionados. Por ejemplo, cuando el análogo de aminoácido a utilizar está estructuralmente relacionado con un aminoácido que se produce de forma natural, la sintetasa debería cargar la molécula de ARNt mutante externa con el análogo de aminoácido deseado con una eficacia al menos sustancialmente equivalente a la de, y más preferiblemente al menos aproximadamente dos veces, 3 veces, 4 veces, 5 veces o más que la del aminoácido que se produce de forma natural. Sin embargo, en los casos en los que no es necesario un producto de proteína bien definido, la sintetasa puede tener una especificidad relajada para cargar los aminoácidos. En una forma de realización de este tipo, se podría producir una mezcla de ARNts mutantes externos, con diversos aminoácidos o análogos.

En determinadas formas de realización, es preferible que la sintetasa tenga actividad tanto para el análogo de aminoácido como para el aminoácido que es codificado por el codón correspondiente de la molécula de ARNt.

Una sintetasa se puede conseguir por una diversidad de técnicas conocidas para un experto en la técnica, incluidas combinaciones de técnicas tales como, por ejemplo, métodos computacionales, métodos de selección y la incorporación de sintetasas de otros organismos (véase más adelante).

En determinadas formas de realización, sintetasas se pueden utilizar o desarrollar para que carguen eficazmente moléculas de ARNt que no están cargadas por sintetasas de la célula huésped. Por ejemplo, generalmente pares adecuados se pueden desarrollar a través de la modificación de sintetasas a partir de organismos distintos de la célula huésped. En determinadas formas de realización, la sintetasa puede ser desarrollada por procesos de selección. En determinadas formas de realización, la sintetasa puede ser diseñado utilizando técnicas computacionales tales como las descritas en Datta *et al.*, *J. Am. Chem. Soc.* 124: 5652-5653, 2002, y en la Patente de EE.UU. N° 7.139.665, incorporada aquí como referencia.

Diseño Computacional de AARS

Específicamente, en una forma de realización, el presente método depende, en parte, del diseño y la ingeniería de AARS naturales para una forma modificada que tiene especificidad de sustrato relajado, de manera que puede absorber análogos de aminoácidos no canónicos como sustrato, y cargar un ARNt modificado (con su anticodón cambiado) con un aminoácido no canónico de este tipo. Las siguientes secciones describen brevemente un método para la generación de tales AARS modificadas, método que se describe con mayor detalle en la Patente de

EE.UU. Nº 7.139.665, cuyos contenidos completos se incorporan aquí como referencia.

Brevemente, los métodos de algunas formas de realización descritos en ese documento se refieren a herramientas computacionales para modificar la especificidad de sustrato de una aminoacil ARNt sintetasa (AARSs) a través de la mutación para permitir que la enzima utilice de manera más eficaz un análogo o análogos de aminoácidos en los sistemas de traducción de proteínas, ya sea *in vitro*, en células enteras, o en otros sistemas de traducción. Una característica de algunas formas de realización incluye rediseñar sistemáticamente el sitio de unión al sustrato de una enzima AARS para facilitar el uso de sustratos antinaturales en la reacción de traducción de péptidos o proteínas que cataliza la enzima.

De acuerdo con un método, se construye un banco de rotámeros para el aminoácido artificial variando sus ángulos de torsión para crear rotámeros que encajarían en el bolsillo de unión para el sustrato natural. La orientación geométrica de la cadena principal del análogo de aminoácido se especifica por la orientación cristalográfica de la cadena principal del sustrato natural en la estructura cristalina. Se puede utilizar la estructura cristalográfica de la aminoácido sintetasa específica para el organismo, o se puede utilizar una estructura homóloga de otro organismo, dependiendo de la semejanza estructural. A los aminoácidos en el bolsillo de unión de la sintetasa que interactúan con la cadena lateral en el análogo se les permite variar en la identidad y la conformación de rotámeros en los subsiguientes cálculos de diseño de proteínas.

Uno de estos protocolos también emplea un método computacional para potenciar las interacciones entre el sustrato y las posiciones de las proteínas. Esto se hace mediante la ampliación de las energías por pares entre el sustrato y los aminoácidos permitidos en las posiciones de diseño a la proteína en los cálculos de energía. En un cálculo de optimización en que se amplían las interacciones proteína-sustrato en comparación con las interacciones entre proteínas, la secuencia de selección se desvía hacia la selección de aquellos aminoácidos que tienen una interacción favorable con el sustrato.

El método descrito ayudó a construir una nueva forma modificada de la fenilalanil-ARNt sintetasa de *E. coli*, basado en la estructura conocida de la PheRS de *Thermus thermophilus* relacionada (tPheRS). La nueva forma modificada de la fenilalanil-ARNt sintetasa de *E. coli* (ePheRS) permite una incorporación eficaz *in vivo* de la funcionalidad reactiva de aril-cetona en proteínas recombinantes. Además, una triptofanil-ARNt sintetasa modificada se modificó de una manera similar y ha demostrado tener la capacidad de incorporar análogos de aminoácidos no naturales en polipéptidos en lugar de triptófano que se produce de forma natural. Los resultados descritos en ese documento también demuestran el poder general del diseño de proteínas computacional en el desarrollo de aminoacil-ARNt sintetasas para la activación y carga de aminoácidos no naturales.

A. Secuencia e Información Estructural Disponibles para ARNt sintetasas

La traducción de proteínas a partir de un molde de ARNm se lleva a cabo por parte de los ribosomas. Durante el proceso de traducción; cada uno de los ARNt se empareja con su aminoácido mucho antes de que llegue al ribosoma. El emparejamiento se realiza mediante una colección de enzimas conocidas como las aminoacil-ARNt sintetasas (AARS). Estas enzimas cargan cada uno de los ARNt con el aminoácido apropiado, permitiendo así que cada uno de los ARNt haga la traducción apropiada del código genético de ADN (y el ARNm transcrito a partir del ADN) en el código de aminoácidos de las proteínas.

La mayoría de las células producen veinte aminoacil-ARNt sintetasas diferentes, una para cada uno de los tipos de aminoácido. Estos veinte enzimas están optimizadas en cada caso para funcionar con su propio aminoácido particular y el conjunto de moléculas de ARNt apropiadas para ese aminoácido. Las aminoacil-ARNt sintetasas deben realizar sus tareas con una alta precisión. Muchas de estas enzimas reconocen sus moléculas de ARNt utilizando el anticodón. Estas enzimas producen alrededor de un error en 10.000. Para la mayoría de los aminoácidos, este nivel de precisión no es demasiado difícil de conseguir, ya que la mayoría de los aminoácidos son bastante diferentes uno de otro.

En el presente método, una descripción precisa del bolsillo de unión de AARS para ARNt es importante para la estrategia de diseño computacional, ya que depende de la estructura cristalina para las descripciones de la cadena principal de la proteína, aunque en muchos casos es perfectamente aceptable utilizar la estructura cristalina de un proteína homóloga (por ejemplo, un homólogo de una especie relacionada) o incluso un dominio conservado para sustituir la descripción de la estructura cristalográfica del bolsillo de unión. La estructura cristalina también define la orientación del sustrato de aminoácido natural en el bolsillo de unión de un sintetasa, así como la posición relativa del sustrato de aminoácido a los residuos sintetasa, especialmente aquellos residuos en y alrededor del bolsillo de unión. Para diseñar el bolsillo de unión para los análogos, se prefiere que estos análogos se unan a la sintetasa en la misma orientación que el sustrato de aminoácido natural, ya que esta orientación puede ser importante para la etapa de adenilación.

Las AARSs pueden ser de cualquier organismo, incluyendo procariotas y eucariotas, siendo posibles las enzimas de bacterias, hongos, extremeófilos tales como las arqueobacterias, gusanos, insectos, peces, anfibios, aves, animales (particularmente mamíferos y especialmente seres humanos) y plantas.

Tal como se describió anteriormente, la mayoría de las células producen veinte diferentes aminoacil-ARNt sintetetasas, una para cada uno de los tipos de aminoácido. Algunos sintetetasas adecuadas son conocidas, incluidas: la fenilalanil-ARNt sintetetasa de levadura (Kwon *et al.*, *J. Am. Chem. Soc.* 125: 7512-7513, 2003); la tirosil-ARNt sintetetasa de *Methanococcus jannaschii* (Wang *et al.*, *Science* 292, 498-500, 2001); y la tirosil-ARNt-sintetasa de levadura (Ohno *et al.*, *J. Biochem.* 130, 417- 423, 2001). De hecho, las estructuras cristalinas de casi todos las 20 diferentes enzimas AARS están actualmente disponibles en el Brookhaven Protein Data Bank (PDB, véase Bernstein *et al.*, *J. Mol. Biol.* 112: 535-542, 1977). Una lista de todas las AARSs con estructuras cristalinas está disponible a partir de abril de 2001 en la página web de PDB. Por ejemplo, la estructura cristalina de la fenilalanil-ARNt sintetetasa de *Thermus Aquaticus* complejada con fenilalanina tiene una resolución de 2,7 Å, y su PDB ID es 1B70.

La base de datos de estructura o la base de datos Molecular Modeling DataBase (MMDB) contiene datos experimentales de determinaciones de la estructura cristalográfica y de RMN. Los datos para MMDB se obtienen del banco de datos Protein Data Bank (PDB). El NCBI (Centro Nacional de Información Biotecnológica) ha cruzado datos estructurales con la información bibliográfica, a las bases de datos de secuencias, y para la taxonomía de NCBI. Cn3D, el visor de estructura 3D de NCBI, se puede utilizar para una visualización interactiva fácil de las estructuras moleculares de Entrez.

La base de datos Entrez Dominios 3D contiene dominios de la proteína a partir de la base de datos de dominio conservado (CDD) de NCBI. Los biólogos computacionales definen dominios conservados sobre la base de patrones o motivos de secuencias recurrentes. La CDD contiene actualmente dominios derivados de dos colecciones populares, Smart y Pfam, además de las contribuciones de colegas en la NCBI, tal como COG. Las bases de datos fuente también proporcionan descripciones y enlaces a las citas. Dado que los dominios conservados corresponden a compactar unidades estructurales, CDs contienen enlaces a la estructura 3D a través Cn3D siempre que sea posible.

Para identificar dominios conservados en una secuencia de proteínas, el servicio CD-Search emplea el algoritmo BLAST específico para la posición inversa. La secuencia de consulta se compara con una matriz de puntuación específica para la posición, preparada a partir de la alineación subyacente del dominio conservado. Las solicitudes pueden mostrarse como una alineación por pares de la secuencia de consultas con una secuencia de dominio representativa o como una alineación múltiple. CD-Search se ejecuta ahora por defecto en paralelo con las búsquedas BLAST de proteínas. Mientras que el usuario espera la cola de BLAST para procesar aún más la petición, la arquitectura del dominio de la consulta ya puede ser estudiada. Además, CDART, la herramienta de recuperación de la arquitectura del dominio conservado, permite al usuario buscar proteínas con arquitecturas de dominio similares. CDART utiliza los resultados de búsqueda de CD-precomputados para identificar rápidamente proteínas con un conjunto de dominios similares a los de la consulta. Para más detalles, véase Marchler-Bauer *et al.*, *Nucleic Acids Research* 31: 383-387, 2003; y Marchler-Bauer *et al.*, *Nucleic Acids Research* 30: 281-283, 2002.

Además, una base de datos de aminoacil-ARNt sintetetasas conocida ha sido publicada por Maciej Szymanski, Marzanna A. Deniziak y Jan Barciszewski, en *Nucleic Acids Res.* 29: 288-290, 2001 (titulado "Base de datos de las aminoacil-ARNt sintetetasas"). Un sitio web correspondiente (rose.man.poznan.pl/aars/seq_main.html) proporciona detalles acerca de todas las AARSs conocidas de diferentes especies. Por ejemplo, de acuerdo con la base de datos, la isoleucil-ARNt sintetetasa para la bacteria *Deinococcus radiodurans* radiorresistente (Nº de Acceso AAF10907) tiene 1078 aminoácidos, y fue publicada por White *et al.*, en *Science* 286: 1571-1577 (1999); la Valil-ARNt sintetetasa para ratón (*Mus musculus*) tiene 1263 aminoácidos (Nº. de Acceso AAD26531), y fue publicada por Snoek M. y van Vugt H: en *Immunogenetics* 49: 468-470 (1999); y están también disponibles las secuencias de fenilalanil-ARNt sintetetasa para seres humanos, *Drosophila*, *S. pombe*, *S. cerevisiae*, *Candida albicans*, *E. coli* y otras numerosas bacterias, incluida *Thermus aquaticus* ssp. *thermophilus*. La base de datos fue actualizada por última vez en noviembre de 2006. Información similar para otras AARSs recientemente identificadas se puede obtener, por ejemplo, realizando una búsqueda BLAST utilizando cualquiera de las secuencias conocidas en la base de datos de AARS como consulta frente a las bases de datos públicas disponibles (tales como la base de datos no redundante en el NCBI, o "nr") o de propiedad privada.

Alternativamente, en determinadas formas de realización, si no se conoce la estructura cristalina exacta de una AARS particular, pero su secuencia de proteínas es similar u homóloga a una secuencia de AARS conocida con una estructura cristalina conocida. En tales casos, se espera que la conformación de las AARS en cuestión sea similar a la estructura cristalina conocida de las AARS homólogas. La estructura conocida puede, por lo tanto, ser

utilizada como la estructura para las AARS de interés o, más preferiblemente, se puede utilizar para predecir la estructura de las AARS de interés (es decir, en "modelado por homología" o "modelado molecular"). Como un ejemplo particular, la base de datos de modelado molecular (MMDB) descrita anteriormente (véase, Wang *et al.*, *Nucl. Acids Res.* 2000, 28: 243-245; Marchler-Bauer *et al.*, *Nucl. Acids Res.* 1999, 27: 240-243) proporciona motores de búsqueda que se pueden utilizar para identificar proteínas y/o ácidos nucleicos que son similares u homólogos a una secuencia de proteínas (a la que se alude como secuencias "vecinas" en la MMDB), incluyendo secuencias vecinas, cuyas estructuras tridimensionales son conocidas. La base de datos proporciona, además, enlaces a las estructuras conocidas junto con herramientas de alineación y visualización, tales como Cn3D (desarrollado por NCBI), RasMol, etc., por lo que las secuencias homólogas y las parentales pueden ser comparadas y se puede obtener una estructura para la secuencia parental basada en tales alineaciones de secuencias y estructuras conocidas.

La secuencia de AARS homóloga con estructura 3D conocida es preferiblemente al menos aproximadamente 60%, o al menos aproximadamente 70%, o al menos aproximadamente 80%, o al menos aproximadamente 90%, o al menos aproximadamente 95% idéntica, o al menos aproximadamente 98% idéntica a las AARS de interés en la región del sitio activo o la región de bolsillo para la unión al sustrato de aminoácido. Un sitio activo o sitio en el bolsillo de este tipo puede no ser continuo en la secuencia primaria de aminoácidos de las AARS, ya que aminoácidos distantes pueden reunirse en la estructura 3D. En este caso, la homología o identidad de la secuencia se puede calcular utilizando, por ejemplo, el programa BLASTp estándar de la NCBI para la proteína utilizando condiciones por defecto, en regiones alineadas juntas (sin inserciones ni deleciones en cualquiera de las dos secuencias que se comparan) e incluyen residuos que se sabe están implicados en la unión de aminoácidos sustrato. Por ejemplo, la subunidad (alfa) catalítica de la fenilalanil-ARNt sintetasa de *Thermus Aquaticus* parece tener una región de "inserto" desde los residuos 156 a 165 en comparación con sus homólogos de otras especies. Esta región puede no tenerse en cuenta en el cálculo de identidad de la secuencia. Alternativamente, la AARS homóloga es preferiblemente aproximadamente 35%, o 40%, o 45%, o 50%, o 55% idéntica en general a las AARS de interés. La subunidad alfa de la fenilalanil-ARNt sintetasa de *E. coli* es aproximadamente 45% idéntica en general, y aproximadamente 80% idéntica en la región del sitio activo a la fenilalanil-ARNt sintetasa de *Thermus Aquaticus*. Las subunidades alfa de fenilalanil-ARNt sintetasa humana son aproximadamente 62%, 60%, 54%, 50% idénticas en general con sus homólogos de *Drosophila*, gusano (*C. elegans*), planta (*Arabidopsis thaliana*), levadura (*S. cerevisiae*) respectivamente.

En los pocos casos en los que la estructura para una secuencia de AARS particular no puede ser conocida o ni estar disponible, es posible determinar la estructura utilizando técnicas experimentales rutinarias (por ejemplo, cristalografía de rayos X y resonancia magnética nuclear (RMN)) y sin una experimentación excesiva. Véase, p. ej., *NMR of Macromolecules: A Practical Approach*, G.C.K. Roberts, Ed., Oxford University Press Inc., Nueva York (1993); Ishima y Torchia, *Nat. Struct. Biol.* 7: 740-743, 2000; Gardner y Kay, *Annu. Rev. Bioph. Biom.* 27: 357-406, 1998; Kay, *Biochem. Cell. Biol.* 75: 1-15, 1997; Dayie *et al.*, *Annu. Rev. Phys. Chem.* 47: 243-282, 1996; Wuthrich, *Acta Crystallogr. D* 51: 249-270, 1995; Kahn *et al.*, *J. Synchrotron Radiat.* 7: 131-138, 2000; Oakley y Wilce, *Clin. Exp. Pharmacol. P.* 27: 145-151, 2000; Fourme *et al.*, *J. Synchrotron Radiat.* 6: 834-844, 1999.

Alternativamente, la estructura tridimensional de una secuencia de AARS puede calcularse a partir de la propia secuencia y utilizando desde el principio técnicas de modelado molecular ya conocidas en la técnica. Véase, p. ej., Smith *et al.*, *J. Comput. Biol.* 4: 217-225, 1997; Eisenhaber *et al.*, *Proteins* 24: 169-179, 1996; Bohm, *Biophys Chem.* 59: 1-32, 1996; Fetrow y Bryant, *BioTechnol.* 11: 479-484, 1993; Swindells y Thorton, *Curr. Opin. Biotech.* 2: 512-519, 1991; Levitt *et al.*, *Annu. Rev. Biochem.* 66: 549-579, 1997; Eisenhaber *et al.*, *Crit. Rev. Biochem. Mol.* 30: 1-94, 1995; Xia *et al.*, *J. Mol. Biol.* 300: 171-185, 2000; Jones, *Curr. Opin. Struc. Biol.* 10: 371-379, 2000. Estructuras tridimensionales obtenidas a partir de modelado desde el principio son típicamente menos fiables que las estructuras obtenidas utilizando técnicas empíricas (p. ej., espectroscopía de RMN o cristalografía de rayos X) o semi-empíricas (p. ej., modelado de homología). Sin embargo, tales estructuras serán generalmente de suficiente calidad, aunque menos preferidas para su uso en algunos de los métodos descritos en esta memoria.

B. Métodos para Predecir la Estructura 3D basados en la Homología de la Secuencia

Para las proteínas AARS que no han sido cristalizadas o que han sido el foco de otras determinaciones estructurales, un modelo molecular generado por ordenador de las AARS y su sitio de unión, se puede generar, sin embargo, utilizando cualquiera de un cierto número de técnicas disponibles en la técnica. Por ejemplo, las posiciones del carbono C α de la secuencia AARS diana se pueden representar en mapa en un determinado patrón de coordenadas de una enzima AARS ("AARS conocida") que tiene una secuencia similar y una estructura deducida utilizando técnicas de modelado de homología, y la estructura de la proteína diana y las velocidades de cada uno de los átomos calculada a una temperatura de simulación (To) a la que se ha de determinar una simulación de acoplamiento con un análogo de aminoácido. Típicamente, un protocolo de este tipo implica principalmente la predicción de las conformaciones de cadena lateral en la proteína AARS diana modelada, al

tiempo que suponiendo una traza de la cadena principal tomada de una estructura terciaria, tal como se proporciona por la proteína AARS conocida. Los programas de ordenador para llevar a cabo rutinas de minimización de la energía se utilizan comúnmente para generar modelos moleculares. Por ejemplo, tanto el algoritmo CHARMM (Brooks et al. (1983) *J Comput Chem* 4: 187-217) como AMBER (Weiner et al. (1981) *J. Comput. Chem.* 106: 765) manejan toda la configuración del sistema molecular, el cálculo del campo de fuerza y el análisis (véase también, Eisenfield et al. (1991) *Am J Physiol* 261: C376-386; Lybrand (1991) *J Pharm Belg* 46: 49-54; Froimowitz (1990) *Biotechniques* 8: 640-644; Burbarn et al. (1990) *Proteins* 7: 99-111; Pedersen (1985) *Environ Health Perspect.* 61: 185-190; y Kini et al. (1991) *J Biomol Struct Dyn* 9: 475-488). En el corazón de estos programas se encuentra un conjunto de subrutinas que, dada la posición de cada uno de los átomos en el modelo, calculan la energía potencial total del sistema y la fuerza en cada átomo. Estos programas pueden utilizar un conjunto de partida de las coordenadas atómicas, los parámetros para los diversos términos de la función de energía potencial, y una descripción de la topología molecular (la estructura covalente). Características comunes de este tipo de métodos de modelado molecular incluyen: provisiones para la manipulación de enlaces hidrógeno y otras fuerzas de restricción; el uso de condiciones de contorno periódicas; y provisiones para ajustar ocasionalmente las posiciones, velocidades, u otros parámetros con el fin de mantener o cambiar la temperatura, la presión, el volumen, las fuerzas de restricción, u otras condiciones controladas externamente.

La mayoría de los métodos de minimización de energía convencionales utilizan los datos de coordenadas entrada y el hecho de que la función de energía potencial es una función explícita, diferenciable de coordenadas cartesianas, para calcular la energía potencial y su gradiente (que da la fuerza sobre cada uno de los átomos) para cualquier conjunto de posiciones atómicas. Esta información se puede utilizar para generar un nuevo conjunto de coordenadas en un esfuerzo por reducir la energía potencial total y, mediante la repetición de este proceso una y otra vez, para optimizar la estructura molecular bajo un conjunto dado de condiciones externas. Estos métodos de minimización de la energía se aplican rutinariamente a moléculas similares a las proteínas AARS objeto.

En general, los métodos de minimización de la energía pueden llevarse a cabo para una temperatura dada, T_i , que puede ser diferente de la temperatura de simulación de acoplamiento, T_o . Tras la minimización de la energía de la molécula en T_i , se computan las coordenadas y velocidades de todos los átomos en el sistema. Adicionalmente, se calculan los modos normales del sistema. Se apreciará por los expertos en la técnica que cada uno de los modos normales es un movimiento periódico colectivo, con todas las partes del sistema en movimiento en fase entre sí, y que el movimiento de la molécula es la superposición de todos los modos normales. Para una temperatura dada, la amplitud media cuadrada de movimiento en un modo particular es inversamente proporcional a la constante de fuerza efectiva para ese modo. A este respecto, las vibraciones de baja frecuencia dominarán a menudo el movimiento de la molécula.

Después de que el modelo molecular ha sido minimizado en energía en T_i , el sistema se "calienta" o "enfía" a la temperatura de simulación, T_o , llevando a cabo una operación de equilibrio, en que las velocidades de los átomos aumentan de una manera escalonada hasta alcanzar la temperatura deseada, T_o . El sistema se equilibra aún más durante un período de tiempo especificado hasta que permanecen constantes determinadas propiedades del sistema tales como la energía cinética media. Las coordenadas y velocidades de cada uno de los átomos se obtienen entonces a partir del sistema equilibrado.

También pueden llevarse a cabo rutinas de minimización de la energía adicionales. Por ejemplo, una segunda clase de métodos implica calcular soluciones aproximadas a la EOM restringida para la proteína. Estos métodos utilizan una estrategia iterativa para resolver los multiplicadores de Lagrange y, típicamente, sólo necesitan unas pocas iteraciones si las correcciones necesarias son pequeñas. El método más popular de este tipo, SHAKE (Ryckaert et al. (1977) *J. Comput. Phys.* 23: 327; y Van Gunsteren et al. (1977) *Mol. Phys.* 34: 1311), es fácil de implementar y aumenta como $O(N)$ a medida que se incrementa el número de restricciones. Por lo tanto, el método es aplicable a macromoléculas tales como proteínas AARS. Un método alternativo, RATTLE (Anderson (1983) *J. Comput. Phys.* 52:24) se basa en la versión de velocidad del algoritmo de Verlet. Al igual que SHAKE, RATTLE es un algoritmo iterativo y se puede utilizar para minimizar en energía el modelo de una proteína AARS objeto.

C. Métodos Alternativos

En otras realizaciones, en lugar de mantener la identidad del análogo de aminoácido constante y variando la estructura de AARS (modelando varias estructuras mutantes diferentes), el presente método se lleva a cabo utilizando el o los modelos moleculares para una AARS modificada sencilla (p. ej., en el que se cambian uno más residuos de aminoácidos de no anclaje) y el muestreo de una diversidad de diferentes análogos de aminoácidos o fragmentos potenciales de los mismos, para identificar análogos que son propensos a interactuar con, y ser sustratos para la enzima AARS modificada. Esta estrategia puede hacer uso de coordinar bancos para análogos de aminoácidos (incluidas variantes de rotámeros) o bancos de los grupos funcionales y espaciadores que se

pueden unir para formar la cadena lateral de un análogo de aminoácido.

Al utilizar tales estrategias según se describe anteriormente, p. ej., el modelado de homología, se puede derivar un conjunto de coordenadas para el sitio de unión para las AARS modificadas.

Existe una diversidad de métodos computacionales que se pueden adaptar fácilmente para identificar la estructura de análogos de aminoácidos que tendrían propiedades estéricas y electrónicas apropiadas para interactuar con el sitio de unión al sustrato de una AARS modificada. Véase, por ejemplo, Cohen *et al.* (1990) *J. Med. Cam.* 33: 883-894; Kuntz *et al.* (1982) *J. Mol. Biol.* 161: 269-288; DesJarlais (1988) *J. Med. Cam.* 31: 722-729; Bartlett *et al.* (1989) (*Spec. Publ., Roy. Soc. Chem.*) 78: 182-196; Goodford *et al.* (1985) *J. Med. Cam.* 29: 849-857; DesJarlais *et al.* *J. Med. Cam.* 29: 2149-2153). Métodos dirigidos generalmente se dividen en dos categorías: (1) el diseño por analogía, en el que las estructuras 3-D de las moléculas conocidas (tales como las de una base de datos cristalográfica) se acoplan a la estructura del sitio de unión AARS y se puntúan en cuanto a la bondad del ajuste; y (2) diseño de novo, en el que el modelo análogo de aminoácido se construye por piezas en el sitio de unión de AARS. El último enfoque, en particular, puede facilitar el desarrollo de nuevas moléculas, de diseño único, para unirse al sitio de unión de AARS modificada objeto.

En una forma de realización ilustrativa, el diseño de análogos de aminoácidos potenciales que pueden funcionar con una AARS modificada particular comienza desde la perspectiva general de la complementariedad de forma para el sitio de unión al sustrato de la enzima, y se emplea un algoritmo de búsqueda que es capaz de rastrear una base de datos de pequeñas moléculas de conocida estructura tridimensional para los candidatos que se ajustan geométricamente en el sitio de unión al sustrato. Tales bancos pueden ser bancos de pequeñas moléculas generales, o pueden ser bancos dirigidos a análogos de aminoácidos o moléculas pequeñas que pueden utilizarse para crear análogos de aminoácidos. No se espera que las moléculas que se encuentran en la búsqueda de forma sean necesariamente conductoras por sí mismas, ya que no se ha de hacer necesariamente una evaluación de la interacción química durante la búsqueda inicial. Más bien, se prevé que tales candidatos puedan actuar como el marco para un diseño adicional, proporcionando esqueletos moleculares a los cuales se pueden sustituciones atómicas apropiadas. Por supuesto, se puede evaluar la complementariedad química de estas moléculas, pero se espera que cambiarán los tipos de átomos para maximizar la electrostática, los enlaces hidrógeno y las interacciones hidrofóbicas con el sitio de unión al sustrato. La mayoría de los algoritmos de este tipo proporcionan un método para encontrar una amplia variedad de estructuras químicas que pueden ser complementarias a la forma del sitio de unión al sustrato de AARS.

Por ejemplo, cada uno de un conjunto de pequeñas moléculas de una base de datos particular, tal como el banco de datos Cambridge Crystallographic Data Bank (CCDB) (Allen *et al.* (1973) *J. Chem. Doc.* 13: 119), está acoplado de forma individual al sitio de unión de las AARS modificadas en un cierto número de orientaciones geométricamente permisibles con el uso de un algoritmo de acoplamiento. En una realización preferida, se puede utilizar un conjunto de algoritmos de ordenador denominado DOCK para caracterizar la forma de invaginaciones y ranuras que forman el sitio de unión. Véase, por ejemplo, Kuntz *et al.* (1982) *J. Mol. Biol.* 161: 269-288. El programa también puede buscar una base de datos de pequeñas moléculas para los moldes, cuyas formas son complementarias a un sitio de unión particular, de las AARS modificadas. Algoritmos ilustrativos que se pueden adaptar para este propósito se describen, por ejemplo, en DesJarlais *et al.* (1988) *J. Med. Chem.* 31: 722-729.

Las orientaciones son evaluadas en cuanto a la bondad de ajuste y las mejores se mantienen para un examen ulterior utilizando programas de mecánica molecular tal como AMBER o CHARMM. Tales algoritmos han demostrado previamente tener éxito en encontrar una diversidad de moléculas que son complementarias en forma a un determinado sitio de unión de un receptor o enzima, y han demostrado tener varias características atractivas. En primer lugar, este tipo de algoritmos puede recuperar una notable diversidad de arquitecturas moleculares. En segundo lugar, las mejores estructuras han demostrado, en aplicaciones previas a otras proteínas, una impresionante complementariedad de la forma con una superficie específica extendida. En tercer lugar, la estrategia general parece ser bastante más robusta con respecto a las pequeñas incertidumbres en el posicionamiento de los átomos candidatos.

En determinadas formas de realización, el presente método puede utilizar un algoritmo descrito por Goodford (1985, *J. Med. Chem.* 28:849-857) y Boobbyer *et al.* (1989, *J. Med. Chem.* 32: 1083-1094). Estos artículos describen un programa de ordenador (GRID), que busca determinar las regiones de alta afinidad por diferentes grupos químicos (denominados sondas) en la superficie molecular del sitio de unión. Por lo tanto, GRID proporciona una herramienta para sugerir modificaciones a ligandos conocidos que pudieran potenciar la unión. Se puede anticipar que algunos de los sitios discernidos por GRID como regiones de alta afinidad corresponden a "patrones farmacofóricos" determinado por inferencia a partir de una serie de ligandos conocidos. Tal como se utiliza en esta memoria, un patrón farmacofórico es una disposición geométrica de características del análogo de aminoácido anticipado que se cree que es importante para la unión. Goodsell y Olson (1990, *Proteins: Struct.*

Funct. Genet. 8: 195-202) han utilizado el algoritmo Metropolis (reasociación simulada) para acoplar un solo ligando conocido en una proteína diana, y su estrategia se puede adaptar para la identificación de análogos de aminoácidos adecuados para el acoplamiento con el sitio de unión de AARS. Este algoritmo puede permitir la flexibilidad a la torsión en la cadena lateral de aminoácidos y el uso de mapas de energía de interacción GRID como tablas de búsqueda rápida para computar energías de interacción aproximadas.

Sin embargo, una forma de realización adicional utiliza un algoritmo de ordenador tal como CLIX que busca en las bases de datos tales como CCDB pequeñas moléculas que puedan ser orientadas en el sitio de unión al sustrato de la AARS de una manera que sea tanto estéricamente aceptable como que tenga una alta probabilidad de lograr interacciones químicas favorables entre la molécula candidata y los residuos de aminoácidos que la rodean. El método se basa en la caracterización del sitio de unión al sustrato en términos de un conjunto de posiciones de unión favorables para diferentes grupos químicos y luego la búsqueda de orientaciones de las moléculas candidatas que provocan una máxima coincidencia espacial de grupos químicos candidatos individuales con los miembros del conjunto. La disponibilidad actual de energía de ordenador dicta que una búsqueda basada en ordenador para nuevos ligandos siga una estrategia "de lo ancho". Una estrategia "de lo ancho" tiene como objetivo reducir progresivamente el tamaño del espacio de búsqueda potencial candidato por la aplicación de criterios cada vez más estrictos, en contraposición a una estrategia de profundidad en el que se lleva a cabo un análisis máximamente detallado de un candidato antes de proceder a la siguiente. CLIX se ajusta a esta estrategia en que su análisis de unión es rudimentario - busca satisfacer las condiciones necesarias de ajuste estérico y de tener grupos individuales en lugares "correctos" para la unión, sin imponer la condición suficiente de que interacciones de unión favorables ocurran en realidad. Se produce una "lista restringida" clasificada de moléculas, en sus orientaciones favorecidas, que puede ser examinada sobre una base de molécula por molécula, usando gráficos por ordenador y técnicas de modelado molecular más sofisticadas. CLIX también es capaz de sugerir cambios en los grupos químicos sustituyentes de las moléculas candidatas que pudieran potenciar la unión. De nuevo, el banco de partida puede ser de análogos de aminoácidos o de moléculas que pueden ser utilizadas para generar la cadena lateral de un análogo de aminoácido.

Los detalles algorítmicos de CLIX se describen en Lawrence *et al.*, (1992) *Proteins* 12: 31-41, y el algoritmo CLIX pueden resumirse como sigue. El programa GRID se utiliza para determinar posiciones discretas favorables de interacción (denominadas sitios diana) en el sitio de unión de la proteína AARS para una amplia diversidad de grupos químicos representativos. Para cada uno de los ligandos candidatos en el CCDB se hace un intento exhaustivo para hacer coincidir, en un sentido espacial en el sitio de unión de la proteína, un par de grupos químicos sustituyentes del candidato con un par de sitios de interacción favorables correspondientes propuestos por GRID. Durante este proceso se consideran todas las posibles combinaciones de pares de grupos de ligandos con pares de sitios GRID. Tras localizar tal coincidencia, el programa hace girar el ligando candidato alrededor de los dos pares de grupos y verifica el impedimento estérico y la coincidencia de otros grupos atómicos candidatos con sitios diana apropiados. Se conservan combinaciones candidato / orientación particulares que son buenos ajustes geométricos en el sitio de unión y muestran una coincidencia suficiente de grupos atómicos con sitios GRID.

En consonancia con la estrategia "de lo ancho", este enfoque implica simplificar suposiciones. La proteína rígida y la geometría de moléculas pequeñas se mantienen a lo largo del mismo. Como primera aproximación la geometría rígida es aceptable como las coordenadas minimizadas en energía del sitio de unión de las AARS modificadas, describen un mínimo de energía para la molécula, aunque uno local.

Otro supuesto implícito en CLIX es que el ligando potencial, cuando se introduce en el sitio de unión al sustrato de las AARS modificadas, no induce cambios en la estereoquímica o la distribución de la carga parcial de la proteína y, así, altera la base sobre la cual se computaron los mapas de energía de interacción GRID. También hay que destacar que los sitios de interacción predichos por GRID se utilizan en un sentido posicional y del tipo solamente, es decir, cuando un grupo atómico candidato se coloca en un sitio predicho tan favorable por GRID, no se hace verificación alguna para asegurarse de que la geometría de unión, el estado de protonación o la distribución de carga parcial favorece una fuerte interacción entre la proteína y ese grupo. Tal análisis detallado debe formar parte de más de un modelado avanzado de candidatos identificados en la lista restringida CLIX.

Todavía otra forma de realización de un método de diseño molecular asistido por ordenador para la identificación de análogos de aminoácidos que pueden ser utilizados por una AARS predeterminada modificada comprende la síntesis *de novo* de inhibidores potenciales por la conexión algorítmica de pequeños fragmentos moleculares que exhibirá la complementariedad estructural y electrostática deseada con el sitio de unión al sustrato de la enzima. La metodología emplea a un gran conjunto de moldes de pequeñas moléculas que es iterativamente reconstruido en un modelo del sitio de unión al sustrato de la AARS'. Cada una de las etapas de crecimiento del ligando se evalúa de acuerdo con una función de la energía basada en la mecánica molecular, que considera interacciones de van der Waals y de Coulomb, la energía de deformación interna del ligando de alargamiento y la desolvatación

tanto del ligando como de la enzima. El espacio de búsqueda puede ser gestionado mediante el uso de un árbol de datos que se mantiene bajo control mediante la poda de acuerdo con los criterios de unión.

Aún en otra forma de realización, análogos de aminoácidos potenciales pueden determinarse utilizando un método basado en un algoritmo de dinámica molecular restringida a la minimización de energía para la determinación de posiciones energéticamente favorables de grupos funcionales en el sitio de unión al sustrato de una enzima AARS modificada. El método puede ayudar al diseño de moléculas que incorporan tales grupos funcionales mediante modificación de aminoácidos y análogos de aminoácidos conocidos o mediante la síntesis *de novo*.

Por ejemplo, el método de búsqueda simultánea de múltiples copias (MCSS) descrito por Miranker *et al.* (1991) *Proteins* 11: 29-34 se puede adaptar para su uso en el método objeto. Para determinar y caracterizar mínimos locales de un grupo funcional en el campo de fuerza de la proteína, múltiples copias de grupos funcionales seleccionados se distribuyen primero en un sitio de unión de interés en la proteína AARS. La minimización de la energía de estas copias mediante mecánica molecular o dinámicas restringidas proporciona los mínimos locales distintos. La vecindad de estos mínimos puede ser explorada por una búsqueda grid o por la minimización restringida. En una forma de realización, el método MCSS utiliza la aproximación de Hartree dependiente del tiempo (TDH) clásica para minimizar simultáneamente o apagar muchos grupos idénticos en el campo de fuerza de la proteína.

La implementación del algoritmo MCSS requiere una elección de grupos funcionales y un modelo de mecánica molecular para cada uno de ellos. Los grupos deben ser lo suficientemente simples como para ser caracterizados y manipulados (3-6 átomos, pocos o ningún grados de libertad de diedros) con facilidad, pero lo suficientemente complejos como para aproximarse a las interacciones estéricas y electrostáticas que el grupo funcional tendría en la unión del sustrato al sitio de la proteína AARS. Un conjunto preferido es, por ejemplo, uno en el que la mayoría de las moléculas orgánicas pueden ser descritas como una colección de tales grupos (*Patai's Guide to the Chemistry of Functional Groups*, ed. S. Patai (Nueva York: John Wiley and Sons, (1989)). Esto incluye fragmentos tales como acetonitrilo, metanol, acetato, metil-amonio, dimetil-éter, metano y acetaldehído.

La determinación de los mínimos de energía local en el sitio de unión requiere que se tomen muestras de muchas posiciones de partida. Esto se puede lograr mediante la distribución de, por ejemplo, 1.000-5.000 grupos al azar dentro de una esfera centrada en el sitio de unión; sólo el espacio no ocupado por la proteína necesita ser considerado. Si la energía de interacción de un grupo particular en un lugar determinado con la proteína es más positivo que un punto de corte dado (p. ej., 5,0 kcal/mol), el grupo se descarta de ese sitio. Teniendo en cuenta el conjunto de las posiciones de partida, todos los fragmentos se minimizan al mismo tiempo mediante el uso de la aproximación TDH (Elber *et al.* (1990) *J. Am. Chem. Soc.* 112: 9161-9175). En este método, las fuerzas sobre cada uno de los fragmentos consiste en sus fuerzas internas y las debidas a la proteína. El elemento esencial de este método es que se omiten las interacciones entre los fragmentos y las fuerzas sobre la proteína se normalizan a las debidas a un solo fragmento. De esta manera, se puede realizar la minimización simultánea o dinámica de cualquier número de grupos funcionales en el campo de una sola proteína

La minimización se realiza sucesivamente en subconjuntos de, p. ej., 100, de los grupos colocadas al azar. Después de un cierto número de intervalos de paso, tales como 1.000 intervalos, los resultados pueden ser examinados para eliminar los grupos que converjan hacia el mismo mínimo. Este proceso se repite hasta que se haya completado la minimización (p. ej., gradiente de RMS de 0,01 kcal/mol/Å). Así, el conjunto de moléculas minimizado en energía resultante comprende lo que equivale a un conjunto de fragmentos inconexos en tres dimensiones que representan posibles cadenas laterales para análogos de aminoácidos.

La siguiente etapa es, entonces, conectar las piezas con espaciadores montados a partir de pequeñas entidades químicas (átomos, cadenas o restos de anillo) para formar análogos de aminoácidos, p. ej., cada uno de los desconectados puede ser vinculado en el espacio para generar una sola molécula utilizando programas de ordenador tales como, por ejemplo, NEWLEAD (Tschinke *et al.* (1993) *J. Med. Chem.* 36: 3863,3870). El procedimiento adoptado por NEWLEAD ejecuta la siguiente secuencia de comandos (1) conectar dos restos aislados, (2) mantener las soluciones intermedias para su posterior procesamiento, (3) repetir las etapas anteriores para cada una de las soluciones intermedias hasta que no se encuentren unidades desconectadas, y (4) salida de las soluciones finales, cada uno de los cuales es una molécula sencilla. Dicho programa puede utilizar, por ejemplo, tres tipos de separadores: separadores de bancos, espaciadores de un solo átomo y separadores de anillo condensado. Los espaciadores de bancos son estructuras optimizadas de moléculas pequeñas tales como etileno, benceno y metilamida. La salida producida por programas tales como NEWLEAD consiste en un conjunto de moléculas que contienen los fragmentos originales ahora conectados por espaciadores. Los átomos que pertenecen a los fragmentos de entrada mantienen sus orientaciones originales en el espacio. Las moléculas son químicamente plausibles porque el simple maquillaje de los espaciadores y grupos funcionales, y energéticamente aceptables debido al rechazo de soluciones con violaciones de radios de van-der Waals.

Además, el orden en que se realizan las etapas de este método es puramente ilustrativo por naturaleza. De hecho, las etapas se pueden realizar en cualquier orden o en paralelo, a menos que se indique lo contrario por la presente descripción.

5 Además de ello, los métodos descritos en esta memoria se pueden realizar en cualquier hardware, software, o cualquier combinación de los mismos, ya que estos términos son conocidos actualmente en la técnica. En particular, el presente método puede llevarse a cabo mediante software, firmware o microcódigo que opera en un ordenador u ordenadores de cualquier tipo. Adicionalmente, el software puede comprender instrucciones de
10 ordenador de cualquier forma (p. ej., código fuente, código objeto, código interpretado, etc.) almacenados en cualquier medio legible por ordenador (p. ej., ROM, RAM, medios magnéticos, cinta o tarjeta perforada, disco compacto (CD) en cualquier forma, DVD, etc.). Además, un software de este tipo también puede ser en forma de una señal de datos de ordenador incorporada en una onda portadora, tal como la que se encuentra dentro de las páginas Web conocidas transferidas entre dispositivos conectados a Internet. En consecuencia, determinadas
15 formas de realización no se limitan a cualquier plataforma en particular, salvo que se indique específicamente lo contrario en la presente descripción.

Medios de hardware informático ilustrativos, adecuados para llevar a cabo determinadas formas de realización puede ser un servidor de Silicon Graphics Power Challenge con 10 R10000 procesadores funcionando en paralelo.
20 Un entorno de desarrollo de software adecuado incluye CERIUS2 de Biosym/Molecular Simulations (San Diego, CA), u otros equivalentes.

El método computacional descrito anteriormente se ha utilizado eficazmente en la modificación de enzimas de la maquinaria de síntesis de proteínas (p. ej., AARS) para permitir la incorporación de aminoácidos antinaturales. El
25 mismo conjunto de herramientas computacionales también se puede aprovechar para diseñar los productos finales (p. ej., anticuerpos monoclonales u otros agentes terapéuticos) en que los aminoácidos antinaturales se incorporarían a fin de potenciar o modificar sus propiedades estructurales o funcionales.

Aunque se han mostrado y descrito realizaciones particulares en esta memoria, resultará evidente para los expertos en la técnica que pueden hacerse cambios y modificaciones sin apartarse del aspecto más amplio y, por lo tanto, las reivindicaciones adjuntas deben abarcar dentro de su alcance todos esos cambios y modificaciones que caigan dentro del verdadero espíritu de esta invención.

Adopción de AARS de diferentes organismos

35 Una segunda estrategia para generar un ARNt mutante externo, una RS modificada o mutante externa, o un par ARNt/RS modificado implica importar un ARNt y/o sintetasa de otro organismo en el sistema de traducción de interés tal como *Escherichia coli*. En este ejemplo particular, las propiedades del candidato sintetasa heteróloga incluyen, p. ej., que no carga ARNt de *Escherichia coli* razonablemente bien (preferiblemente no en todos), y las
40 propiedades del candidato de ARNt heterólogo incluyen, p. ej., que no sea acilado por sintetasa de *Escherichia coli* en una medida razonable (preferiblemente no en todas).

Schimmel *et al.* informó que GlnRS de *Escherichia coli* (EcGlnRS) no acila ARNtGln de *Saccharomyces cerevisiae* (EcGlnRS carece de un dominio de unión a ARN N-terminal que posee GlnRS de *Saccharomyces cerevisiae* (ScGlnRS)). Véase, E. F. Whelihan y P. Schimmel, *EMBO J.*, 16:2968 (1997). Por ejemplo, se analizó el ARNtGln supresor de ámbar de *Saccharomyces cerevisiae* (ScARNtGlnCUA) para determinar si también no es un sustrato para EcGlnRS. Ensayos de aminoacilación *in vitro* demostraron ser este el caso; y estudios de supresión *in vitro* demuestran que el ScARNtGlnCUA es competente en la traducción. Véase, p. ej., Liu y Schultz, *PNAS. U S A*, 96:4780 (1999). Se demostró, además, que ScGlnRS no acila cualquier ARNt de *Escherichia coli*, sólo el
50 ScARNtGlnCUA *in vitro*. El grado al que es capaz ScGlnRS de aminoacilar el ScARNtGlnCUA en *Escherichia coli* también se evaluó utilizando un ensayo de complementación *in vivo*. Una mutación sin sentido ámbar se introdujo en un sitio permisivo en el gen β -lactamasa. La supresión de la mutación por parte de un ARNt supresor ámbar debería producir β -lactamasa de longitud completa y conferir resistencia a la ampicilina a la célula. Cuando sólo se expresa ScARNtGlnCUA, las células presentan una Cl_{50} de 20 μ g/mL de ampicilina, lo que indica virtualmente
55 ninguna acilación por parte de sintetisas endógenas de *Escherichia coli*; cuando se coexpresa ScARNtGlnCUA con ScGlnRS, las células adquieren una Cl_{50} de aproximadamente 500 μ g/mL de ampicilina, demostrando que ScGlnRS acila ScARNtGlnCUA eficazmente en *Escherichia coli*. Véase, Liu y Schultz, *PNAS, U S A*, 96: 4780 (1999).

60 Como otro ejemplo, se sabe que ARNt^{Asp} de *Saccharomyces cerevisiae* es un mutante externo a sintetisas de *Escherichia coli*. Véase, p. ej., Doctor y Mudd, *J. Biol. Chem.*, 238:3677 (1963); y Kwok y Wong, *Can. J. Biochem.*, 58: 213 (1980). Se demostró que un ARNt supresor ámbar derivado del mismo (ScARNt^{Asp}_{CUA}) es también un

mutante externo en *Escherichia coli* utilizando el ensayo de β -lactamasa *in vivo* descrito anteriormente. Sin embargo, el anticodón de ARNt^{Asp} es un elemento de reconocimiento crítico de AspRS, véase, p. ej., Giege, *et al.*, *Biochimie*, 78: 605 (1996), y la mutación del anticodón a CUA resulta en una pérdida de afinidad del supresor para AspRS. Un mutante de *Escherichia coli* AspRS E93K ha demostrado reconocer ARNt^{Asp}_{CUA} supresor ámbar de *Escherichia coli* aproximadamente un orden de magnitud mejor que AspRS de tipo salvaje (wt). Véase, p. ej. Martin, 'Thesis', Universidad Louis Pasteur de Estrasburgo, Francia, 1995. Se especuló que la introducción de la mutación relacionada en AspRS(E188K) de *Saccharomyces cerevisiae* podría restaurar su afinidad por SeARNt^{Asp}_{CUA}. Se determinó que el mutante AspRS(E188K) de *Saccharomyces cerevisiae* no acila ARNts de *Escherichia coli*, pero carga ScARNt^{Asp}_{CUA} con una eficacia moderada tal como se muestra por los experimentos de aminoacilación *in vitro*. Véase, p. ej., Pasternak, *et al.*, *Helv. Chim. Acta*, 83: 2277 (2000).

Una estrategia similar implica el uso de una sintetasa heteróloga como la sintetasa mutante externa y un ARNt iniciador mutante del mismo organismo o de un organismo relacionado como el ARNt modificado. RajBhandary y colaboradores encontraron que un mutante ámbar de ARNt^{Met} iniciador humano es acilado por GlnRS de *Escherichia coli* y actúa como un supresor ámbar en células de levadura sólo cuando EcGlnRS se co-expresa. Véase, Kowal, *et al.*, *PNAS U S A*, 98:2268 (2001). Así, este par representa un par mutante externo para su uso en levaduras. Además, se encontró que un ARNt^{Met} iniciador de *Escherichia coli* mutante ámbar es inactivo hacia cualquier sintetasa de *Escherichia coli*. Se seleccionó un TyrRS de levadura mutante que carga este ARNt mutante, resultando en un par mutante externo en *Escherichia coli*.

Utilizando los métodos descritos en esta memoria, se desarrollan los pares y componentes de pares deseados arriba para generar ARNt y/o RS mutante externo que poseen características deseadas, p. ej., que preferiblemente pueden aminoacilar un O-ARNt con un aminoácido antinatural.

En determinadas formas de realización, el ARNt modificado y la RS modificada pueden derivarse por mutación de un ARNt que se produce de forma natural y RS a partir de una diversidad de organismos. En una forma de realización, el ARNt modificado y/o la RS modificada se derivan de al menos un organismo, en que el organismo es un organismo procariótico, p. ej., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*, *P. furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, o similares. Opcionalmente, el organismo es un organismo eucariota, p. ej. plantas (p. ej., plantas complejas tales como monocotiledóneas o dicotiledóneas.), algas, hongos (p. ej. levaduras, etc.), animales (p. ej., mamíferos, insectos, artrópodos, etc.), insectos, protistas, o similares. Opcionalmente, el ARNt modificado se deriva por mutación de un ARNt que se produce de forma natural a partir de un primer organismo y la RS modificada se deriva por mutación de una RS que se produce de forma natural origen natural de un segundo organismo. En una forma de realización, el ARNt modificado y la RS modificada se pueden derivar de un ARNt mutado y una RS mutada. En determinadas formas de realización, la RS modificada y/o el ARNt modificado de un primer organismo se proporciona a un sistema de traducción de un segundo organismo, que opcionalmente tiene una RS y/o un ARNt endógeno no funcional con respecto a los codones reconocidos por el ARNt modificado o la RS modificada.

El ARNt mutante externo y/o la ARNt sintetasa mutante externa también pueden aislarse, opcionalmente, a partir de una diversidad de organismos. En una forma de realización, el ARNt mutante externo y/o la sintetasa mutante externa se aíslan a partir de al menos un organismo, en que el organismo es un organismo procariótico, p. ej., *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*, *P. furiosus*, *P. horikoshii*, *A. Pernix*, *T. thermophilus*, o similares. Opcionalmente, el organismo es un organismo eucariota, p. ej., plantas (por ejemplo, plantas complejas tales como monocotiledóneas o dicotiledóneas.), algas, hongos (p. ej., levaduras, etc.), animales (p. ej., mamíferos, insectos, artrópodos, etc.), insectos, protistas, o similares. Opcionalmente, el ARNt externo se aísla a partir de un ARNt que se produce de forma natural a partir de un primer organismo y la sintetasa mutante externa se aísla a partir de una RS que se produce de forma natural a partir de un segundo organismo. En una forma de realización, el ARNt mutante externo y/o la ARNt sintetasa mutante externa se pueden aislar a partir de uno o más bancos (que comprende opcionalmente uno o más ARNt y/o RS de uno o más organismos (incluidos los que comprenden procariotas y/o eucariotas).

También se describen en esta memoria métodos para la selección de un par ARNt mutante externo y/o una ARNt sintetasa para su uso en cualquier sistema de traducción. Los métodos incluyen: introducir un gen marcador, un ARNt y/o una aminoacil-ARNt sintetasa (RS) aislados o derivados de un primer organismo en un primer conjunto de células del segundo organismo; introducir el gen marcador y el ARNt o la RS en un conjunto de células duplicadas procedente del segundo organismo; y seleccionar para que sobrevivan las células en el primer conjunto que no sobreviven en el conjunto de células por duplicado, en que el primer conjunto y el conjunto de células por duplicado se hacen crecer en presencia de un agente de selección, y en que las células supervivientes comprenden el ARNt mutante externo y/o la RS para uso en el sistema de traducción. En una forma de realización, comparar y seleccionar incluye un ensayo de complementación *in vivo*. En otra forma de realización, se varía la concentración del agente de selección. El mismo ensayo puede realizarse también en un sistema *in vitro* o *in vivo* basado en el

segundo organismo.

Generación de AARS por mutagénesis y Selección / Rastreo

La mutación o modificación de una AARS a utilizar para la incorporación de un aminoácido no natural en un polipéptido o una proteína diana se puede realizar mediante el uso de mutagénesis dirigida una vez que se han identificado los residuos de aminoácidos de contacto deseados. La identificación de los aminoácidos de contacto se puede realizar utilizando cualquier método que permita el análisis de la estructura de la AARS, incluyendo el análisis cristalográfico, el modelado por ordenador, la resonancia magnética nuclear (RMN), el rastreo de bancos, o una combinación de cualquiera de estos u otros métodos.

Se ha secuenciado un cierto número de moléculas de AARS, y proporcionan una orientación en cuanto a qué aminoácidos son importantes para la unión del aminoácido con el que cargar el ARNt correspondiente. Véanse, por ejemplo, SEQ ID Nos. 48-103.

En determinadas formas de realización, las AARS capaz de cargar un ARNt mutante externo particular con un aminoácido antinatural particular, se pueden obtener por mutagénesis de la AARS para generar un banco de candidatos, seguido por el rastreo y/o la selección de las AARSs candidatas capaces de su deseada función. AARSs mutantes externas y ARNts mutantes externos de este tipo se pueden usar para la producción *in vitro* / *in vivo* de las proteínas deseadas con aminoácidos antinaturales modificados.

Así, se proporcionan en determinadas formas de realización descritas en esta memoria métodos para la generación de componentes de la maquinaria biosintética de proteínas tales como las RSs mutantes externas, ARNts mutante externos y/o pares de ARNt/RS mutantes externos que se pueden utilizar para incorporar un aminoácido antinatural.

En una forma de realización, métodos para producir al menos una aminoacil-ARNt sintetasa mutante externa recombinante comprenden: (a) generar un banco de RSs (opcionalmente mutantes) derivado de al menos una aminoacil-ARNt sintetasa (RS) procedente de un primer organismo, p. ej., un organismo eucariota (tal como una levadura), o un organismo procariota tal como *Methanococcus jannaschii*, *Methanobacterium thermoautotrophicum*, *Halobacterium*, *Escherichia coli*, *A. fulgidus*, *P. furiosus*, *P. horikoshii*, *A. pernix*, *T. thermophilus*, o similares; (b) seleccionar (y/o rastrear) el banco de RSs (opcionalmente RSs mutantes) para miembros que aminoacilan un ARNt mutante externo en presencia de un aminoácido antinatural y un aminoácido natural, proporcionando de ese modo una agrupación de RSs activas (opcionalmente mutantes); y/o, (c) seleccionar (opcionalmente a través de selección negativa) la agrupación para RSs activas (p. ej. RSs mutantes) que preferiblemente aminoacilan el O-ARNt en ausencia del aminoácido antinatural, proporcionando con ello la al menos una sintetasa mutante externa recombinante, en donde la al menos una sintetasa mutante externa recombinante aminoacila preferiblemente el ARNt mutante externo con el aminoácido antinatural. Sintetasas mutantes externas recombinantes, producidas por los métodos también se incluyen en determinadas formas de realización descritas en esta memoria.

En una forma de realización, la RS es una RS inactiva, que puede haber sido generada a partir de la mutación de una RS activa. Por ejemplo, la RS inactiva se puede generar mutando al menos aproximadamente 1, al menos aproximadamente 2, al menos aproximadamente 3, al menos aproximadamente 4, al menos aproximadamente 5, al menos aproximadamente 6, o al menos aproximadamente 10 o más aminoácidos a diferentes aminoácidos, p. ej., alanina.

Bancos de RSs mutantes pueden generarse utilizando diversas técnicas de mutagénesis conocidas en la técnica. Por ejemplo, Las RSs mutantes se pueden generar por mutaciones específicas del sitio, mutaciones aleatorias, mutaciones de recombinación generadoras de diversidad, construcciones quiméricas, y por otros métodos descritos en esta memoria o conocidos en la técnica.

En una forma de realización, seleccionar (y/o rastrear) el banco de RSs (opcionalmente RSs mutantes) para los miembros que son activos, p. ej., que aminoacilan un ARNt mutante externo en presencia de un aminoácido antinatural y un aminoácido natural, incluye: introducir un marcador de selección o rastreo positivo, p. ej., un gen de resistencia a antibióticos, o similares, y el banco de RSs (opcionalmente mutantes) en una pluralidad de células, en donde el marcador de selección y/o rastreo positivo comprende al menos un codón, cuya traducción (opcionalmente de forma condicional) depende de la capacidad de RSs candidatas de cargar el ARNt mutante externo (con un aminoácido natural y/o antinatural); hacer crecer la pluralidad de células en presencia de un agente de selección; identificar las células que sobreviven (o que muestran una respuesta específica) en presencia del agente de selección y/o rastreo mediante la traducción con éxito del codón en el marcador de selección o rastreo positivo, proporcionando de este modo un subconjunto de células seleccionadas positivamente que contiene la agrupación de RSs activas (opcionalmente mutantes). Opcionalmente, se puede variar la concentración del agente

de selección y/o rastreo. En determinadas formas de realización, las células no contienen un ARNt, RS o par RS-ARNt endógeno funcional que pueda ayudar a traducir el codón. El par ARNt/RS endógeno puede ser desactivado por delección de genes y/o inhibidores de la RS.

Dado que muchos genes esenciales de la célula probablemente contengan también un codón que depende de la capacidad de la sintetasa mutante externa de cargar el ARNt modificado en ausencia del par RS/ARNt funcional endógeno, en una forma de realización no se necesitan marcadores de selección positivos extras para el proceso de selección positiva - la supervivencia de la célula se puede utilizar como una lectura del proceso de selección positiva.

En un aspecto, el marcador de selección positiva es un gen cloroanfenicol acetiltransferasa (CAT). Opcionalmente, el marcador de selección positiva es un gen β -lactamasa. En otro aspecto, el marcador de rastreo positivo comprende un marcador de rastreo fluorescente o luminiscente o un marcador de rastreo basado en afinidad (p. ej., un marcador de superficie celular).

En una forma de realización similar, se puede utilizar un sistema *in vitro* exento de células para someter a ensayo la capacidad de la sintetasa mutante externa de cargar el ARNt modificado en un rastreo positivo. Por ejemplo, la capacidad del sistema *in vitro* para traducir un gen de rastreo positivo tal como un gen marcador fluorescente, puede depender de la capacidad de la sintetasa mutante externa de cargar ARNt modificado para leer a través de un codón del gen marcador.

En una forma de realización, la selección o el rastreo negativo de la agrupación para RSs activas (opcionalmente mutantes) que preferiblemente aminoacilan el ARNt mutante en ausencia del aminoácido no natural incluye: introducir un marcador de selección o rastreo negativo con la agrupación de RSs activas (opcionalmente mutantes) procedentes de la selección o rastreo positivo en una pluralidad de sistema de traducción, en donde el marcador de selección o rastreo negativo comprende al menos un codón (p. ej., un codón para un gen marcador tóxico, p. ej., un gen de la ribonucleasa barnasa), cuya traducción depende de la capacidad de una RS candidata de cargar el ARNt mutante externo (con un aminoácido natural); e identificar el sistema de traducción que muestra una respuesta de rastreo específica en un primer medio suplementado con el aminoácido antinatural y un agente de rastreo o selección, pero fracasa en mostrar la respuesta específica en un segundo medio suplementado con el aminoácido natural y el agente de selección o rastreo, proporcionando con ello células supervivientes o células rastreadas con la al menos una RS recombinante.

En un aspecto, se varía la concentración del agente de selección (y/o rastreo). En algunos aspectos los primer y segundo organismos son diferentes. Así, el primer y/o segundo organismo comprende opcionalmente: un procariota, un eucariota, un mamífero, una *Escherichia coli*, un hongo, una levadura, una arqueobacteria, un Eubacterium, una planta, un insecto, un protista, etc. En otras formas de realización, el marcador de rastreo comprende un marcador de rastreo fluorescente o luminiscente (tal como la proteína verde fluorescente) o un marcador de rastreo basado en afinidad.

Además, algunos aspectos incluyen el que el marcador de selección negativa comprende un gen ribonucleasa barnasa (que comprende al menos uno de dichos codones). Otros aspectos incluyen que el marcador de rastreo comprende opcionalmente un marcador de rastreo fluorescente o luminiscente o un marcador de rastreo basado en afinidad. En las realizaciones de esta memoria, los rastreos y/o selecciones incluyen opcionalmente una variación de la rigurosidad de rastreo y/o selección.

En un aspecto, el segundo conjunto de RS mutadas se deriva de al menos una RS recombinante que puede ser generada por mutagénesis, p. ej., mutagénesis aleatoria, mutagénesis específica del sitio, recombinación o una combinación de las mismas.

Los métodos incorporados en esta memoria comprenden opcionalmente el que el aminoácido antinatural se selección, p. ej., de una 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, un fosfontirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina y una isopropil-L-fenilalanina. Una RS recombinante producida por los métodos de esta memoria también se incluye en las realizaciones descritas en esta memoria.

En un aspecto relacionado, métodos para producir un ARNt mutante externo recombinante incluyen: (a) generar un banco de ARNts mutantes derivados de al menos un ARNt, a partir de un primer organismo; (b) seleccionar (p. ej., seleccionar negativamente) o rastrear el banco de ARNts (opcionalmente mutantes) que se aminoacilan mediante una aminoacil-ARNt sintetasa (RS) procedente de un segundo organismo en ausencia de una RS procedente del

primer organismo, proporcionando de ese modo una agrupación de ARNts (opcionalmente mutantes); y, (c) seleccionar o rastrear la agrupación de ARNts (opcionalmente mutantes) en cuanto a miembros que se aminoacilan por una RS mutante externa introducida, proporcionando de este modo al menos un ARNt recombinante; en donde el al menos un ARNt recombinante reconoce un codón de aminoácido no natural y no es reconocido por la eficacia de la RS procedente del segundo organismo y se aminoacila preferiblemente por la RS mutante externa.

Los diversos métodos descritos en esta memoria comprenden opcionalmente, en donde la selección o el rastreo comprende uno o más selecciones o rastreos positivos o negativos, p. ej., un cambio en la permeabilidad de aminoácidos, un cambio en la eficacia de la traducción, y un cambio en la fidelidad de la traducción. Adicionalmente, el uno o más cambios se basan opcionalmente en una mutación en uno o más genes en un organismo en el que se utiliza un par ARNt-ARNt sintetasa mutante externo para producir tales proteínas. La selección y/o el rastreo de esta memoria comprende opcionalmente en el que se utilizan al menos 2 codones dentro de uno o más gen de selección o dentro de uno o más genes de rastreo. Tales múltiples codones se encuentran opcionalmente dentro del mismo gen o dentro de diferentes genes de rastreo/selección. Adicionalmente, los múltiples codones opcionales son opcionalmente codones diferentes o comprenden el mismo tipo de codones.

Kits son una característica adicional de determinadas formas de realización descritas en esta memoria. Por ejemplo, los kits pueden incluir uno o más sistema de traducción como se señaló anteriormente (p. ej., una célula), uno o más ARNt (incluido ARNt modificado o mutado), uno o más AARS (incluida AARS modificada o mutada), uno o más aminoácidos antinaturales, p. ej., con el material de envasado apropiado, recipientes para contener los componentes del kit, materiales de instrucciones para poner en práctica de los métodos de esta memoria y/o similares. Si se proporcionan uno o más AARS y/o uno o más ARNt en un kit, éstos se pueden suministrar como ácidos nucleicos o proteínas, y pueden ser parte de un único vector o pueden estar contenidos en vectores separados. De manera similar, los productos de los sistemas de traducción (p. ej., proteínas tales como análogos de EPO que comprenden aminoácidos antinaturales) pueden proporcionarse en forma de kit, p. ej., con recipientes para contener los componentes del kit, materiales de instrucciones para poner en práctica los métodos de esta memoria y/o similares.

Usos Ilustrativos

Según se ha reseñado, bastantes más de 100 aminoácidos no codificados (todos ribosómicamente aceptables) se han introducido proteínas utilizando otros métodos (véase, p. ej., Schultz *et al.*, *J. Am. Chem. Soc.*, 103: 1563-1567, 1981; Hinsberg *et al.*, *J. Am. Chem. Soc.*, 104: 766-773, 1982; Pollack *et al.*, *Science*, 242: 1038-1040, 1988; Nowak *et al.*, *Science*, 268: 439-442, 1995) todos estos análogos pueden utilizarse en los presentes métodos para una incorporación eficaz de estos análogos en productos de proteínas.

En otra forma de realización preferida, dos o más análogos se pueden utilizar en el mismo sistema de traducción *in vitro* o *in vivo*, cada uno con su ARNt mutante externo o pares de sintetasa mutantes externos. Esto se logra más fácilmente cuando un aminoácido natural es codificado por cuatro o más codones (tales como seis para Leu y Arg). Sin embargo, para los aminoácidos codificados por sólo dos codones, se puede reservar para el aminoácido natural, mientras que el otro era "compartido" por uno o más análogos de aminoácido. Estos análogos pueden asemejarse a los de un solo aminoácido natural (por ejemplo, diferentes análogos de Phe), o pueden asemejarse a diferentes aminoácidos (por ejemplo, análogos de Phe y Tyr).

En determinadas formas de realización, un primer ácido nucleico que codifica una molécula de ARNt mutante externa/modificada que no se carga de manera eficaz por una aminoacil-ARNt sintetasa endógena en la célula / sistema de traducción *in vitro* (IVT), o el propio ARNt mutante externo / modificado. De acuerdo con algunas formas de realización, un segundo ácido nucleico que codifica una aminoacil-ARNt sintetasa (AARS) mutante externa / modificada también se introduce en la célula / IVT. La AARS mutante externa / modificada es capaz de cargar el ARNt mutante externo / modificado con un análogo de aminoácido elegido. El análogo de aminoácido puede entonces ser proporcionado a la célula de modo que se puede incorporar en una o más proteínas dentro de la célula o IVT.

En otras formas de realización, el entorno es una célula. Una diversidad de células (o lisados de las mismas adecuados para IVT) se puede utilizar en determinados métodos, que incluyen, por ejemplo, una célula bacteriana, una célula fúngica, una célula de insecto y una célula de mamífero (p. ej., una célula humana o una célula de mamífero no humano). En una realización, la célula es una célula de *E. coli*, en otra realización, la célula es una célula de *Pseudomonas*.

En determinadas formas de realización, el análogo de aminoácido puede ser proporcionado por contacto directo de la célula o IVT con el análogo, por ejemplo mediante la aplicación de una disolución del análogo a la célula en

cultivo, o mediante la adición directa del análogo al IVT. El análogo también se puede proporcionar introduciendo una o más construcciones adicionales de ácidos nucleicos en la célula / IVT, en donde la o las construcciones de ácido nucleico adicionales codifica una o más proteínas de síntesis de análogo de aminoácidos que son necesarias para la síntesis del análogo deseado.

Determinadas formas de realización implican, además, la introducción de una construcción de ácido nucleico molde en la célula / IVT, codificando el molde una proteína, en el que la construcción de ácido nucleico contiene al menos una secuencia de codón degenerado. Los ácidos nucleicos introducida en la célula / IVT se pueden introducir como una construcción o como una pluralidad de construcciones. En determinadas formas de realización, los diversos ácidos nucleicos se incluyen en la misma construcción. Por ejemplo, los ácidos nucleicos pueden introducirse en cualesquiera vectores adecuados capaces de expresar el ARNt y / o las proteínas codificadas en la célula / IVT. En una forma de realización, las primera y segunda secuencias de ácidos nucleicos se proporcionan en uno o más plásmidos. En otra forma de realización, el vector o los vectores utilizados son vectores virales, que incluyen, por ejemplo, vectores adenovirales y lentivirales. Las secuencias se pueden introducir con una secuencia promotora apropiada para la célula / IVT, o múltiples secuencias que pueden ser inducibles para controlar la expresión de las secuencias.

Para su uso *in vitro*, una o más sintetisas mutante externas se pueden producir de forma recombinante y suministrar a cualquiera de los sistemas disponibles de traducción *in vitro* (tal como PROTEINscript-PRO™ basado en lisado de germen de trigo disponible comercialmente, el sistema de E. coli de Ambion® para la transcripción/traducción *in vitro* acoplada; o el kit Retic Lysate IVT™ basado en lisado de reticulocitos de conejo de Ambion®). Opcionalmente, el sistema de traducción *in vitro* se puede agotar selectivamente de uno o más AARSs naturales (por ejemplo, mediante inmuno-agotamiento utilizando anticuerpos inmovilizados contra AARS natural) y/o aminoácidos naturales, de manera que se puede lograr la incorporación mejorada del análogo. Alternativamente, los ácidos nucleicos que codifican las sintetisas mutantes externas rediseñadas pueden ser suministrados en lugar de las AARSs producidas de forma recombinante. El sistema de traducción *in vitro* también se suministra con los análogos a incorporar en productos proteicos maduros.

Aunque la síntesis *in vitro* de proteínas no puede llevarse a cabo, por lo general, a la misma escala que la síntesis *in vivo*, métodos *in vitro* pueden producir cientos de microgramos de proteína purificada que contienen análogos de aminoácidos. Tales proteínas han sido producidas en cantidades suficientes para su caracterización utilizando el dicroísmo circular (CD), espectrometría de resonancia magnética nuclear (RMN) y cristalografía de rayos X. Esta metodología también se puede utilizar para investigar el papel de hidrofobicidad, envasado, entropía de la cadena lateral y enlaces hidrógeno en la determinación de la estabilidad de proteínas y el plegamiento. También se puede utilizar para sondear un mecanismo catalítico, transducción de señales y la transferencia de electrones en proteínas. Además, las propiedades de las proteínas se pueden modificar utilizando esta metodología. Por ejemplo, se pueden generar proteínas foto-enjauladas que pueden ser activadas por fotólisis, y se han introducido nuevas manijas químicas en proteínas para la incorporación específica de sitio de sondas espectroscópicas ópticas y otras.

El desarrollo de una estrategia general para la incorporación de análogos de aminoácidos en proteínas *in vivo*, directamente desde los medios de crecimiento, mejoraría en gran medida el poder de mutagénesis de aminoácido antinaturales. Por ejemplo, la capacidad de sintetizar grandes cantidades de proteínas que contienen átomos pesados facilitaría la determinación de la estructura de la proteína, y la capacidad de sustituir de una manera específica del sitio fluoróforos o grupos fotoescindibles en proteínas en células vivas proporcionaría herramientas poderosas para el estudio de la función de proteínas *in vivo*. Alternativamente, se podrían mejorar las propiedades de las proteínas, proporcionando bloques de construcción con nuevos grupos funcionales tales como un aminoácido que contiene cetó.

Para el uso *in vivo*, una o más AARS se pueden suministrar a una célula huésped (procariota o eucariota) en forma de materiales genéticos tales como secuencias codificadoras en plásmidos o vectores virales, que opcionalmente pueden integrarse en el genoma del huésped y expresar de forma constitutiva o inducible las AARSs re-diseñadas. Una proteína heteróloga o endógena de interés se puede expresar en células huésped de este tipo, en presencia de análogos de aminoácidos suministrados. Los productos de proteína se pueden purificar después utilizando cualquier técnica de purificación de proteínas reconocida en la técnica, o técnicas especialmente diseñadas para la proteína de interés.

Estos son unos pocos medios posibles para generar un transcrito que codifica un polipéptido. En general, cualquier medio conocido en la técnica para la generación de transcritos se puede emplear para sintetizar proteínas con análogos de aminoácidos. Por ejemplo, cualquier sistema de transcripción *in vitro* o sistemas de transcripción / traducción acopladas se pueden utilizar para generar un transcrito de interés, que luego sirve como molde para la síntesis de proteínas. Alternativamente, para generar una transcripción se puede utilizar cualquier célula, célula /

línea celular modificada o componentes funcionales (lisados, fracciones de membrana, etc.) que sean capaces de expresar proteínas a partir de materiales genéticos. Estos medios para generar un transcrito incluirán típicamente componentes tales como la ARN polimerasa (T7, SP6, etc.) y co-factores, nucleótidos (ATP, CTP, GTP, UTP), factores de transcripción necesarios, y las condiciones tampón apropiadas, así como al menos un molde de ADN adecuado, pero también se pueden añadir otros componentes para condiciones de reacción optimizadas. Un experto en la técnica podría fácilmente imaginar otras formas de realización similares a las descritas en esta memoria.

Restos químicos

En determinadas formas de realización, el o los aminoácidos antinaturales y/o la molécula terapéutica comprenden un resto químicamente reactivo. El resto puede ser fuertemente electrofílico o nucleofílico y por lo tanto estará disponible para reaccionar directamente con la molécula terapéutica o el anticuerpo o fragmento del mismo. Alternativamente, el resto puede ser un electrófilo o nucleófilo más débil y, por lo tanto, requiere la activación antes de la conjugación con la molécula terapéutica o el anticuerpo o fragmento del mismo. Esta alternativa sería deseable cuando fuese necesario retrasar la activación del resto químicamente reactivo hasta que se añade un agente a la molécula con el fin de prevenir la reacción del agente con el resto. En cualquier escenario, el resto es químicamente reactivo, los escenarios difieren (en la reacción con el escenario de anticuerpo) por si después de la adición de un agente, el resto se hace reaccionar directamente con un anticuerpo o fragmento del mismo o se hace reaccionar primero con uno o más productos químicos para hacer que el resto sea capaz de reaccionar con un anticuerpo o fragmento del mismo. En determinadas formas de realización, el resto químicamente reactivo incluye un grupo amino, un grupo sulfhidrilo, un grupo hidroxilo, un grupo que contiene carbonilo, o un grupo lábil alquilo.

Determinadas formas de realización pueden emplear la química click, que incluye, pero no se limita a, la cicloadición 1,3-dipolar de Huisgen, en particular la variante escalonada catalizada por Cu(I), la reacción de Diels-Alder, sustitución nucleófila, especialmente a los pequeños anillos tensados tales como compuestos epoxi y aziridina, formación de ureas y amidas similar a la química del carbonilo, reacciones de adición a dobles enlaces carbono-carbono tal como epoxidación y dihidroxilación.

Por lo tanto, además de o en lugar de la glicosilación de polipéptidos de las formas de realización descritas en esta memoria, se pueden añadir otros restos químicos (incluyendo poli(etilen)glicol), enlazado, unido o conjugado o incorporado de otra manera a los polipéptidos modificados. La PEGilación es un proceso para fijar covalentemente oligosacáridos y polímeros sintéticos tales como polietilenglicol (PEG) de forma específica del sitio a moléculas de proteínas terapéuticas. La PEGilación puede mejorar significativamente la semi-vida de las proteínas mediante el blindaje del polipéptido frente a las enzimas proteolíticas y aumentar el tamaño aparente de la proteína, reduciendo así las tasas de aclaramiento. Además, conjugados de PEG pueden mejorar la solubilidad de la proteína y tienen efectos beneficiosos sobre la biodistribución. Las propiedades físicas y farmacológicas de proteínas PEGiladas se ven afectadas por el número y el tamaño de las cadenas de PEG fijadas al polipéptido, la ubicación de los sitios de PEG y la química utilizada para la PEGilación.

Ejemplos de conjugación de PEG a proteínas incluyen reacciones de PEGs derivatizados con éster de N-hidroxisuccinimidilo con lisina, reacciones de 1,4-adición de maleimida y PEGs derivatizados con vinilsulfona con cisteína, y condensación de hidrazida que contiene PEGs con aldehídos generados por la oxidación de las glicoproteínas. Cuando está presente más de un sitio reactivo en una proteína (p. ej., múltiples grupos amino o tiol) o se utilizan electrófilos reactivos, puede producirse la fijación no selectiva de una o múltiples moléculas de PEG, lo cual conduce a la generación de una mezcla heterogénea que es difícil de separar. La falta de selectividad y de un control posicional en la fijación de cadenas de PEG puede conducir a pérdidas significativas en la actividad biológica y posiblemente a una inmunogenicidad potenciada de la proteína conjugada. De hecho, históricamente, la pérdida de actividad biológica y la heterogeneidad del producto han sido los dos problemas más comunes con los que se topa en el desarrollo de productos farmacéuticos de proteínas de acción prolongada utilizando técnicas estándares de PEGilación. La modificación de proteínas con PEGs reactivos con aminos resulta típicamente en una pérdida drástica de la actividad biológica debido a la modificación de los residuos lisina localizados en regiones de la proteína importantes para la actividad biológica. En determinadas situaciones, la bioactividad de hormonas de crecimiento se puede reducir 400 veces o más. Por ejemplo, la bioactividad de GCSF se reduce 1.000 veces cuando las proteínas se modifican utilizando tecnologías convencionales de PEGilación de aminos (Clark *et al.*, *J. Biol. Chem.* 271: 21969, 1996; Bowen *et al.*, *Exp. Hematol.* 27, 425, 1999). Así, existe la necesidad de un método que permite la fijación completamente específica del sitio e irreversible de cadenas de PEG a proteínas.

Sería ventajoso utilizar tecnologías avanzadas de ingeniería de proteínas para crear productos farmacéuticos de proteínas humanas, de acción prolongada, "agradables para el paciente", por ejemplo incorporando de aminoácidos antinaturales en una proteína de fármacos, de manera que la proteína de fármacos modificada

pueden lograr una semi-vida y/o una actividad biológica más larga y/o sostenida. En aras de este fin, pueden utilizarse determinadas formas de realización descritas en esta memoria para superar problemas tales como la heterogeneidad y la pérdida de la actividad inherente a las técnicas estándares de PEGilación con aminos. La incorporación de aminoácidos antinaturales proporcionará sitios únicos, predeterminados lejos de la unión o el sitio catalítico de la proteína diana, en que las moléculas de PEG pueden ser conjugadas de una forma específica del sitio. Además, las moléculas de PEG pueden fijarse a aminoácidos antinaturales a través de técnicas distintas a la PEGilación con aminos, evitando así los grupos amina primarios de lisinas una PEGilación indeseable. Estas técnicas se pueden utilizar para potenciar la semi-vida, la eficacia y/o la seguridad de los productos bio-farmacéuticos en todos los ámbitos, incluyendo el campo específico de cáncer, endocrinología, enfermedades infecciosas, y la inflamación, etc.

Como un ejemplo ilustrativo, la Química Click o la cicloadición puede utilizarse para formar un enlace triazol. Un ejemplo particular de cicloadición es una [3 + 2] cicloadición de Huisgen mediada por cobre (Tornoe *et al.*, *J. Org. Chem.* 67: 3057, 2002; Rostovtsev *et al.*, *Angew. Chem., Ed. Int.* 41: 596, 2002; y Wang *et al.*, *J. Am. Chem. Soc.* 125: 3192, 2003) de una azida y un alquino es mutante externo para todos los grupos funcionales encontrados en las proteínas, y forma un enlace triazol estable, esta reacción se puede utilizar para la PEGilación selectiva de proteínas. Por ejemplo, Deiters *et al.* (*Bioorg. Med. Chem. Lett.* 14(23): 5743- 5745, 2004) informan de una metodología de PEGilación de aplicación general basada en la incorporación específica del sitio de para-azidofenilalanina en proteínas en levaduras. El grupo azido se utilizó en una reacción de [3 + 2] cicloadición suave con un reactivo de PEG derivatizado con alquino para proporcionar selectivamente proteína PEGilada. Esta estrategia debería ser útil para la generación de proteínas PEGiladas selectivamente para aplicaciones terapéuticas.

En determinadas formas de realización, el polipéptido es una proteína terapéutica, de diagnóstico, u otra proteína seleccionada de: Alfa-1-antitripsina, angiotensina, Factor antihemolítico, anticuerpos (incluido un anticuerpo o un fragmento funcional o derivado del mismo seleccionado de: Fab, Fab', F(ab)2, Fd, Fv, scFv, diacuerpo, tricuerpo, tetracuerpo, dímero, trímero o minicuerpo), moléculas angiogénicas, moléculas angiostáticas, apolipoproteína, apoproteína, factor natriurético auricular, polipéptido natriurético auricular, asparaginasa, adenosina desaminasa, hirudina, factor neurotrófico ciliar, factor morfogénico óseo (cualquiera y todas las BMPs), péptidos atriales, C-X-C quimioquinas (p. ej., T39765, NAP-2, ENA-78, Gro-a, Gro-b, Groc, IP-10, GCP-2, NAP-4, SDF-1, PF4, MIG), calcitonina, quimioquinas CC (por ejemplo, proteína-1 quimioattractiva de monocitos, proteína-2 quimioattractiva de monocitos, proteína-3 quimioattractiva de monocitos, proteína-1 alfa inflamatoria de monocitos, proteína-1 beta inflamatoria de monocitos, RANTES, I309, R83915, R91733, HCC1, T58847, D31065, T64262), ligando CD40, calcitonina, ligando C-kit, colágeno, factor estimulante de colonias (CSF), péptido natriurético tipo C (CNP), factor del complemento 5a, inhibidor del complemento, receptor 1 del complemento, citoquinas, (p. ej. péptido-78 activador de neutrófilos epitelial, GRO α /MGSa, GRO β , GRO γ , MIP-1 α , MIP-1 δ , MCP-1), ácidos desoxirribonucleicos, Factor de Crecimiento Epidérmico (EGF), eritropoyetina, toxinas exfoliantes A y B, Factor IX, Factor VII, Factor VIII, Factor X, Factor de Crecimiento de Fibroblastos (FGF), fibrinógeno, fibronectina, el factor estimulante de colonias de granulocitos (G-CSF), factor estimulante de colonias de granulocitos y macrófagos (GM-CSF), folitropina, glucocerebrosidasa, gonadotropina, glucagón, GLP-1, factores de crecimiento, proteínas Hedgehog (p. ej., Sonic, India, Desert), hormona del crecimiento humano, hemoglobina, factor de crecimiento de hepatocitos (HGF), virus de la hepatitis, hirudina, albúmina de suero humano, insulina, factor de crecimiento similar a la insulina (IGF), interferones (p. ej., IFN- α , IFN- β , IFN- γ , IFN- ϵ , IFN- ζ , IFN- η , IFN- κ , IFN- λ , IFN- τ , IFN- ς , IFN- ω), interleucinas (p. ej., IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, etc.), factor de crecimiento de queratinocitos (KGF), lactoferrina, factor inhibidor de la leucemia, luciferasa, hormona luteinizante, neurturina, factor inhibidor de neutrófilos (NIF), oncostatina M, proteína osteogénica, hormona paratiroidea, PD-ECSF, PDGF, hormonas peptídicas (p. ej., hormona de crecimiento humana), pletotropina, proteína A, proteína G, fenilalanina hidroxilasa, parathormona (PTH), prolactina, exotoxinas pirogénicas A, B y C, relaxina, renina, ácidos ribonucleicos, SCF, receptor del complemento soluble I, Soluble I-CAM 1, receptores solubles de interleucina (IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15), receptor de TNF soluble, somatomedina, somatostatina, somatotropina, estreptoquinasa, superantígenos, es decir, las enterotoxinas estafilocócicas (SEA, SEB, SEC1, SEC2, SEC3, SED, SEE), superóxido dismutasa (SOD), toxina del síndrome de choque tóxico (TSST-1), timosina alfa 1, activador tisular del plasminógeno, factor de necrosis tumoral beta (TNF beta), receptor de factor de necrosis tumoral (TNFR), factor de necrosis tumoral-alfa (TNF alfa), factor de necrosis tumoral relacionada con ligando inductor de apoptosis (TRAIL), factor de crecimiento endotelial vascular (VEGEF), uroquinasa; un modulador de la transcripción que modula el crecimiento celular, la diferenciación o la regulación, en donde el modulador transcripcional es de procariotas, virus o eucariotas, incluidos hongos, plantas, levaduras, insectos y animales, incluidos mamíferos; activador de la expresión seleccionado de citoquinas, moléculas inflamatorias, factores de crecimiento, sus receptores, productos de oncogenes, interleucinas (p. ej. IL-1, IL-2, IL-8, etc.), interferones, FGF, IGF-I, IGF-II, FGF, PDGF, TNF, TGF- α , TGF- β , EGF, KGF, SCF/c-Kit, CD40L/CD40, VLA-4/VCAM-1, ICAM-1/LFA-1 y hialurina/CD44; moléculas de transducción de señales y los correspondientes productos de oncogenes, p. ej., Mos, Ras, Raf y Met; activadores y

supresores de la transcripción, p. ej., p53, Tat, Fos, Myc, Jun, Myb, Rel; receptores de hormonas esteroides seleccionados de receptores para el estrógeno, progesterona, testosterona, aldosterona, LDL, o corticosterona; o una enzima seleccionada de: amidasas, aminoácido racemasas, acilasas, deshalogenasas, dioxigenasas, diarilpropano peroxidasas, epimerasas, epóxido hidrolasas, esterasas, isomerasas, quinasas, glucosa isomerasas, glicosidasas, glicosil transferasas, haloperoxidasas, monooxigenasas (p. ej., P450s), lipasas, lignina peroxidasas, nitrilo hidratasas, nitrilasas, proteasas, fosfatasas, subtilisinas, transaminasas o nucleasas.

En el caso de que la proteína o molécula de interés a ser modificada sea un anticuerpo o fragmento de anticuerpo, el o los residuos de aminoácidos no naturales se pueden colocar en cualquier lugar o posición en la estructura del anticuerpo, dependiendo del objetivo deseado. Por ejemplo, el residuo de aminoácido no natural se puede colocar en la región variable de Fab, la región Fc, o en otra ubicación que interactúa con la región Fc del anticuerpo. En otras formas de realización, el residuo de aminoácido no natural se puede colocar en la interfaz de unión del anticuerpo, o de la región VH. En determinadas formas de realización, el anticuerpo modificado exhibe un aumento o disminución en su capacidad para exterminar a una o más dianas. En particular, puede ser deseable un anticuerpo con una mayor capacidad de exterminar una o más dianas, o con efectos secundarios reducidos.

En otras formas de realización, el o los aminoácidos no naturales confieren una afinidad de unión mejorada a un receptor de Fc y/o a C1q del sistema del complemento. En particular, un anticuerpo modificado puede tener una afinidad alterada (p. ej., potenciada) y/o una especificidad para un antígeno o un participante en la unión de proteínas (p. ej., C1q del complemento y/o el receptor de Fc en macrófagos, etc.). Por ejemplo, la modificación de una molécula puede aumentar o disminuir su función de citotoxicidad mediada por células dependiente del anticuerpo (ADCC), o complementar la actividad de fijación. En otros ejemplos, la modificación de una molécula particular puede aumentar o disminuir su capacidad de unirse a otra molécula de la estructura antagonista natural (tal como un anticuerpo).

Glicosilación a través de Aminoácidos Antinaturales

La modificación post-traducciona de proteínas mediante glicosilación puede afectar el plegamiento y la estabilidad de proteínas, modificar la actividad intrínseca de las proteínas y modular sus interacciones con otras biomoléculas. Véase, p. ej., Varki, *Glycobiology* 3: 97-130, 1993. Glicoproteínas naturales están a menudo presentes como una población de muchas glicoformas diferentes, lo que dificulta el análisis de la estructura de glicano y el estudio de los efectos de glicosilación sobre la estructura y función de proteínas. Por lo tanto, se necesitan métodos para la síntesis de proteínas glicosiladas homogéneamente naturales y antinaturales para la comprensión sistemática de la función de glicano, y para el desarrollo de terapias mejoradas con glicoproteínas.

Una estrategia previamente conocida para la fabricación de proteínas que tienen patrones de glicosilación deseados hace uso de glicosidasas para convertir una glicoproteína natural heterogénea a un simple núcleo homogéneo, sobre la que se pueden injertar sacáridos secuencialmente con glicosil transferasas. Véase, p. ej., Witte *et al.*, *J. Am. Chem. Soc.* 119: 2114-2118, 1997. Una limitación de esta estrategia es que los sitios de glicosilación primarios están predeterminados por la línea celular en la que se expresa la proteína. Alternativamente, un glicopéptido que contiene la estructura de glicano deseada se puede sintetizar mediante síntesis de péptidos en fase sólida. Este glicopéptido puede acoplarse a otros péptidos o fragmentos de proteínas recombinantes para producir una glicoproteína grande por ligamiento químico nativo (véase, p. ej., Shin *et al.*, *J. Am. Chem. Soc.* 121: 11684-11689, 1999), el ligamiento de la proteína expresada (véase, p. ej., Tolbert y Wong, *J. Am. Chem. Soc.* 122: 5421-5428, 2000), o con proteasas modificadas por ingeniería (véase, p. ej., Witte *et al.*, *J. Am. Chem. Soc.* 120: 1979-1989, 1998). Tanto el ligamiento químico nativo como el ligamiento de proteínas expresado son lo más eficaces con proteínas pequeñas, y requieren un residuo cisteína en el extremo N del glicopéptido.

Cuando se utiliza una proteasa para ligar péptidos juntos, el sitio de ligamiento se debe colocar alejado del sitio de glicosilación para los buenos rendimientos de acoplamiento. Véase, p. ej., Witte *et al.*, *J. Am. Chem. Soc.* 120: 1979-1989, 1998. Una tercera estrategia consiste en modificar proteínas con sacáridos directamente utilizando métodos químicos. Se puede conseguir una buena selectividad con derivados de sacáridos de haloacetamida, que están acoplados al grupo tiol de la cisteína (véase, p. ej., Davis y Flitsch, *Tetrahedron Lett.* 32: 6793-6796, 1991, y Macmillan *et al.*, *Org. Lett.* 4: 1467-1470, 2002). Sin embargo, este método puede ser problemático con proteínas que tienen más de un residuo cisteína.

Determinadas formas de realización proporcionadas en esta memoria describen métodos para la síntesis de glicoproteínas. Estos métodos implican, en algunas formas de realización, la incorporación en una proteína de un aminoácido antinatural que comprende un primer grupo reactivo; y poner en contacto la proteína con un resto de sacárido que comprende un segundo grupo reactivo, en donde el primer grupo reactivo reacciona con el segundo grupo reactivo, formando de este modo un enlace covalente que fija el resto de sacárido al aminoácido antinatural

de la proteína. Las glicoproteínas producidas por estos métodos también se incluyen en determinadas formas de realización.

El primer grupo reactivo es, en algunas formas de realización, un resto electrófilo (p. ej., un resto ceto, un resto aldehído, y/o similares), y el segundo grupo reactivo es un resto nucleófilo. En algunas formas de realización, el primer grupo reactivo es un resto nucleófilo y el segundo grupo reactivo es un resto electrófilo (p. ej., un resto ceto, un resto aldehído, y/o similares). Por ejemplo, un resto electrófilo está fijado al resto sacárido y el resto nucleófilo está fijado al aminoácido antinatural. El resto sacárido puede incluir un único resto carbohidrato, o el resto de sacárido puede incluir dos o más restos carbohidrato.

En algunas formas de realización, los métodos implican, además, poner en contacto el resto sacárido con una glicosil-transferasa, un resto donante de azúcar, y otros reaccionantes requeridos para la actividad glicosil-transferasa durante un tiempo suficiente y bajo condiciones apropiadas para transferir un azúcar desde el resto donante de azúcar al resto sacárido. El producto de esta reacción puede, si se desea, ser contactado por al menos una segunda glicosil-transferasa, junto con el resto donante de azúcar apropiado.

En determinadas formas de realización, el método comprende, además, poner en contacto el resto sacárido con una o más de una β 1-4N-acetilglucosaminil transferasa, una α 1,3-fucosil transferasa, una α 1,2-fucosil transferasa, una α 4fucosil transferasa, una β 1,4-galactosil transferasa, una sialil transferasa, y/o similares, para formar una estructura de oligosacárido biantenarico o triantenarico. En una forma de realización, el resto sacárido comprende un GlcNAc terminal, el resto donante de azúcar es UDP-Gal y la glicosil transferasa es una β 1,4-galactosil transferasa.

En una realización, el resto sacárido comprende un GlcNAc terminal, el resto donante de azúcar es UDP-GlcNAc y la glicosiltransferasa es una β 1,4-acetilglucosaminil transferasa.

Opcionalmente, los algunos métodos comprenden, además, poner en contacto el producto de la reacción de N-acetilglucosaminil transferasa con un β 1,4-manosil transferasa y GDP-manosa para formar un resto sacárido que comprende $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc. Opcionalmente, el método comprende, además, poner en contacto el resto $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc con una α 1,3-manosil transferasa y GDP-manosa para formar un resto sacárido que comprende $\text{Man}\alpha$ 1-3 $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc. Opcionalmente, el método comprende, además, poner en contacto el resto $\text{Man}\alpha$ 1-3 $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc con una α 1-6 manosiltransferasa y GDP-manosa para formar un resto sacárido que comprende $\text{Man}\alpha$ 1-6($\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc. Opcionalmente, el método comprende, además, poner en contacto el resto $\text{Man}\alpha$ 1-6($\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc con una β 1-2N-acetilglucosaminil transferasa y UDP-GlcNAc para formar un resto sacárido que comprende $\text{Man}\alpha$ 1-6(GlcNAc β 1-2 $\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc. Opcionalmente, el método comprende, además, poner en contacto el resto $\text{Man}\alpha$ 1-6(GlcNAc β 1-2 $\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc con una β 1-2N-acetilglucosaminil transferasa y UDP-GlcNAc para formar un resto sacárido que comprende GlcNAc β 1-2 $\text{Man}\alpha$ 1-6(GlcNAc β 1-2 $\text{Man}\alpha$ 1-3) $\text{Man}\beta$ 1-4GlcNAc β 1-4GlcNAc.

La etapa de incorporar en una proteína un aminoácido antinatural que comprende un primer grupo reactivo, en algunas formas de realización, comprende utilizar un ARNt mutante externo, una RS mutante externa o un par ARNt/RS mutante externa. En tales casos, el ARNt mutante externo reconoce preferiblemente un codón degenerado para el ARNt de tipo salvaje, e incorpora el aminoácido antinatural en la proteína en respuesta al codón degenerado, y en donde la sintetasa mutante externa aminoacila preferiblemente el ARNt mutante externo con el aminoácido antinatural. En algunas formas de realización, el aminoácido no natural se incorpora en el polipéptido *in vivo*.

Se conoce una amplia diversidad de grupos reactivos adecuados por los expertos en la técnica. Tales grupos reactivos adecuados pueden incluir, por ejemplo, amino, hidroxilo, carboxilo, carboxilato, carbonilo, alqueno, alquino, aldehído, éster, éter (p. ej., tio-éter), amida, amina, nitrilo, vinilo, sulfuro, sulfonilo, fosforilo, o grupos químicamente reactivos de manera similar. Grupos reactivos adecuados adicionales incluyen, pero no se limitan a, maleimida, N-hidroxisuccinimida, sulfo-N-hidroxisuccinimida, ácido nitrilotriacético, hidroxilo activado, haloacetilo (p. ej., bromoacetilo, yodoacetilo), carboxilo activado, hidrazida, epoxi, aziridina, cloruro de sulfonilo, trifluorometildiaziridina, disulfuro de piridilo, N-acil-imidazol, carbamato de imidazol, vinilsulfona, carbonato de succinimidilo, arilazida, anhídrido, diazoacetato, benzofenona, isotiocianato, isocianato, imidoéster, fluorobenceno, biotina y avidina.

En algunas formas de realización, uno de los grupos reactivos es un resto electrófilo y el segundo grupo reactivo es un resto nucleófilo. El resto nucleófilo o el resto electrófilo se pueden fijar a la cadena lateral del aminoácido no natural; el grupo correspondiente se fija luego al resto sacárido.

Restos electrófilos adecuados que reaccionan con restos nucleófilos para formar un enlace covalente son conocidos por los expertos en la técnica. En determinadas formas de realización, tales restos electrófilos incluyen, pero no se limitan a, p. ej., un grupo carbonilo, un grupo sulfonilo, un grupo aldehído, un grupo cetona, un grupo éster estéricamente impedido, un grupo tioéster, un grupo imina estable, un grupo epóxido, un grupo aziridina, etc.

Restos nucleófilos adecuados que pueden reaccionar con un resto electrófilo son conocidos por los expertos en la técnica. En determinadas formas de realización, tales nucleófilos incluyen, por ejemplo, aminas alifáticas o aromáticas, tales como etilendiamina. En determinadas formas de realización, los restos nucleófilos incluyen, pero no se limitan a, p. ej., $-NR_1-NH_2$ (hidrazida), $-NR_1(C=O)NR_2NH_2$ (semicarbazida), $-NR_1(C=S)NR_2NH_2$ (tiosemicarbazida), $-(C=O)NR_1NH_2$ (carbonilhidrazida), $-(C=S)NR_1NH_2$ (tiocarbonilhidrazida), $-(SO_2)NR_1NH_2$ (sulfonilhidrazida), $-NR_1NR_2(C=O)NR_3NH_2$ (carbazida), $NR_1NR_2(C=S)NR_3NH_2$ (tiocarbazida), $-O-NH_2$ (hidroxilamina), y similares, en donde cada uno de R_1 , R_2 y R_3 es independientemente H, o alquilo que tiene 1-6 carbonos, preferiblemente H. En determinadas formas de realización, el grupo reactivo es una hidrazida, hidroxilamina, semicarbazida, carbohidrazida, una sulfonilhidrazida, o similares.

El producto de la reacción entre el nucleófilo y el resto electrófilo incorpora típicamente los átomos presentes originalmente en el resto nucleófilo. Enlaces típicos obtenidos por reacción de los aldehídos o cetonas con los restos nucleófilos incluyen productos de reacción tales como una oxima, una amida, una hidrazona, una hidrazona reducida, una carbohidrazona, una tiocarbohidrazona, un sulfonilhidrazona, una semicarbazona, una tiosemicarbazona, o una funcionalidad similar, dependiendo del resto nucleófilo utilizado y el resto electrófilo (p. ej., aldehído, cetona, y/o similares) que se hace reaccionar con el resto nucleófilo. A enlaces con ácidos carboxílicos se les alude típicamente como carbohidrazidas o ácidos hidroxámicos. A enlaces con ácidos sulfónicos se les alude típicamente como sulfonilhidrazidas o N-sulfonilhidroxilaminas. El enlace resultante puede ser posteriormente estabilizado por reducción química.

Estos métodos pueden implicar, además, poner en contacto el resto sacárido con una glicosil transferasa, un resto donante de azúcar, y otros reaccionantes requeridos para la actividad glicosil transferasa durante un tiempo suficiente y bajo condiciones apropiadas para transferir un azúcar desde el resto donante de azúcar al resto sacárido. En determinadas formas de realización, el método comprende, además, poner en contacto el producto de la reacción de glicosil transferasa con al menos una segunda glicosiltransferasa y un segundo resto donante de azúcar. En otras palabras, determinadas formas de realización descritas en esta memoria proporcionan métodos en los que un resto sacárido enlazado a aminoácido o un aminoácido antinatural que incluye un resto sacárido se glicosila adicionalmente. Estas etapas de glicosilación se llevan a cabo preferiblemente (aunque no necesariamente) enzimáticamente utilizando, por ejemplo, una glicosiltransferasa, glicosidasa, u otra enzima conocida para los expertos en la técnica. En algunas formas de realización, se llevan a cabo una pluralidad de etapas enzimáticas en una única mezcla de reacción que contiene dos o más glicosil transferasas diferentes. Por ejemplo, se puede llevar a cabo una etapa de galactosilación y una etapa de sialilación simultáneamente incluyendo tanto sialil transferasa como galactosil transferasa en la mezcla de reacción.

Para síntesis enzimáticas de sacáridos que implican reacciones de glicosil transferasa, las células recombinantes contienen opcionalmente al menos un gen heterólogo que codifica una glicosil transferasa. Se conocen muchas glicosil transferasas, como lo son sus secuencias de polinucleótidos. Véase, p. ej., "The WWW Guide To Cloned Glycosyl transferases," (disponible en la World Wide Web). Secuencias de aminoácidos glicosil transferasa y secuencias de nucleótidos que codifican glicosil transferasas de las que pueden deducirse las secuencias de aminoácidos también se encuentran en diversas bases de datos disponibles públicamente, incluidas GenBank, Swiss-Prot, EMBL, y otras.

En determinadas formas de realización, una glicosil transferasa incluye, pero no se limita a, p. ej., una galactosil transferasa, una fucosil transferasa, una glucosil transferasa, una N-acetilgalactosaminil transferasa, una N-acetilglucosaminil transferasa, una glucuronil transferasa, una sialil transferasa, una manosil transferasa, una transferasa de ácido glucurónico, una transferasa de ácido galacturónico, una oligosacaril transferasa, y similares. Glicosil transferasas adecuadas incluyen las obtenidas de eucariotas o procariotas.

Un aceptor para las glicosil transferasas estará presente en la glicoproteína que ha de ser modificada por los métodos descritos en esta memoria. Aceptores adecuados incluyen, por ejemplo, aceptores de galactosilo tales como Gal β 1,4GalNAc-; Gal β 1,3GalNAc-; lacto-N-tetraosa-; Gal β 1,3GlcNAc-; Gal β 1,4GlcNAc-; Gal β 1,3Ara-; Gal β 1,6GlcNAc-; y Gal β 1,4Glc- (lactosa). Otros aceptores conocidos por los expertos en la técnica (véase, p. ej., Paulson *et al.*, *J. Biol. Chem.* 253: 5617-5624, 1978). Típicamente, los aceptores forman parte de un resto sacárido que está fijado a la glicoproteína.

En una forma de realización, el resto sacárido comprende un GlcNAc terminal, el resto donante de azúcar es UDP-GlcNAc y la glicosil transferasa es una β 1-4N-acetilglucosaminil transferasa. En otra realización, el resto sacárido

comprende un GlcNAc terminal, el resto donante de azúcar es UDP-Gal y la glicosil transferasa es una β 1-4-galactosil transferasa. Asimismo se pueden añadir azúcares adicionales.

Las reacciones de glicosilación incluyen, además de la glicosil transferasa adecuada y aceptor, un azúcar nucleótido activado que actúa como un donante de azúcar para la glicosil transferasa. Las reacciones también pueden incluir otros ingredientes que facilitan la actividad glicosil transferasa. Estos ingredientes pueden incluir un catión divalente (p. ej., Mg^{2+} o Mn^{2+}), materiales necesarios para la regeneración de ATP, iones fosfato, y disolventes orgánicos. Las concentraciones o cantidades de los diversos reaccionantes utilizados en los procedimientos dependen de numerosos factores que incluyen condiciones de reacción tales como la temperatura y el valor del pH, y la elección y cantidad de sacáridos aceptores a ser glicosilados. El medio de reacción también puede comprender detergentes solubilizantes (p. ej. Triton o SDS) y disolventes orgánicos tales como metanol o etanol, si es necesario.

También proporcionadas por parte de determinadas formas de realización para la modificación de una glicoproteína son composiciones que incluyen un sistema de traducción que puede o puede no incluir una célula huésped, un ARNt mutante externo, una RS mutante externa, o cualquiera o la totalidad de estos.

Tal como se utiliza en esta memoria, la expresión "resto sacárido" se refiere a restos de azúcares naturales y antinaturales (es decir, un resto de azúcar que se produce de forma no natural, p. ej., un resto azúcar que es modificado, p. ej., en una o más posiciones de hidroxilo o amino, p. ej., deshidroxilado, desaminado, esterificado, etc., p. ej., 2-desoxiGal es un ejemplo de un resto azúcar antinatural).

El término "carbohidrato" tiene la fórmula general $(CH_2O)_n$, e incluye, pero no se limita a, p. ej., monosacáridos, disacáridos, oligosacáridos y polisacáridos. Los oligosacáridos son cadenas compuestas de unidades sacárido, que se conocen alternativamente como azúcares. Unidades sacárido se pueden organizar en cualquier orden y el enlace entre dos unidades sacárido puede producirse en cualquiera de aproximadamente diez maneras diferentes. Las siguientes abreviaturas se utilizan en esta memoria: Ara = arabinosilo; Fru = fructosilo; Fuc = fucosilo; Gal = galactosilo; GalNAc = N-acetilgalactosaminilo; Glc = glucosilo; GlcNAc = N-acetilglucosaminilo; Man = manosilo; y NeuAc = sialilo (típicamente N-acetilneuraminilo).

Se considera que los oligosacáridos tienen un extremo reductor y un extremo no reductor, sea o no el sacárido en el extremo reductor, de hecho, un azúcar reductor. De acuerdo con la nomenclatura aceptada, los oligosacáridos se representan en esta memoria con el extremo no reductor a la izquierda y el extremo reductor a la derecha. Todos los oligosacáridos descritos en esta memoria se describen con el nombre o abreviatura para el sacárido no reductor (p. ej., Gal), seguido por la configuración del enlace glicosídico (α o β), el enlace del anillo, la posición del anillo del sacárido reductor implicado en el enlace, y luego el nombre o la abreviatura del sacárido reductor (p. ej. GlcNAc). El enlace entre dos azúcares puede expresarse, por ejemplo, como 2,3; 2 $\rightarrow$ 3; 2-3; o (2,3). Enlaces naturales y no naturales (p. ej., 1-2; 1-3; 1-4; 1-6; 2-3; 2-4; 2-6; etc.) entre dos azúcares están incluidos en determinadas realizaciones. Cada uno de los sacáridos es una piranosa

La expresión "ácido siálico" (abreviado "Sia") se refiere a cualquier miembro de una familia de azúcares carboxilados de nueve carbonos. El miembro más común de la familia del ácido siálico es ácido N-acetilneuramínico (ácido 2-ceto-5-acetamido-3,5-didesoxi-D-glicero-D-galactononulopiranos-1-ónico) (a menudo abreviado como Neu5Ac, NeuAc o NANA). Un segundo miembro de la familia es ácido N-glicolil-neuramínico (Neu5Gc o NeuGc), en el que el grupo N-acetilo de NeuAc está hidroxilado. Un tercer miembro de la familia del ácido siálico es el ácido 2-ceto-3-desoxi-nonulosónico (KDN) (Nadano *et al.*, *J. Biol. Chem.* 261: 11550-11557, 1986; Kanamori *et al.*, *J. Biol. Chem.* 265: 21811-21819, 1990). También se incluyen ácidos siálicos 9-sustituidos tales como 9-O-acilo C1-C6-Neu5Ac tal como 9-O-lactil-Neu5Ac o 9-O-acetil-Neu5Ac, 9-desoxi-9-fluoro-Neu5Ac y 9-azido-9-desoxi-Neu5Ac. Para una revisión de la familia del ácido siálico, véase, p. ej., Varki, *Glycobiology* 2: 25-40, 1992; Sialic Acids: Chemistry, Metabolism and Function, R. Schauer, comp. (editorial Springer, Nueva York (1992). La síntesis y el uso de compuestos de ácido siálico en un proceso de sialilación se describen, por ejemplo, en la solicitud internacional WO 92/16640 (cuyo contenido completo se incorpora en esta memoria como referencia).

Sustratos donantes para glicosil transferasas son azúcares nucleótidos activados. Tales azúcares activados consisten generalmente en uridina y guanosina difosfato, y citidina monofosfato, los derivados de los azúcares en los que el nucleósido difosfato o monofosfato sirve como un grupo lábil. Sistemas bacterianos, vegetales y de hongos a veces pueden utilizar otros azúcares nucleótidos activados.

La incorporación de un aminoácido no natural, p. ej., un aminoácido no natural que comprende un resto en el que puede estar fijado un resto sacárido, o un aminoácido antinatural que incluye un resto sacárido, se puede hacer para, p. ej., adaptar los cambios en la estructura y/o función de proteínas, p. ej. para cambiar el tamaño, acidez,

nucleofilia, enlaces hidrógeno, hidrofobia, accesibilidad de sitios diana de proteasa, acceso diana a un resto de proteína, etc. Las proteínas que incluyen un ácido amino antinatural, p. ej., un aminoácido no natural que comprende un resto en que se puede fijar un resto sacárido, o un aminoácido antinatural que incluye un resto sacárido, pueden tener propiedades catalíticas o físicas mejoradas o incluso totalmente nuevas.

5 Por ejemplo, las siguientes propiedades se modifican opcionalmente por la inclusión de un aminoácido antinatural, p. ej., un aminoácido antinatural que comprende un resto amino, en que se puede fijar un resto sacárido, o un aminoácido antinatural que incluye un resto sacárido en una proteína: la toxicidad, biodistribución, propiedades estructurales, propiedades espectroscópicas, propiedades químicas y/o fotoquímicas, la capacidad catalítica, la
10 semi-vida (p. ej., la semi-vida en suero), la capacidad de reaccionar con otras moléculas, p. ej. de forma covalente o no covalente, y similares. Las composiciones que incluyen proteínas que incluyen al menos un aminoácido antinatural, p. ej., un aminoácido antinatural que comprende un resto al que puede fijarse un resto sacárido, o un aminoácido no natural que incluye un resto sacárido son útiles, p. ej., para nuevos productos terapéuticos, de diagnóstico, enzimas catalíticas, enzimas industriales, proteínas de unión (p. ej., anticuerpos) y, p. ej., el estudio
15 de la estructura y función de proteínas. Véase, p. ej., Dougherty, *Curr. Opin. en Chem. Biol.*, 4: 645- 652 (2000).

En un aspecto, una composición incluye al menos una proteína con al menos uno, p. ej., al menos aproximadamente dos, tres, cuatro, cinco, seis, siete, ocho, nueve, o al menos aproximadamente diez o más aminoácidos antinaturales, p. ej., un aminoácido no natural que comprende un resto al que se puede fijar un resto
20 sacárido, o un aminoácido antinatural que incluye un resto sacárido, y/o que incluyen otro aminoácido antinatural. Los aminoácidos antinaturales pueden ser iguales o diferentes, p. ej., puede haber 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más sitios diferentes en la proteína que comprenden 1, 2, 3, 4, 5, 6, 7, 8, 9 ó 10 o más aminoácidos antinaturales diferentes. En otro aspecto, una composición incluye una proteína con al menos uno, pero menos que todos, de un aminoácido particular presente en la proteína sustituida con el aminoácido antinatural, p. ej. un aminoácido
25 antinatural que comprende un resto al que se puede fijar un resto sacárido, o un aminoácido antinatural que incluye un resto sacárido. Para una proteína dada con más de un aminoácido antinatural, los aminoácidos no naturales pueden ser idénticos o diferentes (p. ej., la proteína puede incluir dos o más tipos diferentes de aminoácidos antinaturales, o puede incluir dos de los mismos aminoácidos antinaturales). Para una proteína dada con más de dos aminoácidos antinaturales, los aminoácidos antinaturales pueden ser el mismo, diferente, o una
30 combinación de múltiples aminoácidos antinaturales del mismo tipo con al menos un aminoácido antinatural diferente.

Esencialmente, cualquier proteína (o parte de la misma) que incluya un aminoácido antinatural, p. ej., un aminoácido antinatural que comprende un resto al que se fija un resto sacárido, tal como un aminoácido aldehído-
35 o ceto-derivatizado, o un aminoácido antinatural que incluye un resto sacárido (y cualquier ácido nucleico codificante correspondiente, p. ej., que incluye uno o más codones selectores) se puede producir utilizando las composiciones y los métodos de esta memoria. No se intenta identificar los cientos de miles de proteínas conocidas, cualquiera de las cuales se puede modificar para incluir uno o más aminoácidos antinaturales, p. ej., adaptando cualquier método de mutación disponible para que incluya uno o más codones degenerados apropiados
40 en un sistema de traducción relevante. Repositorios de secuencias comunes de proteínas conocidas incluyen GenBank EMBL, DDBJ y el NCBI. Otros repositorios pueden ser fácilmente identificados por búsquedas en internet.

Típicamente, las proteínas son, por ejemplo, al menos aproximadamente 60%, 70%, 75%, 80%, 90%, 95%, o al
45 menos aproximadamente 99% o más idénticas a cualquier proteína disponible (p. ej., una proteína terapéutica, una proteína de diagnóstico, una enzima industrial, o parte de la misma, y similares), y comprenden uno o más aminoácidos antinaturales.

Además de modificar uno o más residuos aminoácidos de la proteína, se puede modificar la composición de
50 carbohidratos de la proteína, es decir, a través de la glicosilación. La modificación post-traducción de las proteínas por glicosilación puede afectar al plegamiento y la estabilidad de proteínas, modificar la actividad intrínseca de las proteínas y modular sus interacciones con otras biomoléculas. Véase, p. ej., Varki, *Glycobiology* 3: 97-130, 1993, se incorpora aquí como referencia en su totalidad. Glicoproteínas naturales están a menudo presentes como una población de muchas glicoformas diferentes, lo que dificulta el análisis de la estructura de glicano y el estudio de
55 los efectos de glicosilación sobre la estructura y función de proteínas. Por lo tanto, se necesitan métodos para la síntesis de proteínas glicosiladas homogéneamente naturales y antinaturales para la comprensión sistemática de la función de glicano, y para el desarrollo de productos terapéuticos de glicoproteína mejorados.

Una clase de proteínas que se pueden hacer utilizando determinadas composiciones y métodos descritos en esta
60 memoria incluyen moduladores de la transcripción, enzimas, o una porción del mismo. Moduladores transcripcionales ilustrativos incluyen genes y proteínas moduladoras de la transcripción que modulan el crecimiento celular, la diferenciación, la regulación, o similares. Moduladores de la transcripción se encuentran en

procariotas, virus y eucariotas, incluidos hongos, plantas, levaduras, insectos y animales, incluidos mamíferos, proporcionando una amplia gama de dianas terapéuticas. Se apreciará que activadores de la expresión y la transcripción regulan la transcripción mediante muchos mecanismos, p. ej., mediante la unión a receptores, estimulando una cascada de transducción de señales, regulando la expresión de factores de transcripción, uniéndose a promotores y potenciadores, uniéndose a proteínas que se unen a promotores y potenciadores, desenrollando ADN, cortando y empalmado pre-ARNm, poliadenilando ARN y degradando ARN. Algunos ejemplos de enzimas incluyen, pero no se limitan a, p. ej., amidasas, aminoácido racemasas, acilasas, deshalogenasas, dioxigenasas, diarilpropano peroxidasas, epimerasas, epóxido hidrolasas, esterases, isomerases, quinasas, glucosa isomerases, glicosidasas, glicosil transferasas, haloperoxidasas, monooxigenasas (p. ej., p450s), lipasas, lignina peroxidasas, nitrilo hidratasas, nitrilasas, proteasas, fosfatasa, subtilisinas, transaminasas y nucleasas.

Algunos de los polipéptidos que pueden ser modificados de acuerdo con determinadas formas de realización descritas en esta memoria están comercialmente disponibles (véase, p. ej., el catálogo de Sigma BioSciences y la lista de precios), y las correspondientes secuencias de proteínas y genes y, típicamente, muchas variantes de los mismos, son bien conocidos (véase, p. ej. GenBank).

Ejemplos de propiedades terapéuticamente relevantes que pueden ser manipuladas o modificadas por cualquiera de las realizaciones descritas en esta memoria (incluyendo la glicosilación y/o pegilación, y/o la incorporación de aminoácidos no naturales) incluyen la semi-vida sérica, la semi-vida útil, estabilidad, inmunogenicidad, actividad terapéutica, detectabilidad (p. ej., mediante la inclusión de grupos informadores (p. ej., etiquetas o sitios de unión de etiquetas) en los aminoácidos no naturales, la especificidad, la reducción de la DL50 u otros efectos secundarios, la capacidad de penetrar en el cuerpo a través del tracto gástrico (p. ej., disponibilidad oral) o similares. Ejemplos de propiedades de diagnóstico relevantes incluyen semi-vida útil, estabilidad, actividad diagnóstica, detectabilidad, especificidad, o similares. Ejemplos de propiedades enzimáticas relevantes incluyen semi-vida útil, estabilidad, especificidad, actividad enzimática, capacidad de producción, o similares.

Una diversidad de otras proteínas también se puede modificar para incluir uno o más aminoácidos antinaturales de acuerdo con determinadas formas de realización descritas en esta memoria. Por ejemplo, las proteínas procedentes de hongos infecciosos, p. ej., *Aspergillus*, especies de *Candida*; bacterias, particularmente *E. coli*, que sirve de modelo para bacterias patógenas, así como bacterias médicamente importantes tales como *Staphylococci* (p. ej., aureus), o *Streptococci* (p. ej., *pneumoniae*); protozoos tales como sporozoa (p. ej., *Plasmodia*), rizópodos (p. ej., *Entamoeba*) y flagelados (*Trypanosoma*, *Leishmania*, *Trichomonas*, *Giardia*, etc.); virus tales como virus de (+)ARN (ejemplos incluyen Poxvirus, p. ej., vaccinia; picornavirus, p. ej., polio; Togavirus, p. ej., rubéola; Flavivirus, p. ej., VHC, y Coronavirus), virus de (-)ARN (p. ej., Rhabdovirus, p. ej., VSV; Paramyxovirus, p. ej., RSV; Orthomyxovirus, p. ej., influenza; Bunyavirus; y Arenavirus), virus ADNds (Reovirus, por ejemplo), virus de ARN a ADN, es decir, Retrovirus, p. ej., VIH y HTLV, y determinados virus de ADN a ARN tales como hepatitis B.

Proteínas relacionadas con la agricultura tales como las proteínas de resistencia a los insectos (p. ej., las proteínas Cry), enzimas productoras de almidón y de lípidos, toxinas de plantas e insectos, proteínas de resistencia a la toxina, proteínas de desintoxicación de micotoxinas, enzimas de crecimiento de las plantas (p. ej., ribulosa 1,5-bisfosfato carboxilasa/oxigenasa, "RUBISCO"), lipoxigenasa (LOX) y fosfoenolpiruvato (PEP) carboxilasa son también dianas adecuadas para la modificación por parte de determinadas formas de realización descritas en esta memoria.

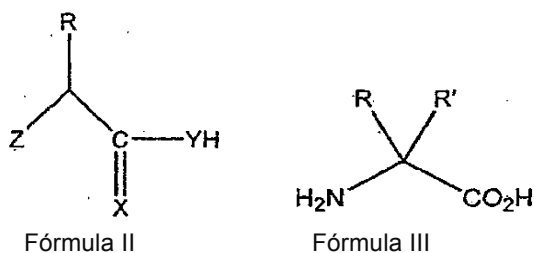
En determinadas formas de realización, la proteína o polipéptido de interés (o parte del mismo) en los métodos y/o composiciones descritos en esta memoria es codificado por un ácido nucleico. Típicamente, el ácido nucleico comprende al menos un codón degenerado, al menos aproximadamente dos, tres, cuatro, cinco, seis, siete, ocho, nueve, o al menos aproximadamente diez o más codones degenerados.

Así, los polipéptidos y polinucleótidos artificiales descritos anteriormente (p. ej., hecho por el hombre y que no se produce de forma natural) son también características de determinadas formas de realización descritas en esta memoria. Un polinucleótido artificial puede incluir, p. ej., (a) un polinucleótido que comprende una secuencia de nucleótidos que codifica un polipéptido artificial; (b) un polinucleótido que es complementario a o que codifica una secuencia de polinucleótidos de (a); (c) un ácido nucleico que se hibrida a un polinucleótido de (a) o (b) bajo condiciones rigurosas, sustancialmente a lo largo de toda la longitud del ácido nucleico; (d) un polinucleótido que es al menos aproximadamente 95%, preferiblemente al menos aproximadamente 98% idéntico a un polinucleótido de (a), (b) o (c); y (e) un polinucleótido que comprende una variación conservativa de (a), (b), (c) o (d).

Aminoácidos antinaturales se describen generalmente arriba. De particular interés para la producción de glicoproteínas como se describe en esta memoria son aminoácidos antinaturales en que R en la Fórmula I incluye un resto que puede reaccionar con un grupo reactivo que está unido a un resto sacárido, para enlazar el resto

sacárido a una proteína que incluye el aminoácido antinatural. Grupos R adecuados incluyen, por ejemplo, ceto-, azido-, hidroxil-, hidrazina, ciano-, halo-, aminooxi-, alqueno-, alquino-, carbonilo, éter, tiol, seleno-, sulfonil-, borato, boronato, fosfo-, fosfono-, fosfina, heterocíclico, enona, imina, (aldehído, éster, tioácido, tioéster, éster estéricamente impedido, hidroxilamina, amina, y similares, o cualquier combinación de los mismos. En algunas formas de realización, los aminoácidos antinaturales tienen un reticulante fotoactivable.

Además de aminoácidos antinaturales que contienen nuevas cadenas laterales, los aminoácidos antinaturales también comprenden opcionalmente estructuras de cadena principal modificada, p. ej., como se ilustra por las estructuras de las Fórmulas II y III:



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

Por ejemplo, muchos aminoácidos antinaturales se basan en aminoácidos naturales tales como tirosina, glutamina, fenilalanina y similares. Análogos de tirosina incluyen tirosinas para-sustituidas, tirosinas orto-sustituidas y meta-sustituidas, en donde la tirosina sustituida comprende un grupo acetilo, un grupo benzoilo, un grupo amino, una hidrazina, una hidroxiamina, un grupo tiol, un grupo carboxi, un grupo isopropilo, un grupo metilo, un hidrocarburo C6-C20 de cadena lineal o ramificada, un hidrocarburo saturado o insaturado, un grupo O-metilo, un grupo poliéter, un grupo nitro, o similares. Además, también se contemplan anillos de arilo polisustituidos. Análogos de glutamina incluyen, pero no se limitan a derivados α-hidroxi, derivados γ-sustituidos, derivados cíclicos, y derivados de glutamina amida-sustituidos. Análogos de fenilalanina ilustrativos incluyen, pero no se limitan a fenilalaninas meta-sustituidas, orto-sustituidas y/o para-sustituidas, en donde el sustituyente comprende un grupo hidroxilo, un grupo metoxi, un grupo metilo, un grupo alilo, un aldehído o grupo ceto, o similares.

Ejemplos específicos de aminoácidos antinaturales incluyen, pero no se limitan a, p-acetil-L-fenilalanina, O-metil-L-tirosina, una L-3-(2-naftil)alanina, una 3-metil-fenilalanina, una O-4-alil-L-tirosina, una 4-propil-L-tirosina, una tri-O-acetil-GlcNAcβ-serina, β-O-GlcNAc-L-serina, una tri-O-acetil-GlcNAc-α-treonina, una α-GalNAc-L-treonina, 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, un fosfonotirosina, una p-yodo-fenilalanina, una p-bromofenilalanina, una p-amino-L-fenilalanina, una isopropil-L-fenilalanina, las listadas más adelante o en otra parte en esta memoria, y similares.

Aminoácidos antinaturales adecuados para su uso en algunos de los métodos descritos en esta memoria también incluyen los que tienen un resto sacárido fijado a la cadena lateral del aminoácido. En una forma de realización, un aminoácido antinatural con un resto sacárido incluye un aminoácido serina o treonina con un resto Man, GalNAc, Glc, Fuc o Gal. Ejemplos de aminoácidos antinaturales que incluyen un resto sacárido incluyen, pero no se limitan a, p. ej., una tri-O-acetil-GlcNAcβ-serina, una β-O-GlcNAc-L-serina, una tri-O-acetil-GalNAc-α-treonina, una α-GalNAc-L-treonina, una O-Man-L-serina, una tetra-acetil-O-Man-L-serina, una O-GalNAc-L-serina, una tri-O-acetil-GalNAc-L-serina, una Glc-L-serina, una tetraacetil-Glc-L-serina, una fuc-L-serina, una tri-acetil-fuc-L-serina, una O-Gal-L-serina, una tetra-acetil-O-Gal-L-serina, una beta-O-GlcNAc-L-treonina, un tri-acetil-beta-GlcNAc-L-treonina, una O-Man-L-treonina, una tetra-acetil-O-Man-L-treonina, una O-GalNAc-L-treonina, una tri-acetil-O-GalNAc-L-treonina, una Glc-L-treonina, una tetraacetil-Glc-L-treonina, una fuc-L-treonina, una tri-acetil-fuc-L-treonina, una O-Gal-L-treonina, una tetra-acetil-O-Gal-L-serina, y similares. Determinadas formas de realización también incluyen

formas de lo anterior no protegidas y acetiladas.

En algunas formas de realización, el diseño de aminoácidos antinaturales es desviado por información conocida sobre los sitios activos de sintetetasas, p. ej., ARNt sintetetasas mutantes externas utilizados para aminoacilar un ARNt mutante externo. Por ejemplo, se proporcionan tres clases de análogos de glutamina, incluidos derivados sustituidos en el nitrógeno de la amida (1), un grupo metilo en la posición γ (2), y un derivado N-C γ cíclico (3). En base a la estructura cristalina de rayos x de GlnRS de *E. coli*, en el que los residuos del sitio de unión clave son homólogos a GlnRS de levadura, los análogos fueron diseñados para complementar una disposición de mutaciones de la cadena lateral de residuos dentro de un intervalo de 10 Å de la cadena lateral de glutamina, p. ej. una mutación del sitio activo Phe233 a un pequeño aminoácido hidrófobo puede ser complementado por el aumento de volumen estérico en la posición de C γ de Gln.

Por ejemplo, 1,5-anhídrido N-ftaloil-L-glutámico (compuesto número 4 en la FIG. 23 del documento WO 02/085923) se utiliza opcionalmente para sintetizar análogos de glutamina con sustituyentes en el nitrógeno de la amida. Véase, p. ej., King y Kidd, *J. Chem. Soc.*, 3315-3319, 1949; Friedman y Chatterji, *J. Am. Chem. Soc.* 81, 3750-3752, 1959; Craig *et al.*, *J. Org. Chem.* 53, 1167-1170, 1988; y Azoulay *et al.*, *Eur. J. Med. Chem.* 26, 201-5, 1991. El anhídrido se prepara típicamente a partir de ácido glutámico, primero mediante protección de la amina como la ftalimida, seguido por calentamiento a reflujo en ácido acético. El anhídrido se abre luego con un cierto número de aminas, resultando en una gama de sustituyentes en la amida. La desprotección del grupo ftaloilo con hidrazina proporciona un aminoácido libre tal como se muestra en la FIG. 23 del documento WO 2002/085923.

La sustitución en la posición γ se consigue típicamente a través de alquilación de ácido glutámico. Véase, p. ej., Koskinen y Rapoport, *J. Org. Chem.* 54, 1859-1866, 1989. Un aminoácido protegido, p. ej. tal como se ilustra por el número de compuesto 5 en la FIG. 24 del documento WO 02/085923 se prepara opcionalmente primero por alquilación del resto amino con 9-bromo-9-fenilfluoreno (PhflBr) (véase, p. ej. Christie y Rapoport, *J. Org. Chem.* 1989, 1859-1866, 1985) y, a continuación, esterificación del resto ácido utilizando O-terc-butil-N,N'-diisopropilisourea. La adición de KN(Si(CH₃)₃)₂ desprotona de forma regioselectiva en la posición β del éster metílico para formar el enolato, que luego se alquila opcionalmente con una gama de yoduros de alquilo. La hidrólisis del éster t-butílico y el grupo Phfl dio el análogo de γ -metil-glutamina deseado (Compuesto número 2 en la FIG. 24 del documento WO 02/085923).

Determinadas otras formas de realización incluyen un inmunoconjugado que comprende un anticuerpo (o su fragmento funcional) específico para una diana (p. ej., una célula diana), el anticuerpo (o fragmento o equivalente funcional del mismo) conjugado, en posiciones específicas predeterminadas, con dos o moléculas terapéuticas, en el que cada una de las posiciones comprenden un aminoácido antinatural. En determinadas formas de realización, los fragmentos de anticuerpo son fragmentos F(ab')₂, Fab', Fab o Fv.

En determinadas formas de realización, las dos o más moléculas terapéuticas son las mismas. En determinadas formas de realización, las dos o más moléculas terapéuticas son diferentes.

En determinadas formas de realización, las moléculas terapéuticas están conjugadas con los mismos aminoácidos antinaturales. En determinadas formas de realización, las moléculas terapéuticas se conjugan con diferentes aminoácidos antinaturales.

En determinadas formas de realización, la naturaleza o química del aminoácido antinatural/enlace de molécula terapéutica permite la escisión del enlace bajo determinadas condiciones tales como condiciones de carácter ácido suaves o débiles (p. ej., aproximadamente pH 4-6, de preferencia aproximadamente pH 5), entorno reductor (p. ej., la presencia de un agente reductor), o cationes divalentes, y está opcionalmente acelerada por el calor.

En determinadas formas de realización, la molécula terapéutica está conjugada a un anticuerpo a través de un enlazador / espaciador (p. ej., una o más repeticiones de metileno (-CH₂-), metilenoxi (-CH₂-O-), metilencarbonilo (-CH₂-CO-), aminoácidos, o combinaciones de los mismos).

Complejos multiproteicos

Los aminoácidos antinaturales se pueden también utilizar para unir dos o más proteínas o sub-unidades de proteínas con funcionalidades únicas. Por ejemplo, los anticuerpos biespecíficos se pueden generar mediante el enlace de dos anticuerpos (o partes funcionales de los mismos o sus derivados, tales como fragmentos Fab, Fab', Fd, Fv, scFv, etc.) a través de aminoácidos antinaturales incorporados en el mismo.

Así, determinadas formas de realización en esta memoria proporcionan métodos para la síntesis de conjugados de multiproteicos. Estos métodos implican, en algunas formas de realización, la incorporación en una primera proteína (p. ej., un primer anticuerpo) un primer aminoácido no natural que comprende un primer grupo reactivo; y

poner en contacto la primera proteína con una segunda proteína (p. ej., un segundo anticuerpo) que comprende un segundo aminoácido antinatural que comprende un segundo grupo reactivo, en donde el primer grupo reactivo reacciona con el segundo grupo reactivo, formando de este modo un enlace covalente que fija la segunda proteína a la primera proteína.

El primer grupo reactivo es, en algunas realizaciones, un resto electrófilo (p. ej., un resto ceto, un resto aldehído, y/o similares), y el segundo grupo reactivo es un resto nucleófilo. En algunas formas de realización, el primer grupo reactivo es un resto nucleófilo y el segundo grupo reactivo es un resto electrófilo (p. ej., un resto ceto, un resto aldehído, y/o similares). Por ejemplo, un resto electrófilo está fijado al aminoácido antinatural del primer anticuerpo y el resto nucleófilo está fijado al aminoácido antinatural del segundo anticuerpo.

Diferentes dominios funcionales de proteínas diferentes pueden estar unidos entre sí de manera similar para crear nuevas proteínas con nuevas funciones (p. ej., nuevos factores de transcripción con una combinación única de unión a ADN y dominios de activación de la transcripción; nuevas enzimas con nuevos dominios reguladores, etc.).

Unión sensible al pH

Muchas interacciones de proteínas son sensibles al pH, en el sentido de que la afinidad de unión de una proteína a su participante en la unión habitual puede cambiar a medida que cambia el pH del entorno. Por ejemplo, muchos ligandos (tales como insulina, interferones, hormona de crecimiento, etc.) se unen a sus respectivos receptores de la superficie celular para provocar la transducción de señales. El complejo ligando-receptor será entonces internalizado por endocitosis mediada por receptor, y pasa por una serie sucesiva de endosomas más y más ácidos. Finalmente, la interacción ligando-receptor se debilita a un cierto pH de carácter ácido (p. ej., aproximadamente pH 5,0), y el ligando se disocia del receptor. Algunos receptores (y quizás algunos ligandos) pueden ser reciclados de nuevo a la superficie celular. Allí, pueden ser capaces de unirse a sus respectivos participantes en la unión normales.

Si la unión sensible al pH puede ser modulada de tal manera que el complejo ligando-receptor puede disociarse a un pH relativamente más alto, entonces determinados ligandos pueden ser disociados antes de sus receptores, y son reciclados preferentemente a la superficie celular en lugar de ser degradados. Esto resultará en una semi-vida incrementada *in vivo* de tales ligandos, lo que podría ser deseable, ya que puede ser necesaria menos insulina para una eficacia igual (o mejor) en pacientes de diabetes. En otras situaciones, puede ser deseable modular la unión sensible al pH, favoreciendo la unión a un pH más bajo.

Por ejemplo, los anticuerpos monoclonales son generalmente muy específicos para sus dianas. Sin embargo, en muchas aplicaciones, tales como en la terapia del cáncer, tienden a provocar determinados efectos secundarios, por ejemplo, uniéndose a tejidos no tumorales. Una razón podría ser que las dianas de tumores contra los que se dirigen anticuerpos monoclonales no se expresan específicamente en células tumorales, sino que también se expresan (aunque puede ser en menor número) en algunas células sanas. Estos efectos secundarios son generalmente indeseables, y hay una necesidad de anticuerpos con una especificidad mejorada.

El pH de la sangre humana está altamente regulado y se mantiene en el intervalo de aproximadamente 7,6-7,8. Por otra parte, las células tumorales tienen un pH extracelular de 6,3-6,5, debido a la acumulación de ácidos metabólicos que se borran de manera ineficaz debido a la vascularización deficiente del tumor. Si la interacción entre un antígeno tumoral y su anticuerpo terapéutico puede ser modulada de tal forma que a pH bajo, se ve favorecida la unión, el tumor-anticuerpo puede tener una especificidad / afinidad / selectividad añadida para esos antígenos tumorales, a pesar de que los mismos antígenos tumorales también se encuentran, en ocasiones, en los tejidos normales.

De hecho, dichos anticuerpos modificados pueden ser deseables no sólo para la terapia del cáncer, sino también deseables para cualquier unión antígeno-anticuerpo que pueda producirse a un nivel inferior al normal de pH.

Técnicas generales

La práctica de las formas de realización descritas en esta memoria empleará, a menos que se indique lo contrario, técnicas convencionales de biología molecular, biología celular, cultivo celular, microbiología y ADN recombinante, que están dentro de la experiencia de la técnica. Tales técnicas se explican completamente en la bibliografía. Véase, por ejemplo, *Molecular Cloning: A Laboratory Manual*, 2ª Ed., comp. por Sambrook, Fritsch y Maniatis (Cold Spring Harbor Laboratory Press: 1989); *DNA Cloning*, Volúmenes I y II (D.N. Glover comp., 1985.); *Oligonucleotide Synthesis* (M.J. Gait comp., 1984.); Mullis *et al.*; patente de EE.UU. N°: 4.683.195; *Nucleic Acid Hybridization* (B.D. Harnes y S.J. Higgins comps. 1984); *Transcription And Translation* (B.D. Harnes y S.J. Higgins comps. 1984); B. Perbal, *A Practical Guide To Molecular Cloning* (1984); el tratado, *Methods in Enzymology* (Academic Press, Inc.,

N.Y.); *Methods in Enzymology*, Vols. 154 y 155 (Wu *et al.* comps.), *Immunochemical Methods In Cell And Molecular Biology* (Mayer y Walker, comps., Academic Press, Londres, 1987).

Además de ello, los textos generales que describen la clonación, la mutación, el cultivo de células en general y similares, incluyen Berger y Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology* vol. 152 Academic Press, Inc., San Diego, Calif. (Berger); Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual* (3ª Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y., 2000 ("Sambrook") y *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, comps., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (suplementado a lo largo de 2002) ("Ausubel")) están todos incorporados como referencia en su totalidad. Estos textos describen mutagénesis, el uso de vectores, promotores y muchos otros temas relevantes relacionados, p. ej., con la generación de ARNt mutante externo, sintetasas mutantes externas, y pares de los mismos.

Diversos tipos de mutagénesis se utilizan en determinadas formas de realización, p. ej., para producir nuevas sintetasas o ARNts. Estos incluyen, pero no se limitan a dirigido al sitio (tal como a través del uso del codón Ámbar, Ocre, Ocre Oscuro u otro codón de terminación), a través de mutagénesis de codones por bamboleo, mutagénesis puntual aleatoria, recombinación homóloga (barajeo de ADN), mutagénesis utilizando moldes que contienen uracilo, mutagénesis dirigida a oligonucleótidos, mutagénesis de ADN modificada con fosforotioato, mutagénesis utilizando ADN dúplex con huecos, o similares. Métodos adecuados adicionales incluyen la reparación del punto de emparejamiento erróneo, la mutagénesis utilizando cepas huéspedes deficientes en la reparación, restricción-selección y restricción-purificación, mutagénesis por delección, mutagénesis por síntesis génica total, reparación de rotura de doble hebra, y similares. La mutagénesis, p. ej., que implica construcciones quiméricas, también se incluyen en determinadas formas de realización. En una forma de realización, la mutagénesis puede ser guiada por la información conocida de la molécula que se produce de forma natural o la molécula que se produce de forma natural alterada o mutada, p. ej., la secuencia, la comparación de secuencias, las propiedades físicas, estructura cristalina o similares.

Los textos y ejemplos anteriores que se encuentran en esta memoria describen estos procesos, así como las siguientes publicaciones y referencias citadas en los mismos: Sieber, *et al.*, *Nature Biotechnology*, 19: 456-460 (2001); Ling *et al.*, *Approaches to DNA mutagenesis: an overview, Anal. Biochem.* 254 (2): 157-178 (1997); Date *et al.*, *Methods Mol. Biol.* 57: 369-374 (1996); I. A. Lorimer, I. Pastan, *Nucleic Acids Res.* 23, 3067-8 (1995); W. P. C. Stemmer, *Nature* 370, 389-91 (1994); Arnold, *Curr. Opin. in Biotech.* 4:450-455 (1993); Bass *et al.*, *Science* 242: 240-245 (1988); Fritz *et al.*, *Nucl. Acids Res.* 16: 6987-6999 (1988); Kramer *et al.*, *Nucl. Acids Res.* 16: 7207 (1988); Sakamar y Khorana, *Nucl. Acids Res.* 14: 6361-6372 (1988); Sayers *et al.*, *Nucl. Acids Res.* 16:791-802 (1988); Sayers *et al.*, *Nucl. Acids Res.* 16: 803-814 (1988); Carter, *Methods in Enzymol.* 154: 382-403 (1987); Kramer y Fritz *Methods in Enzymol.* 154:350-367 (1987); Kunkel, *Nucleic Acids & Mol. Biol.* (Eckstein, F. y Lilley, D. M. J. comps., Springer Verlag, Berlín) (1987); Kunkel *et al.*, *Methods in Enzymol.* 154, 367-382 (1987); Zoller y Smith, *Methods in Enzymol.* 154:329-350 (1987); Carter, *Biochem. J.* 237:1-7 (1986); Eghtedarzadeh y Henikoff, *Nucl. Acids Res.* 14: 5115 (1986); Mandeck, *PNAS, USA*, 83:7177-7181 (1986); Nakamaye y Eckstein, *Nucl. Acids Res.* 14: 9679-9698 (1986); Wells *et al.*, *Phil. Trans. R. Soc. Lond. A* 317: 415-423 (1986); Botstein y Shortle, *Science* 229:1193-1201(1985); Carter *et al.*, *Nucl. Acids Res.* 13: 4431-4443 (1985); Grundström *et al.*, *Nucl. Acids Res.* 13: 3305-3316 (1985); Kunkel, *PNAS, USA* 82:488-492 (1985); Smith, *Ann. Rev. Genet.* 19:423-462(1985); Taylor *et al.*, *Nucl. Acids Res.* 13: 8749-8764 (1985); Taylor *et al.*, *Nucl. Acids Res.* 13: 8765-8787 (1985); Wells *et al.*, *Gene* 34:315-323 (1985); Kramer *et al.*, *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer *et al.*, *Cell* 38:879-887 (1984); Nambiar *et al.*, *Science* 223: 1299-1301 (1984); Zoller y Smith, *Methods in Enzymol.* 100:468-500 (1983); y Zoller y Smith, *Nucl. Acids Res.* 10:6487-6500 (1982). Detalles adicionales sobre muchos de los métodos anteriores se pueden encontrar en *Methods in Enzymology* Volumen 154, que también describe controles útiles para los problemas de resolución de problemas con diversos métodos de mutagénesis.

Los oligonucleótidos, p. ej., para uso en mutagénesis en determinadas formas de realización, p. ej., mutando bancos de sintetasas, o alterando ARNts, se sintetizan típicamente por vía química de acuerdo con el método del triéster de fosforoamidita en fase sólida descrito por Beaucage y Caruthers, *Tetrahedron Letts.* 22 (20): 1859-1862, (1981), p. ej., utilizando un sintetizador automatizado, según se describe en Needham-VanDevanter *et al.*, *Nucl. Acids Res.*, 12: 6159-6168 (1984).

Además, esencialmente cualquier ácido nucleico puede ser ordenado de forma personalizada o estandarizada de cualquiera de una diversidad de fuentes comerciales, como The Midland Certified Reagent Company, The Great American Gene Company, ExpressGen Inc., Operon Technologies Inc. (Alameda, Calif.) y muchas otras.

Todas las formas de realización descritas en esta memoria están destinadas a ser capaces de ser combinadas con una o más otras formas de realización, incluso para las que se describen en diferentes secciones de la descripción.

Todas las anteriores patentes de EE.UU., publicaciones de solicitudes de patente de EE.UU., solicitudes de patente de EE.UU., las patentes extranjeras, solicitudes de patente extranjeras y publicaciones no patente a las que se alude en esta memoria y/o se listan en la Hoja de Datos de Solicitud, se incorporan a esta memoria como referencia en su totalidad.

5

EJEMPLOS

Estos ejemplos ilustran la incorporación de un análogo de aminoácido en proteínas en las posiciones codificadas por codones que normalmente codifican específicamente fenilalanina (Phe) o codifican específicamente triptófano (Trp). Un diagrama esquemático se muestra en la Figura 1. Se pueden utilizar estrategias similares para cualesquiera otros análogos.

Phe es codificada por dos codones, UUC y UUU. Ambos codones son leídos por un único ARNt que está equipado con la secuencia de anticodón GAA. Por lo tanto, el codón CUU es reconocido a través del apareamiento de bases estándar de Watson-Crick entre codón y anticodón; UUU se lee a través de un par de bases de bamboleo G-U en la primera posición del anticodón (Crick, *J. Mol. Biol.* 19: 548, 1966; Soll y Rajbhandary, *J. Mol. Biol.* 29: 113, 1967). La desnaturalización térmica de los dúplex de ARN ha proporcionado estimaciones de las energías libres de Gibbs de fusión de pares de bases G-U, G-C, A-U, y A-C como 4,1, 6,5, 6,3 y 2,6 kcal/mol, respectivamente, a 37°C. Así, el par de bases de bamboleo, G-U es menos estable que el par de bases de Watson-Crick, A-U. Un ARNt^{Phe} modificado, equipado con el anticodón AAA (ARNt^{Phe}_{AAA}) fue diseñado para leer el codón UUU, y se predijo que leía tales codones de una manera más rápida que el ARNt^{Phe}_{GAA} de tipo salvaje.

La dihidrofolato reductasa murina (mDHFR), que contiene nueve residuos Phe, fue elegida como la proteína de ensayo. El plásmido de expresión pQE16 codifica mDHFR bajo el control de un promotor de bacteriófago T5; la proteína está equipada con una etiqueta hexahistidina C-terminal (HIS₆) para facilitar la purificación a través de cromatografía de afinidad con metal inmovilizada.

Los PheRS de levadura modificada (mu-yPheRS) se preparó mediante la introducción de una mutación Thr415Gly o Thr415Ala en la subunidad α de la sintetasa (Datta et al., *J. Am. Chem. Soc.*, 124: 5652, 2002). La cinética de activación de Nal y Phe por parte de mu-yPheRS se analizaron *in vitro* a través del ensayo de intercambio de adenosina trifosfato-pirofosfato. Se encontró que la constante de especificidad (k_{cat}/K_M) para la activación de Nal por parte de mu-yPheRS era $1,55 \times 10^{-3} \text{ (s}^{-1} \text{ M}^{-1}\text{)}$, 8 veces mayor que la de Phe. Por lo tanto, cuando la relación de Nal a Phe en el medio de cultivo es alta, yARNt^{Phe}_{AAA} debería ser cargado predominantemente con Nal. Además, el mutante T415G fue generado por mutagénesis de cuatro cebadores.

Tanto la sintetasa de *E. coli* como de levaduras son $\alpha_2\beta_2$ hetero-tetrameros y los pesos moleculares para cada una de las subunidades son bastante diferentes $\alpha(\text{ePheRS}) = 37 \text{ kDa}$; $\alpha(\text{yPheRS}) = 57 \text{ kDa}$; $\beta(\text{ePheRS}) = 87 \text{ kDa}$; y $\beta(\text{yPheRS}) = 67,5$, todos aproximadamente.

Por lo tanto, los siguientes ejemplos se proporcionan como modo de ilustración y no a modo de limitación.

EJEMPLO 1

Con el fin de alterar la capacidad de una aminoacil ARNt sintetasa de levadura, el gen yPheRS se amplificó a partir del plásmido molde pUC-ASab2 que codifica subunidades alfa y beta del gen PheRS. La amplificación se llevó a cabo con una secuencia intergénica de 14 pares de bases que contiene un sitio de reiniciación de la traducción aguas arriba del código de inicio ATG del gen de la subunidad beta.

Se utilizaron los siguientes oligo-cebadores para la PCR: 5'-CGA TTT TCA CAC AGG ATC CAG ACC ATG ATT CTA G-3' (SEQ ID NO: 7) (cebador 1 con el sitio de restricción *Bam*HI) y 5'-GAC GGC CAG TGA ATT CGA GCT CGG TAC-3' (SEQ ID NO: 8) (cebador 2 con el sitio de restricción *Kpn*II). El producto de ADN resultante se introdujo en los sitios *Bam*HI y *Kpn*II de pQE32 para dar pQE32- yFRS. El polinucleótido yPheRS mutante fue generado mediante el uso de mutagénesis del cebador mediante técnicas estándares.

En síntesis, se sintetizaron dos oligonucleótidos complementarios: 5'-CTA CCT ACA ATC CTT ACG GCG AGC CAT CAA TGG AAA TC-3' (SEQ ID NO: 9) (cebador 3) y 5'-GAT TTC CAT TGA TGG CTC GCC GTA AGG ATT GTA GGT AG-3' (SEQ ID NO: 10) (cebador 4) para llevar la mutación específica en la posición 415 de la subunidad alfa del polinucleótido yPheRS.

EJEMPLO 2

El plásmido pQE32-yFRS, y pQE32-T415G, pQE32-T415A fueron cada uno de ellos transformados en la cepa de células huésped BLR de *E. coli* (de NOVAGEN®) para formar las cepas de expresión BLR(pQE32-yFRS_ y BLR(pQE32-T415G). Las células se cultivaron en medio LB, a una concentración de 0,6 a una DO 600. Entonces la expresión se indujo con IPTG 1 mM durante 4 horas. Se recogieron las células y los polipéptidos se purificaron por medio de una columna de afinidad de níquel-ácido nitrilotriacético bajo condiciones nativas de acuerdo con el protocolo del fabricante (QIAGEN®). El imidazol en el tampón de elución se separó mediante columna de desalación, y los polipéptidos se eluyeron en un tampón que contiene Tris-HCl 50 mM (pH = 7,5), DTT 1 mM. Partes alícuotas de polipéptidos se almacenaron a -80°C con glicerol al 50%.

EJEMPLO 3

Se utilizó la reacción de intercambio ATP-PP_i dependiente del aminoácido para evaluar la activación de aminoácidos no naturales por parte de yPheRS. El ensayo se realizó en 200 microlitros de tampón de reacción que contiene HEPES 50 mM (pH = 7,6), MgCl₂ 20 mM, DTT 1 mM, ATP 2 mM y [³²P] -PP_i 2 mM, con una actividad específica de 0,2 - 0,5 TBq/mol. Dependiendo de la actividad de los diversos aminoácidos no naturales con la sintetasa, la concentración de aminoácidos varió de 10 microM a 5 mM, y la concentración de enzimas varió de 10 nM a 100 nM. Partes alícuotas de 20 microlitros fueron retiradas de la disolución de reacción en diversos instantes y fueron inactivadas en 500 microlitros de disolución tampón que contiene NaPP_i 200 mM, HClO₄ al 7% p/v y carbón vegetal activado al 3% p/v. El carbón vegetal se centrifugó y se lavó dos veces con 500 microlitros de NaPP_i 10 mM y disolución de HClO₄ al 0,5%. El ATP radiomarcado absorbido por el carbón vegetal se cuantificó a través de métodos de centelleo líquido. Las constantes de especificidad se calcularon por ajuste de regresión no lineal de los datos a un modelo de Michaelis-Menten. Los parámetros cinéticos para el intercambio ATP-PP_i de aminoácidos por las yPheRS (T415G), yPheRS de tipo salvaje y variante yPheRS_naph se muestran en la siguiente Tabla.

Amino-ácido	Enzima	K _m (μM)	K _{cat} (s ⁻¹)	K _{cat} /K _m (s ⁻¹ M ⁻¹)	K _{cat} /K _m (relativa)
Phe	T415G	55 ± 14	0,202 ± 0,11	3512 ± 1134	1 ^a
Trp	T415G	2,83 ± 1,6	0,153 ± 0,003	63190 ± 34590	18 ^a
2Nal	T415G	7,03 ± 0,14	0,208 ± 0,04	29535 ± 5848	8,4 ^a
Phe	tipo salvaje	3,85 ± 0,99	0,181 ± 0,011	50994 ± 22655	15 ^a
Phe	naf	11010 ± 2688	0,0095 ± 0,0021	0,855 ± 0,007	1 ^b
Trp	naf	1424 ± 597	0,0035 ± 0,0009	2,52 ± 0,44	2,9 ^b
2Nal	naf	2030 ± 691	0,030 ± 0,018	14,54 ± 4,22	17 ^b

EJEMPLO 4

El plásmido de expresión, pQE16 (QIAGEN®) se utilizó con dihidrofolato reductasa murina (mDHFR) de polipéptido marcador con un gen de la etiqueta de hexa-histidina C-terminal bajo el control de un promotor del bacteriófago T5 y el terminador t₀.

Un codón Ámbar (TAG) se colocó en la 38ª posición de mDHFR utilizando un kit de mutagénesis QUICK-CHANGE®. Se utilizaron dos oligo-cebadores complementarios (5'-CCG CTC AGG AAC GAG TAG AAG TAC TTC CAA AGA ATG-3' (SEQ ID NO: 11) y 5'-CAT TCT TTG GAA GTA CTT CTC CTA GTT CCT GAG CGG-3' (SEQ ID NO: 12)) para producir pQE16am. El gen yPheRS mutante T415G se amplificó a partir pQE32-T415G y un promotor *tac* constitutivo con un sitio de unión al represor *lac* abolido se añadió aguas arriba del codón de inicio del gen.

Toda la casete de expresión de T415G se insertó en el sitio *PvuII* de pQE16 para producir pQE16am-T415G. El ARNt supresor de levadura mutante (*mutARNt*^{Phe}(CUA)) se expresó constitutivamente bajo el control del promotor *lpp*. La casete de expresión de *mutARNt*^{Phe}(CUA) se insertó en el plásmido represor pREP4 para formar pREP4-ARNt utilizando métodos conocidos.

Una cepa bacteriana auxotrófica de fenilalanina (Phe); AF (K10, Hfr(Cavalli) *pheS93rel-1 tonA22 thi T2^R pheA18*) se utilizó como una cepa huésped. Una doble cepa doble auxotrófica Phe/Trp AFW (K10, Hfr (Cavalli) *pheS13rel-1tonA22 thi T2^R pheA18, trpB114*) y una cepa triple auxotrófica *Phe/Trp/Lys* AFWK (K10, Hfr (Cavalli) *pheS13rel-1 tonA22 thi T2^R pheA18, trpB114, lysA*) se prepararon mediante transducción mediada por el fago P1 con los transposones *trpB::Tn10* y *lysA::Tn10*.

EJEMPLO 5

Las cepas de células huésped auxotróficas AF, AFW y AFWK se transforman cada una de ellas con el plásmido pQE16am-T415G y pREP4-ARNt para proporcionar las cepas de expresión AF[pQE16am-T415G/pREP4-ARNt] y AFW[pQE16am-T415G/pREP4-ARNt], respectivamente. Las células se cultivaron en medio mínimo M9 suplementado con glucosa, tiamina, MgSO₄, CaCl₂, 20 aminoácidos (20 mg/L), antibióticos (kanamicina y ampicilina). Cuando las células alcanzaron una lectura a una DO₆₀₀ de 1,0, se sedimentaron por centrifugación, se lavaron dos veces con NaCl frío al 0,9% y se cambiaron a medio M9 suplementado que contenía 17 aminoácidos (20 mg/L), aminoácido 3 mM no natural de interés, y las concentraciones indicadas de Phe, Trp y Lys. La expresión de proteínas se indujo mediante la adición de IPTG (1 mM). Después de 4 horas, las células se sedimentaron y la proteína se purificó por medio de una etiqueta de hexa-histidina C-terminal y una columna de centrifugación Nickel-NTA de acuerdo con las instrucciones del fabricante (QIAGEN®).

EJEMPLO 6

mDHFR mutante se purificó en condiciones desnaturalizantes y se eluyó con tampón estándar (urea 8 M, NaH₂PO₄ 100 mM, Tris 10 mM, pH 4,5). Los polipéptidos fueron digeridos con tripsina con 10 microlitros de la disolución diluida en 90 microlitros de (NH₄)₂CO₃ 75 mM y el pH se ajustó a 8. Se añadieron dos microlitros de tripsina modificada (0,2 microgramos/microlitro) y la muestra se incubó a temperatura ambiente durante la noche. Los polipéptidos se digirieron con endoproteinasa con Lys-C, 10 microlitros de disolución diluida en 90 microlitros de Tris-HCl 25 mM, pH 8 y EDTA 1 mM. A continuación, se añadieron 2 microlitros de Lys-C (0,2 microgramos/microlitro; CALBIOCHEM®) y la reacción se incubó a 37° durante 10 horas. La reacción de digestión se detuvo mediante la adición de 2 microlitros de ácido trifluoroacético (TFA). La disolución se purificó por medio de ZIPTIP_{C18}® (MILLIPORE®) y los péptidos digeridos se eluyeron con 3 microlitros de CH₃CN al 50%, TFA al 0,1%. Se utilizó un microlitro para el análisis por espectrometría de masas desorción/ionización láser asistida por matriz (MALDI-MS) con ácido alfa-ciano-4-hidroxicinámico y ácido 2,5-dihidroxibenzoico como matriz. El análisis se realizó utilizando un espectrómetro de masas Voyager DE PRO MALDI-TOF de PERSEPTIVE BIOSYSTEMS® en modo lineal y de iones positivos.

Se realizó un análisis LC-MS/MS de péptidos digerido con proteasa en espectrometría de masas con trampa de iones FINNIGAN® LCQ con bomba de HPLC y sonda ESI. La secuenciación de masas en tándem se llevó a cabo mediante fragmentación del ion precursor con m/z correspondiente a fragmento digerido con proteasa que incluye el residuo en la posición 38 de mDHFR mutante.

EJEMPLO 7

Los plásmidos se construyeron para yPheRS de tipo salvaje e yPheRS (T415G) según se describe en el Ejemplo 1 en esta memoria. Además, el gen *lysS* de *E. coli* se amplificó mediante PCR a partir del plásmido molde pXLLysKS1, utilizando los siguientes cebadores: 5'-GCA CTG ACC ATG GCT GAA CAA CAC GCA CAG-3' (SEQ ID NO: 13) (con el sitio de restricción *NcoI*) y 5'-GGA CTT CGG ATC CTT TCT GTG GGC GCA TCG C-3' (SEQ ID NO: 14) (con el sitio de restricción *BamHI*). El ADN resultante se introdujo en los sitios *NcoI* y *BamHI* de pQE60 para producir pQE60-eLysS. Las enzimas clonadas contienen etiquetas hexaHistidina N-terminal o C-terminal para facilitar la purificación de proteínas.

En las dos primeras reacciones, (cebador 1 y cebador 4) y (cebador 2 y cebador 3) se añadieron en tubos individuales y se generaron dos fragmentos de ADN a partir de estas dos reacciones PCR. Con la mezcla de dos productos de reacción y cebadores externos adicionales, se obtuvo un fragmento de ADN de 3400 pb. El fragmento se purificó por métodos estándares y se digirió con *BamHI* y *KpnI* y se insertó en pQE32 para producir pQE32-T415G. Las enzimas PheRS clonadas contenían una etiqueta de la secuencia hexa-histidina N-terminal conocida para la purificación. El gen yPheRS completo se secuenció en ADN para su verificación.

EJEMPLO 8

El plásmido pQE32-T415A y pQE60-eLysS se co-transformaron individualmente con un plásmido represor pREP4 en una cepa BLR de *E. coli* para formar cepas de expresión BLR (pQE32-yFRS), BLR (pQE32-T415G), BLR (pQE32-T415A) y BLR (pQE60-eLysS). La sobre-expresión se llevó a cabo en medios 2x YT con 100 microgramos/mL de ampicilina y 35 microgramos/mL de kanamicina. A una DO₆₀₀ = 0,6, la expresión de variantes yPheRS y la lisil-ARNt sintetasa de *E. coli* codificada por el gen *lysS* (eLysS) se indujeron con IPTG 1 mM. Después de expresión durante 4 horas, se recogieron las células y las proteínas se purificaron sobre una columna de afinidad de níquel-ácido nitrilotriacético bajo condiciones nativas de acuerdo con el protocolo del fabricante (QIAGEN®). El imidazol en el tampón de elución se separó mediante columna de desalación y los polipéptidos se eluyeron en un tampón que contenía Tris-HCl 50 mM (pH 7,5), DTT 1 mM. Partes alícuotas de los polipéptidos se almacenaron a -80°C con glicerol al 50%. Las concentraciones de variantes de yPheRS y eLysS se determinaron por absorbancia UV a 280 nm.

EJEMPLO 9

El péptido 38 (residuos 26-39; codónK NGDLPWPPLRNEámbar) (SEQ ID NO: 15) contiene el codón ámbar en la posición número 38. Los péptidos (K38S, K38L), el péptido W38 y el péptido pBrF (Z)38 fueron separados y detectados por MS. Los polipéptidos se sintetizaron en células huésped auxótrofas triples con (a) ARNt^{Phe}_{CUA} e yPheRS (T415G); (b) ARNt^{Phe}_{CUA_UG} e yPheRS (T415G); (c) ARNt^{Phe}_{CUA} e yPheRS (T415A); (d) ARNt^{Phe}_{CUA_UG} e yPheRS (T415A) o (e) en una sola cepa auxotrófica con ARNt^{Phe}_{CUA_UG} e yPheRS (T415A). Los medios mínimos de expresión se suplementaron con pBrF 6,0 mM, Trp 0,01 mM, Lys 1,0 mM, Phe 0,03 mM (a y b) o Phe 0,01 mM (c, d y e) y 25 mg/L de 17 aminoácidos, los resultados se muestran en la Figura 9.

EJEMPLO 10

Se utilizó la reacción de intercambio de ATP-PP_i dependiente de aminoácidos para evaluar la activación de análogos de aminoácidos mediante yPheRS según se describe en los Ejemplos anteriores. En síntesis, una parte alícuota de 200 microlitros de tampón de reacción contiene HEPES 50 mM (pH 7,6), MgCl₂ 20 mM, DTT 1 mM, ATP 2 mM y ³²P-pirofosfato (PP_i) 2 mM con actividad específica de 10-50 Ci/mol. Dependiendo de la actividad de análogos por parte de la sintetasa, la concentración de aminoácidos varió de 10 microM a 2,5 mM y la concentración de enzimas varió de 10nM a 100 nM. Partes alícuotas de 20 microlitros fueron retiradas de la disolución de reacción en diversos instantes y se inactivaron en 500 microlitros de solución tampón que contiene NaPP_i 200 mM, HClO₄ al 7% p/v y carbón vegetal activado al 3% p/v. El carbón vegetal se centrifugó y se lavó dos veces con 500 microlitros de NaPP_i 10 mM y disolución de HClO₄ al 0,5%. El ATP radiomarcado absorbido por el carbón vegetal se cuantificó a través de métodos de centelleo líquido. Las constantes de especificidad se calcularon por ajuste de regresión no lineal de los datos a un modelo de Michaelis Menten.

Los resultados de los parámetros cinéticos se muestran en la Tabla I. Las sustituciones en el anillo de indol (especialmente en la posición 6^a) eran muy favorables para algunos análogos (8-10).

Tabla I: Parámetros cinéticos para el intercambio ATP-PP_i de aminoácidos ilustrativos (1-11) por parte de PheRS de levadura mutante externa

Amino-ácido	Enzima	Km (μm)	kcat (s ⁻¹)	Kcat/Km (s ⁻¹ M ⁻¹)	Kcat/Km (relativa)*
1	T415G	264 ± 42	0,05 ± 0,002	184 ± 30	1
2	T415G	22 ± 3	0,03 ± 0,001	1,538 ± 228	8
3	T415G	12 ± 2	0,05 ± 0,001	4,365 ± 797	24
4	T415G	11 ± 3	0,05 ± 0,002	4,558 ± 1,186	25
5	T415G	757 ± 149	0,4 ± 0,003	48 ± 10	1\4
6	T415G	20 ± 5	0,30 ± 0,006	15,000 ± 4,063	82
7	T415G	27 ± 2	0,04 ± 0,001	1,550 ± 125	8
8	T415G	20 ± 8	0,20 ± 0,018	10,256 ± 4,562	56
9	T415G	8 ± 4	0,55 ± 0,097	70,876 ± 34,843	385
10	T415G	31 ± 18	0,06 ± 0,005	1,939 ± 1,149	10
11	T415G	94 ± 50	0,05 ± 0,006	533 ± 293	3
1	T415G	68 ± 20	0,52 ± 0,093	7,627 ± 2,664	41

En que los aminoácidos se indican como en la Figura 2.

EJEMPLO 11

El ARNt supresor ámbar de levadura mutante (yARNt^{Phe}_{CUA}) fue constitutivamente expresado bajo el control del promotor *lpp*. La casete de expresión de yARNt^{Phe}_{CUA} se insertó en el plásmido represor pREP4 para formar pREP4-yARNt como se describe anteriormente en los Ejemplos en esta memoria. El supresor yARNt^{Phe}_{CUA_30U40G} de levadura mutante (yARNt^{Phe}_{CUA_UG}) se construyó a partir de yARNt^{Phe}_{CUA} mediante el uso de un kit de mutagénesis QUICK-CHANGE®. Dos oligonucleótidos complementarios, designados como cebador UG-f (5'-GAA CAC AGG ACC TCC ACA TTT AGA GTA TGG CGC TCT CCC-3') (SEQ ID NO: 16) para el cebador directo y el cebador UG-r (5'-GGG AGA GCG CCA TAC AAA TGT TCT GGA GGT CCT GTG TTC-3') (SEQ ID NO: 17) para el cebador inverso se sintetizaron para portar la mutación específica, ya sea en la posición 30 o la posición 40 del ARNt supresor de levadura mutante. El plásmido resultante que porta el gen que codifica yARNt^{Phe}_{CUA_UG} se designa pREP4-yARNt_{UG}. Con el fin de construir los plásmidos para la transcripción *in vitro* de yARNt^{Phe}, los genes yARNt^{Phe}_{CUA} e yARNt^{Phe}_{CUA_UG} se amplificaron a partir del plásmido molde pREP4-yARNt y pREP4-yARNt_{UG}, respectivamente. Al final de la secuencia de ARNt, se insertó un sitio *Bst*NI para producir el

transcrito exacto de yARNt^{Phe}. Se añadió una secuencia promotora T7 para la transcripción *in vitro* de yARNt^{Phe} por parte de una T7 ARN polimerasa. Para la PCR se utilizaron los siguientes cebadores: 5'-CTG GGT AAG CTT CGC TAA GGA TCT GTG CTG GCC CGA ACT CTG-3' (SEQ ID NO: 18) (con sitios de restricción *HindIII* y *BstNI*) y 5'-GAT TAC GGA TTC CTA ATA CGA CTC ACT ATA GCG GAC TTA GCT C-3' (SEQ ID NO: 19) (con el sitio de restricción *EcoRI* y una secuencia de promotor T7). El ADN resultante se introdujo en los sitios *HindIII* y *EcoRI* de pUC18 para proporcionar pUC18-yARNt^{Phe}_{CUA} y pUC18-yARNt^{Phe}_{CUA_UG}.

Con el fin de facilitar la manipulación del ADN, se eliminó un sitio de escisión *BstNI* cerca de la secuencia del promotor T7 de pUC18-yARNt^{Phe}_{CUA} para aumentar el tamaño del fragmento de ADN que contiene el gen yARNt^{Phe}_{CUA} de 180 pb a 500 pb después de la digestión con *BstNI*. Dos oligonucleótidos complementarios, 5'-CGG AAG CAG AAA GTG TAA AGA GCG GGG TGC CTA ATT AGT G-3' (SEQ ID NO: 20) para el cebador directo y 5'-CAC TCA TTA GGC ACC CCG CTC TTT ACA CTT TCC GCT TAT G-3' (SEQ ID NO: 21) para el cebador inverso, se sintetizaron para portar la mutación específica.

EJEMPLO 12

El ADN linearizado se preparó mediante digestión con *BstNI* de pUC18-yARNt^{Phe}_{CUA} y pUC18-yARNt^{Phe}_{CUA_UG} como se ha descrito previamente (Véase Sampson, Uhlenbeck, *PNAS USA* 85: 1033-1037 (1988)). La transcripción *in vitro* de moldes de ADN linearizado y la purificación de transcritos se realizaron como se ha descrito previamente (Véase Nowak, *et al.*, *Ion Channels* pt. B 293: 504-529 (1998)). La transcripción *in vitro* del ADN linearizado para producir transcritos de ARNt 76rnero se realizó con el kit T7-MEGASHORTSCRIPT® de AMBION®. Los transcritos se aislaron con una extracción 25:24:1 fenol:CHCl₃:alcohol isoamílico (PCI). La capa orgánica se volvió a extraer con agua y se realizó una extracción 24:1 CHCl₃:alcohol isoamílico (CI) en las capas acuosas. A continuación, la capa acuosa se mezcló con un volumen igual de isopropanol, se precipitó durante la noche a -20°C, se sedimentó, se secó y se volvió a disolver en agua. Los nucleótidos que no habían reaccionado en la disolución de ARNt fueron eliminados utilizando CHROMA SPIN-30® DEPC-H₂O (BD Bioscience®). Las concentraciones de los transcritos se determinaron mediante absorbancia UV a 260 nm. La aminoacilación de yARNt^{Phe}_{GAA} de tipo salvaje con Phe y Trp mediante variantes de yPheRS se realizó como se ha descrito previamente (Véase Sampson, Uhlenbeck, *PNAS USA* 85: 1033-1037 (1988)). Se llevaron a cabo reacciones de aminoacilación en el tampón que contiene HEPES 30 mM (pH 7,45), MgCl₂ 15 mM, DTT 4 mM, KCl 25 mM KCl y ATP 2 mM a 30°C, en volúmenes de reacción de 100 microlitros. En el ensayo se utilizó ARNt total de levadura purificado a una concentración final de 4 mg/mL (concentración de yARNt^{Phe}_{GAA} aproximadamente 2,24 microM). Para la aminoacilación de Phe, se utilizaron [³H]-Phe 13,3 microM (5,3 Ci/mmol) y variantes de yPheRS 80 nM; para la aminoacilación de Trp, se utilizaron [³H]-Trp 3,3 microM (30,0 Ci/mmol) y variantes de yPheRS 160 nM. La aminoacilación de transcritos de yARNt^{Phe} se realizó en volúmenes de reacción de 100 microlitros en tampón que contenía potasio-HEPES 100 mM (pH 7,4), MgCl₂ 10 mM, DTT 1 mM, EDTA 0,2 mM, ATP 2 mM, y 4 unidades/mL de pirofosfatasa inorgánica de levadura a 37°C para eLysS. Para la aminoacilación de Lys, se utilizaron 4 microM de transcrito de yARNt^{Phe}, [³H] Lys 1,1microM (91 Ci/mmol) y eLysS 80 nM. Los ARNts fueron asociados antes de su uso por calentamiento hasta 85°C durante 4 minutos en el tampón de reasociación (Tris 60 mM, pH 7,8, MgCl₂ 2 mM) seguido de enfriamiento lento a temperatura ambiente. Las reacciones se iniciaron mediante la adición de la enzima y se inactivaron partes alícuotas de 10 microlitros mediante la detección en discos de filtro Whatman empapados con TCA al 5%. Los filtros se lavaron durante tres periodos de 10 minutos en TCA al 5% enfriado con hielo, se lavaron en etanol al 95% enfriado con hielo, y se contaron mediante métodos de centelleo líquido.

EJEMPLO 13

La construcción del plásmido para la incorporación *in vivo* de un aminoácido no natural se realizó en una cepa doble auxotrófica Phe/Trp, AFW, y una cepa triple auxotrófica Phe/Trp/Lys, AFWK. Las cepas auxotróficas se construyeron a partir de una cepa auxotrófica Phe, AF (K10, Hfr(Cavalli) *pheS13rel-1 tonA22 thi T2^R pheA18, trpB114*) (Véase Furter, *Prot. Sci.*, 7: 419-426 (1998)) mediante transducción del trasposón mediada por el fago P1. Un vector pQE16 (QIAGEN®) se eligió como el plásmido de expresión, que codifica una proteína marcadora dihidrofolato reductasa murina (mDHFR) con el gen de la etiqueta hexa-histidina C-terminal bajo el control de un promotor de bacteriófago T5 y un terminador t₀. Se utilizó un kit de mutagénesis Quick-change para colocar un codón ámbar (TAG) en la 38ª posición de mDHFR con dos oligonucleótidos complementarios (5'-CCG CTC GAG AGG AAC TAG AAG TAC TTC CAA AGA ATG-3' (SEQ ID NO: 11); 5'-CAT TCT TTG GAA GTA CTT CTC CTA GTT CCT GAG CGG-3') (SEQ ID NO: 12) para producir pQE16am. Los genes de yPheRS mutantes T415G y T415A se amplificaron a partir de pQE32-T415G y pQE32-T415A y se añadió un promotor *tac* constitutivo con un sitio de unión al represor *lac* abolido aguas arriba del codón de inicio del gen. Toda la casete de expresión de T415G y T415A se insertó en el sitio *PvuII* de pQE16am-T415G y pQE16am-T415A.

EJEMPLO 14

Las cepas bacterianas auxotróficas AF, AFW y AFWK se transformaron con el plásmido pQE16am que contiene variantes de yPheRS y vectores pREP4-yARNt que contienen variantes de yARNt para investigar incorporación de pBrF. Las cepas de expresión de *E. coli* se cultivaron en medio mínimo M9 suplementado con glucosa, tiamina, MgSO₄, CaCl₂, 20 aminoácidos (a 25 mg/L) antibióticos (35 microgramos/mL de kanamicina y 100 microgramos/mL de ampicilina). Cuando las células alcanzaron una DO₆₀₀ de 0,8-1,0, se sedimentaron por centrifugación, se lavaron dos veces con NaCl al 0,9% frío y se desplazaron a medios de expresión suplementados con 17 aminoácidos (a 20 mg/L), 6 mM de pBrF (p-bromo-fenilalanina) o plodoF (p-yodo-fenilalanina), y las concentraciones indicadas de fenilalanina, triptófano y lisina. La expresión de proteínas fue inducida por la adición de IPTG 1 mM. Después de cuatro horas de expresión, las células se sedimentaron por centrifugación, y la proteína se purificó en virtud de la etiqueta de hexa-histidina C-terminal a través de una columna de centrifugación de nickel-NTA de acuerdo con las instrucciones del fabricante (QIAGEN®). Después de la purificación, los niveles de expresión de mDHFR se determinaron mediante absorbancia UV a 280 nm

EJEMPLO 15

Se realizó un análisis LC-MS/MS de digestos tripticos de mDHFR en una espectrometría de masas con trampa de iones Finnigan LCQ con bomba de HPLC y sonda ESI. La mDHFR mutante purificada en condiciones desnaturalizantes se encontraba en tampón de elución (urea 8 M, NaH₂PO₄ 100 mM, Tris 10 mM, pH 4,5). Para la digestión de tripsina, 10 microL de la disolución se diluyeron en 90 microL de (NH₄)₂CO₃ 75 mM. Se añadió un microlitro de tripsina modificada (0,2 microgramos/microlitro). La muestra se incubó a 37° C durante 2 a 6 horas. La reacción de digestión se detuvo mediante la adición de 12 microL de disolución de TFA al 5%. La disolución de péptido digerido se sometió a desalación con una columna C18 Vydac Microspin (el grupo Nest) y se eluyó con 50 microL de 80% de acetonitrilo y 20% de ácido fórmico al 0,1% p/v. Se secó la disolución de péptido digerido eluida a partir de la columna Microspin, se redisolvió en 10% de acetonitrilo y 90% de disolución de TFA 0,1%, y se inyectó en una bomba de HPLC. Se separaron los péptidos mediante la columna Magic C18 (Michrom, 300 Å, 0,3 x 150 mm) y se eluyó a un caudal de 30 microL/min utilizando un gradiente de 10-95% de disolvente A (90% de acetonitrilo y 10% de disolución de ácido acético 0,1 M) y disolvente B (2% de acetonitrilo y 98% de disolución de ácido acético 0,1 M) durante 30 minutos. Se detectó el flujo de eluyente de la columna a la fuente de electroproyección y cada una de las señales de digesto triptico. La secuenciación de masas en tándem se llevó a cabo de forma simultánea mediante la fragmentación del ion precursor con m/z correspondiente a fragmento digerido por proteasa que incluye el residuo en la posición 38 de mDHFR mutante. Por lo tanto, los polipéptidos de DHFR fueron sintetizados en una célula huésped auxotrófica triple con (a), (b) ARNt^{Phe}_{CUA} de levadura PheRS de levadura (T415G); (c) ARNt^{Phe}_{CUA_UG} de levadura y PheRS de levadura (T415G); (d) ARNt^{Phe}_{CUA} de levadura y PheRS levadura (T415A); (e) ARNt^{Phe}_{CUA_UG} de levadura y PheRS de levadura (T415A) o (f) en una sola cepa auxotrófica con ARNt^{Phe}_{CUA_UG} de levadura y PheRS de levadura (T415A), cuyos resultados se muestran en la Figura 12.

EJEMPLO 17

Se rastreó un banco de PheRS de levadura (utilizando la proteína fluorescente verde o GFP) para identificar mutaciones de PheRS que permiten la incorporación de Nal. Posiciones de las secuencias de aminoácidos específicas que estaban mutadas incluyen los números de residuos 412, 415, 418 y 437, que están situadas en el sitio de unión y contactan con el aminoácido. Tal como se indica en la Figura 13, *p*-etnil-fenilalanina se incorporó en una proteína de ensayo utilizando la PheRS de levadura modificada.

En síntesis, GFP se ligó en un vector que contenía el gen T415G mutante o PheRS de levadura de tipo salvaje de acuerdo con procesos estándares. El ARNt supresor ámbar de levadura mutante (ytRNA^{Phe}_{AAA}) se expresó de forma constitutiva bajo el control del promotor *lpp* y se transformó en una cepa auxotrófica doble de Phe/Trp AFW de *E. coli*. Las células se cultivaron en medio mínimo M9 suplementado con glucosa, tiamina, MgSO₄, CaCl₂, 20 aminoácidos (a 25 mg/L), antibióticos (35 µg/mL de kanamicina y 100 µg/mL de ampicilina). Cuando las células alcanzaron una DO₆₀₀ de 0,8-1,0, las células se sedimentaron y lavaron dos veces con NaCl al 0,9% enfiada con hielo y se cambiaron a medio de expresión suplementado con 18 aminoácidos (a 20 mg/L) y diversas concentraciones de fenilalanina, triptófano y 2Nal. La expresión de proteínas se indujo mediante IPTG.

EJEMPLO 18

Se realizaron mutagénesis de los cuatro residuos aminoácidos seleccionados (N412, T145, S418 y S437) mediante mutación por PCR en dos etapas. En síntesis, se realiza una serie de mutagénesis por PCR en el gen GFP_{UV} en un plásmido pQE9_GFP_{UV} (STRATAGENE®), utilizando cuatro pares complementarios de cebadores (F64LS65T_f: 5'-CTT GTC ACT ACT CTG ACC TAT GGT GTT CAA TGC TTC TCC CGT-3' (SEQ ID NO: 22); F64LS65T_r: 5'-ACG GGA GAA GCA TTG AAC ACC ATA GGT CAG AGT AGT GAC AAG-3' (SEQ ID NO: 23); S99F_f: 5'-GTA CAG GAA CGC ACT ATA TTC TTC AAA GAT GAC GGG AAC-3' (SEQ ID NO: 24); S99F_r: 5'-GTT CCC GTC

ATC TTT GAA GAA TAT AGT GCG TTC CTG TAC-3' (SEQ ID NO: 25); T153M_f: 5'- CAC AAT GTA TAC ATC ATG GCA GAC AAA CAA AAG AAT GGA-3' (SEQ ID NO: 26); T153M_r: 5'-TCC ATT CTT TTG TTT GTC TGC CAT GAT GTA TAC ATT GTG -3' (SEQ ID NO: 27)).

5 Los mutantes de GFP se generaron según se describe en esta memoria. En síntesis, una GFP3 tiene 12 residuos Phe, de los cuales cinco son codificados por los codones de bamboleo de Phe (UUU). Se prepararon una GFP5 y una variante de GFP6 reemplazando los codones UUC por UUU, utilizando reacciones PCR de dos etapas, seguido de ligamiento. Se preparó una GFP5 reemplazando cuatro codones UUC y un codón Leu en los residuos F8, L64, F84, F99 y F165 por codones UUU utilizando doce cebadores (1: 5'-GTG CCA CCT GAC GTC TAA GAA ACC ATT ATT ATC ATG ACA TTA ACC-3' (SEQ ID NO: 28) 2: 5'-GAG TAA AGG AGA AGA ACT TTT TAC TGG AGT TGT CCC AAT TC-3' (SEQ ID NO: 29) 3: 5'- GAA TTG GGA CAA CTC CAG TAA AAA GTT CTT CTC CTT TAC TC-3' (SEQ ID NO: 30) 4: 5'- GGC CAA CAC TTG TCA CTA CTT TTA CCT ATG GTG TTC AAT GCT T-3' (SEQ ID NO: 31) 5: 5'- AAG CAT TGA ACA CCA TAG GTA AAA GTA GTG ACA AGT GTT GGC C-3' (SEQ ID NO: 32) 6:5'- CAT ATG AAA CGG CAT GAC TTT TTT AAG AGT GCC ATG CCC GAA G-3' (SEQ ID NO:33) 7: 5'- CTT CGG GCA TGG CAC TCT TAA AAA AGT CAT GCC GTT TCA TAT G (SEQ ID NO: 34) 8: 5'- GTT ATG TAC AGG AAC GCA CTA TAT TTT TCA AAG ATG ACG GGA ACT ACA A-3' (SEQ ID NO: 35) 9:5'- TTG TAG TTC CCG TCA TCT TTG AAA AAT ATA GTG CGT TCC TGT ACA TAA C-3' (SEQ ID NO: 36) 10: 5'- ACA AAA GAA TGG AAT CAA AGC TAA CTT TAA AAT TCG CCA CAA CAT TGA AGA TG-3' (SEQ ID NO: 37) 11: 5'- CAT CTT CAA TGT TGT GGC GAA TTT TAA AGT TAG CTT TGA TTC CAT TCT TTT GT-3' (SEQ ID NO: 38); 12: 5'- CGC CAA GCT AGC TTG GAT TCT CAC CAA TAA AAA ACG CCC-3' (SEQ ID NO: 39). Se obtuvieron cinco fragmentos de solapamiento parcialmente solapantes de casetes de expresión GFP3 mediante cinco reacciones PCR con cinco conjuntos de cebadores (1 y 3; 2 y 5; 4 y 7; 6 y 9; 8 y 10).

Estos productos de la PCR se purificaron mediante electroforesis en gel de agarosa, seguido de extracción de gel. Se preparó una GFP6, de la cual todos los residuos Phe son codificados por UUU reemplazando dos codones Phe (F71 y F99) de GFP5 por codones UUU utilizando seis cebadores. (1 y 12 son los mismos de arriba; 13: 5'- TAC CTA TGG TGT TCA ATG CTT TTC CCG TTA TCC GGA TCA TAT G-3' (SEQ ID NO: 40); 14: 5'-CAT ATG ATC CGG ATA ACG GGA AAA GCA TTG AAC ACC ATA GGT A-3' (SEQ ID NO: 41); 15: 5'- GTT ATG TAC AGG AAC GCA CTA TAT TTT TTA AAG ATG ACG GGA ACT ACA AG-3' (SEQ ID NO: 42); 16: 5'- CTT GTA GTT CCC GTC ATC TTT AAA AAA TAT AGT GCG TTC CTG TAC ATA AC-3' (SEQ ID NO: 43)).

EJEMPLO 19

Se realiza también una construcción de banco en dos etapas. En síntesis, la mutagénesis por saturación en cuatro residuos (N412, T415, S418 y S437) se consiguió con mutagénesis por PCR en dos etapas. Primero, se introdujeron codones degenerados en S437 mediante mutagénesis por PCR con dos cebadores complementarios (437_f: 5'-GTC GAA ATC GGT AAC NNK GGT ATG TTC AGA CCA GAA ATG CTC G-3' (SEQ ID NO: 44); 437_r: 5'- C GAG CAT TTC TGG TCT GAA CAT ACC MNN GTT AC C GAT TTC GAC-3' (SEQ ID NO: 45)). Después de digestión durante 1 h del producto de la PCR con *DpnI*, se transformó el producto de la PCR en el huésped de clonación XL-1 blue. Los plásmidos de la 437^a posición se saturaron y aislaron y se utilizaron como un molde para la 2^a mutagénesis por PCR para introducir una mutación en los residuos N412, T415 y S418. La 2^a mutagénesis por PCR se realizó con otro par de cebadores complementarios (412_418_f: 5'-C AAG CCT ACC TAC NNK CCT TAC NNK GAG CCA NNK ATG GAA ATC TTT T-3' (SEQ ID NO: 46); 412_418_r: 5'-A AAA GAT TTC CAT MNN TGG CTC MNN GTA AGG MNN GTA GG T AGG CTT G-3' (SEQ ID NO: 47)). Después de la PCR, los productos se digirieron con *DpnI* durante 1 h, se limpió y se concentró mediante columna de centrifugación. El producto eluido se electroporó en una célula electrocompetente ElectroTen-Blue (Stratagene) de acuerdo con el protocolo del fabricante. Se obtuvieron ocho millones de transformantes. El plásmido del banco se expandió en cultivo y se digirió con *NsiI* y *BglII*. Después de la purificación de estos insertos, éstos se ligaron con fragmentos grandes de pQE9_GFP6_yPheRS (T415G) y pQE9_GFP9_yPheRS (T415G) obtenidos mediante digestión con *NsiI* y *BglII*.

El banco se transformó en células de *E. coli* AFW y DHF químicamente competentes. Estas células se inocularon luego en medio 2x YT con kanamicina y se cultivaron durante la noche. Cuando las células alcanzaron una DO₆₀₀ de 0,8, las células se sedimentaron y se resuspendieron en agua destilada. Los stocks de glicerol del banco se expresaron como es costumbre en el técnica.

después de la expresión de GFP durante 3 horas, 1 mL de células (basado en una DO₆₀₀ de 1,0) se lavó con PBS y se diluyó en agua destilada, luego se sometió a análisis por citometría de flujo (aparato MoFlo cell sorter®, DakoCytomation, Fort Collins, CO), utilizando una longitud de onda de excitación de 488 nm, emisión de 525 nm y un filtro de corte 495 nm. En cada una de las mediciones se recogieron al menos 20.000 eventos. Los datos se analizaron con el software Summit (DakoCytomation). El rastreo de bancos se hizo tanto de forma positive como negative, es decir, las variantes de yPheRS que permiten una elevada incorporación de 2NaI o cualesquiera otros aminoácidos naturales, excepto Phe, en los codones UUU desplegarán GFP y son menos brillantes y, de esta

forma, se recogen células de baja fluorescencia. Las variantes de yPheRS que no permiten la incorporación de cualesquiera otros aminoácidos natural, excepto Phe, en los codones UUU no afectarán al plegamiento de GFP y son brillantes. Por lo tanto, se recogen células brillantes. La Figura 14 ilustra histogramas del rastreo de bancos de yPheRS de GFP.

5

LISTADO DE SECUENCIAS

- <110> California Institute of Technology
Wang, Pin
5 Kwon, Inchan
Son, Soojin
Tang, Yi
Tirrel, David A.
- 10 <120> INCORPORACIÓN ESPECÍFICA DEL SITIO DE AMINOÁCIDOS EN MOLÉCULAS
- <130> 110197 412 PC
- <140> PCT
15 <141> 05-03-2007
- <150> US 60/779.375
<151> 03-03-2006
- 20 <150> US 60/779.376
<151> 03-03-2006
- <160> 104
- 25 <170> FastSEQ para Windows versión 4.0
- <210> 1
<211> 7790
<212> ADN
30 <213> Secuencia Artificial
- <220>
<223> Gen dihidrofolato reductasa de ratón modificado y gen MetRS de E. coli
- 35 <400> 1

```

ctcgagaaat cataaaaaat ttattttgctt tgtgagcgga taacaattat aatagattca 60
attgtgagcg gataacaatt tcacacagaa ttcatataag aggagaaatt aactatgaga 120
ggatcgcatc accatcacca tcacggatcc ggcacatggt ttcgaccatt gaactcgatc 180
gtcgccgtgt cccaaaatat ggggattggc aagaacggag acctaccctg gcctccgctc 240
aggaacgagt tcaagtactt ccaaagaatg accacaacct cttcagtgga aggtaaacag 300
aatctgggtga ttatgggttag gaaaacctgg ttctccattc ctgagaagaa tcgaccttta 360
aaggacagaa ttaatatagt tctcagtaga gaactcaaag aaccaccacg aggagctcat 420
tttcttgcca aaagtttgga tgatgcctta agacttattg aacaaccgga attggcaagt 480
aaagtagaca tggtttggtat agtcggaggc agttctgttt accaggaagc catgaatcaa 540
ccaggccacc ttagactctt tgtgacaagg atcatgcagg aatttgaaaag tgacacgttt 600
ttcccagaaa ttgatttggg gaaatataaa cttctcccag aatacccagg cgtcctctct 660
gaggtccagg aggaaaaagg catcaagtat aagtttgaaag tctacgagaa gaaaggttgg 720
aagatcttaa gcttaattag ctgagcttgg actcctgttg atagatccag taatgacctc 780
agaactccat ctggatttgt tcagaaacgt cgggttgccg cgggcgtttt ttattgggtga 840
gaatccaagc tagctctaga gacgtccggc cggagctcca ccgcggttgg ggcgctctta 900
gagtcactta cttaacattt tcccatttgg tactatctaa ccccttttca ctattaaagaa 960
gtaatgccta ctatgactca agtcgcgaag aaaattcttg tgacgtgcgc actgcccgtac 1020
gctaaccggt caatccacct cggccatatg ctggagcaca tccaggetga tgcctgggtc 1080
cgttaccagc gaatgcgcgg ccacgaggtc aacttcatct gcgcccagca tgcccacggt 1140
acaccgatca tgctgaaagc tcagcagctt ggtatcacc cggagcagat gattggcgaa 1200

```

atgagtcagg	agcatcagac	tgatttcgca	ggctttaaca	tcagctatga	caactatcac	1260
tcgacgcaca	gcgaagagaa	cgcagctt	tcagaactta	tctactctcg	cctgaaagaa	1320
aacgggttta	ttaaaaaccg	caccatctct	cagctgtacg	atccggaaaa	aggcatgttc	1380
ctgccggacc	gttttgtgaa	aggcacctgc	ccgaaatgta	aatccccgga	tcaatacggc	1440
gataactgcg	aagtctcgcg	cgcgacctac	agccccgactg	aactgatcga	gccgaaatcg	1500
gtgggtttctg	gcgctacgcc	ggtaatgcgt	gattctgaac	acttcttctt	tgatctgccc	1560
tctttcagcg	aaatgttgca	ggcatggacc	cgcagcgggtg	cgttgagga	gcagggtggca	1620
aataaaatgc	aggagtgggt	tgaatctggc	ctgcaacagt	gggatatctc	ccgcgacgcc	1680
ccttacttcg	gttttgaaat	tccgaacgcg	ccgggcaaat	atctctacgt	ctggctggac	1740
gcaccgattg	gctacatggg	ttctttcaag	aatctgtgcg	acaagcgcgg	cgacagcgta	1800
agcttcgatg	aatactggaa	gaaagactcc	accgccgagc	tgtaccactt	catcggtaaa	1860
gatattgttt	acttccacag	cctgttctgg	cctgccatgc	tggaaggcag	caacttccgc	1920
aagccgtcca	acctgtttgt	tcatggctat	gtgacggtga	acggcgcaaa	gatgtccaag	1980
tctcggggca	cctttattaa	agccagcacc	tggtgaatc	atcttgacgc	agacagcctg	2040
cgttactact	acactgcgaa	actctcttcg	cgcattgatg	atatcgatct	caacctggaa	2100
gatttcgttc	agcgtgtgaa	tgccgatata	gttaacaaag	tggttaacct	ggctgccgt	2160
aatgcgggct	ttatcaacaa	gcgttttgac	ggcgtgctgg	caagcgaact	ggctccccg	2220
cagttgtaca	aaaccttcac	tgatgccgct	gaagtgattg	gtgaagcgtg	ggaaagccgt	2280
gaatttggtg	aagccgtgcg	cgaaatcatg	gcgctggctg	atctggctaa	ccgctatgtc	2340
gatgaacagg	ctccgtgggt	ggtggcgaaa	caggaaaggcc	gcgatgccga	cctgcaggca	2400
atttgctcaa	tgggcatcaa	cctgttccgc	gtgctgatga	cttacctgaa	gccgggtactg	2460
ccgaaactga	ccgagcgtgc	agaagcatte	ctcaatacgg	aactgacctg	ggatggtatc	2520
cagcaaccgc	tgctgggcca	caaagtgaat	ccgttcaagg	cgtgtataaa	ccgcatacgt	2580
atgaggcagg	ttgaagcact	ggtggaagcg	tctaaatgag	aagtataaag	ccgtgccgcg	2640
ccggttaactg	gcccgcgtgg	agatgatccg	atccaggaaa	ccatcacctt	tgacgacttc	2700
gctaaagtgt	acctgcgcgt	ggcgcgtgatt	gaaaacgcag	agtttgttga	aggttctgac	2760
aaactgctgc	gcctgacgct	ggatctcggc	ggtgaaaaac	gcaatgtctt	ctccggtatt	2820
cgttctgctt	acccggatcc	gcaggcactg	attggctcgtc	acaccattat	ggtggctaac	2880
ctggcaccac	gtaaaatgcg	cttcgggtatc	tctgaaggca	tggtgatggc	tgccggtcct	2940
ggcgggaaag	atattttctt	gctaagcccg	gatgccgggtg	ctaaaccggg	tcatcagggtg	3000
aaataatccc	cattcaaggc	gctgcatacg	cagccttttg	ctttataaat	ctctaaagtt	3060
gttttcttgc	gattttgtct	ctctctaacc	cgcataaata	ctggtagcat	ctgcattcaa	3120
ctggataaaa	ttacagggat	gcagaatgag	acactttatc	tatcaggacg	aaaaatcaca	3180
taaattcagg	gcagttgagc	aacagggaaa	cgagttgcat	atcagttggg	gaaaagttgg	3240
caccaaaggc	aaagccagat	aaaaagtttt	tcagatgctg	cggcagcggc	aaaagcggag	3300
cccgaacctcg	agggggggcc	cggtagcccg	ccggacgtct	ctagagctag	cttggcgaga	3360
ttttcaggag	ctaaggaagc	taaaatggag	aaaaaaatca	ctggatatac	caccgttgat	3420
atatcccaat	ggcatcgtaa	agaacatttt	gaggcatttc	agtcagttgc	tcaatagacc	3480
tataaccaga	ccgttcagct	ggatattacg	gcctttttta	agaccgtaaa	gaaaaataag	3540
cacaagtttt	atccggcctt	tattcacatt	cttgcgccgc	tgatgaatgc	tcatccggaa	3600
tttcgtagtg	caatgaaaga	cggtgagctg	gtgatatggg	atagtgttca	cccttggtac	3660
accgttttcc	atgagcaaac	tgaaaagctt	tcatcgctct	ggagtgaata	ccacgacgat	3720
ttccggcagt	ttctacacat	atattcgcaa	gatgtggcgt	gttacgggtg	aaacctggcc	3780
tatttcccta	aagggtttat	tgagaatatg	tttttctctt	cagccaatcc	ctgggtgagt	3840
ttcaccagtt	ttgattttaa	cgtggccaat	atggacaact	tcttcgcccc	cgttttcacc	3900
atgggcaaat	attatacgca	aggcgacaag	gtgctgatgc	cgtcggcgat	tcaggttcat	3960
catgccgttt	gtgatggctt	ccatgtcggc	agaatgctta	atgaattaca	acagtactgc	4020
gatgagtggc	agggcggggc	gtaatttttt	taaggcagtt	attggtgccc	ttaaacgcct	4080
ggggtaatga	ctctctagct	tgaggcatca	aataaaacga	aaggctcagt	cgaagactg	4140
ggcctttcgt	tttatctggt	gtttgtcggc	gaacgctctc	ctgagtagga	caaattccgc	4200
ctctagatta	cgtgcagtcg	atgataagct	gtcaaacatg	agaattgtgc	ctaattgagt	4260
agctaactta	cattaattgc	gttgcgctca	ctgcccgctt	tccagtcggg	aaacctgtcg	4320
tgccagctgc	attaatgaat	cggccaacgc	gcggggagag	gcgggttgcg	tattgggcgc	4380
caggggtggt	tttcttttca	ccagtgcagc	gggcaacagc	tgattgcccc	tcaccgcctg	4440
gccctgagag	agttgcagca	agcggccac	gctgggtttgc	cccagcaggc	gaaaatccctg	4500
tttgatgggtg	gttaacggcg	ggatataaca	tgagctgtct	tcgggtatcgt	cgtatcccac	4560
taccgagata	tccgcaccaa	cgcgcagccc	ggactcggta	atggcgcgca	ttgcgcccag	4620
cgccatctga	tcggtggcaa	ccagcatcgc	agtgggaacg	atgccctcat	tcagcatttg	4680

catgggtttgt	tgaaaaaccgg	acatggcact	ccagtcgect	tcccgttccg	ctatcggctg	4740
aatttgattg	cgagtggagat	atztatgcc	gccagccaga	cgagagcg	ccgagacaga	4800
acttaattggg	cccgttaaca	gcgcgatttg	ctggtgaccc	aatgagacca	gatgctccac	4860
gcccagtcgc	gtaccgtctt	catgggagaa	aataatactg	ttgatgggtg	tctggtcaga	4920
gacatcaaga	aataacgccg	gaacattagt	gcaggcagct	tccacagcaa	tggcatcctg	4980
gtcatccagc	ggatagttaa	tgatcagccc	actgacgcgt	tgcgcgagaa	gattgtgcac	5040
cgccgcttta	caggcttcga	cgccgcttcg	ttctaccatc	gacaccacca	cgctggcacc	5100
cagttgatcg	gcgcgagatt	taatcgccgc	gacaatttgc	gacggcgctg	gcagggccag	5160
actggaggtg	gcaacgccaa	tcagcaacga	ctgtttgccc	gccagttggt	gtgccacgcg	5220
gttgggaatg	taattcagct	ccgccatcgc	cgcttccact	ttttcccgcg	ttttccgaga	5280
aacgtggctg	gectggttca	ccacgcggga	aacggtctga	taagagacac	cggcatactc	5340
tgcgacatcg	tataacgtta	ctggtttcac	attcaccacc	ctgaattgac	tctcttcccg	5400
gcgctatcat	gccataccgc	gaaagggttt	gcaccattcg	atggtgtcgg	aatttcgggc	5460
agcgttggtg	cctggccacg	ggtgcgcag	atctagagct	gcctcgcgcg	tttcgggtgat	5520
gacggtgaaa	acctctgaca	catgcagctc	ccggagacgg	tcacagcttg	tctgtaagcg	5580
gatgccggga	gcagacaagc	ccgtcagggc	gcgtcagcgg	gtgttgggcg	gtgtcggggc	5640
gcagccatga	cccagtcacg	tagcgatagc	ggagtgtata	ctggcttaac	tatgcggcat	5700
cagagcagat	tgtactgaga	gtgcaccata	tgcggtgtga	aataaccgcac	agatgcgtaa	5760
ggagaaaaata	ccgcatacag	cgctcttccg	cttcctcgct	cactgactcg	ctgcgctcgg	5820
tcgttcggct	gcgccgagcg	gtatcagctc	actcaaaggc	ggtaatacgg	ttatccacag	5880
aatcagggga	taacgcagga	aagaacatgt	gagcaaaaag	ccagcaaaaag	ccagggaacc	5940
gtaaaaaaggc	cgcggttgctg	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	6000
aaaatcgacg	ctcaagtcag	aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	6060
ttccccctgg	aagctccctc	gtgcgctctc	ctgttccgac	cctgcccgtt	accggatacc	6120
tgctccgctt	tctcccttcg	ggaagcggtg	cgctttctca	tagctcacgc	tgtaggtatc	6180
tcagttcggt	gtaggtcggt	cgctccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	6240
ccgaccgctg	cgcccttatcc	ggtaactatc	gtcttgagtc	caacccggta	agacacgact	6300
tatcgccact	ggcagcagcc	actggtaaca	ggattagcag	agcgagggtat	gtaggcgggtg	6360
ctacagagtt	cttgaagtgg	tggcctaact	acggctacac	tagaaggaca	gtatttggtta	6420
tctgcgctct	gctgaagcca	gttaccttcg	gaaaaagagt	tggtagctct	tgatccggca	6480
aacaaaccac	cgctggtagc	ggtgggtttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	6540
aaaaaggatc	tcaagaagat	cctttgatct	tttctacggg	gtctgacgct	cagtgggaacg	6600
aaaactcacg	ttaagggatt	ttggtcatga	gattatcaaa	aaggatcttc	acctagatcc	6660
tttttaaatta	aaaatgaagt	tttaaatcaa	tctaaagtat	atatgagtaa	acttgggtctg	6720
acagttacca	atgcttaate	agtgaggcac	ctatctcagc	gatctgtcta	tttcggttcac	6780
ccatagttgc	ctgactcccc	gtcgtgtaga	taactacgat	acgggagggc	ttaccatctg	6840
gccccagtg	tgcaatgata	ccgcgagacc	cacgctcacc	ggctccagat	ttatcagcaa	6900
tacaccagcc	agccgggaagg	gccgagcgca	gaagtgggtcc	tgcaacttta	tccgcctcca	6960
tccagttctat	taattgtttg	cggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	7020
gcaacgttgt	tgccattgct	acaggcatcg	tggtgtcacg	ctcgctcgctt	ggtatggctt	7080
cattcagctc	cggttcccaa	cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	7140
aagcgggttag	ctccttcggt	cctccgatcg	ttgtcagaag	taagttggcc	gcagtggttat	7200
cactcatggt	tatggcagca	ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	7260
tttctgtgac	tggtgagtac	tcaaccaagt	cattctgaga	atagtgtatg	cggcgaccga	7320
ggtgctcttg	cccggcgctca	atacgggata	ataccgcgcc	acatagcaga	actttaaaag	7380
tgctcatcat	tggaaaacgt	tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	7440
gatccagttc	gatgtaaccc	actcgtgcac	ccaactgac	ttcagcatct	tttactttca	7500
ccagcgtttc	tgggtgagca	aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	7560
cgacacggaa	atgttgaaata	ctcatactct	tcctttttca	atattattga	agcatttatc	7620
aggggttattg	tctcatgagc	ggatacatat	ttgaatgtat	ttagaaaaat	aaacaaatag	7680
gggttcccg	cacatttccc	cgaaaagtgc	cacctgacgt	ctaagaaacc	attattatca	7740
tgacattaac	ctataaaaaat	aggcgtatca	cgaggccctt	tcgtcttcac		7790

<210> 2

<211> 8294

<212> ADN

5 <213> gen fenilalanina aminoacilsintetasa de levadura

<400> 2

ctcgagaaat	cataaaaaat	ttatttgcct	tgtgagcgga	taacaattat	aatagattca	60
attgtgagcg	gataacaatt	tcacacagaa	ttcattaaag	aggagaaatt	aactatgaga	120
ggatcgcatc	accatcacca	tcacggatcc	gtcgacctgc	agccccgggt	accggtagaa	180
aaaatgagta	aaggagaaga	actttttact	ggagtgtgcc	caattcttgt	tgaattagat	240
ggtgatgtta	atgggacaaa	attttctgtc	agtggagagg	gtgaagggtga	tgcaacatac	300
ggaaaactta	cccttaaatt	tatttgcact	actggaaaac	tacctgttcc	atggccaaca	360
cttgcacta	cttttaccta	tgggtgttcaa	tgtttttccc	gttatccgga	tcatatgaaa	420
cggcatgact	tttttaagag	tgccatgccc	gaaggttatg	tacaggaacg	cactatattt	480
tttaagatg	acgggaacta	caagacgcgt	gctgaagtca	agtttgaagg	tgataccctt	540
gttaatcgta	tcgagttaaa	agggtattgat	tttaagaaag	atggaaacat	tctcggaac	600
aaactcgagt	acaactataa	ctcacacaat	gtatacatca	tggcagacaa	acaaaagaat	660
ggaatcaaag	ctaactttta	aattcgccac	aacattgaag	atggatccgt	tcaactagca	720
gaccattatc	aacaaaaatc	tccaattggc	gatggccctg	tccttttacc	agacaacat	780
tacctgtcga	cacaatctgc	cccttcgaaa	gatcccaacg	aaaagcgtga	ccacatggtc	840
cttcttgagt	ttgtaactgc	tgctgggatt	acacatggca	tggatgagct	ctacaaactg	900
cagccaagct	taattagctg	aagcttggac	tcctgttgat	agatccagta	atgacctcag	960
aactccatct	ggatttgttc	agaacgcctg	gttgccgccc	ggcgtttttt	attggtgaga	1020
atccaagcta	gcttgccgag	attttcagga	gctaaggaaag	ctaaaatgga	gaaaaaaatc	1080
actggatata	ccaccgttga	tatatcccaa	tggcatcgta	aagaacattt	tgaggcattt	1140
cagtcagttg	ctcaatgtac	ctataaccag	accgttcagc	tggcacgaca	ggtttcccga	1200
ctggaaagcg	ggcagtgagc	gcaacgcaat	taatgtgagt	tagctcactc	attaggcacc	1260
ccaggcttta	cacttttatgc	ttccggctcg	tatgttgtgt	ggaattgtga	gcggaataca	1320
atttcacaca	ggaaacagct	atgaccatga	ttacgccaag	cttgcatgcc	tcaggttgac	1380
aattaatcat	cggctcgtat	aatggatcca	attgtgagcg	gaatcgattt	tcacacagga	1440
aacagaccat	gaatctagag	atgtctgact	tccaattaga	aattctaaag	aaactagatg	1500
aattggatga	gatcaagtcc	acactggcaa	ctttccctca	gcaéggctót	caagatgttc	1560
tttccgcttt	gaactccttg	aaagcccaca	acaagttaga	gttttccaag	gtcgacacgg	1620
ttacgtatga	cttgaccaaa	gaaggtgttc	aaattttgaa	tgaaggttcg	tacgaaaatta	1680
aactagtcaa	gctcatccaa	gagttgggtc	aacttcaa	caaagatgtg	atgtccaaac	1740
taggccctca	agttggtaag	gtcggtcagg	ctagagcttt	caagaacggc	tggatcgcca	1800
aaaacgcctc	aaacgagctt	gaactctccg	caaaattgca	aaataccgat	ttaaatgagc	1860
ttactgatga	aacgcaatct	attctagcgc	aaatcaagaa	caactcgcac	ctggatagca	1920
ttgacgcca	gattttgaac	gacttgaaga	aaagaaagtt	aattgctcaa	ggtaaaaatca	1980
cagatttcag	tgtcacccaa	gggccagagt	tctcgaccga	cctcaccaaa	ttggaaaaccg	2040
atcttacctc	cgacatggtc	tccaccaatg	catacaagga	cttgaagttc	aagccttaca	2100
atttcaatcc	tcaaggtgtg	caaatatctt	caggtgtctc	tcaccctcta	aacaaagtca	2160
gagaggaatt	tagacaaatt	ttcttttcca	tgggattcac	agagatgcc	tcgaaccaat	2220
acgtcgagac	aggtttctgg	aacttcgatg	ccctttacgt	cccacaacag	catcctgttc	2280
gtgacctgca	agacactttc	tacatcaagg	accactaac	cgctgagttg	cccgatgaca	2340
agacatacat	ggacaatata	aaagccgttc	acgaacaggg	gagattcggg	tccatcgggt	2400
atcgttacaa	ctggaagcca	gaagaatgtc	aaaaattggg	cttgagaact	cactccacag	2460
ccatctctgc	cagaatgctg	cacgatttgg	ccaaagatcc	aaagcccacc	agattgtttt	2520
ctatcgaccg	tgttttccgt	aacgaagcag	ttgacgccac	ccatttggcc	gaattccacc	2580
agggtggaagg	tgttcttgcc	gactacaaca	ttactctggg	tgacctgac	aagtctcatg	2640
aagagttttt	cgaaagaatg	ggtgtcaccg	gtttgagatt	caagcctacc	tacaatcctt	2700
acaccgagcc	atcaatggaa	atcttttctt	ggcacgaagg	tttgcaaaaa	tgggtcgaaa	2760
tcggtaactc	tgggtatgtc	agaccagaaa	tgtctgagtc	catgggtcta	ccaaaggatc	2820
taagagtcc	tgggtggggg	ttatccttgg	aaagacctac	catgatcaaa	tataagggtc	2880
aaaacatcag	agaactgtta	ggtcataaag	tctctttgga	ctttatcgaa	accaatcctg	2940
ctgctagatt	ggacgaagac	ttgtacgaat	aaggcaggaa	tagattatgc	ctaccgtctc	3000
cgtgaacaag	cagcaattat	ttgatcttct	aggcaaagac	tacacttccc	aagagtctga	3060
cgaattatgt	tttgaattcg	gtatggaaat	ggacgaagac	accacagaag	aggccttgaa	3120
aaccggggag	gagccggaat	tgaagcttga	tatcagtgcc	aatcgttacg	atttgccttg	3180
tatcgaaggt	atctcacagt	cgctgaacga	atacttggaa	cgtaaagaaa	gacctgacta	3240
taaattaagc	aagccaacca	ctaagttgat	catcgacaaa	tcaacggagc	aaattagacc	3300
ttttgctacc	gctgctgtat	tgagaaatat	caagcttaac	gaaaaatcct	acgcttcttt	3360
tattgccttg	caagataaat	tacatgccaa	tctatgtaga	aacagaagct	tggttgccat	3420

gggtactcac	gatttagatt	caattgaagg	tccattccat	tacagagctc	taccaccaa	3480
ggacatcaag	ttcgtaccat	tgaatcaaac	ccaagagttt	actggtgaca	aattgatcga	3540
gttttataaa	tctccagaac	agaaaaacaa	catagggaga	tacgttcaca	ttattgagga	3600
ttctccagtc	ttcccagtta	ttatggacag	caaagatcgt	gtttgctccc	tgccaccatt	3660
aatcaatagt	gaacattcga	agatctctgt	gaacacccgt	aacattttga	ttgatataac	3720
cgccaccgat	aagaccaaag	ccgagatcgt	tttgaacata	ttactacaa	tgttctcacg	3780
ttattgtgac	gaaccattca	cgggtgagcc	tgtagaaatt	gtctctgaac	acaatggcca	3840
atcccgtttg	gcgcaaaact	tcaacgatag	aattatggat	gtctccatta	agtacatcaa	3900
ctctgtcttt	ggcctagatc	aatccgctga	tgaatttgct	cattgtctaa	aaaagatgtc	3960
gttgcatgcc	gttcaatcaa	aggaagacaa	ggacatcttg	cacgttgaca	ttccggtaac	4020
tagacctgat	attttgacag	cttgatgat	aatggaagat	gccgctgtcg	gttatggttt	4080
caataatttg	ccaaaggggtg	agaaattatc	caatgccaac	ttcattgcca	aaccattacc	4140
aatcaacaag	gtttctgata	ttttcagagt	tgcatcctct	caagccacgt	gggttgaggt	4200
tttccatttg	accttatgtt	cgcacgatga	aaacttttaa	tttctaagac	aatccgacaa	4260
tgatgattta	gctgtcaa	tggccaaccc	aaagactttg	gaataccaag	ttgttagaac	4320
cactttattg	cctgggatct	taaagacagt	caaggaaaac	agaaaacatt	ccttaccat	4380
caaagtcttt	gaaaccggtg	acgttgat	taaagacgac	aaactagaaa	ggaaggcgta	4440
caatgaacgt	cactgggctg	ccatctacgt	gggtaagaat	tctgggtttg	aaatcattca	4500
agggttattg	ggtaaaatca	tgcaaacctt	tgaacagag	tggttgag	actacgggtg	4560
tgctgcttct	ggcagaggtt	actggattga	agaagacgat	tctgtgaaa	cctatctccc	4620
aggtagaggt	gccaaagtca	tggtcagatc	caaagaaggc	gctgagccaa	gctaaatcgg	4680
ccacttggtg	gtcttgcatc	ctgaagtc	gatgaatttc	gacgttccat	tcgctgcac	4740
ctttgtagag	gttaatgccg	aagtcttct	ataatgta	gttctaaca	aaatctttac	4800
tgatttataa	aacttatata	gatagataga	catatatata	tatctatata	tagaaacaca	4860
actaaagttt	accatgtttt	atatagggt	ccgagctcga	attcactggc	cgtcgtttta	4920
caacgtcgtg	actgggaaaa	cccggtttac	gtgcagtcga	tgataagctg	tcaaacatga	4980
gaattgtgcc	taatgagtga	gctaacttac	attaattgctg	ttgcgctcac	tgcccgcttt	5040
ccagtcggga	aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgcy	cggggagag	5100
cggtttgctg	attggggcgc	agggtggttt	ttcttttcac	cagtgcagcy	ggcaacagct	5160
gattgcccct	caccgcctgg	ccctgagaga	gttgacgcaa	gcggtccacg	ctgggtttgc	5220
ccagcagcgg	aaaatcctgt	ttgatgggtg	ttaacggcgg	gatataacat	gagctgtctt	5280
cggatcgtc	gtatcccact	accgagatat	ccgcaccaac	gcgcagcccg	gactcggtta	5340
tggcgcgcat	tgcccccagc	gccatctgat	cgttggaac	cagcatcgca	gtgggaacga	5400
tgcctctcatt	cagcatttgc	atgggttgtt	gaaaaccgga	catggcactc	cagtcgcctt	5460
cccgttccgc	talcggctga	atltgatggc	gagtgcagata	tttatgccag	ccagccagac	5520
gcagacgcgc	cgagacagaa	cttaattggc	ccgctaacag	cgcgatttgc	tggtgacca	5580
atgcgaccag	atgctccacg	cccagtcgcy	taccgtcttc	atgggagaaa	ataatactgt	5640
tgatgggtgt	ctggtcagag	acatcaagaa	ataacgccgg	aacattagtg	caggcagctt	5700
ccacagcaat	ggcatcctgg	tcacccagcy	gatagttaat	gatcagccca	ctgacgcgtt	5760
gcgcgagaag	attgtgcacc	gccgctttac	aggcttcgac	gccgcttcgt	tctaccatcg	5820
acaccaccac	gctggcaccc	agttgatcgg	cgcgagattt	aatcgccgcy	acaatttgcg	5880
acggcgcggtg	cagggccaga	ctggaggtgg	caacgccaat	cagcaacgac	tgtttgcccg	5940
ccagttgttg	tgccacgcgg	ttgggaatgt	aattcagctc	cgccatcgcc	gcttccactt	6000
tttcccgctg	tttcgcagaa	acgtggctgg	cctggttcac	cacgcgggaa	acgggtctgat	6060
aagagacacc	ggcatactct	gcgacatcgt	ataacgttac	tggtttcaca	ttcaccaccc	6120
tgaattgact	ctcttccggg	cgctatcatg	ccataccgcy	aaaggttttg	caccattcga	6180
tggtgtcgga	atttcgggca	gcgttggttc	ctggccacgg	gtgcgcatga	cttaagcgg	6240
gtgaaatacc	gcacagatgc	gtaaggagaa	aataccgcat	caggcgctct	tccgcttctt	6300
cgctcactga	ctcgctgcgc	tcggtctgtc	ggctgcggcg	agcggtatca	gctcactcaa	6360
aggcggtaat	acggttatcc	acagaatcag	gggataacgc	aggaaagaac	atgtgagcaa	6420
aaggccagca	aaaggccagg	aaccgtaaaa	aggccgcgtt	gctggcggtt	ttccataggc	6480
tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	tcagaggtgg	cgaaacccga	6540
caggactata	aagataccag	gcgtttcccc	ctggaagctc	cctcgctgcg	tctcctgttc	6600
cgacctgtcc	gcttaccgga	tacctgtccg	cctttctccc	ttcggaagc	gtggcgcttt	6660
ctcaatgctc	acgctgtagg	tatctcagtt	cggtgttaggt	cgttcgctcc	aagctgggct	6720
gtgtgcacga	accccccggt	cagcccgacc	ctgtgcctt	atccggtaac	tatcgtcttg	6780
agtccaaccc	ggtaagacac	gacttatcgc	cactggcagc	agccactggt	aacagatta	6840
gcagagcgag	gtatgtaggc	ggtgctacag	agttcttgaa	gtgggtggcct	aactacggct	6900

```

acactagaag gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa 6960
gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcgggtgg ttttttgttt 7020
gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta 7080
cgggggtctga cgctcagtg aacgaaaact cacgttaagg gatttttggtc atgagattat 7140
caaaaaggat cttcacctag atccttttaa attaaaaatg aagtttttaa tcaatctaaa 7200
gtatatatga gtaaacctgg tctgacagtt accaatgctt aatcagtgag gcacctatct 7260
cagcgatctg tctatttcgt tcatccatag ctgcctgact ccccgctcgt tagataacta 7320
cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgcga gaccacgct 7380
caccggctcc agatttatca gcaataaac agccagccgg aagggccgag cgcagaagtg 7440
gtcctgcaac tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa 7500
gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc atcgtggtgt 7560
cacgctcgtc gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta 7620
catgatcccc catgttggtg aaaaaagcgg tttagctcct cggctcctcg atcgttgctca 7680
gaagtaagtt ggcgcagtg ttatcactca tggttatggc agcactgcat aattctctta 7740
ctgtcatgcc atccgtaaga tgcttttctg tgactgggtg gtactcaacc aagtcattct 7800
gagaatagtg tatgcggcga ccgagttgct cttgcccggc gtcaatacgg gataataccg 7860
cgccacatag cagaacttta aaagtgtca tcattggaaa acgttcttcg gggcgaaaac 7920
tctcaaggat cttaccgctg ttgagatcca gttcagatga acccactcgt gcacccaact 7980
gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa 8040
atgccgcaaa aaagggata agggcgacac ggaaatgttg aatactcata ctcttccttt 8100
ttcaatatta ttgaagcatt tatcagggtt attgtctcat gageggatag atatttgaat 8160
gtatttagaa aaataaacia ataggggttc cgcgcacatt tccccgaaa gtgccacctg 8220
acgtctaaga aaccattatt atcatgacat taacctataa aaataggcgt atcacgaggc 8280
cctttcgtct tcac 8294

```

<210> 3

<211> 1512

<212> ADN

5 <213> Secuencia Artificial

<220>

<223> subunidad alfa del gen fenilalanina aminoacilsintetasa de levadura

10 <220>

<221> característica misc.

<222> 1234, 1235, 1236, 1243, 1244, 1245, 1252, 1253, 1254, 1309, 1310, 1311

<223> n = A, T, C ó G

15

<400> 3

```

atgtctgact tccaattaga aattctaaag aaactagatg aattggatga gatcaagtcc 60
acactggcaa ctttcctca gcacggetct caagatgttc tttccgcttt gaactctttg 120
aaagcccaca acaagttaga gttttccaag gtcgacacgg ttacgtatga cttgaccaa 180
gaaggtgctc aaattttgaa tgaaggttcg tacgaaatta aactagtcaa gctcatccaa 240
gagttgggtc aacttcaaat caaagatgtg atgtccaaac taggccccta agttggtaag 300
gtcggtcagg ctagagcttt caagaacggc tggatcgcca aaaacgcctc aaacgagctt 360
gaactctccg caaaattgca aaataccgat ttaaattgagc ttactgatga aacgcaatct 420
attctagcgc aaatcaagaa caactcgcat ctggatagca ttgacgcaa gattttgaac 480
gacttgaaga aaagaaagtt aattgctcaa ggtaaaatca cagatttcag tgtcaccaaa 540
gggccagagt tctcgaccga cctcaccaaa ttggaaaccg atcttacctc cgacatgggtc 600
tccaccaatg catacaagga cttgaagttc aagccttaca atttcaattc tcaaggtgtg 660
caaatatctt cagggtctct tcacccctta aacaaagtca gagaggaatt tagacaaatt 720
ttcttttcca tgggattcac agagatgccc tcgaaccaat acgtcgagac aggtttctgg 780
aacttcgatg ccttttacgt ccacaaacag catcctgctc gtgacctgca agacactttc 840
tacatcaagg acccataac cgctgagttg cccgatgaca agacatacat ggacaatatc 900
aaagccgttc acgaacaggg gagattcggg tccatcggtt atcgttacaa ctggaagcca 960
gaagaatgtc aaaaattgggt cttgagaact cactccacag ccattctctg cagaatgctg 1020

```

```

caccgatttgg ccaaagatcc aaagcccacc agattggttt ctatcgaccg tgttttccgt 1080
aacgaagcag ttgacgccac ccatttggcc gaattccacc aggtggaagg tgttcttgcc 1140
gactacaaca ttactctggg tgacctgacg aagttcatgg aagagttttt cgaaagaatg 1200
gggtgtcaccg gtttgagatt caagcctacc tacnnncctt acnnngagcc annnatggaa 1260
atcttttctt ggcacgaagg tttgcaaaaa tgggtcgaaa tcggtaacnn nggtatgttc 1320
agaccagaaa tgctcgagtc catgggtcta ccaaaggatc taagagtcct tgggtggggg 1380
ttatccttgg aaagacctac catgatcaaa tataagggtc aaaacatcag agaactgtta 1440
ggtcataaag tctctttgga ctttatcgaa accaatcctg ctgctagatt ggacgaagac 1500
ttgtacgaat aa                                     1512

```

<210> 4

<211> 13

5 <212> PRT

<213> Secuencia Artificial

<220>

<223> Derivado artificial de hidrofolato reductasa

10

<400> 4

```

      Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Lys
      1              5              10

```

15 <210> 5

<211> 14

<212> PRT

<213> Secuencia Artificial

20 <220>

<223> Derivado artificial de hidrofolato reductasa

<220>

<221> VARIANTE

25 <222> 13

<223> Xaa = aminoácido supresor

<400> 5

```

      Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Xaa Lys
      1              5              10

```

30

<210> 6

<211> 8

<212> PRT

35 <213> Secuencia Artificial

<220>

<223> Derivado artificial de hidrofolato reductasa

40 <400> 6

```

      Pro Trp Pro Pro Leu Arg Asn Glu
      1              5

```

<210> 7

<211> 34

45 <212> ADN

<213> Secuencia Artificial

<220>

<223> Cebador de oligonucleótidos

50

<400> 7
 cgatttcac acaggatcca gaccatgatt ctag 34

5 <210> 8
 <211> 27
 <212> ADN
 <213> Secuencia Artificial

10 <220>
 <223> Cebador de oligonucleótidos

<400> 8
 gacggccagt gaattcgagc tcggtac 27

15 <210> 9
 <211> 38
 <212> ADN
 <213> Secuencia Artificial

20 <220>
 <223> Cebador de oligonucleótidos

<400> 9
 ctacctacaa tccttacggc gagccatcaa tggaaatc 38

25 <210> 10
 <211> 38
 <212> ADN
 <213> Secuencia Artificial

30 <220>
 <223> Cebador de oligonucleótidos

<400> 10
 gatttcatt gatggctcgc cgtaaggatt gtaggtag 38

35 <210> 11
 <211> 36
 <212> ADN
 <213> Secuencia Artificial

40 <220>
 <223> Cebador de oligonucleótidos

45 <400> 11
 ccgctcagga acgagtagaa gtacttcaa agaatg 36

<210> 12
 <211> 36
 <212> ADN
 <213> Secuencia Artificial

50 <220>
 <223> Cebador de oligonucleótidos

55 <400> 12
 cattctttgg aagtacttct actcgttcct gagcgg 36

<210> 13
 <211> 30
 <212> ADN

60

<213> Secuencia Artificial

<220>

<223> Cebador de oligonucleótidos

5 <400> 13
gcactgacca tggctgaaca acacgcacag 30

<210> 14

10 <211> 31
<212> ADN
<213> Secuencia Artificial

<220>

15 <223> Cebador de oligonucleótidos

<400> 14
ggacttcgga tcctttctgt gggcgcatcg c 31

<210> 15

20 <211> 14
<212> PRT
<213> Secuencia Artificial

<220>

25 <223> subunidad alfa del gen fenilalanina aminoacilsintetasa de levadura

<220>

30 <221> VARIANTE
<222> 13
<223> Xaa = aminoácido supresor

<400> 15

35 Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Xaa Lys
1 5 10

<210> 16

<211> 39

<212> ADN

40 <213> Secuencia Artificial

<220>

<223> Cebador de oligonucleótidos

<400> 16

45 gaacacagga cctccacatt tagagtatgg cgctctccc 39

<210> 17

<211> 39

<212> ADN

50 <213> Secuencia Artificial

<220>

<223> Cebador de oligonucleótidos

<400> 17

55 gggagagcgc catactctaa atgtggaggt cctgtgttc 39

<210> 18

<211> 42
 <212> ADN
 <213> Secuencia Artificial

5 <220>
 <223> Cebador de oligonucleótidos

<400> 18
 ctgggtaagc ttcgctaagg atctgccctg gtgcgaactc tg 42

10 <210> 19
 <211> 43
 <212> ADN
 <213> Secuencia Artificial

15 <220>
 <223> Cebador de oligonucleótidos

20 <400> 19
 gattacggat tcctaatacg actcactata gcggacttag ctc 43

25 <210> 20
 <211> 40
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

30 <400> 20
 cggaagcaga aagtgtaaag agcgggggtgc ctaatgagtg 40

35 <210> 21
 <211> 40
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

40 <400> 21
 cactcattag gcaccccgct cttacactt tatgcttcg 40

45 <210> 22
 <211> 42
 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

50 <400> 22
 ctgtcacta ctctgacctt tgggtgtcaa tgcttctccc gt 42

55 <210> 23
 <211> 42
 <212> ADN
 <213> Secuencia Artificial

60 <220>
 <223> Cebador de oligonucleótidos

<400> 23

5 acgggagaag cattgaacac cataggtcag agtagtgaca ag 42

<210> 24
<211> 39
<212> ADN
<213> Secuencia Artificial

10 <220>
<223> Cebador de oligonucleótidos

15 <400> 24

gtacaggaac gcactatatt cttcaaagat gacgggaac 39

<210> 25
<211> 39
<212> ADN
<213> Secuencia Artificial

20 <220>
<223> Cebador de oligonucleótidos

25 <400> 25

gttcccgta tctttgaaga atatagtgcg ttctgtac 39

<210> 26
<211> 39
<212> ADN
<213> Secuencia Artificial

30 <220>
<223> Cebador de oligonucleótidos

35 <400> 26

cacaatgtat acatcatggc agacaaaca aagaatgga 39

40 <210> 27
<211> 39
<212> ADN
<213> Secuencia Artificial

45 <220>
<223> Cebador de oligonucleótidos

50 <400> 27

tccattcttt tgtttgtctg ccatgatgta tacattgtg 39

<210> 28
<211> 45
<212> ADN
<213> Secuencia Artificial

55 <220>
<223> Cebador de oligonucleótidos

60 <400> 28

gtgccacctg acgtctaaga aaccattatt atcatgacat taacc 45

<210> 29
<211> 41
5 <212> ADN
<213> Secuencia Artificial

<220>
<223> Cebador de oligonucleótidos

10 <400> 29
gagtaaagga gaagaacitt ttactggagt tgtccaatt c 41

<210> 30
15 <211> 41
<212> ADN
<213> Secuencia Artificial

<220>
20 <223> Cebador de oligonucleótidos

<400> 30
gaattgggac aactccagta aaaagttctt ctctttact c 41

25 <210> 31
<211> 43
<212> ADN
<213> Secuencia Artificial

30 <220>
<223> Cebador de oligonucleótidos

<400> 31
ggccaacact tgcactact ttacctatg ggttcaatg ctt 43

35 <210> 32
<211> 43
<212> ADN
<213> Secuencia Artificial

40 <220>
<223> Cebador de oligonucleótidos

<400> 32
45 aagcatgaa caccataggt aaaagtagtg acaagtgttg gcc 43

<210> 33
<211> 43
<212> ADN
50 <213> Secuencia Artificial

<220>
<223> Cebador de oligonucleótidos

55 <400> 33
catatgaaac ggcattgactt ttttaagagt gccatgcccg aag 43

<210> 34
<211> 43
60 <212> ADN
<213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

5 <400> 34
 ctfcgggcat ggcactctta aaaaagtcac gccgtttcat atg 43

<210> 35
 <211> 49
 10 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

15 <400> 35
 gttatgtaca ggaacgcact atattttca aagatgacgg gaactacaa 49

<210> 36
 <211> 49
 20 <212> ADN
 <213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

25 <400> 36
 ttgtagtcc cgtcatcttt gaaaaatata gtgcgttcct gtacataac 49

30 <210> 37
 <211> 53
 <212> ADN
 <213> Secuencia Artificial

35 <220>
 <223> Cebador de oligonucleótidos

<400> 37
 acaaaagaat ggaatcaaag ctaactttaa aattcgccac aacattgaag atg 53

40 <210> 38
 <211> 53
 <212> ADN
 <213> Secuencia Artificial

45 <220>
 <223> Cebador de oligonucleótidos

<400> 38
 50 catcttcaat gttgtggcga attttaaagt tagctttgat tccattcttt tgt 53

<210> 39
 <211> 39
 <212> ADN
 55 <213> Secuencia Artificial

<220>
 <223> Cebador de oligonucleótidos

60 <400> 39

ccccaagcta gcttgattc tcaccaataa aaaacgccc 39
 <210> 40
 <211> 43
 5 <212> ADN
 <213> Secuencia Artificial
 <220>
 <223> Cebador de oligonucleótidos
 10
 <400> 40
 tacctatggt gtccaatgct ttcccgta tccgcatcat atg 43
 <210> 41
 15 <211> 43
 <212> ADN
 <213> Secuencia Artificial
 <220>
 20 <223> Cebador de oligonucleótidos
 <400> 41
 catatgatcc ggataacggg aaaagcattg aacaccatag gta 43
 25 <210> 42
 <211> 50
 <212> ADN
 <213> Secuencia Artificial
 30 <220>
 <223> Cebador de oligonucleótidos
 <400> 42
 35 gttatgtaca ggaacgcact atattttta aagatgacgg gaactacaag 50
 <210> 43
 <211> 50
 <212> ADN
 <213> Secuencia Artificial
 40 <220>
 <223> Cebador de oligonucleótidos
 <400> 43
 45 ctgtagttc ccgtcatctt taaaaaatat agtgcgttcc tgtacataac 50
 <210> 44
 <211> 43
 <212> ADN
 50 <213> Secuencia Artificial
 <220>
 <223> Cebador de oligonucleótidos
 55 <220>
 <221> característica misc
 <222> 16, 17
 <223> n = A, T, C o G
 60 <400> 44
 gtcgaaatcg gtaacnnkkg tatgttcaga ccagaaatgc tcg 43

<210> 45
 <211> 43
 <212> ADN
 5 <213> Secuencia Artificial

 <220>
 <223> Cebador de oligonucleótidos

 10 <220>
 <221> característica misc
 <222> 27, 28
 <223> n = A, T, C o G

 15 <400> 45
 cgagcatttc tggctgaac atacmnnngt taccgatttc gac 43

 <210> 46
 <211> 47
 20 <212> ADN
 <213> Secuencia Artificial

 <220>
 <223> Cebador de oligonucleótidos

 25 <220>
 <221> característica misc
 <222> 14, 15, 23, 24, 32, 33
 <223> n = A, T, C o G

 30 <400> 46
 caagcctacc tacnnkcctt acnnkgagcc annkatggaa atctttt 47

 <210> 47
 35 <211> 47
 <212> ADN
 <213> Secuencia Artificial

 <220>
 40 <223> Cebador de oligonucleótidos

 <220>
 <221> característica misc
 <222> 15, 16, 24, 25, 33, 34
 45 <223> n = A, T, C o G

 <400> 47
 aaaagatttc catmnnntggc tcmnngtaag gmnngtaggt aggcttg 47

 50 <210> 48
 <211> 432
 <212> PRT
 <213> Thermus thermophilus

 55 <400> 48

Met	Ala	Gly	Thr	Gly	His	Thr	Pro	Glu	Glu	Ala	Leu	Ala	Leu	Leu	Lys
1				5					10					15	
Arg	Gly	Ala	Glu	Glu	Ile	Val	Pro	Glu	Glu	Glu	Leu	Leu	Ala	Lys	Leu
			20					25					30		
Lys	Glu	Gly	Arg	Pro	Leu	Thr	Val	Lys	Leu	Gly	Ala	Asp	Pro	Thr	Arg
		35					40					45			
Pro	Asp	Leu	His	Leu	Gly	His	Ala	Val	Val	Leu	Arg	Lys	Met	Arg	Gln
	50					55					60				
Phe	Gln	Glu	Leu	Gly	His	Lys	Val	Val	Leu	Ile	Ile	Gly	Asp	Phe	Thr
65					70					75					80
Gly	Met	Ile	Gly	Asp	Pro	Ser	Gly	Arg	Ser	Lys	Thr	Arg	Pro	Pro	Leu
				85					90					95	
Thr	Leu	Glu	Glu	Thr	Arg	Glu	Asn	Ala	Lys	Thr	Tyr	Val	Ala	Gln	Ala
			100					105					110		
Gly	Lys	Ile	Leu	Arg	Gln	Glu	Pro	His	Leu	Phe	Glu	Leu	Arg	Tyr	Asn
		115				120						125			
Ser	Glu	Trp	Leu	Glu	Gly	Leu	Thr	Phe	Lys	Glu	Val	Val	Arg	Leu	Thr
	130					135					140				
Ser	Leu	Met	Thr	Val	Ala	Gln	Met	Leu	Glu	Arg	Glu	Asp	Phe	Lys	Lys
145					150					155					160
Arg	Tyr	Glu	Ala	Gly	Ile	Pro	Ile	Ser	Leu	His	Glu	Leu	Leu	Tyr	Pro
				165					170					175	
Phe	Ala	Gln	Ala	Tyr	Asp	Ser	Val	Ala	Ile	Arg	Ala	Asp	Val	Glu	Met
			180					185					190		
Gly	Gly	Thr	Asp	Gln	Arg	Phe	Asn	Leu	Leu	Val	Gly	Arg	Glu	Val	Gln
		195					200					205			
Arg	Ala	Tyr	Gly	Gln	Ser	Pro	Gln	Val	Cys	Phe	Leu	Met	Pro	Leu	Leu
	210					215						220			
Val	Gly	Leu	Asp	Gly	Arg	Glu	Lys	Met	Ser	Lys	Ser	Leu	Asp	Asn	Tyr
225					230					235					240
Ile	Gly	Leu	Thr	Glu	Pro	Pro	Glu	Ala	Met	Phe	Lys	Lys	Leu	Met	Arg
				245					250					255	
Val	Pro	Asp	Pro	Leu	Leu	Pro	Ser	Tyr	Phe	Arg	Leu	Leu	Thr	Asp	Leu
		260						265					270		
Glu	Glu	Glu	Glu	Ile	Glu	Ala	Leu	Leu	Lys	Ala	Gly	Pro	Val	Pro	Ala
		275					280					285			
His	Arg	Val	Leu	Ala	Arg	Leu	Leu	Thr	Ala	Ala	Tyr	Ala	Leu	Pro	Gln

```

      290      295      300
Ile Pro Pro Arg Ile Asp Arg Ala Phe Tyr Glu Ser Leu Gly Tyr Ala
305      310      315      320
Trp Glu Ala Phe Gly Arg Asp Lys Glu Ala Gly Pro Glu Glu Val Arg
      325      330      335
Arg Ala Glu Ala Arg Tyr Asp Glu Val Ala Lys Gly Gly Ile Pro Glu
      340      345      350
Glu Ile Pro Glu Val Thr Ile Pro Ala Ser Glu Leu Lys Glu Gly Arg
      355      360      365
Ile Trp Val Ala Arg Leu Phe Thr Leu Ala Gly Leu Thr Pro Ser Asn
      370      375      380
Ala Glu Ala Arg Arg Leu Ile Gln Asn Arg Gly Leu Arg Leu Asp Gly
385      390      395      400
Glu Val Leu Thr Asp Pro Met Leu Gln Val Asp Leu Ser Arg Pro Arg
      405      410      415
Ile Leu Gln Arg Gly Lys Asp Arg Phe Val Arg Val Arg Leu Ser Asp
      420      425      430

```

<210> 49
 <211> 372
 <212> PRT
 <213> Homo sapiens

<400> 49

```

Met Gly Asp Ala Pro Ser Pro Glu Glu Lys Leu His Leu Ile Thr Arg
1      5      10      15
Asn Leu Gln Glu Val Leu Gly Glu Glu Lys Leu Lys Glu Ile Leu Lys
      20      25      30
Glu Arg Glu Leu Lys Ile Tyr Trp Gly Thr Ala Thr Thr Gly Lys Pro
      35      40      45
His Val Ala Tyr Phe Val Pro Met Ser Lys Ile Ala Asp Phe Leu Lys
      50      55      60
Ala Gly Cys Glu Val Thr Ile Leu Phe Ala Asp Leu His Ala Tyr Leu
65      70      75      80
Asp Asn Met Lys Ala Pro Trp Glu Leu Leu Glu Leu Arg Val Ser Tyr
      85      90      95
Tyr Glu Asn Val Ile Lys Ala Met Leu Glu Ser Ile Gly Val Pro Leu
      100      105      110
Glu Lys Leu Lys Phe Ile Lys Gly Thr Asp Tyr Gln Leu Ser Lys Glu
      115      120      125
Tyr Thr Leu Asp Val Tyr Arg Leu Ser Ser Val Val Thr Gln His Asp
      130      135      140
Ser Lys Lys Ala Gly Ala Glu Val Val Lys Gln Val Glu His Pro Leu
145      150      155      160
Leu Ser Gly Leu Leu Tyr Pro Gly Leu Gln Ala Leu Asp Glu Glu Tyr
      165      170      175
Leu Lys Val Asp Ala Gln Phe Gly Gly Ile Asp Gln Arg Lys Ile Phe
      180      185      190
Thr Phe Ala Glu Lys Tyr Leu Pro Ala Leu Gly Tyr Ser Lys Arg Val
      195      200      205
His Leu Met Asn Pro Met Val Pro Gly Leu Thr Gly Ser Lys Met Ser
      210      215      220
Ser Ser Glu Glu Glu Ser Lys Ile Asp Leu Leu Asp Arg Lys Glu Asp
225      230      235      240
Val Lys Lys Lys Leu Lys Lys Ala Phe Cys Glu Pro Gly Asn Val Glu
      245      250      255
Asn Asn Gly Val Leu Ser Phe Ile Lys His Val Leu Phe Pro Leu Lys

```

```

                260                265                270
Ser Glu Phe Val Ile Leu Arg Asp Glu Lys Trp Gly Gly Asn Lys Thr
                275                280                285
Tyr Thr Ala Tyr Val Asp Leu Glu Lys Asp Phe Ala Ala Glu Val Val
                290                295                300
His Pro Gly Asp Leu Lys Asn Ser Val Glu Val Ala Leu Asn Lys Leu
305                310                315                320
Leu Asp Pro Ile Arg Glu Lys Phe Asn Thr Pro Ala Leu Lys Lys Leu
                325                330                335
Ala Ser Ala Ala Tyr Pro Asp Pro Ser Lys Gln Lys Pro Met Ala Lys
                340                345                350
Gly Pro Ala Lys Asn Ser Glu Pro Glu Glu Val Ile Leu Glu His His
                355                360                365
His His His His
                370

```

<210> 50
 <211> 424
 <212> PRT
 <213> Escherichia coli

<400> 50

```

Met Ala Ser Ser Asn Leu Ile Lys Gln Leu Gln Glu Arg Gly Leu Val
 1                5                10                15
Ala Gln Val Thr Asp Glu Glu Ala Leu Ala Glu Arg Leu Ala Gln Gly
                20                25                30
Pro Ile Ala Leu Tyr Cys Gly Phe Asp Pro Thr Ala Asp Ser Leu His
                35                40                45
Leu Gly His Leu Val Pro Leu Leu Cys Leu Lys Arg Phe Gln Gln Ala
 50                55                60
Gly His Lys Pro Val Ala Leu Val Gly Gly Ala Thr Gly Leu Ile Gly
65                70                75                80
Asp Pro Ser Phe Lys Ala Ala Glu Arg Lys Leu Asn Thr Glu Glu Thr
                85                90                95
Val Gln Glu Trp Val Asp Lys Ile Arg Lys Gln Val Ala Pro Phe Leu
                100                105                110
Asp Phe Asp Cys Gly Glu Asn Ser Ala Ile Ala Ala Asn Asn Tyr Asp
                115                120                125
Trp Phe Gly Asn Met Asn Val Leu Thr Phe Leu Arg Asp Ile Gly Lys
                130                135                140
His Phe Ser Val Asn Gln Met Ile Asn Lys Glu Ala Val Lys Gln Arg
145                150                155                160
Leu Asn Arg Glu Asp Gln Gly Ile Ser Phe Thr Glu Phe Ser Tyr Asn
                165                170                175
Leu Leu Gln Gly Tyr Asp Phe Ala Cys Leu Asn Lys Gln Tyr Gly Val
                180                185                190
Val Leu Gln Ile Gly Gly Ser Asp Gln Trp Gly Asn Ile Thr Ser Gly
                195                200                205
Ile Asp Leu Thr Arg Arg Leu His Gln Asn Gln Val Phe Gly Leu Thr
                210                215                220
Val Pro Leu Ile Thr Lys Ala Asp Gly Thr Lys Phe Gly Lys Thr Glu
225                230                235                240
Gly Gly Ala Val Trp Leu Asp Pro Lys Lys Thr Ser Pro Tyr Lys Phe
                245                250                255
Tyr Gln Phe Trp Ile Asn Thr Ala Asp Ala Asp Val Tyr Arg Phe Leu
                260                265                270
Lys Phe Phe Thr Phe Met Ser Ile Glu Glu Ile Asn Ala Leu Glu Glu

```

```

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

```

<210> 51
 <211> 442
 <212> PRT
 <213> Drosophila melanogaster

 <400> 51

```

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

```

```

                245                250                255
Val Trp Leu Asp Gly Asn Lys Thr Ser Pro Phe Ala Leu Tyr Gln Phe
                260                265                270
Phe Leu Arg Met Pro Asp Ser Glu Val Glu Lys Leu Leu Lys Leu Phe
                275                280                285
Thr Phe Ile Pro Leu Pro Gln Val Glu Gln Leu Met Arg Glu His Thr
                290                295                300
Lys Glu Pro Glu Lys Arg Lys Ala Gln Thr Leu Leu Ala Glu Asp Val
305                310                315                320
Thr Leu Leu Val His Gly Glu Ser Gly Leu Lys Gln Ala Glu Arg Val
                325                330                335
Thr Asn Ala Leu Tyr Lys Gly Asn Val Glu Gly Leu Ala Glu Leu Asn
                340                345                350
Leu Ser Glu Ile Gln Gln Thr Phe Gln Gly Ala Thr Met Val Asn Leu
                355                360                365
Leu Thr Glu Pro Gly Met Ser Ile Leu Glu Leu Ala Met Lys Ala Lys
                370                375                380
Cys Phe Pro Thr Glu Thr Asp Ala Val Arg Ile Ile Asn Ala Gly Gly
385                390                395                400
Phe Tyr Val Asn Gln Lys Arg Val Gln Asn Ile Ala Glu Val Leu Thr
                405                410                415
Thr Gly Val His Ile Leu Arg Asn Gly Ile Ser Leu Leu Arg Val Gly
                420                425                430
Lys Arg Asn Phe Tyr Ile Val Arg Trp Gln
                435                440

```

<210> 52
 <211> 492
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 52

```

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

```

```

      195              200              205
Arg His Ile Arg Val Gln Ser Met Leu Ala Arg Asp Ser Ile Ser Ser
 210              215              220
Arg Leu Gln Thr Lys Asn Gly Leu Gly Phe Asn Glu Phe Thr Tyr Gln
225              230              235              240
Val Leu Gln Ala Tyr Asp Phe Tyr His Leu Tyr Lys Glu Glu Asn Val
      245              250              255
Thr Ile Gln Val Gly Gly Asn Asp Gln Trp Gly Asn Ile Thr Ala Gly
 260              265              270
Ile Asp Leu Ile Asn Arg Ile Gln Pro Ile Lys Asn Lys Gly Leu Pro
 275              280              285
Phe Gly Ile Thr Val Pro Leu Leu Thr Thr Ala Thr Gly Glu Lys Phe
 290              295              300
Gly Lys Ser Ala Gly Asn Ala Val Phe Ile Asp Pro Ser Ile Asn Thr
305              310              315              320
Ala Tyr Asp Val Tyr Gln Phe Phe Tyr Asn Thr Leu Asp Ala Asp Val
      325              330              335
Pro Lys Phe Leu Lys Ile Phe Thr Phe Leu Asn Ser Ser Glu Ile Lys
 340              345              350
Lys Ile Val Glu Thr His Ile Lys Ser Pro Ser Leu Arg Tyr Gly Gln
 355              360              365
Thr Leu Leu Ala Lys Glu Val Thr Asp Met Leu Tyr Gly Val Gly Ser
 370              375              380
Gly Ser Asp Ser Glu Ala Leu Ser Asn Ile Ile Phe Gly Arg Tyr Asp
385              390              395              400
Gly Thr Leu Ser Ala Ala Lys Leu Val Asp Leu Cys Lys Lys Ala Arg
      405              410              415
Ile Leu Gln Tyr Ala Asp Arg Glu Ile Asp Leu Ile Lys Leu Ile Cys
 420              425              430
Lys Leu Val Asn Cys Ser Val Ser Glu Ala Arg Arg Lys Leu Ser Gln
 435              440              445
Gly Ser Val Tyr Leu His His Ser Lys Ser Lys Val Asn Glu Asn Ile
 450              455              460
Ser Asn Leu Ala Pro Phe Leu Ile Asp Asp Arg Val Leu Ile Leu Arg
465              470              475              480
Ile Gly Lys Gln Lys Cys Phe Ile Ile Glu Met Arg
      485              490

```

<210> 53

<211> 445

<212> PRT

<213> Schizosaccharomyces pombe

<400> 53

```

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

```

			100					105					110			
Ser	Tyr	Ala	Gln	Asp	Cys	Asn	Tyr	Pro	Phe	Ser	Gln	Met	Pro	Ser	Ser	
		115					120					125				
Ser	Gln	Trp	Ser	Ile	Val	Arg	Asn	Ser	Ser	Trp	Tyr	Glu	Asn	Leu	Lys	
		130					135					140				
Leu	Leu	Lys	Phe	Leu	Ser	Ser	Val	Gly	Pro	His	Val	Arg	Val	Ser	Gln	
145					150					155					160	
Met	Leu	Ala	Arg	Asp	Ser	Val	Thr	Thr	Arg	Leu	Gln	Ser	Pro	Ser	Gly	
			165						170						175	
Leu	Ser	Phe	Ala	Glu	Leu	Thr	Tyr	Gln	Leu	Leu	Gln	Ala	Tyr	Asp	Tyr	
			180					185						190		
Ser	Tyr	Leu	Tyr	Glu	Asn	His	Ser	Val	Asn	Leu	Gln	Ile	Gly	Gly	Ser	
		195					200					205				
Asp	Gln	Trp	Gly	Asn	Ile	Thr	Ala	Gly	Thr	Asp	Leu	Val	Arg	Arg	Thr	
		210				215					220					
His	Pro	Asn	Ala	Asn	Val	Tyr	Ala	Leu	Thr	Thr	Pro	Leu	Leu	Thr	Ser	
225					230					235					240	
Ser	Ser	Gly	Gln	Lys	Leu	Gly	Lys	Ser	Ala	Gly	Asn	Ala	Ile	Trp	Leu	
			245						250					255		
Asp	Pro	Lys	Leu	Thr	Asp	Ser	Tyr	Ser	Leu	Tyr	Gln	Tyr	Phe	Ile	Ser	
			260					265						270		
Ala	Pro	Asp	Asp	Leu	Ala	Cys	Lys	Cys	Leu	Asp	Met	Leu	Thr	Leu	Leu	
		275					280					285				
Pro	Leu	Glu	Gln	Leu	Glu	Gln	Ile	Lys	Ala	Glu	His	Glu	Lys	Asp	Pro	
		290				295					300					
Ser	Gln	Arg	Ile	Val	His	Lys	Tyr	Leu	Ala	Ser	Asn	Val	Val	Arg	Met	
305					310					315					320	
Val	His	Gly	Lys	Lys	Ala	Leu	Glu	Leu	Ala	Gln	Ile	Gln	Thr	Lys	Leu	
			325						330					335		
Leu	His	Gly	Ala	His	Gln	Ala	Pro	Phe	Gly	Phe	Tyr	Ser	Glu	Ala	Pro	
			340					345						350		
Gln	Gln	Gly	Asp	Ser	Phe	Pro	Ser	Leu	Pro	Glu	Ile	Arg	Ala	Leu	Phe	
		355					360					365				
Lys	Asp	Cys	Lys	Phe	Tyr	Arg	Thr	Ile	Asp	Ser	Ser	Ile	Lys	Asp	Gln	
		370				375					380					
Pro	Phe	Ser	Arg	Leu	Leu	Arg	Thr	Leu	Gln	Ile	Tyr	Thr	Ser	Arg	Lys	
385					390					395					400	
Glu	Ala	Thr	Glu	His	Ile	Leu	Ser	Gly	Ala	Val	Ser	Leu	Gly	His	Lys	
			405						410					415		
Pro	Ile	Leu	Asp	Ser	Asn	Tyr	Lys	Phe	Pro	Asp	Asn	Ser	Leu	Phe	Val	
			420					425								

<210> 54
<211> 528
<212> PRT
<213> *Xenopus laevis*

<400> 54

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

50	55	60
Ala Gly Cys Glu Val Thr Ile Leu Phe Ala Asp Leu His Ala Tyr Leu		
65	70	75
Asp Asn Met Lys Ala Pro Trp Asp Leu Leu Glu Leu Arg Thr Arg Tyr		80
	85	90
Tyr Glu Gln Val Ile Gln Ala Met Leu Gln Ser Ile Gly Val Pro Leu		95
	100	105
Glu Arg Leu Arg Phe Ile Arg Gly Thr Glu Phe Gln Leu Ser Lys Glu		110
	115	120
Tyr Thr Leu Asp Val Tyr Arg Leu Ser Ser Val Val Thr Gln His Asp		125
	130	135
Ala Lys Lys Ala Gly Ala Glu Val Val Lys Gln Val Glu His Pro Leu		140
	145	150
Leu Ser Gly Leu Leu Tyr Pro Gly Leu Gln Ala Leu Asp Glu Glu Tyr		155
	165	170
Leu Lys Val Asp Ala Gln Phe Gly Gly Val Asp Gln Arg Lys Ile Phe		175
	180	185
Thr Phe Ala Glu Lys Tyr Leu Pro Ala Leu Gly Tyr Ala Lys Arg Ile		190
	195	200
His Leu Met Asn Pro Met Val Pro Gly Leu Thr Gly Ala Lys Met Ser		205
	210	215
Ser Ser Glu Glu Glu Ser Lys Ile Asp Leu Leu Asp Ser Pro Ala Asp		220
	225	230
Val Lys Lys Lys Leu Lys Lys Ala Phe Cys Glu Pro Gly Asn Val Glu		235
	245	250
Asn Asn Gly Val Leu Ser Phe Val Arg His Val Leu Phe Pro Leu Lys		255
	260	265
Ser Glu Phe Val Val Leu Arg Asp Glu Lys Phe Gly Gly Asn Lys Thr		270
	275	280
Tyr Thr Asp Phe Glu Thr Leu Glu Lys Asp Phe Ala Glu Glu Leu Val		285
	290	295
His Pro Gly Asp Leu Lys Ala Ser Val Glu Lys Ala Leu Asn Lys Leu		300
	305	310
Leu His Pro Ile Arg Glu Lys Phe Asn Ser Pro Glu Met Lys Lys Leu		315
	325	330
Ser Asn Asp Ala Tyr Pro Asp Ala Ser Lys Gln Lys Ser Val Pro Lys		335
	340	345
Gly Ser Thr Lys Asn Ser Gly Thr Glu Glu Ile Asp Pro Ser Leu Leu		350
	355	360
Asp Leu Arg Val Gly Lys Ile Leu Ser Val Ser Gln His Pro Asp Ala		365
	370	375
Asp Ser Leu Tyr Val Glu Ser Val Asp Val Gly Glu Ala Asn Pro Arg		380
	385	390
Cys Val Val Ser Gly Leu Val Gln Tyr Val Pro Ser Asp Gln Leu Leu		395
	405	410
Gly Arg Ser Val Val Leu Leu Cys Asn Leu Lys Pro Gln Lys Met Arg		415
	420	425
Gly Ile Glu Ser Gln Gly Met Leu Leu Cys Ala Ser Thr Glu Gly Glu		430
	435	440
Gln Lys Gln Val Glu Pro Leu Asp Pro Pro Ser Gly Ser Ala Pro Gly		445
	450	455
Glu Arg Ile Tyr Ile Glu Gly Tyr Glu Asn Gly Glu Pro Glu Gly Glu		460
	465	470
Leu Lys Pro Lys Lys Lys Val Phe Glu Lys Leu Gln Val Asp Phe Arg		475
	485	490
Ile Ser Asp Asp Leu Cys Ala Gln Trp Lys Gly Lys Asn Phe Leu Thr		495
	500	505
Lys Leu Gly Ser Val Thr Cys Lys Thr Leu Arg Gly Gly Ser Ile Gly		510

515

520

525

<210> 55
<211> 402

<212> PRT
 <213> *Helicobacter pylori*

<400> 55

```

Met Glu Gln Lys Ile Ala Ile Ala Leu Lys Glu Ile Ala Arg Gly Thr
 1      5      10      15
Asn Glu Ile Ile Gly Leu Glu Tyr Ile Glu Lys Leu Val Arg Lys Tyr
      20      25      30
Tyr Glu Thr Asn Glu Arg Phe Ile Val Lys Ala Gly Phe Asp Pro Thr
      35      40      45
Ala Pro Asp Leu His Leu Gly His Thr Val Leu Ile Gln Lys Leu Ala
      50      55      60
Leu Leu Gln Gln Tyr Gly Ala Arg Val Lys Phe Leu Ile Gly Asp Phe
      65      70      75      80
Thr Ala Met Ile Gly Asp Pro Thr Gly Lys Asn Glu Thr Arg Lys Pro
      85      90      95
Leu Asn Arg Glu Gln Val Leu Glu Asn Ala Lys Thr Tyr Glu Gln
      100      105      110
Ile Tyr Lys Ile Leu Asp Glu Lys His Thr Glu Val Cys Phe Asn Ser
      115      120      125
Thr Trp Leu Asp Ala Leu Gly Ala Lys Gly Met Ile Glu Leu Cys Ala
      130      135      140
Lys Phe Ser Val Ala Arg Met Leu Glu Arg Asp Asp Phe Thr Lys Arg
      145      150      155      160
Tyr Lys Glu Asn Arg Pro Ile Ser Ile Val Glu Phe Leu Tyr Pro Leu
      165      170      175
Leu Gln Gly Tyr Asp Ser Val Ala Met Asp Ala Asp Ile Glu Leu Gly
      180      185      190
Gly Asn Asp Gln Lys Phe Asn Leu Leu Val Gly Arg Phe Leu Gln Arg
      195      200      205
Ala Tyr Gly Leu Asn Lys Glu Gln Ser Val Ile Thr Met Pro Leu Leu
      210      215      220
Glu Gly Leu Asp Gly Val Gln Lys Met Ser Lys Ser Leu Gly Asn Tyr
      225      230      235      240
Val Gly Ile Thr Glu Glu Pro Asn Ala Met Phe Gly Lys Ile Met Ser
      245      250      255
Val Ser Asp Asp Leu Met Trp Arg Tyr Thr Leu Leu Ser Thr Lys
      260      265      270
Thr Leu Glu Glu Ile Glu Asp Leu Lys His Gly Ile Leu Asn Gln Thr
      275      280      285
Leu His Pro Lys Ala Val Lys Glu Asp Leu Ala Ser Glu Ile Val Ala
      290      295      300
Arg Tyr Tyr Asp Asn Asp Gln Ala Ile Lys Ala Lys Glu Gln Phe Ser
      305      310      315      320
Lys Val Phe Ser Ala Asn Leu Leu Pro Glu Ile Leu Ser Glu Ser Asp
      325      330      335
Phe Asp Glu Gly Val Gly Ile Leu Asp Val Leu Lys Gln Ile Gly Phe
      340      345      350
Cys Pro Ser Thr Ser Gln Ala Arg Asp Ile Gln Gly Gly Gly Val
      355      360      365
Lys Ile Asn Gln Glu Val Ile Lys Asn Glu Ser Tyr Arg Phe Val Lys
      370      375      380
Gly Asn Tyr Val Ile Gln Leu Gly Lys Lys Arg Phe Met Lys Leu Asn
      385      390      395      400
Ile Asn

```

<210> 56
 <211> 447
 <212> PRT
 <213> *Caenorhabditis elegans*

<400> 56

Met	Phe	Asn	Ser	Lys	Arg	Ile	Gly	Asn	Val	Ala	Leu	Lys	Thr	Val	Arg
1				5					10					15	
Ala	Pro	Arg	Glu	Ser	Ser	Phe	Val	Asp	Tyr	Ile	Thr	Asp	Leu	Asn	Ala
			20					25					30		
Arg	Lys	Gln	Leu	Gln	His	Ser	Tyr	Pro	Thr	Asp	Leu	Leu	Ser	Lys	Cys
		35					40					45			
Ser	Glu	Asp	Leu	Arg	Gln	Leu	Pro	Pro	Tyr	Val	Tyr	Ala	Gly	Phe	Asp
	50					55					60				
Pro	Thr	Ala	Glu	Ser	Leu	His	Ile	Gly	Asn	Leu	Ile	Leu	Val	Asn	
65					70					75				80	
Leu	Ile	Arg	Ala	Gln	Gln	Phe	Gly	Ile	Arg	Pro	Ile	Ala	Leu	Ile	Gly
				85					90					95	
Glu	Phe	Thr	Ala	Ser	Ile	Gly	Asp	Pro	Ser	Gly	Lys	Lys	Ser	Glu	Arg
			100					105					110		
Gly	Leu	Leu	Ala	Gly	Asp	Val	Ile	Met	His	Asn	Ser	Arg	Lys	Val	Thr
			115					120					125		
Asp	Gln	Ile	Cys	Lys	Ile	Phe	Glu	Asn	Ala	Pro	Gly	Ser	Ser	Glu	Lys
	130					135					140				
Pro	Ile	Ile	Val	Asn	Asn	Asn	Asp	Trp	Leu	Gly	Lys	Ile	Ser	Leu	Arg
145				150						155				160	
Asp	Phe	Leu	Arg	Glu	Cys	Lys	Asn	Met	Gln	Val	Gly	Lys	Met	Leu	Arg
				165					170					175	
Met	Asn	Thr	Ile	Lys	Asn	Arg	Leu	Glu	Val	Gly	Leu	Ser	Tyr	Thr	Glu
			180					185					190		
Phe	Ser	Tyr	Gln	Thr	Met	Gln	Ala	Phe	Asp	Trp	Tyr	Thr	Leu	Ser	Glu
	195						200					205			
Lys	Tyr	Gly	Cys	Arg	Phe	Gln	Leu	Gly	Gly	Tyr	Asp	Gln	Leu	Gly	His
	210					215					220				
Leu	Asp	Phe	Gly	Ala	His	Tyr	Ile	Lys	Lys	Met	Met	Asn	Gln	Ala	Phe
225					230					235				240	
Ala	Ala	Gly	Val	Cys	Phe	Pro	Ile	Leu	Thr	Asp	Ser	Thr	Gly	Ala	Lys
				245						250				255	
Leu	Gly	Lys	Ser	Glu	Gly	Gly	Gly	Ala	Leu	Trp	Leu	Asp	Ala	Thr	Lys
			260					265					270		
Thr	Ser	Pro	Phe	His	Phe	Tyr	Gln	Phe	Phe	Ala	Gln	Leu	His	Asp	Asp
	275						280						285		
Lys	Ala	Glu	Glu	Leu	Leu	Leu	Leu	Phe	Ser	Leu	Gln	Asp	Ile	Glu	His
	290					295					300				
Ile	Arg	Asp	Val	Leu	Lys	Asn	His	Arg	Ser	Asn	Leu	Gly	Gln	Trp	Ile
305					310					315				320	
Ala	Gln	Arg	Glu	Leu	Ala	Ala	Glu	Ile	Thr	Arg	Ile	Val	His	Gly	Lys
				325						330				335	
Glu	Gly	Leu	Glu	Val	Ala	Met									

<210> 57
<211> 412
<212> PRT
<213> *Pseudomonas aeruginosa*

<400> 57

Met	Ser	Val	Pro	Thr	His	Gln	Gln	Asp	Leu	Ile	Ala	Leu	Leu	Glu	Glu
1				5					10					15	
Arg	Gly	Phe	Val	His	Gln	Cys	Thr	Asp	Arg	Asp	Gly	Leu	Ala	Ala	His
			20					25					30		
Leu	Ala	Ala	Gly	Pro	Ala	Thr	Ala	Tyr	Leu	Gly	Phe	Asp	Ala	Thr	Ala
		35					40				45				
Asp	Ser	Leu	His	Val	Gly	His	Leu	Gln	Gly	Leu	Met	Leu	Met	Arg	Trp
	50					55					60				
Leu	Gln	Lys	Ala	Gly	His	Arg	Pro	Leu	Leu	Ile	Gly	Gly	Ala	Thr	
65					70					75				80	
Thr	Arg	Ile	Gly	Asp	Pro	Ser	Phe	Arg	Asp	Ser	Ser	Arg	Pro	Ile	Leu
			85						90					95	
Thr	Glu	Ala	Gln	Ile	Gln	Ala	Asn	Ile	Asp	Gly	Ile	Ala	Arg	Val	Phe
		100						105					110		
Ser	Arg	Tyr	Val	Glu	Leu	His	Asp	Asp	Ser	Leu	Val	Asn	Asn	Ala	Glu
		115					120					125			
Trp	Leu	Asp	Gly	Val	Gly	Tyr	Leu	Glu	Phe	Leu	Asp	Arg	Val	Gly	Arg
	130					135					140				
His	Phe	Ser	Ile	Asn	Arg	Leu	Leu	Thr	Phe	Asp	Ala	Ile	Arg	Gln	Arg
145				150						155				160	
Leu	Asp	Arg	Glu	His	Ser	Leu	Ser	Phe	Leu	Glu	Phe	Gly	Tyr	Thr	Leu
			165					170						175	
Leu	Gln	Ala	Tyr	Asp	Phe	Val	Glu	Leu	Ser	Arg	Arg	Arg	Gly	Cys	Thr
		180					185						190		
Leu	Gln	Leu	Gly	Gly	Ala	Asp	Gln	Trp	Ala	Asn	Ile	Ile	Asn	Gly	Val
	195						200					205			
Glu	Leu	Ser	Arg	Arg	Gln	Gly	Gly	Ala	Gln	Leu	Phe	Gly	Leu	Thr	Met
	210					215					220				
Pro	Leu	Leu	Ala	Thr	Ser	Asp	Gly	Arg	Lys	Met	Gly	Lys	Ser	Ala	Gln
225				230						235				240	
Gly	Ala	Val	Trp	Leu	Asn	Ala	Glu	Arg	Leu	Ala	Pro	Phe	Asp	Phe	Trp
			245						250					255	
Gln	Phe	Trp	Arg	Asn	Cys	Asp	Asp	Arg	Asp	Val	Gly	Arg	Phe	Leu	Ala
		260					265						270		
Leu	Phe	Ser	Glu	Leu	Pro	Met	Asp	Glu	Val	Arg	Arg	Leu	Gly	Ala	Leu
		275					280					285			
Gln	Gly	Ala	Glu	Leu	Asn	Glu	Ala	Lys	Val	Val	Leu	Ala	Asn	Ala	Ala
	290					295					300				
Thr	Ala	Leu	Ala	His	Gly	Glu	His	Ala	Ala	Arg	Ser	Ala	Ala	Asp	Ala
305				310						315				320	
Ala	Arg	Gly	Val	Phe	Ala	Asp	Gly	Thr	Arg	Asp	Ser	Gly	Leu	Pro	Val
				325					330					335	
Met	Lys	Leu	Ser	Arg	Ala	Arg	Leu	Ala	Gln	Gly	Leu	Ser	Leu	Thr	Asp
			340					345					350		
Leu	Leu	Leu	Glu	His	Ala	Ile	Gln	Pro	Ser	Arg	Ser	Ala	Val	Arg	Arg
		355					360					365			
Leu	Ala	Ala	Gly	Gly	Gly	Leu	Arg	Leu	Asp	Gly	Thr	Pro	Val	Ser	Asp
	370					375					380				
Pro	Asp	Thr	Pro	Leu	Ala	Gly	Glu	Val	Asp	Gly	Leu	Arg	Leu	Ser	Leu
385				390						395				400	
Gly	Lys	Lys	Gln	His	Leu	His	Leu	Arg	Leu	Glu	Asp				
			405						410						

<210> 58

<211> 399

<212> PRT

<213> Pseudomonas aeruginosa

<400> 58

```

Met Lys Ser Val Glu Glu Gln Leu Ala Leu Ile Gln Arg Gly Ala Asp
 1      5      10      15
Glu Ile Leu Val Glu Ala Glu Leu Val Ala Lys Leu Lys Arg Gly Gln
 20      25      30
Pro Leu Arg Ile Lys Ala Gly Phe Asp Pro Thr Ala Pro Asp Leu His
 35      40      45
Leu Gly His Thr Val Leu Ile Asn Lys Leu Arg Gln Phe Gln Asp Leu
 50      55      60
Gly His Gln Val Ile Phe Leu Ile Gly Asp Phe Thr Gly Met Ile Gly
 65      70      75      80
Asp Pro Ser Gly Lys Ser Val Thr Arg Pro Pro Leu Thr Arg Glu Gln
 85      90      95
Val Leu Glu Asn Ala Glu Thr Tyr Lys Ser Gln Val Phe Lys Ile Leu
100      105      110
Asp Pro Ala Lys Thr Glu Val Ala Phe Asn Ser Thr Trp Met Asp Gln
115      120      125
Leu Thr Pro Ala Asp Phe Ile Arg Leu Ala Ser Gln Tyr Thr Val Ala
130      135      140
Arg Met Leu Glu Arg Asp Asp Phe Ser Lys Arg Tyr Ala Ser Asn Gln
145      150      155      160
Pro Ile Ala Ile His Glu Phe Leu Tyr Pro Leu Val Gln Gly Tyr Asp
165      170      175
Ser Val Ala Leu Lys Ala Asp Val Glu Leu Gly Gly Thr Asp Gln Lys
180      185      190
Phe Asn Leu Leu Met Gly Arg Glu Leu Gln Arg Ala Tyr Gly Gln Glu
195      200      205
Ala Gln Val Ile Leu Thr Met Pro Leu Leu Glu Gly Leu Asp Gly Val
210      215      220
Lys Lys Met Ser Lys Ser Leu Gly Asn Tyr Ile Gly Ile Gln Glu Ala
225      230      235      240
Pro Gly Val Met Tyr Ser Lys Leu Val Ser Ile Pro Asp Thr Leu Met
245      250      255
Trp Arg Tyr Phe Glu Leu Leu Ser Phe Arg Ser Leu Asp Glu Ile Asp
260      265      270
Ser Phe Arg Lys Asp Val Glu Ala Gly Ala Asn Pro Arg Asp Ile Lys
275      280      285
Ile Lys Leu Ala Glu Glu Ile Val Ala Arg Phe His Gly Glu Glu Ala
290      295      300
Ala Ala Ser Ala His Lys Ser Ala Gly Asn Arg Leu Lys Glu Gly Glu

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

```

<210> 59
 <211> 334
 <212> PRT
 <213> Escherichia coli

 <400> 59

Met	Thr	Lys	Pro	Ile	Val	Phe	Ser	Gly	Ala	Gln	Pro	Ser	Gly	Glu	Leu
1				5					10					15	
Thr	Ile	Gly	Asn	Tyr	Met	Gly	Ala	Leu	Arg	Gln	Trp	Val	Asn	Met	Gln
			20					25					30		
Asp	Asp	Tyr	His	Cys	Ile	Tyr	Cys	Ile	Val	Asp	Gln	His	Ala	Ile	Thr
		35					40					45			
Val	Arg	Gln	Asp	Ala	Gln	Lys	Leu	Arg	Lys	Ala	Thr	Leu	Asp	Thr	Leu
						55					60				
Ala	Leu	Tyr	Leu	Ala	Cys	Gly	Ile	Asp	Pro	Glu	Lys	Ser	Thr	Ile	Phe
65					70					75					80
Val	Gln	Ser	His	Val	Pro	Glu	His	Ala	Gln	Leu	Gly	Trp	Ala	Leu	Asn
			85						90					95	
Cys	Tyr	Thr	Tyr	Phe	Gly	Glu	Leu	Ser	Arg	Met	Thr	Gln	Phe	Lys	Asp
			100					105					110		
Lys	Ser	Ala	Arg	Tyr	Ala	Glu	Asn	Ile	Asn	Ala	Gly	Leu	Phe	Asp	Tyr
		115					120					125			
Pro	Val	Leu	Met	Ala	Ala	Asp	Ile	Leu	Leu	Tyr	Gln	Thr	Asn	Leu	Val
			130			135					140				
Pro	Val	Gly	Glu	Asp	Gln	Lys	Gln	His	Leu	Glu	Leu	Ser	Arg	Asp	Ile
145					150					155					160
Ala	Gln	Arg	Phe	Asn	Ala	Leu	Tyr	Gly	Glu	Ile	Phe	Lys	Val	Pro	Glu
			165						170					175	
Pro	Phe	Ile	Pro	Lys	Ser	Gly	Ala	Arg	Val	Met	Ser	Leu	Leu	Glu	Pro
			180					185					190		
Thr	Lys	Lys	Met	Ser	Lys	Ser	Asp	Asp	Asn	Arg	Asn	Asn	Val	Ile	Gly
		195					200					205			
Leu	Leu	Glu	Asp	Pro	Lys	Ser	Val	Val	Lys	Lys	Ile	Lys	Arg	Ala	Val
						215					220				
Thr	Asp	Ser	Asp	Glu	Pro	Pro	Val	Val	Arg	Tyr	Asp	Val	Gln	Asn	Lys
225					230					235					240
Ala	Gly	Val	Ser	Asn	Leu	Leu	Asp	Ile	Leu	Ser	Ala	Val	Thr	Gly	Gln
				245					250					255	
Ser	Ile	Pro	Glu	Leu	Glu	Lys	Gln	Phe	Glu	Gly	Lys	Met	Tyr	Gly	His
			260					265					270		
Leu	Lys	Gly	Glu	Val	Ala	Asp	Ala	Val	Ser	Gly	Met	Leu	Thr	Glu	Leu
			275				280					285			
Gln	Glu	Arg	Tyr	His	Arg	Phe	Arg	Asn	Asp	Glu	Ala	Phe	Leu	Gln	Gln
			290			295				300					
Val	Met	Lys	Asp	Gly	Ala	Glu	Lys	Ala	Ser	Ala	His	Ala	Ser	Arg	Thr
305						310				315					320
Leu	Lys	Ala	Val	Tyr	Glu	Ala	Ile	Gly	Phe	Val	Ala	Lys	Pro		
						325				330					

<210> 60
<211> 326
<212> PRT
<213> Helicobacter pylori

<400> 60

Met	His	Lys	Lys	Arg	Val	Phe	Ser	Gly	Ile	Gln	Pro	Thr	Gly	Gln	Ile
1				5					10					15	
His	Leu	Gly	Asn	Tyr	Leu	Gly	Ala	Ile	Lys	His	Trp	Val	Glu	Met	Gln
			20					25					30		
Asp	Glu	Tyr	Glu	Asn	Leu	Phe	Cys	Val	Val	Asn	Ser	His	Ala	Ile	Thr
		35					40					45			
Leu	Pro	Ile	Asp	Pro	Ala	Phe	Leu	Lys	Ser	Gln	Ser	Tyr	Glu	Leu	Val
	50					55					60				
Lys	Leu	Leu	Leu	Ala	Cys	Gly	Ile	Asp	Pro	Lys	Gln	Ser	Gly	Leu	Phe
65					70					75					80
Ile	Gln	Ser	Glu	Val	Asp	Glu	His	Pro	Ala	Leu	Ala	Trp	Leu	Leu	Asn
				85					90					95	
Cys	Gln	Val	Ser	Met	Gly	Glu	Met	Gln	Arg	Met	Thr	Gln	Phe	Lys	Asp
			100					105					110		
Lys	Ser	Leu	Lys	Asn	Pro	Lys	Ser	Val	Asn	Val	Gly	Leu	Phe	Asn	Tyr
		115					120					125			
Pro	Ile	Leu	Met	Ala	Ser	Asp	Ile	Leu	Leu	Tyr	Gln	Ser	Asp	Leu	Val
	130					135					140				
Pro	Val	Gly	Glu	Asp	Gln	Lys	Gln	His	Leu	Glu	Leu	Thr	Arg	Asn	Ile
145					150					155					160
Ala	Glu	Lys	Phe	Asn	Arg	Asp	Phe	Gly	Asn	Cys	Phe	Lys	Val	Pro	Glu
				165					170					175	
Pro	Leu	Ile	Ala	Lys	Val	Gly	Ala	Arg	Val	Met	Gly	Leu	Asp	Asp	Pro
			180					185					190		
Lys	Val	Lys	Met	Ser	Lys	Ser	His	Gln	Gly	Ala	Asn	His	Ala	Ile	Phe
	195						200					205			
Leu	Leu	Asp	Glu	Pro	Asp	Ile	Ile	Val	Lys	Lys	Ile	Lys	Lys	Ala	Ala
	210					215					220				
Thr	Asp	Ser	Met	Gly	Val	Ile	Ala	Phe	Asp	Glu	Lys	Arg	Glu	Gly	Ile
225					230					235					240
Phe	Asn	Leu	Leu	Asn	Ile	Tyr	Met	Leu	Leu	Ser	Asn	Glu	Ser	Pro	Glu
				245					250					255	
Asn	Ile	Glu	Glu	Arg	Phe	Lys	Asn	Lys	Gly	Tyr	Gly	Asp	Phe	Lys	Lys
			260					265					270		
Glu	Leu	Ala	Glu	Val	Met	Ile	Gln	Ala	Leu	Lys	Pro	Ile	Gln	Glu	Arg
		275					280					285			
Tyr	Lys	Glu	Ile	Ser	Asp	Asp	Glu	Val	Lys	Ala	Ile	Leu	Asn	Gly	Gly
	290					295					300				
Ala	Glu	Lys	Ala	Arg	Pro	Leu	Ala	Arg	Ala	Thr	Tyr	Gln	Lys	Ala	Lys
305					310					315					320
Glu	Leu	Met	Gly	Leu	Ile										
					325										

<210> 61
 <211> 361
 <212> PRT
 <213> Schizosaccharomyces pombe

<400> 61

```

Met Ala Lys Leu Pro Lys Ile Thr Ser Leu Leu Pro His Ser Arg Val
 1          5          10          15
Val Ser Gly Ile Gln Pro Thr Gly Ile Pro His Ile Gly Asn Tyr Leu
 20          25          30
Gly Ser Leu Lys Gln Trp Val Gln Leu Gln Glu Glu Ala Ala Arg Thr
 35          40          45
Pro Phe Ser Lys Cys Phe Phe Phe Val Ala Asp Leu His Ala Leu Thr
 50          55          60
Val Pro Gln Asp Pro Leu Lys Phe Arg Gln Ala Arg Leu Asp Met Leu
 65          70          75          80
Ala Ala Leu Leu Ala Ile Gly Ile Asn Pro Gln Lys Ser Thr Leu Phe
 85          90          95
Phe Gln Ser Asp Val Ala Gln His Ser Glu Leu Ala Trp Leu Ala
100          105          110
Cys Ser Thr Ser Met Gly Gln Leu Asn Arg Met Thr Gln Trp Lys Ser
115          120          125
Lys Leu His Leu His Asp His Asp Asp Leu Ser Phe Leu Asp Ala Ser
130          135          140
Ala Thr Ser Ser Thr Arg Phe Asn Leu Gly Leu Phe Ser Tyr Pro Val
145          150          155          160
Leu Gln Ala Ala Asp Ile Leu Leu Tyr Gly Ala Thr His Ile Pro Val
165          170          175
Gly Lys Asp Gln Ser Gln His Val Glu Leu Thr Arg Ser Ile Ala Arg
180          185          190
Ser Phe Asn Ser Ser Tyr Lys Glu Lys Ile Leu Thr Val Pro Asp Ile
195          200          205
Ile Leu Asn Ser Ser Ser Ser Ile Met Ala Leu Cys Gln Pro Glu Lys
210          215          220
Lys Met Ser Lys Ser Asp Ile Asn Ser Lys Asn Tyr Ile Leu Leu Ser
225          230          235          240
Asp Ser Thr Gly Glu Ile Arg Lys Lys Ile Ser Arg Ala Gln Thr Asp
245          250          255
Asn Ile Lys Gly Ile Thr Tyr Gly Asp Ser Asn Arg Pro Gly Ile Asn
260          265          270
Asn Leu Ile Asn Ile Phe Ala Ala Ile Ser Asp Ser Thr Pro Ser Asp
275          280          285
Ile Ala Gln Ala Asn Ala Ser Cys Ser Asn Ala Glu Phe Lys Glu Lys
290          295          300
Val Ser Ser Ala Ile Ile Arg Cys Leu Gln Pro Ile Ser Thr Ser Phe
305          310          315          320
Asn Glu Trp Arg Gln Asn Arg Glu Leu Leu Arg Asp Ile Ala Lys Lys
325          330          335
Gly Ala Glu Glu Ala Val Ala Glu Ala Ser Ser Cys Met His Lys Leu
340          345          350
Lys Thr Leu Thr Gly Leu Ser Val Tyr
355          360

```

<210> 62

<211> 665

<212> PRT

<213> Drosophila melanogaster

<400> 62

Met Phe Arg Phe Gly Lys Leu His Arg Ser Leu Ser Lys Val Ala Lys

1					5					10					15
Lys	Gly	Gln	Thr	Ala	Val	Tyr	Ser	Gln	Arg	Arg	Cys	Ile	Gln	Ala	Arg
			20					25					30		
Ser	Phe	Leu	Pro	Gly	Leu	Ala	Ser	Arg	Thr	Ala	Gly	Thr	Pro	Ser	Ala
		35					40					45			
Thr	Ala	Gln	Val	Thr	Gly	Gly	Asn	Lys	Gln	His	Val	His	Pro	His	Gly
		50				55					60				
Asn	Ser	Val	Asn	Ser	Gly	His	Glu	Glu	His	Asn	Thr	Arg	Trp	Pro	Arg
65					70					75					80
Lys	Val	Phe	Ser	Gly	Ile	Gln	Pro	Thr	Gly	Ser	Leu	His	Leu	Gly	Asn
				85					90					95	
Tyr	Leu	Gly	Ala	Val	Arg	Lys	Trp	Val	Gln	Leu	Gln	Asn	Ala	Arg	Asp
			100					105					110		
Asp	Val	Thr	Val	Cys	Ile	Val	Asp	Leu	His	Ser	Ile	Thr	Met	Pro	His
		115					120					125			
Asn	Pro	Pro	Leu	Leu	Arg	Glu	Asn	Ile	Phe	Thr	Met	Ala	Ala	Thr	Leu
		130				135					140				
Leu	Ala	Cys	Gly	Ile	Asp	Pro	Thr	Lys	Ser	Thr	Leu	Phe	Val	Gln	Ser
145					150					155					160
Ala	Val	Ala	Glu	His	Ala	Glu	Phe	Asn	Trp	Ile	Leu	Ser	Ser	Leu	Thr
				165					170					175	
Thr	Met	Pro	Arg	Leu	Ala	Gln	Leu	Pro	Gln	Phe	Arg	Glu	Lys	Ser	Arg
			180					185					190		
Leu	Leu	Lys	Asp	Val	Pro	Leu	Gly	Leu	Tyr	Val	Tyr	Pro	Val	Leu	Gln
		195					200					205			
Ala	Ala	Asp	Ile	Met	Leu	Tyr	Lys	Ser	Thr	His	Val	Pro	Val	Gly	Ala
		210				215					220				
Asp	Gln	Ile	Gln	His	Ile	Gln	Leu	Ala	Gln	His	Leu	Ala	Arg	Ile	Tyr
225					230					235					240
Asn	Gly	Arg	Tyr	Gly	Glu	Thr	Phe	Pro	Val	Cys	Thr	Ala	Ile	Ile	Glu
				245					250					255	
Asp	Gly	Asp	Ala	Ser	Arg	Val	Leu	Ser	Leu	Arg	Asp	Pro	Ser	Lys	Lys
			260					265					270		
Met	Ser	Lys	Ser	Glu	Ala	Asn	Pro	Lys	Ala	Thr	Ile	Asn	Leu	Cys	Asp
		275					280					285			
Ser	Pro	Asp	Leu	Ile	Thr	Gln	Lys	Ile	Lys	Lys	Ala	Val	Thr	Asp	Phe
		290				295					300				
Thr	Ser	Asp	Ile	Thr	Tyr	Asn	Pro	Gly	Lys	Arg	Ala	Gly	Val	Ser	Asn
305					310					315					320
Leu	Val	Asn	Ile	His	Ala	Gln	Val	Thr	Gly	Gln	Ser	Ile	Lys	Thr	Val
				325					330					335	
Val	Asn	Glu	Ala	Ala	Thr	Leu	Asp	Thr	Ala	Lys	Tyr	Lys	Asp	Arg	Val
			340					345					350		
Ala	Glu	Ala	Val	Val	Glu	His	Leu	Arg	Pro	Ile	Arg	Glu	Gln	Ile	His
		355					360					365			

```

465          470          475          480
Ile Arg Arg Ala Pro Ile Val Pro Thr Gln Ser Met Arg Val Glu Lys
          485          490          495
Gln Leu Gly Ser Thr Arg Thr Arg His Val Phe Gln Met Asp Thr Cys
          500          505          510
Asn Ala Pro His Ile Gly Phe Lys Pro Pro Pro Gly Phe Ser Ala Ser
          515          520          525
Met Val Thr Gly Ala Leu Ala Lys Asn Lys Arg Ala Thr Val Lys Pro
          530          535          540
Val Thr Gly Asn Leu Gly Phe Gly Gln Gln Ser Gln Gly Gly Ser Tyr
545          550          555          560
Lys His Ser Gln Asn Val Asn Pro Ser Ala Val Ser Thr Ile Ala Arg
          565          570          575
Glu Val Asn Ala Ala Lys Lys Gln Glu Phe Thr Cys Asn Pro Ala Leu
          580          585          590
Tyr Gly Ser Ala Leu Glu Ser Leu Ser Asn Ser Ser Arg Ser Ser Glu
          595          600          605
Lys Thr Asn Ser Thr Ala Gly Asp Ser Glu Phe Ile Ile Val Ala Ser
          610          615          620
Glu Glu Glu Glu Gln Ala Ala Ser Ser Glu Glu Gln Arg Glu Glu Ala
625          630          635          640
Glu Glu Glu Glu Ser Gly Glu Arg Glu Arg Glu Lys Gly Glu Arg Asn
          645          650          655
Ala Thr Glu Ala Thr Lys Val Glu Ile
          660          665

```

<210> 63
 <211> 360
 <212> PRT
 <213> Homo sapiens

<400> 63

```

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

```

```

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

```

<210> 64

<211> 379

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 64

```

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

```

225		230		235		240									
Pro	Glu	Lys	Lys	Met	Ser	Lys	Ser	Asp	Pro	Asn	His	Asp	Ser	Val	Ile
		245		250		255									
Phe	Leu	Asn	Asp	Glu	Pro	Lys	Ala	Ile	Gln	Lys	Lys	Ile	Arg	Lys	Ala
		260		265		270									
Leu	Thr	Asp	Ser	Ile	Ser	Asp	Arg	Phe	Tyr	Tyr	Asp	Pro	Val	Glu	Arg
		275		280		285									
Pro	Gly	Val	Ser	Asn	Leu	Ile	Asn	Ile	Val	Ser	Gly	Ile	Gln	Arg	Lys
		290		295		300									
Ser	Ile	Glu	Asp	Val	Val	Glu	Asp	Val	Ser	Arg	Phe	Asn	Asn	Tyr	Arg
305				310		315									320
Asp	Phe	Lys	Asp	Tyr	Val	Ser	Glu	Val	Ile	Ile	Glu	Glu	Leu	Lys	Gly
				325		330									335
Pro	Arg	Thr	Glu	Phe	Glu	Lys	Tyr	Ile	Asn	Glu	Pro	Thr	Tyr	Leu	His
				340		345									350
Ser	Val	Val	Glu	Ser	Gly	Met	Arg	Lys	Ala	Arg	Glu	Lys	Ala	Ala	Lys
				355		360									365
Asn	Leu	Ala	Asp	Ile	His	Lys	Ile	Met	Gly	Phe					
				370		375									

<210> 65

<211> 475

<212> PRT

<213> Xenopus laevis

<400> 65

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

```

                245                250                255
Val Lys Val Gln Lys His Val Thr Phe Asn Gln Val Lys Gly Ile Phe
                260                265                270
Gly Phe Thr Asp Ser Asp Cys Ile Gly Lys Ile Ser Phe Pro Ala Ile
                275                280                285
Gln Ala Ala Pro Ser Phe Ser Ser Ser Phe Pro Glu Ile Phe Lys Gly
                290                295                300
Arg Lys Asp Ile Gln Cys Leu Ile Pro Cys Ala Ile Asp Gln Asp Pro
305                310                315                320
Tyr Phe Arg Met Thr Arg Asp Val Ala Pro Arg Ile Asn Tyr Pro Lys
                325                330                335
Pro Ala Leu Met His Ser Thr Phe Phe Pro Ala Leu Gln Gly Ala Gln
                340                345                350
Thr Lys Met Ser Ala Ser Asp Pro Asn Ser Ser Ile Phe Leu Thr Asp
                355                360                365
Thr Ala Lys Gln Ile Lys Ser Lys Ile Asn Lys His Ala Phe Ser Gly
                370                375                380
Gly Lys Asp Thr Ile Glu Glu His Arg Gln Phe Gly Gly Asn Trp Glu
385                390                395                400
Val Asp Val Ser Tyr Met Tyr Leu Thr Phe Phe Leu Glu Asp Asp Glu
                405                410                415
Arg Leu Glu Gln Ile Lys Gln Asp Tyr Ser Ser Gly Ala Leu Leu Thr
                420                425                430
Gly Asp Leu Lys Lys Ile Leu Thr Glu Thr Leu Gln Pro Met Ile Ser
                435                440                445
Ala His Gln Glu Arg Arg Lys His Ile Thr Glu Glu Thr Val Lys Gln
                450                455                460
Phe Met Met Pro Arg Lys Leu Ala Phe Asp Phe
465                470                475

```

<210> 66

<211> 417

<212> PRT

<213> Caenorhabditis elegans

<400> 66

```

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

```

```

      165      170      175
Asp Pro Thr Lys Thr Phe Ile Phe Asn Asn Phe Asp Tyr Met Cys Pro
      180      185      190
Pro Phe Tyr Glu Asn Ile Val Lys Ile Trp Lys Val Val Asn Thr Asn
      195      200      205
Gln Ala Arg Ala Ile Phe Gly Phe Thr Pro Glu Asp Cys Leu Gly Lys
      210      215      220
Ala Ala Phe Pro Ala Val Glu Ala Ala Pro Cys Phe Ala Ser Ser Phe
      225      230      235
Pro Gln Ile Phe Gly Lys Arg Asn Asp Ile Pro Cys Leu Ile Pro Cys
      245      250      255
Ala Ile Asp Gln Asp Pro Phe Phe Arg Met Thr Arg Asp Val Ala Pro
      260      265      270
Arg Leu Lys Ala Ser Lys Pro Ser Leu Ile Phe Ser Thr Phe Leu Pro
      275      280      285
Ala Leu Thr Gly Ala Gln Thr Lys Met Ser Ala Ser Glu Pro Asn Thr
      290      295      300
Cys Ile Phe Leu Ser Asp Thr Ala Lys Gln Ile Lys Asn Lys Ile Asn
      305      310      315
Lys Tyr Ala Phe Ser Gly Gly Gln Gln Thr Val Gln Glu His Arg Glu
      325      330      335
Lys Gly Gly Asn Cys Asp Val Asp Ile Ser Tyr Gln Phe Leu Arg Phe
      340      345      350
Phe Leu Asp Asp Asp Glu Lys Leu Ala Glu Ile Arg Glu Asn Tyr Thr
      355      360      365
Lys Gly Glu Met Leu Ser Gly Glu Leu Lys Ala Leu Ala Thr Gln Lys
      370      375      380
Val Gln Glu Ile Val Leu Glu Met Gln Glu Arg Arg Lys Leu Val Thr
      385      390      395
Asp Glu Thr Val Glu Glu Phe Val Lys Val Arg Pro Leu Ala Tyr Lys
      405      410      415
Tyr

```

<210> 67

<211> 337

<212> PRT

<213> *Salmonella enterica* subsp. *enterica* serovar Typhi8

<400> 67

```

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

```

```

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

```

<210> 68
 <211> 448
 <212> PRT
 <213> Pseudomonas aeruginosa

<400> 68

```

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

```

```

      180      185      190
Asn Val Ala Thr Leu Pro Gly Leu Asp Gly Arg Lys Met Ser Lys Ser
      195      200      205
Tyr Asp Asn Thr Ile Pro Leu Phe Ser Pro Ser Arg Gln Leu Lys Asp
      210      215      220
Ala Ile Ala Arg Ile Val Thr Asp Ser Arg Ala Pro Gly Glu Pro Lys
      225      230      235      240
Asp Pro Asp Ser Ser His Leu Phe Leu Leu Tyr Ser Ala Phe Ala Ser
      245      250      255
Ala Glu Gln Val Ala Ala Phe Arg Gln Glu Leu Leu Glu Gly Leu Ala
      260      265      270
Trp Gly Glu Ala Lys Gln Arg Leu Phe Gln Leu Leu Asp Asn Glu Leu
      275      280      285
Gly Glu Ala Arg Glu Arg Tyr Gln Ala Leu Ile Ala Lys Pro Asp Asp
      290      295      300
Ile Glu Asp Ile Leu Leu Ala Gly Ala Ala Lys Ala Arg Arg Ile Ala
      305      310      315      320
Thr Pro Phe Ile Ala Glu Leu Arg Glu Ala Val Gly Leu Arg Ser Leu
      325      330      335
Arg Glu Pro Leu Lys Ser Ala Glu Ser Gly Lys Lys Lys Ala Ala Lys
      340      345      350
Ala Ala Arg Leu Val Ser Phe Arg Asp Asp Asp Gly Ser Phe Arg Phe
      355      360      365
Arg Leu Leu Asp Ala Ala Gly Glu Gln Leu Leu Leu Ser Arg Ala Phe
      370      375      380
Ala Asp Gly Lys Ala Ala Gly Ala Val Ser Lys Arg Leu Leu Ala Gly
      385      390      395      400
Glu Thr Ala Asp Leu Arg Ala Glu Gly Asn Ala Phe Gly Leu Trp Leu
      405      410      415
Asp Gly Glu Ala Val Ala Gln Ser Pro Ala Phe Ala Asp Ala Ala Ala
      420      425      430
Arg Asp Ala Ala Ile Glu Arg Thr Arg Glu Ala Leu Ala Pro Gln Glu
      435      440      445

```

<210> 69

<211> 350

<212> PRT

<213> Thermus thermophilus

<400> 69

```

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

```

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

<210> 70

<211> 350

<212> PRT

<213> Thermus thermophilus

<400> 70

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

Pro	Phe	Arg	Ile	Val	Val	Pro	Gly	Arg	Val	Phe	Arg	Phe	Glu	Gln	Thr
		195					200						205		
Asp	Ala	Thr	His	Glu	Ala	Val	Phe	His	Gln	Leu	Glu	Gly	Leu	Val	Val
		210					215					220			
Gly	Glu	Gly	Ile	Ala	Met	Ala	His	Leu	Lys	Gly	Ala	Ile	Tyr	Glu	Leu
225						230				235					240
Ala	Gln	Ala	Leu	Phe	Gly	Pro	Asp	Ser	Lys	Val	Arg	Phe	Gln	Pro	Val
				245					250					255	
Tyr	Phe	Pro	Phe	Val	Glu	Pro	Gly	Ala	Gln	Phe	Ala	Val	Trp	Trp	Pro
			260					265					270		
Glu	Gly	Gly	Lys	Trp	Leu	Glu	Leu	Gly	Gly	Ala	Gly	Met	Val	His	Pro
		275					280					285			
Lys	Val	Phe	Gln	Ala	Val	Asp	Ala	Tyr	Arg	Glu	Arg	Leu	Gly	Leu	Pro
		290				295					300				
Pro	Ala	Tyr	Arg	Gly	Val	Thr	Gly	Phe	Ala	Phe	Gly	Leu	Gly	Val	Glu
305					310					315					320
Arg	Leu	Ala	Met	Leu	Arg	Tyr	Gly	Ile	Pro	Asp	Ile	Arg	Tyr	Phe	Phe
				325				330						335	
Gly	Gly	Arg	Leu	Lys	Phe	Leu	Glu	Gln	Phe	Lys	Gly	Val	Leu		
			340					345					350		

<210> 71

<211> 344

<212> PRT

<213> Bacillus subtilis

<400> 71

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

```

225      230      235      240
Gln Asp Arg Glu Ile Arg Leu Ala Pro Ser Phe Phe Pro Phe Thr Glu
      245      250      255
Pro Ser Val Glu Val Asp Val Thr Cys Phe Lys Cys Gly Gly Asn Gly
      260      265      270
Cys Ser Val Cys Lys Gly Thr Gly Trp Ile Glu Ile Leu Gly Ala Gly
      275      280      285
Met Val His Pro Asn Val Leu Lys Met Ala Gly Phe Asp Pro Lys Glu
      290      295      300
Tyr Gln Gly Phe Ala Phe Gly Met Gly Val Glu Arg Ile Ala Met Leu
305      310      315      320
Lys Tyr Gly Ile Asp Asp Ile Arg His Phe Tyr Thr Asn Asp Val Arg
      325      330      335
Phe Ile Ser Gln Phe Lys Gln Ala
      340

```

<210> 72
 <211> 328
 <212> PRT
 <213> Helicobacter pylori

<400> 72

```

Met His Thr Leu Ile Glu Arg Leu Glu Lys Val Thr Asn Ser Lys Glu
 1      5      10      15
Leu Glu Glu Ala Arg Leu Asn Ala Leu Gly Lys Lys Gly Val Phe Ala
      20      25      30
Asp Lys Phe Asn Gln Leu Lys His Leu Asn Gly Glu Glu Lys Asn Ala
      35      40      45
Phe Ala Lys Glu Ile His His Tyr Lys Gln Ala Phe Glu Lys Ala Phe
      50      55      60
Glu Trp Lys Lys Lys Ala Ile Ile Glu Leu Glu Leu Glu Arg Leu
      65      70      75      80
Lys Lys Glu Lys Ile Asp Val Ser Leu Phe Asn Ala Ile Lys Thr Ser
      85      90      95
Ser Ser His Pro Leu Asn Tyr Thr Lys Asn Lys Ile Ile Glu Phe Phe
      100      105      110
Thr Pro Leu Gly Tyr Lys Leu Glu Ile Gly Ser Leu Val Glu Asp Asp
      115      120      125
Phe His Asn Phe Ser Ala Leu Asn Leu Pro Pro Tyr His Pro Ala Arg
      130      135      140
Asp Met Gln Asp Thr Phe Tyr Phe Lys Asp His Lys Leu Leu Arg Thr
      145      150      155      160
His Thr Ser Pro Val Gln Ile His Thr Met Gln Glu Gln Thr Pro Pro
      165      170      175
Ile Lys Met Ile Cys Leu Gly Glu Thr Phe Arg Arg Asp Tyr Asp Leu
      180      185      190
Thr His Thr Pro Met Phe His Gln Ile Glu Gly Leu Val Val Asp Gln
      195      200      205
Lys Gly Asn Ile Arg Phe Thr His Leu Lys Gly Val Ile Glu Asp Phe
      210      215      220
Leu His Tyr Phe Phe Gly Gly Val Lys Leu Arg Trp Arg Ser Ser Phe
      225      230      235      240
Phe Pro Phe Thr Glu Pro Ser Ala Glu Val Asp Ile Ser Cys Val Phe
      245      250      255
Cys Lys Gln Glu Gly Cys Arg Val Cys Ser His Thr Gly Trp Leu Glu
      260      265      270
Val Leu Gly Cys Gly Met Val Asn Asn Ala Val Phe Glu Ala Ile Gly

```

```

      275      280      285
Tyr Glu Asn Val Ser Gly Phe Ala Phe Gly Met Gly Ile Glu Arg Leu
    290      295      300
Ala Met Leu Thr Cys Gln Ile Asn Asp Leu Arg Ser Phe Phe Glu Thr
    305      310      315      320
Asp Leu Arg Val Leu Glu Ser Phe
          325

```

<210> 73
 <211> 327
 <212> PRT
 <213> Escherichia coli

<400> 73

```

Met Ser His Leu Ala Glu Leu Val Ala Ser Ala Lys Ala Ala Ile Ser
  1      5      10      15
Gln Ala Ser Asp Val Ala Ala Leu Asp Asn Val Arg Val Glu Tyr Leu
      20      25      30
Gly Lys Lys Gly His Leu Thr Leu Gln Met Thr Thr Leu Arg Glu Leu
      35      40      45
Pro Pro Glu Glu Arg Pro Ala Ala Gly Ala Val Ile Asn Glu Ala Lys
      50      55      60
Glu Gln Val Gln Gln Ala Leu Asn Ala Arg Lys Ala Glu Leu Glu Ser
      65      70      75      80
Ala Ala Leu Asn Ala Arg Leu Ala Ala Glu Thr Ile Asp Val Ser Leu
      85      90      95
Pro Gly Arg Arg Ile Glu Asn Gly Gly Leu His Pro Val Thr Arg Thr
      100      105      110
Ile Asp Arg Ile Glu Ser Phe Phe Gly Glu Leu Gly Phe Thr Val Ala
      115      120      125
Thr Gly Pro Glu Ile Glu Asp Tyr His Asn Phe Asp Ala Leu Asn
      130      135      140
Ile Pro Gly His His Pro Ala Arg Ala Asp His Asp Thr Phe Trp Phe
      145      150      155      160
Asp Thr Thr Arg Leu Leu Arg Thr Gln Thr Ser Gly Val Gln Ile Arg
      165      170      175
Thr Met Lys Ala Gln Gln Pro Pro Ile Arg Ile Ile Ala Pro Gly Arg
      180      185      190
Val Tyr Arg Asn Asp Tyr Asp Gln Thr His Thr Pro Met Phe His Gln
      195      200      205
Met Glu Gly Leu Ile Val Asp Thr Asn Ile Ser Phe Thr Asn Leu Lys
      210      215      220
Gly Thr Leu His Asp Phe Leu Arg Asn Phe Phe Glu Glu Asp Leu Gln
      225      230      235      240
Ile Arg Phe Arg Pro Ser Tyr Phe Pro Phe Thr Glu Pro Ser Ala Glu
      245      250      255
Val Asp Val Met Gly Lys Asn Gly Lys Trp Leu Glu Val Leu Gly Cys
      260      265      270
Gly Met Val His Pro Asn Val Leu Arg Asn Val Gly Ile Asp Pro Glu
      275      280      285
Val Tyr Ser Gly Phe Ala Phe Gly Met Gly Met Glu Arg Leu Thr Met
      290      295      300
Leu Arg Tyr Gly Val Thr Asp Leu Arg Ser Phe Phe Glu Asn Asp Leu
      305      310      315      320
Arg Phe Leu Lys Gln Phe Lys
          325

```

<210> 74
 <211> 453
 <212> PRT
 <213> Drosophila melanogaster

<400> 74

```

Met Leu Leu Thr Leu Arg Val Gln Gly Ala Arg His Trp Leu Lys Ser
 1      5      10      15
Thr Arg Cys Leu Ala Ser Ser Ala Ala Pro Ala Lys Ser Pro Ser Ser
 20      25      30
Pro Pro Gln Leu Glu Val Ser Gly Ser Thr Tyr Ala Thr Asp Gly Trp
 35      40      45
Thr Asn Val Thr Pro Lys Ile Leu Ser Tyr Val Gly Ala Asn Lys His
 50      55      60
Leu Gln Thr Asp His Pro Leu Ser Ile Ile Arg Gln Arg Ile Val Asn
 65      70      75      80
Tyr Phe Tyr Gly Ala Tyr Arg Asn Gln Arg Gly Asn Pro Leu Phe Ser
 85      90      95
Val Tyr Asp Gln Met Asn Pro Val Val Thr Val Gln Gln Asn Phe Asp
100      105      110
Asn Leu Leu Ile Pro Ala Asp His Val Ser Arg Gln Lys Ser Asp Cys
115      120      125
Tyr Tyr Ile Asn Gln Gln His Leu Leu Arg Ala His Thr Thr Ala His
130      135      140
Gln Val Glu Leu Ile Ser Gly Gly Leu Asp Asn Phe Leu Val Val Gly
145      150      155      160
Glu Val Tyr Arg Arg Asp Glu Ile Asp Ser Thr His Tyr Pro Val Phe
165      170      175
His Gln Ala Asp Ala Val Arg Leu Val Thr Lys Asp Lys Leu Phe Glu
180      185      190
Arg Asn Pro Gly Leu Glu Leu Phe Glu Glu Thr Trp Ser Gly Thr Leu
195      200      205
Ala Asp Pro Lys Leu Ile Leu Pro Ser Ser Lys Phe Met Asp Gln Thr
210      215      220
Lys Gln Pro Cys His Thr Leu Glu Ala Val Lys Leu Met Glu His Glu
225      230      235      240
Met Lys His Val Leu Val Gly Leu Thr Lys Asp Leu Phe Gly Pro Arg
245      250      255
Ile Lys Tyr Arg Trp Val Asp Thr Tyr Phe Pro Phe Thr Gln Pro Ser
260      265      270
Trp Glu Leu Glu Ile Tyr Phe Lys Asp Asn Trp Leu Glu Val Leu Gly
275      280      285
Cys Gly Ile Met Arg His Glu Ile Leu Gln Arg Ser Gly Val His Gln
290      295      300
Ser Ile Gly Tyr Ala Phe Gly Val Gly Leu Glu Arg Leu Ala Met Val
305      310      315      320
Leu Phe Asp Ile Pro Asp Ile Arg Leu Phe Trp Ser Asn Asp Ser Gly
325      330      335
Phe Leu Ser Gln Phe Ser Glu Lys Asp Leu His Asn Leu Pro Lys Tyr
340      345      350
Lys Pro Ile Ser His Tyr Pro Gln Cys Thr Asn Asp Leu Ser Phe Trp
355      360      365
Leu Pro Gln Asp Ile Glu Val Asp Ala Gly Phe Ser Pro Asn Asp Phe
370      375      380
Tyr Asp Leu Val Arg Ser Val Ala Gly Asp Met Val Glu Gln Ile Ser
385      390      395      400
Leu Val Asp Lys Phe Lys His Pro Lys Thr Gly Lys Ser Ser Val Cys

```

```

          405          410          415
Phe Arg Ile Val Tyr Arg His Met Glu Arg Thr Leu Thr Gln Ala Glu
          420          425          430
Val Asn Glu Ile His Lys Gln Ile Ala Ser Ala Ser Val Asp Ser Phe
          435          440          445
Asn Val Gln Ile Arg
          450

```

<210> 75

<211> 339

<212> PRT

<213> Chlamydomophila pneumoniae

<400> 75

```

Met Glu Met Lys Glu Glu Ile Glu Ala Val Lys Gln Gln Phe His Ser
 1          5          10          15
Glu Leu Asp Gln Val Asn Ser Ser Gln Ala Leu Ala Asp Leu Lys Val
 20          25          30
Arg Tyr Leu Gly Lys Lys Gly Ile Phe Arg Ser Phe Ser Glu Lys Leu
 35          40          45
Lys Gln Cys Thr Asp Lys Ala Lys Leu Gly Ser Leu Ile Asn Asp Phe
 50          55          60
Lys Thr Tyr Val Glu Asp Leu Leu Gln Glu Lys Ser Leu Val Leu Leu
 65          70          75          80
Ala Ser Glu Gln Ala Glu Ala Phe Ser Lys Glu Lys Ile Asp Ser Ser
 85          90          95
Leu Pro Gly Asp Ser Gln Pro Ser Gly Gly Arg His Ile Leu Lys Ser
100          105          110
Ile Leu Asp Asp Val Val Asp Ile Phe Val His Leu Gly Phe Cys Val
115          120          125
Arg Glu Ala Pro Asn Ile Glu Ser Glu Ala Asn Asn Phe Thr Leu Leu
130          135          140
Asn Phe Thr Glu Asp His Pro Ala Arg Gln Met His Asp Thr Phe Tyr
145          150          155          160
Leu Asn Ala Thr Thr Val Leu Arg Thr His Thr Ser Asn Val Gln Ala
165          170          175
Arg Glu Leu Lys Lys Gln Gln Pro Pro Ile Lys Val Val Ala Pro Gly
180          185          190
Leu Cys Phe Arg Asn Glu Asp Ile Ser Ala Arg Ser His Val Leu Phe
195          200          205
His Gln Val Glu Ala Phe Tyr Val Asp His Asn Val Thr Phe Ser Asp
210          215          220
Leu Thr Ala Ile Leu Ser Ala Phe Tyr His Ser Phe Phe Gln Arg Lys
225          230          235          240
Thr Glu Leu Arg Phe Arg His Ser Tyr Phe Pro Phe Val Glu Pro Gly
245          250          255
Ile Glu Val Asp Val Ser Cys Glu Cys Cys Gly Lys Gly Cys Ala Leu
260          265          270
Cys Lys His Thr Gly Trp Leu Glu Val Ala Gly Ala Gly Met Ile His
275          280          285
Pro Gln Val Leu Arg Asn Gly Asn Val Asp Pro Glu Ile Tyr Ser Gly
290          295          300
Tyr Ala Val Gly Met Gly Ile Glu Arg Leu Ala Met Leu Lys Tyr Gly
305          310          315          320
Val Ser Asp Ile Arg Leu Phe Ser Glu Asn Asp Leu Arg Phe Leu Gln
325          330          335
Gln Phe Ser

```

<210> 76

<211> 451

<212> PRT

<213> Homo sapiens

<400> 76

```

Met Val Gly Ser Ala Leu Arg Arg Gly Ala His Ala Tyr Val Tyr Leu
1      5      10      15
Val Ser Lys Ala Ser His Ile Ser Arg Gly His Gln His Gln Ala Trp
20      25      30
Gly Ser Arg Pro Pro Ala Ala Glu Cys Ala Thr Gln Arg Ala Pro Gly
35      40      45
Ser Val Val Glu Leu Leu Gly Lys Ser Tyr Pro Gln Asp Asp His Ser
50      55      60
Asn Leu Thr Arg Lys Val Leu Thr Arg Val Gly Arg Asn Leu His Asn
65      70      75      80
Gln Gln His His Pro Leu Trp Leu Ile Lys Glu Arg Val Lys Glu His
85      90      95
Phe Tyr Lys Gln Tyr Val Gly Arg Phe Gly Thr Pro Leu Phe Ser Val
100      105      110
Tyr Asp Asn Leu Ser Pro Val Val Thr Thr Trp Gln Asn Phe Asp Ser
115      120      125
Leu Leu Ile Pro Ala Asp His Pro Ser Arg Lys Lys Gly Asp Asn Tyr
130      135      140
Tyr Leu Asn Arg Thr His Met Leu Arg Ala His Thr Ser Ala His Gln
145      150      155      160
Trp Asp Leu Leu His Ala Gly Leu Asp Ala Phe Leu Val Val Gly Asp
165      170      175
Val Tyr Arg Arg Asp Gln Ile Asp Ser Gln His Tyr Pro Ile Phe His
180      185      190
Gln Leu Glu Ala Val Arg Leu Phe Ser Lys His Glu Leu Phe Ala Gly
195      200      205
Ile Lys Asp Gly Glu Ser Leu Gln Leu Phe Glu Gln Ser Ser Arg Ser
210      215      220
Ala His Lys Gln Glu Thr His Thr Met Glu Ala Val Lys Leu Val Glu
225      230      235      240
Phe Asp Leu Lys Gln Thr Leu Thr Arg Leu Met Ala His Leu Phe Gly
245      250      255
Asp Glu Leu Glu Ile Arg Trp Val Asp Cys Tyr Phe Pro Phe Thr His
260      265      270
Pro Ser Phe Glu Met Glu Ile Asn Phe His Gly Glu Trp Leu Glu Val
275      280      285
Leu Gly Cys Gly Val Met Glu Gln Gln Leu Val Asn Ser Ala Gly Ala
290      295      300
Gln Asp Arg Ile Gly Trp Ala Phe Gly Leu Gly Leu Glu Arg Leu Ala
305      310      315      320
Met Ile Leu Tyr Asp Ile Pro Asp Ile Arg Leu Phe Trp Cys Glu Asp
325      330      335
Glu Arg Phe Leu Lys Gln Phe Cys Val Ser Asn Ile Asn Gln Lys Val
340      345      350
Lys Phe Gln Pro Leu Ser Lys Tyr Pro Ala Val Ile Asn Asp Ile Ser
355      360      365
Phe Trp Leu Pro Ser Glu Asn Tyr Ala Glu Asn Asp Phe Tyr Asp Leu
370      375      380
Val Arg Thr Ile Gly Gly Asp Leu Val Glu Lys Val Asp Leu Ile Asp

385      390      395      400
Lys Phe Val His Pro Lys Thr His Lys Thr Ser His Cys Tyr Arg Ile
405      410      415
Thr Tyr Arg His Met Glu Arg Thr Leu Ser Gln Arg Glu Val Arg His
420      425      430
Ile His Gln Ala Leu Gln Glu Ala Ala Val Gln Leu Leu Gly Val Glu
435      440      445
Gly Arg Phe
450

```

<210> 77

<211> 341

<212> PRT

<213> Mycobacterium tuberculosis

<400> 77

```

Met Leu Ser Pro Glu Ala Leu Thr Thr Ala Val Asp Ala Ala Gln Gln
 1           5           10           15
Ala Ile Ala Leu Ala Asp Thr Leu Asp Val Leu Ala Arg Val Lys Thr
 20           25           30
Glu His Leu Gly Asp Arg Ser Pro Leu Ala Leu Ala Arg Gln Ala Leu
 35           40           45
Ala Val Leu Pro Lys Glu Gln Arg Ala Glu Ala Gly Lys Arg Val Asn
 50           55           60
Ala Ala Arg Asn Ala Ala Gln Arg Ser Tyr Asp Glu Arg Leu Ala Thr
 65           70           75           80
Leu Arg Ala Glu Arg Asp Ala Ala Val Leu Val Ala Glu Gly Ile Asp
 85           90           95
Val Thr Leu Pro Ser Thr Arg Val Pro Ala Gly Ala Arg His Pro Ile
100          105          110
Ile Met Leu Ala Glu His Val Ala Asp Thr Phe Ile Ala Met Gly Trp
115          120          125
Glu Leu Ala Glu Gly Pro Glu Val Glu Thr Glu Gln Phe Asn Phe Asp
130          135          140
Ala Leu Asn Phe Pro Ala Asp His Pro Ala Arg Gly Glu Gln Asp Thr
145          150          155          160
Phe Tyr Ile Ala Pro Glu Asp Ser Arg Gln Leu Leu Arg Thr His Thr
165          170          175
Ser Pro Val Gln Ile Arg Thr Leu Leu Ala Arg Glu Leu Pro Val Tyr
180          185          190
Ile Ile Ser Ile Gly Arg Thr Phe Arg Thr Asp Glu Leu Asp Ala Thr
195          200          205
His Thr Pro Ile Phe His Gln Val Glu Gly Leu Ala Val Asp Arg Gly
210          215          220
Leu Ser Met Ala His Leu Arg Gly Thr Leu Asp Ala Phe Ala Arg Ala
225          230          235          240
Glu Phe Gly Pro Ser Ala Arg Thr Arg Ile Arg Pro His Phe Phe Pro
245          250          255
Phe Thr Glu Pro Ser Ala Glu Val Asp Val Trp Phe Ala Asn Lys Ile
260          265          270
Gly Gly Ala Ala Trp Val Glu Trp Gly Gly Cys Gly Met Val His Pro
275          280          285
Asn Val Leu Arg Ala Thr Gly Ile Asp Pro Asp Leu Tyr Ser Gly Phe
290          295          300
Ala Phe Gly Met Gly Leu Glu Arg Thr Leu Gln Phe Arg Asn Gly Ile
305          310          315          320
Pro Asp Met Arg Asp Met Val Glu Gly Asp Val Arg Phe Ser Leu Pro

```

```

          325
Phe Gly Val Gly Ala
          340

```

330

335

<210> 78

<211> 458

<212> PRT

<213> *Caenorhabditis elegans*

<400> 78

Met	Arg	Phe	Leu	Ser	Thr	Tyr	Gln	Thr	Gly	Ala	Arg	Arg	Cys	Cys	Thr
1				5					10					15	
Met	Leu	Phe	Gln	Lys	Asn	Ala	Lys	Ser	Ile	Arg	Ala	Thr	Leu	Met	Ser
			20					25					30		
Thr	Thr	Ala	Ala	Glu	Glu	Ser	Arg	Lys	Ile	Val	Lys	Glu	Glu	Val	
		35					40				45				
Phe	Glu	Leu	Asp	Gly	Arg	Lys	Tyr	Thr	Pro	Asp	Ala	Leu	Tyr	Asn	Leu
	50					55				60					
Ser	Pro	Gly	Val	Arg	Arg	Leu	Leu	Asp	Arg	Arg	Ile	Leu	Gln	Glu	Ser
65					70					75				80	
Ser	Asn	Pro	Leu	Asn	Leu	Leu	Lys	Arg	Arg	Ile	Val	Asp	Tyr	Val	His
				85					90					95	
Gln	Thr	Tyr	Arg	Lys	Pro	Gly	Asn	Arg	Ser	Pro	Leu	Phe	Thr	Ile	Cys
			100					105					110		
Glu	Ser	Glu	Pro	Arg	Val	Val	Thr	Thr	Tyr	Gln	Asn	Phe	Asp	Ser	Leu
		115					120					125			
Leu	Thr	Pro	Glu	Asp	His	Val	Ser	Arg	Arg	Pro	Ser	Asp	Thr	Tyr	Tyr
		130				135					140				
Val	Asn	His	Glu	His	Cys	Leu	Arg	Ala	His	Thr	Ser	Ala	His	Gln	His
145					150					155				160	
Asn	Leu	Met	Gln	Ser	Gly	Leu	Asp	Ala	Phe	Leu	Val	Ile	Gly	Asp	Val
			165						170					175	
Tyr	Arg	Arg	Asp	Glu	Val	Asp	Arg	Thr	His	Tyr	Pro	Cys	Phe	His	Gln
			180					185					190		
Ile	Glu	Gly	Val	Arg	Leu	Tyr	Ser	Lys	Asp	Glu	Leu	Leu	Gly	Lys	Lys
		195					200					205			
Pro	Asp	Gly	Lys	Asn	Val	Ala	Glu	Leu	Phe	Ser	Thr	Ala	Ala	Ser	Ala
		210				215						220			
Ala	Thr	Glu	Arg	Ser	Pro	Glu	Lys	Gln	Glu	Lys	His	Thr	Leu	Asp	Ala
225					230					235				240	
Thr	Lys	Ala	Ala	Glu	Ile	Gln	Leu	Lys	Gln	Phe	Leu	Glu	Asn	Leu	Cys
			245						250				255		
Asp	Glu	Leu	Phe	Gly	Lys	Asp	Ala	Glu	Lys	Arg	Trp	Val	Asp	Ala	Tyr
		260						265					270		
Phe	Pro	Phe	Thr	His	Pro	Ser	Trp	Glu	Leu	Glu	Val	Phe	Tyr	Asn	Gly
		275					280					285			
Gln	Trp	Leu	Glu	Val	Leu	Gly	Cys	Gly	Ile	Met	Glu	Gln	Lys	Leu	Leu
		290				295					300				
Glu	Ser	Ala	Gly	Val	Thr	Asp	Lys	Ile	Gly	Trp	Ala	Phe	Gly	Ile	Gly
305					310					315				320	
Leu	Glu	Arg	Ile	Ala	Met	Val	Leu	Tyr	Gly	Ile	Pro	Asp	Ile	Arg	Leu
			325						330				335		
Phe	Trp	Ser	Lys	Asp	Thr	Gly	Phe								

<210> 79
<211> 469
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 79

```

Met Phe Leu Asn Arg Met Met Lys Thr Arg Thr Gly Leu Tyr Arg Leu
 1           5           10           15
Tyr Ser Thr Leu Lys Val Pro His Val Glu Ile Asn Gly Ile Lys Tyr
      20           25           30
Lys Thr Asp Pro Gln Thr Thr Asn Val Thr Asp Ser Ile Ile Lys Leu
      35           40           45
Thr Asp Arg Ser Leu His Leu Lys Glu Ser His Pro Val Gly Ile Leu
      50           55           60
Arg Asp Leu Ile Glu Lys Lys Leu Asn Ser Val Asp Asn Thr Phe Lys
65           70           75           80
Ile Phe Asn Asn Phe Lys Pro Val Val Thr Thr Met Glu Asn Phe Asp
      85           90           95
Ser Leu Gly Phe Pro Lys Asp His Pro Gly Arg Ser Lys Ser Asp Thr
      100          105          110
Tyr Tyr Ile Asn Glu Thr His Leu Leu Arg Thr His Thr Ser Ala His
      115          120          125
Glu Leu Glu Cys Phe Gln Lys Ile Arg Asn Asp Ser Asp Asn Ile Lys
      130          135          140
Ser Gly Phe Leu Ile Ser Ala Asp Val Tyr Arg Arg Asp Glu Ile Asp
145          150          155          160
Lys Thr His Tyr Pro Val Phe His Gln Met Glu Gly Ala Thr Ile Trp
      165          170          175
Lys Arg Thr Lys Ala Asp Val Gly Val Lys Glu Pro Met Tyr Ile Glu
      180          185          190
Lys Ile Arg Glu Asp Ile Arg Gln Val Glu Asn Leu Leu Asn Lys Glu
      195          200          205
Asn Val Lys Ile Thr Val Asp Asp Thr Ile Pro Leu Lys Glu Asn
      210          215          220
Asn Pro Lys Gln Glu Tyr Met Ser Asp Leu Glu Val Asp Leu Cys Ser
225          230          235          240
Gln His Leu Lys Arg Ser Ile Glu Leu Ile Val Ser Glu Val Phe Asn
      245          250          255
Lys Lys Ile Ser Ser Met Ile Lys Asn Lys Ala Asn Asn Thr Pro Lys
      260          265          270
Glu Leu Lys Val Arg Trp Ile Asn Ala Tyr Phe Pro Trp Thr Ala Pro
      275          280          285
Ser Trp Glu Ile Glu Val Trp Trp Gln Gly Glu Trp Leu Glu Leu Cys
      290          295          300
Gly Cys Gly Leu Ile Arg Gln Asp Val Leu Leu Arg Ala Gly Tyr Lys

305          310          315          320
Pro Ser Glu Thr Ile Gly Trp Ala Phe Gly Leu Gly Leu Asp Arg Ile
      325          330          335
Ala Met Leu Leu Phe Glu Ile Pro Asp Ile Arg Leu Leu Trp Ser Arg
      340          345          350
Asp Glu Arg Phe Ser Arg Gln Phe Ser Lys Gly Leu Ile Thr Ser Phe
      355          360          365
Lys Pro Tyr Ser Lys His Pro Gly Ser Phe Arg Asp Val Ala Phe Trp
      370          375          380
Leu Pro Glu Asp Lys Pro Asp Ile His Gln Val His Glu Asn Asp Leu
385          390          395          400
Met Glu Ile Ile Arg Asn Ile Ala Gly Asp Leu Val Glu Ser Val Lys
      405          410          415
Leu Val Asp Ser Phe Thr His Pro Lys Thr Gly Arg Lys Ser Met Cys
      420          425          430
Tyr Arg Ile Asn Tyr Gln Ser Met Asp Arg Asn Leu Thr Asn Ala Glu
      435          440          445
Val Asn Thr Leu Gln Asp Met Val Cys Ser Lys Leu Val Lys Glu Tyr
      450          455          460
Ser Val Glu Leu Arg
465

```

<210> 80

<211> 499

<212> PRT

<213> Schizosaccharomyces pombe

<400> 80

```

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

```

```

225          230          235          240
Gly Phe Glu Glu Met Pro Thr Asn Asn Phe Val Glu Ser Gly Phe Trp
          245          250          255
Asn Phe Asp Ala Leu Phe Val Pro Gln Gln His Ser Ala Arg Asp Ala
          260          265          270
Gln Asp Thr Phe Phe Leu Lys Val Pro Ala Ser Thr Asp Lys Leu Pro
          275          280          285
Asp Pro Glu Tyr Val Ala Arg Val Lys Ala Thr His Glu Asn Gly Gly
          290          295          300
Glu Thr Lys Gly Ile Gly Tyr Arg Ala Pro Phe Ser Leu Glu Glu Thr
305          310          315          320
Arg Lys Leu Val Leu Arg Thr His Thr Thr Ala Val Ser Ala Asn Met
          325          330          335
Leu Tyr Lys Leu Ala Gln Asn Gly Phe His Pro Ala Lys Tyr Phe Ser
          340          345          350
Ile Asp Arg Val Phe Arg Asn Glu Thr Val Asp Ala Thr His Leu Ala
          355          360          365
Glu Phe His Gln Val Glu Gly Val Ile Cys Asp Arg Asn Ile Thr Leu
          370          375          380
Gly Asp Leu Ile Gly Phe Leu Glu Val Phe Phe Gly Lys Met Asn Val
385          390          395          400
Lys Asn Leu Arg Phe Lys Pro Ala Tyr Asn Pro Tyr Thr Glu Pro Ser
          405          410          415
Leu Glu Val Phe Ser Tyr His Glu Lys Leu Gly Lys Trp Val Glu Val
          420          425          430
Gly Asn Ser Gly Met Phe Arg Pro Glu Met Leu Glu Pro Met Gly Leu
          435          440          445
Pro Lys Asp Val Arg Cys Leu Gly Phe Gly Leu Ser Leu Glu Arg Pro
          450          455          460
Thr Met Ile Lys Tyr Gly Val Ala Asp Ile Arg Gln Leu Ile Gly Pro
465          470          475          480
Lys Val Asn Leu Asp Leu Ile Glu Ala Ser Pro Ala Val Arg Leu Asp
          485          490          495
Lys Glu Glu

```

<210> 81

<211> 338

<212> PRT

<213> Pseudomonas aeruginosa

<400> 81

```

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

```

```

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

```

<210> 82

<211> 785

<212> PRT

<213> Thermus thermophilus

<400> 82

```

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

```

Leu	His	Ala	Leu	Gly	Tyr	Ala	Leu	Val	Glu	Pro	Glu	Ala	Ala	Leu	Lys
			180					185					190		
Ala	Glu	Ala	Leu	Pro	Leu	Pro	Phe	Ala	Leu	Lys	Val	Glu	Asp	Pro	Glu
		195					200					205			
Gly	Ala	Pro	His	Phe	Thr	Leu	Gly	Tyr	Ala	Phe	Gly	Leu	Arg	Val	Ala
	210					215					220				
Pro	Ser	Pro	Leu	Trp	Met	Gln	Arg	Ala	Leu	Phe	Ala	Ala	Gly	Met	Arg
225				230						235					240
Pro	Ile	Asn	Asn	Val	Val	Asp	Val	Thr	Asn	Tyr	Val	Met	Leu	Glu	Arg
			245						250					255	
Ala	Gln	Pro	Met	His	Ala	Phe	Asp	Leu	Arg	Phe	Val	Gly	Glu	Gly	Ile
		260					265					270			
Ala	Val	Arg	Arg	Ala	Arg	Glu	Gly	Glu	Arg	Leu	Lys	Thr	Leu	Asp	Gly
		275					280					285			
Val	Glu	Arg	Thr	Leu	His	Pro	Glu	Asp	Leu	Val	Ile	Ala	Gly	Trp	Arg
	290					295					300				
Gly	Glu	Glu	Ser	Phe	Pro	Leu	Gly	Leu	Ala	Gly	Val	Met	Gly	Gly	Ala
305				310						315					320
Glu	Ser	Glu	Val	Arg	Glu	Asp	Thr	Glu	Ala	Ile	Ala	Leu	Glu	Val	Ala
			325						330						335
Cys	Phe	Asp	Pro	Val	Ser	Ile	Arg	Lys	Thr	Ala	Arg	Arg	His	Gly	Leu
		340						345					350		
Arg	Thr	Glu	Ala	Ser	His	Arg	Phe	Glu	Arg	Gly	Val	Asp	Pro	Leu	Gly
		355					360					365			
Gln	Val	Pro	Ala	Gln	Arg	Arg	Ala	Leu	Ser	Leu	Leu	Gln	Ala	Leu	Ala
	370					375						380			
Gly	Ala	Arg	Val	Ala	Glu	Ala	Leu	Leu	Glu	Ala	Gly	Ser	Pro	Lys	Pro
385				390						395					400
Pro	Glu	Ala	Ile	Pro	Phe	Arg	Pro	Glu	Tyr	Ala	Asn	Arg	Leu	Leu	Gly
			405						410					415	
Thr	Ser	Tyr	Pro	Glu	Ala	Glu	Gln	Ile	Ala	Ile	Leu	Lys	Arg	Leu	Gly
		420					425						430		
Cys	Arg	Val	Glu	Gly	Glu	Gly	Pro	Thr	Tyr	Arg	Val	Thr	Pro	Pro	Ser
	435						440					445			
His	Arg	Leu	Asp	Leu	Arg	Leu	Glu	Glu	Asp	Leu	Val	Glu	Glu	Val	Ala
	450					455					460				
Arg	Ile	Gln	Gly	Tyr	Glu	Thr	Ile	Pro	Leu	Ala	Leu	Pro	Ala	Phe	Phe
465				470						475				480	
Pro	Ala	Pro	Asp	Asn	Arg	Gly	Val	Glu	Ala	Pro	Tyr	Arg	Lys	Glu	Gln
			485						490					495	
Arg	Leu	Arg	Glu	Val	Leu	Ser	Gly	Leu	Gly	Phe	Gln	Glu	Val	Tyr	Thr
		500						505					510		
Tyr	Ser	Phe	Met</												

```

625          630          635          640
Arg Val Leu Val Glu Gly Glu Glu Val Gly Phe Leu Gly Ala Leu His
          645          650          655
Pro Glu Ile Ala Gln Glu Leu Glu Leu Pro Pro Val His Leu Phe Glu
          660          665          670
Leu Arg Leu Pro Leu Pro Asp Lys Pro Leu Ala Phe Gln Asp Pro Ser
          675          680          685
Arg His Pro Ala Ala Phe Arg Asp Leu Ala Val Val Val Pro Ala Pro
          690          695          700
Thr Pro Tyr Gly Glu Val Glu Ala Leu Val Arg Glu Ala Ala Gly Pro
705          710          715          720
Tyr Leu Glu Ser Leu Ala Leu Phe Asp Leu Tyr Gln Gly Pro Pro Leu
          725          730          735
Pro Glu Gly His Lys Ser Leu Ala Phe His Leu Arg Phe Arg His Pro
          740          745          750
Lys Arg Thr Leu Arg Asp Glu Glu Val Glu Glu Ala Val Ser Arg Val
          755          760          765
Ala Glu Ala Leu Arg Ala Arg Gly Phe Gly Leu Arg Gly Leu Asp Thr
          770          775          780
Pro
785

```

<210> 83
 <211> 598
 <212> PRT
 <213> Arabidopsis thaliana

<400> 83

```

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

```

```

225                230                235                240
Ala His Ser Ala Ile Thr Leu Gln Thr Lys Asn Val Phe Ile Glu Cys
                245                250                255
Thr Ala Thr Asp Leu Thr Lys Ala Lys Ile Val Leu Asn Thr Met Val
                260                265                270
Thr Thr Phe Ser Glu Phe Cys Ala Arg Lys Phe Glu Ile Glu Pro Val
                275                280                285
Glu Val Thr Tyr Asp Asp Gly Lys Ser Tyr Ile Tyr Pro Asp Leu Ala
                290                295                300
Val Tyr Asp Met Glu Val Pro Leu Ser Phe Ile Thr Asp Ser Ile Gly
305                310                315                320
Val Ser Leu Lys Val Glu Gln Val Thr Ser Leu Leu Thr Arg Met Gln
                325                330                335
Leu Gln Ala Glu Gln Ala Lys Ser Ser Asp Asn Gln Cys Ala Ile Lys
                340                345                350
Val His Val Pro Pro Ser Arg Ser Asp Val Leu His Pro Cys Asp Val
                355                360                365
Met Glu Asp Val Ala Ile Ala Tyr Gly Phe Asn Asn Ile Pro Thr Arg
                370                375                380
Lys Pro Ala Ser Ile Lys Pro Leu Thr Leu Asn Glu Leu Thr Asp Leu
385                390                395                400
Leu Arg Ile Glu Ile Ala Met Cys Val Tyr Thr Glu Val Val Thr Trp
                405                410                415
Leu Leu Cys Ser His Lys Glu Asn Phe Ala Met Leu Asn Arg Glu Asp
                420                425                430
Val Asn Ser Ala Val Ile Val Gly Asn Pro Arg Ser Ala Asp Phe Glu
                435                440                445
Ala Met Arg Arg Ala Leu Met Pro Gly Leu Leu Lys Thr Val Gly His
                450                455                460
Asn Asn Lys Tyr Pro Lys Pro Ile Lys Ile Phe Glu Ile Ser Asp Val
465                470                475                480
Val Met Leu Asp Glu Ser Lys Asp Val Gly Ala Ser Asn Arg Arg His
                485                490                495
Leu Ala Ala Leu Tyr Cys Gly Ala Thr Ser Gly Phe Glu Leu Ile His
                500                505                510
Gly Leu Val Asp Arg Ile Met Glu Val Met Ala Ile Pro Phe Leu Thr
                515                520                525
Ile His Glu Asn Asn Val Pro Ile Asn Glu Lys Asp Gly Tyr Tyr Val
                530                535                540
Lys Leu Ser Gln Glu Pro Glu Phe Leu Pro Gly Arg Gln Ala Ser Ile
545                550                555                560
Ile Val Arg Gly Lys His Ile Gly Asn Phe Gly Ile Val His Pro Glu
                565                570                575
Val Leu Asn Asn Phe Asp Ile Pro Asp Pro Cys Ser Tyr Leu Glu Leu
                580                585                590
Asp Ile Glu Ala Ile Leu
                595

```

<210> 84

<211> 589

<212> PRT

<213> Drosophila melanogaster

<400> 84

```

Met Pro Thr Ile Gly Val Lys Arg Asp Leu Leu Phe Glu Ala Leu Gly
1          5          10          15
Lys Thr Tyr Thr Asp Asp Glu Phe Gln Asp Leu Cys Phe Ala Phe Gly

```

				20				25				30				
Leu	Glu	Leu	Asp	Glu	Val	Thr	Thr	Glu	Lys	Gln	Met	Leu	Thr	Lys	Glu	
35				40				45								
Gln	Gly	Asp	Val	Ala	Ala	Ala	Ala	Asn	Ala	Ser	Glu	Glu	Ile	Ile	Tyr	
50				55				60								
Arg	Ile	Asp	Ile	Pro	Ala	Asn	Arg	Tyr	Asp	Leu	Leu	Cys	Leu	Glu	Gly	
65					70					75					80	
Leu	Val	Thr	Gly	Leu	Leu	Val	Phe	Gln	Gly	Lys	Leu	Lys	Pro	Pro	Lys	
				85					90					95		
Phe	Gln	Phe	Val	Glu	Leu	Ala	Lys	Arg	Gln	Val	Leu	Lys	Ile	Asp	Pro	
100				105				110								
Ser	Thr	Ala	Gln	Ile	Arg	Pro	Tyr	Ala	Val	Ala	Ala	Val	Leu	Arg	Asn	
115				120				125								
Val	Thr	Phe	Thr	Gln	Ala	Ser	Tyr	Asn	Ser	Phe	Ile	Asp	Leu	Gln	Asp	
130				135				140								
Lys	Leu	His	Gln	Asn	Ile	Cys	Arg	Lys	Arg	Thr	Leu	Val	Ala	Ile	Gly	
145					150					155					160	
Thr	His	Asp	Leu	Asp	Thr	Leu	Gln	Gly	Pro	Phe	Ser	Tyr	Glu	Ala	Leu	
				165					170					175		
Ala	Pro	Asp	Gln	Ile	Lys	Phe	Lys	Pro	Leu	Asn	Gln	Thr	Lys	Glu	Met	
180				185				190								
Thr	Gly	Ser	Glu	Leu	Met	Asp	Phe	Tyr	Ser	Thr	His	Ala	Gln	Leu	Lys	
195				200				205								
Gln	Tyr	Leu	Pro	Ile	Ile	Arg	Glu	Ser	Pro	Val	Tyr	Pro	Val	Ile	Tyr	
210				215				220								
Asp	Ala	Asn	Arg	Val	Val	Leu	Ser	Leu	Pro	Pro	Ile	Ile	Asn	Gly	Asp	
225					230					235					240	
His	Ser	Lys	Ile	Thr	Leu	Lys	Thr	Lys	Asn	Val	Phe	Ile	Glu	Cys	Thr	
				245					250					255		
Ala	Thr	Asp	Arg	Thr	Lys	Ala	Lys	Val	Val	Leu	Asp	Thr	Ile	Val	Cys	
260				265				270								
Leu	Phe	Ser	Glu	His	Cys	Ala	Gln	Lys	Phe	Thr	Val	Glu	Pro	Cys	Asp	
275				280				285								
Val	Val	Gln	Pro	Asp	Gly	Ser	Val	Ile	Ser	Tyr	Pro	Glu	Leu	Glu	Val	
290				295				300								
Arg	Glu	Glu	Arg	Ile	Ser	Val	Lys	Arg	Ala	Asn	Ala	Tyr	Ile	Gly	Ile	
305					310					315					320	
Asp	Glu	Pro	Ala	Glu	Lys	Leu	Ala	Asp	Met	Leu	Thr	Arg	Met	Tyr	Leu	
				325					330					335		
Glu	Ala	Lys	Val	Asp	Gly	Asp	Ser	Leu	Val	Val	Lys	Ile	Pro	Pro	Thr	
340				345				350								
Arg	His	Asp	Val	Ile	His	Ala	Cys	Asp	Ile	Tyr	Glu	Asp	Val	Ala	Ile	
355				360				365								
Ala	Tyr	Gly	Tyr	Asn	Asn	Ile	Lys	Lys	Ser	Leu	Pro	Ala	Phe	Met	Gln	
370				375				380								
Ile	Ala	Lys	Gln	Phe	Pro	Leu	Asn	Lys	Leu	Thr	Glu	Gln	Leu	Arg	Glu	
385					390					395					400	
Gln	Val	Ala	Gln	Ala	Gly	Phe	Thr	Glu	Ala	Leu	Thr	Phe	Thr	Leu	Cys	
				405					410					415		
Ser	Arg	Asp	Asp	Ile	Gly	Arg	Lys	Leu	Asn	Lys	Asn					

```

                485                490                495
Ala Val Asn Cys Asn Lys Thr Ala Gly Phe Glu Val Val His Gly Leu
                500                505                510
Leu Asp Arg Val Met Gln Leu Leu Ser Val Pro Trp Lys Ser Ala Ser
                515                520                525
Gly Thr Lys Gly Tyr Tyr Leu Gln Ala Thr Glu Asp Pro Ser Tyr Phe
                530                535                540
Pro Gly Arg Cys Ala Asn Val Met Tyr Asp Gly Val Val Ile Gly Lys
545                550                555                560
Ile Gly Val Leu His Pro Thr Val Leu Gln Ala Phe Glu Leu Thr Thr
                565                570                575
Pro Cys Ser Ala Val Glu Phe Thr Ile Glu Pro Phe Val
                580                585

```

<210> 85
 <211> 591
 <212> PRT
 <213> Caenorhabditis elegans

<400> 85

```

Met Pro Thr Val Gly Ile Lys Lys Val Ile Leu Asp Lys His Phe Lys
 1                5                10                15
Arg Val Tyr Ser Glu Lys Glu Phe Asp Glu Leu Cys Phe Glu Tyr Gly
                20                25                30
Leu Glu Leu Asp Glu Ile Thr Ser Glu Lys Ala Ala Val Glu Lys Glu
                35                40                45
Gln Gly Thr Arg Ala Ala Ser Asp Leu Asn Asp Gln Glu Val Tyr Lys
 50                55                60
Ile Asp Ile Pro Ala Asn Arg Tyr Asp Leu Leu Ser Val Glu Gly Leu
65                70                75                80
Ala Arg Ala Ile Arg Ile Phe Lys Gln Glu Ile Pro Ser Pro Ala Tyr
                85                90                95
Lys Tyr Ala Asp Val Pro Lys Thr Gly Leu Gln Lys Ile Ile Val Lys
                100                105                110
Lys Glu Thr Ala Gln Val Arg Pro Phe Val Val Gly Ala Val Leu Arg
                115                120                125
Asp Ile Ser Phe Asp Ala Asp Ser Tyr Ala Ser Phe Ile Asp Leu Gln
                130                135                140
Asp Lys Leu His Gln Asn Ile Cys Arg Lys Arg Thr Leu Val Ala Ile
145                150                155                160
Gly Thr His Asp Leu Asp Thr Ile Gln Gly Pro Phe Glu Tyr Arg Ala
                165                170                175
Glu Ala Pro Lys Asp Ile Lys Phe Lys Pro Leu Asn Gln Thr Lys Glu
                180                185                190
Tyr Thr Ala Glu Glu Leu Met Thr Leu Tyr Ser Thr Asp Ser His Leu
                195                200                205
Lys Ala Tyr Leu Pro Ile Ile Gln Asn His Pro Val Tyr Pro Val Ile
                210                215                220
Tyr Asp Lys Asn Gly Val Val Cys Ser Met Pro Pro Ile Ile Asn Gly
225                230                235                240
Glu His Ser Lys Ile Thr Leu Asn Thr Lys Asn Val Phe Ile Glu Ala
                245                250                255
Thr Ala Thr Asp Lys Gln Lys Ala Phe Val Val Leu Asp Thr Ile Val
                260                265                270
Thr Leu Phe Ser Gln Tyr Cys Ala Lys Pro Phe Thr Ile Glu Gln Val
                275                280                285
Glu Val Val Tyr Glu Glu Thr Gly Val Lys Glu Leu Tyr Pro Leu Leu

```

```

      290      295      300
Ser Tyr Arg Glu Met Thr Val Thr Thr Pro Glu Ile Asn Thr Lys Ile
305      310      315      320
Gly Ile Asn Leu Lys Asp Glu Glu Met Ala Thr Leu Leu Asn Lys Met
      325      330      335
Ser Leu Lys Ala Glu Val Ala Ala Lys Glu Thr Leu Lys Ile Val Val
      340      345      350
Pro Pro Thr Arg His Asp Ile Leu His Ala Cys Asp Ile Ala Glu Asp
      355      360      365
Val Gly Val Ala Phe Gly Tyr Asn Asn Leu Ile Thr Lys Leu Pro Glu
      370      375      380
Ser Asn Thr Val Ala Val Ala Phe Pro Ile Asn Lys Leu Cys Asp Asn
385      390      395      400
Leu Arg Ile Glu Ile Ala Ala Ala Gly Trp Thr Glu Ala Leu Asn Phe
      405      410      415
Ala Leu Cys Ser Arg Asp Asp Ile Ser Ser Lys Leu Arg Gln Pro Asp
      420      425      430
Ala Leu Ser His Ala Val His Ile Gly Asn Pro Lys Thr Leu Glu Phe
      435      440      445
Gln Val Ala Arg Thr Ser Leu Leu Pro Gly Leu Leu Lys Thr Leu Ser
      450      455      460
Ser Asn Arg Asp Met Pro Leu Pro Leu Lys Leu Phe Glu Leu Gln Asp
465      470      475      480
Val Ile Val Lys Asp Ser Asn Thr Asp Val Gly Ala Arg Asn Glu Arg
      485      490      495
Arg Leu Ala Ala Val Tyr Tyr Asn Arg Ala Ala Gly Phe Glu Ile Ile
      500      505      510
Gln Gly Phe Leu Asp Arg Ile Met Arg Met Leu Asn Val Asn Pro Ala
      515      520      525
Arg Asp Gly Thr Gly Tyr Tyr Ile Glu Ala Asp Glu Asn Ser Thr Tyr
      530      535      540
Phe Pro Gly Arg Cys Ala Lys Ile Ile Gly Pro Lys Gly Val Val Leu
545      550      555      560
Gly His Ile Gly Ala Leu His Pro Glu Val Ile Thr Ser Phe Gly Leu
      565      570      575
Thr Leu Pro Cys Gly Ala Val Glu Ile Asn Val Glu Pro Phe Leu
      580      585      590

```

<210> 86
 <211> 804
 <212> PRT
 <213> Bacillus subtilis

<400> 86

```

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

```

			100					105				110			
Gly	Glu	Glu	Ser	Asn	Gly	Met	Ile	Cys	Ser	Leu	Gln	Glu	Leu	Gly	Ile
		115					120					125			
Glu	Ser	Lys	Leu	Val	Ala	Lys	Glu	Tyr	Ala	Glu	Gly	Ile	Phe	Val	Phe
	130						135				140				
Pro	Asn	Asp	Ala	Glu	Thr	Gly	Ser	Asp	Ala	Leu	Ala	Ala	Leu	Gln	Leu
145					150					155					160
Asp	Asp	Ala	Ile	Leu	Glu	Leu	Gly	Leu	Thr	Pro	Asn	Arg	Ala	Asp	Ala
			165						170					175	
Met	Asn	Met	Leu	Gly	Val	Ala	Tyr	Glu	Val	Ala	Ala	Ile	Leu	Asp	Thr
			180						185					190	
Glu	Val	Lys	Leu	Pro	Gln	Thr	Asp	Tyr	Pro	Ala	Ala	Ser	Glu	Gln	Ala
		195					200					205			
Ser	Asp	Tyr	Ile	Ser	Val	Lys	Ile	Glu	Asp	Gln	Glu	Ala	Asn	Pro	Leu
	210					215					220				
Tyr	Thr	Ala	Lys	Ile	Ile	Lys	Asn	Val	Thr	Ile	Ala	Pro	Ser	Pro	Leu
225					230						235				240
Trp	Met	Gln	Thr	Lys	Leu	Met	Asn	Ala	Gly	Ile	Arg	Pro	His	Asn	Asn
				245						250				255	
Val	Val	Asp	Ile	Thr	Asn	Phe	Val	Leu	Leu	Glu	Tyr	Gly	Gln	Pro	Leu
		260						265					270		
His	Ala	Phe	Asp	Tyr	Asp	Arg	Phe	Gly	Ser	Lys	Glu	Val	Val	Val	Arg
		275					280					285			
Lys	Ala	Ala	Glu	Asn	Glu	Met	Ile	Val	Thr	Leu	Asp	Asp	Gln	Glu	Arg
	290					295					300				
Lys	Leu	Ser	Ala	Asp	His	Leu	Val	Ile	Thr	Asn	Gly	Thr	Lys	Ala	Gln
305					310					315					320
Ala	Val	Ala	Gly	Val	Met	Gly	Gly	Ala	Glu	Ser	Glu	Val	Gln	Glu	Asp
				325						330				335	
Thr	Lys	Thr	Ile	Leu	Leu	Glu	Ala	Ala	Tyr	Phe	Asn	Gly	Gln	Lys	Val
			340					345					350		
Arg	Lys	Ala	Ser	Lys	Asp	Leu	Gly	Leu	Arg	Ser	Glu	Ser	Ser	Val	Arg
		355					360					365			
Phe	Glu	Lys	Gly	Ile	Asp	Pro	Ala	Arg	Val	Arg	Leu	Ala	Ala	Glu	Arg
	370					375					380				
Ala	Ala	Gln	Leu	Ile	His	Leu	Tyr	Ala	Gly	Gly	Glu	Val	Leu	Ala	Gly
385					390					395					400
Thr	Val	Glu	Glu	Asp	His	Leu	Thr	Ile	Glu	Ala	Asn	Asn	Ile	His	Val
				405					410					415	
Ser	Ala	Asp	Lys	Val	Ser	Ser	Val	Leu	Gly	Leu	Thr	Ile	Ser	Lys	Glu
			420					425					430		
Glu	Leu	Ile	Ser	Ile	Tyr	Lys	Arg	Leu	Gly	Phe	Thr	Val	Gly	Glu	Ala
		435					440					445			
Asp	Asp	Leu	Leu	Val	Val	Thr	Val	Pro							

				565					570					575	
Ala	Leu	Tyr	Glu	Val	Gly	Ser	Val	Phe	Leu	Thr	Lys	Glu	Glu	Asp	Thr
			580					585					590		
Lys	Pro	Val	Glu	Thr	Glu	Arg	Val	Ala	Gly	Ala	Val	Thr	Gly	Leu	Trp
		595					600					605			
Arg	Lys	Gln	Leu	Trp	Gln	Gly	Glu	Lys	Lys	Pro	Val	Asp	Phe	Phe	Val
	610				615					620					
Val	Lys	Gly	Ile	Val	Glu	Gly	Leu	Leu	Asp	Lys	Leu	Asn	Val	Leu	Asp
625				630					635						640
Ser	Ile	Glu	Phe	Val	Gln	Ser	Glu	Arg	Lys	Gln	Leu	His	Pro	Gly	Arg
			645					650					655		
Thr	Ala	Asn	Ile	Leu	Leu	Asn	Gly	Ser	Leu	Ile	Gly	Phe	Ile	Gly	Gln
		660					665				670				
Val	His	Pro	Ser	Leu	Glu	Lys	Glu	Leu	Asp	Ile	Lys	Glu	Thr	Tyr	Val
	675					680					685				
Phe	Glu	Leu	Asp	Leu	His	Ala	Leu	Leu	Ala	Ala	Glu	Thr	Ala	Pro	Leu
	690				695					700					
Val	Tyr	Thr	Ala	Ile	Pro	Lys	Tyr	Pro	Ser	Val	Thr	Arg	Asp	Ile	Ala
705				710						715					720
Leu	Val	Thr	Asp	Lys	Thr	Val	Thr	Ser	Gly	Gln	Leu	Glu	Ser	Val	Ile
			725					730						735	
Lys	Glu	Ala	Gly	Gly	Lys	Leu	Leu	Lys	Glu	Val	Thr	Val	Phe	Asp	Val
		740					745					750			
Tyr	Glu	Gly	Glu	His	Met	Glu	Glu	Gly	Lys	Lys	Ser	Val	Ala	Phe	Ser
	755				760						765				
Leu	Gln	Tyr	Val	Asn	Pro	Glu	Gln	Thr	Leu	Thr	Glu	Glu	Glu	Val	Thr
	770				775					780					
Lys	Ala	His	Ser	Lys	Val	Leu	Lys	Ala	Leu	Glu	Asp	Thr	Tyr	Gln	Ala
785				790					795						800
Val	Leu	Arg	Gly												

<210> 87

<211> 595

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 87

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

```

145          150          155          160
Ala Leu Pro Pro Lys Asp Ile Lys Phe Val Pro Leu Asn Gln Thr Gln
165          170          175
Glu Phe Thr Gly Asp Lys Leu Ile Glu Phe Tyr Lys Ser Pro Glu Gln
180          185          190
Lys Asn Asn Ile Gly Arg Tyr Val His Ile Ile Glu Asp Ser Pro Val
195          200          205
Phe Pro Val Ile Met Asp Ser Lys Asp Arg Val Cys Ser Leu Pro Pro
210          215          220
Leu Ile Asn Ser Glu His Ser Lys Ile Ser Val Asn Thr Arg Asn Ile
225          230          235
Leu Ile Asp Ile Thr Ala Thr Asp Lys Thr Lys Ala Glu Ile Val Leu
245          250          255
Asn Ile Leu Thr Thr Met Phe Ser Arg Tyr Cys Asp Glu Pro Phe Thr
260          265          270
Val Glu Pro Val Glu Ile Val Ser Glu His Asn Gly Gln Ser Arg Leu
275          280          285
Ala Pro Asn Phe Asn Asp Arg Ile Met Asp Val Ser Ile Lys Tyr Ile
290          295          300
Asn Ser Cys Leu Gly Leu Asp Gln Ser Ala Asp Glu Ile Ala His Cys
305          310          315
Leu Lys Lys Met Ser Leu His Ala Val Gln Ser Lys Glu Asp Lys Asp
325          330          335
Ile Leu His Val Asp Ile Pro Val Thr Arg Pro Asp Ile Leu His Ala
340          345          350
Cys Asp Ile Met Glu Asp Ala Ala Val Gly Tyr Gly Phe Asn Asn Leu
355          360          365
Pro Lys Gly Glu Lys Leu Ser Asn Ala Asn Phe Ile Ala Lys Pro Leu
370          375          380
Pro Ile Asn Lys Val Ser Asp Ile Phe Arg Val Ala Ser Ser Gln Ala
385          390          395
Thr Trp Val Glu Val Leu Pro Leu Thr Leu Cys Ser His Asp Glu Asn
405          410          415
Phe Lys Phe Leu Arg Gln Ser Asp Asn Gly Asp Leu Ala Val Lys Leu
420          425          430
Ala Asn Pro Lys Thr Leu Glu Tyr Gln Val Val Arg Thr Thr Leu Leu
435          440          445
Pro Gly Ile Leu Lys Thr Val Lys Glu Asn Arg Lys His Ser Leu Pro
450          455          460
Ile Lys Val Phe Glu Thr Gly Asp Val Val Phe Lys Asp Asp Lys Leu
465          470          475
Glu Arg Lys Ala Tyr Asn Glu Arg His Trp Ala Ala Ile Tyr Val Gly
485          490          495
Lys Asn Ser Gly Phe Glu Ile Ile Gln Gly Leu Leu Gly Lys Ile Met
500          505          510
Gln Thr Phe Arg Thr Glu Trp Ile Ala Asp Tyr Gly Ala Ala Ser
515          520          525
Gly Arg Gly Tyr Trp Ile Glu Glu Asp Asp Ser Val Lys Thr Tyr Phe
530          535          540
Pro Gly Arg Gly Ala Lys Val Met Phe Arg Ser Lys Glu Gly Ala Glu
545          550          555
Pro Lys Gln Ile Gly His Leu Gly Val Leu His Pro Glu Val Met Met
565          570          575
Asn Phe Asp Val Pro Phe Ala Ala Ser Phe Val Glu Val Asn Ala Glu
580          585          590
Val Phe Leu
595

```

<210> 88

<211> 792

<212> PRT

<213> Chlamydophila pneumoniae

<400> 88

```

Met Arg Ile Pro Ile Thr Leu Leu Gln Thr Tyr Phe Ser Glu Pro Leu
 1           5           10           15
Ser Thr Lys Glu Ile Leu Glu Ala Cys Asp His Ile Gly Ile Glu Ala
          20           25           30
Glu Ile Glu Asn Thr Thr Leu Tyr Ser Phe Ala Ser Val Ile Thr Ala
          35           40           45
Lys Ile Leu His Thr Ile Pro His Pro Asn Ala Asp Lys Leu Arg Val
          50           55           60
Ala Thr Leu Thr Asp Gly Glu Lys Glu His Gln Val Val Cys Gly Ala
65          70          75          80
Pro Asn Cys Glu Ala Gly Leu Ile Val Ala Leu Ala Leu Pro Gly Ala
          85          90          95
Lys Leu Phe Asp Ser Glu Gly Gln Ala Tyr Thr Ile Lys Lys Ser Lys
          100         105         110
Leu Arg Gly Val Glu Ser Gln Gly Met Cys Cys Gly Ala Asp Glu Leu
          115         120         125
Gly Leu Asp Glu Leu Gln Ile Gln Glu Arg Ala Leu Leu Glu Leu Pro
130         135         140
Glu Ala Thr Pro Leu Gly Glu Asp Leu Ala Thr Val Leu Gly Asn Thr
145         150         155         160
Ser Leu Glu Ile Ser Leu Thr Pro Asn Leu Gly His Cys Ala Ser Phe
          165         170         175
Leu Gly Leu Ala Arg Glu Ile Cys His Val Thr Gln Ala Asn Leu Val
          180         185         190
Ile Pro Lys Glu Phe Ser Phe Glu Asn Leu Pro Thr Thr Ala Leu Asp
          195         200         205
Met Gly Asn Asp Pro Asp Ile Cys Pro Phe Phe Ser Tyr Val Val Ile
210         215         220
Thr Gly Ile Ser Ala Gln Pro Ser Pro Ile Lys Leu Gln Glu Ser Leu
225         230         235         240
Gln Ala Leu Lys Gln Lys Pro Ile Asn Ala Ile Val Asp Ile Thr Asn
          245         250         255
Tyr Ile Met Leu Ser Leu Gly Gln Pro Leu His Ala Tyr Asp Ala Ser
          260         265         270
His Val Ala Leu Asp Ser Leu Arg Val Glu Lys Leu Ser Thr Pro Glu
          275         280         285
Ser Leu Thr Leu Leu Asn Gly Glu Thr Val Leu Leu Pro Ser Gly Val
          290         295         300
Pro Val Val Arg Asp Asp His Ser Leu Leu Gly Leu Gly Gly Val Met
305         310         315         320
Gly Ala Lys Ala Pro Ser Phe Gln Glu Thr Thr Thr Thr Thr Val Ile
          325         330         335
Lys Ala Ala Tyr Phe Leu Pro Glu Ala Leu Arg Ala Ser Gln Lys Leu
          340         345         350
Leu Pro Ile Pro Ser Glu Ser Ala Tyr Arg Phe Thr Arg Gly Ile Asp
          355         360         365
Pro Gln Asn Val Val Pro Ala Leu Gln Ala Ala Ile His Tyr Ile Leu
          370         375         380
Glu Ile Phe Pro Glu Ala Thr Ile Ser Pro Ile Tyr Ser Ser Gly Glu
385         390         395         400
Ile Cys Arg Glu Leu Lys Glu Val Ala Leu Arg Pro Lys Thr Leu Gln

```

Arg	Ile	Leu	Gly	Lys	Ser	Phe	Ser	Ile	Glu	Ile	Leu	Ser	Gln	Lys	Leu
			420					425					430		
Gln	Ser	Leu	Gly	Phe	Ser	Thr	Thr	Pro	Gln	Glu	Thr	Ser	Leu	Leu	Val
		435					440					445			
Lys	Val	Pro	Ser	Tyr	Arg	His	Asp	Ile	Asn	Glu	Glu	Ile	Asp	Leu	Val
	450					455				460					
Glu	Glu	Ile	Cys	Arg	Thr	Glu	Ser	Trp	Asn	Ile	Glu	Thr	Gln	Asn	Pro
465					470					475					480
Val	Ser	Cys	Tyr	Thr	Pro	Ile	Tyr	Lys	Leu	Lys	Arg	Glu	Thr	Ala	Gly
				485					490					495	
Phe	Leu	Ala	Asn	Ala	Gly	Leu	Gln	Glu	Phe	Phe	Thr	Pro	Asp	Leu	Leu
			500					505					510		
Asp	Pro	Glu	Thr	Val	Ala	Leu	Thr	Arg	Lys	Glu	Lys	Glu	Glu	Ile	Ser
		515					520					525			
Leu	Gln	Gly	Ser	Lys	His	Thr	Thr	Val	Leu	Arg	Ser	Ser	Leu	Leu	Pro
	530					535					540				
Gly	Leu	Leu	Lys	Ser	Ala	Ala	Thr	Asn	Leu	Asn	Arg	Gln	Ala	Pro	Ser
545					550					555					560
Val	Gln	Ala	Phe	Glu	Ile	Gly	Thr	Val	Tyr	Ala	Lys	His	Gly	Glu	Gln
				565					570					575	
Cys	Gln	Glu	Thr	Gln	Thr	Leu	Ala	Ile	Leu	Leu	Thr	Glu	Asp	Gly	Glu
			580					585					590		
Ser	Arg	Ser	Trp	Leu	Pro	Lys	Pro	Ser	Leu	Ser	Phe	Tyr	Ser	Leu	Lys
		595					600					605			
Gly	Trp	Val	Glu	Arg	Leu	Leu	Tyr	His	His	His	Leu	Ser	Ile	Asp	Ala
	610					615					620				
Leu	Thr	Leu	Glu	Ser	Ser	Ala	Leu	Cys	Glu	Phe	His	Pro	Tyr	Gln	Gln
				630						635					640
Gly	Val	Leu	Arg	Ile	His	Lys	Gln	Ser	Phe	Ala	Thr	Leu	Gly	Gln	Val
				645					650					655	
His	Pro	Glu	Leu	Ala	Lys	Lys	Ala	Gln	Ile	Lys	His	Pro	Val	Phe	Phe
			660					665					670		
Ala	Glu	Leu	Asn	Leu	Asp	Leu	Leu	Cys	Lys	Met	Leu	Lys	Lys	Thr	Thr
		675					680					685			
Lys	Leu	Tyr	Lys	Pro	Tyr	Ala	Ile	Tyr	Pro	Ser	Ser	Phe	Arg	Asp	Leu
	690					695					700				
Thr	Leu	Thr	Val	Pro	Glu	Asp	Ile	Pro	Ala	Asn	Leu	Leu	Arg	Gln	Lys
				710						715					720
Leu	Leu	His	Glu	Gly	Ser	Lys	Trp	Leu	Glu	Ser	Val	Thr	Ile	Ile	Ser
				725						730				735	
Ile	Tyr	Gln	Asp	Lys	Ser	Leu	Glu	Thr	Arg	Asn	Lys	Asn	Val	Ser	Leu
			740					745					750		
Arg	Leu	Val	Phe	Gln	Asp										

<210> 89

<211> 831

<212> PRT

<213> Mycobacterium tuberculosis

<400> 89

Met Arg Leu Pro Tyr Ser Trp Leu Arg Glu Val Val Ala Val Gly Ala

1	5	10	15
Ser Gly Trp Asp Val Thr Pro Gly Glu Leu Glu Gln Thr Leu Leu Arg			
	20	25	30
Ile Gly His Glu Val Glu Glu Val Ile Pro Leu Gly Pro Val Asp Gly			
	35	40	45
Pro Val Thr Val Gly Arg Val Ala Asp Ile Glu Glu Leu Thr Gly Tyr			
	50	55	60
Lys Lys Pro Ile Arg Ala Cys Ala Val Asp Ile Gly Asp Arg Gln Tyr			
65	70	75	80
Arg Glu Ile Ile Cys Gly Ala Thr Asn Phe Ala Val Gly Asp Leu Val			
	85	90	95
Val Val Ala Leu Pro Gly Ala Thr Leu Pro Gly Gly Phe Thr Ile Ser			
	100	105	110
Ala Arg Lys Ala Tyr Gly Arg Asn Ser Asp Gly Met Ile Cys Ser Ala			
	115	120	125
Ala Glu Leu Asn Leu Gly Ala Asp His Ser Gly Ile Leu Val Leu Pro			
	130	135	140
Pro Gly Ala Ala Glu Pro Gly Ala Asp Gly Ala Gly Val Leu Gly Leu			
145	150	155	160
Asp Asp Val Val Phe His Leu Ala Ile Thr Pro Asp Arg Gly Tyr Cys			
	165	170	175
Met Ser Val Arg Gly Leu Ala Arg Glu Leu Ala Cys Ala Tyr Asp Leu			
	180	185	190
Asp Phe Val Asp Pro Ala Ser Asn Ser Arg Val Pro Pro Leu Pro Ile			
	195	200	205
Glu Gly Pro Ala Trp Pro Leu Thr Val Gln Pro Glu Thr Gly Val Arg			
	210	215	220
Arg Phe Ala Leu Arg Pro Val Ile Gly Ile Asp Pro Ala Ala Val Ser			
225	230	235	240
Pro Trp Trp Leu Gln Arg Arg Leu Leu Leu Cys Gly Ile Arg Ala Thr			
	245	250	255
Cys Pro Ala Val Asp Val Thr Asn Tyr Val Met Leu Glu Leu Gly His			
	260	265	270
Pro Met His Ala His Asp Arg Asn Arg Ile Ser Gly Thr Leu Gly Val			
	275	280	285
Arg Phe Ala Arg Ser Gly Glu Thr Ala Val Thr Leu Asp Gly Ile Glu			
	290	295	300
Arg Lys Leu Asp Thr Ala Asp Val Leu Ile Val Asp Asp Ala Ala Thr			
305	310	315	320
Ala Ala Ile Gly Gly Val Met Gly Ala Ala Ser Thr Glu Val Arg Ala			
	325	330	335
Asp Ser Thr Asp Val Leu Leu Glu Ala Ala Ile Trp Asp Pro Ala Ala			
	340	345	350
Val Ser Arg Thr Gln Arg Arg Leu His Leu Pro Ser Glu Ala Ala Arg			
	355	360	365
Arg Tyr Glu Arg Thr Val Asp Pro Ala Ile Ser Val Ala Ala Leu Asp			
	370	375	380
Arg Cys Ala Arg Leu Leu Ala Asp Ile Ala Gly Gly Glu Val Ser Pro			
385	390	395	400
Thr Leu Thr Asp Trp Arg Gly Asp Pro Pro Cys Asp Asp Trp Ser Pro			
	405	410	415
Pro Pro Ile Arg Met Gly Val Asp Val Pro Asp Arg Ile Ala Gly Val			
	420	425	430
Ala Tyr Pro Gln Gly Thr Thr Ala Arg Arg Leu Ala Gln Ile Gly Ala			
	435	440	445
Val Val Thr His Asp Gly Asp Thr Leu Thr Val Thr Pro Pro Ser Trp			
	450	455	460
Arg Pro Asp Leu Arg Gln Pro Ala Asp Leu Val Glu Glu Val Leu Arg			

```

465          470          475          480
Leu Glu Gly Leu Glu Val Ile Pro Ser Val Leu Pro Pro Ala Pro Ala
          485          490          495
Gly Arg Gly Leu Thr Ala Gly Gln Gln Arg Arg Arg Thr Ile Gly Arg
          500          505          510
Ser Leu Ala Leu Ser Gly Tyr Val Glu Ile Leu Pro Thr Pro Phe Leu
          515          520          525
Pro Ala Gly Val Phe Asp Leu Trp Gly Leu Glu Ala Asp Asp Ser Arg
          530          535          540
Arg Met Thr Thr Arg Val Leu Asn Pro Leu Glu Ala Asp Arg Pro Gln
545          550          555          560
Leu Ala Thr Thr Leu Leu Pro Ala Leu Leu Glu Ala Leu Val Arg Asn
          565          570          575
Val Ser Arg Gly Leu Val Asp Val Ala Leu Phe Ala Ile Ala Gln Val
          580          585          590
Val Gln Pro Thr Glu Gln Thr Arg Gly Val Gly Leu Ile Pro Val Asp
          595          600          605
Arg Arg Pro Thr Asp Asp Glu Ile Ala Met Leu Asp Ala Ser Leu Pro
          610          615          620
Arg Gln Pro Gln His Val Ala Ala Val Leu Ala Gly Leu Arg Glu Pro
625          630          635          640
Arg Gly Pro Trp Gly Pro Gly Arg Pro Val Glu Ala Ala Asp Ala Phe
          645          650          655
Glu Ala Val Arg Ile Ile Ala Arg Ala Ser Arg Val Asp Val Thr Leu
          660          665          670
Arg Pro Ala Gln Tyr Leu Pro Trp His Pro Gly Arg Cys Ala Gln Val
          675          680          685
Phe Val Gly Glu Ser Ser Val Gly His Ala Gly Gln Leu His Pro Ala
          690          695          700
Val Ile Glu Arg Ser Gly Leu Pro Lys Gly Thr Cys Ala Val Glu Leu
705          710          715          720
Asn Leu Asp Ala Ile Pro Cys Ser Ala Pro Leu Pro Ala Pro Arg Val
          725          730          735
Ser Pro Tyr Pro Ala Val Phe Gln Asp Val Ser Leu Val Val Ala Ala
          740          745          750
Asp Ile Pro Ala Gln Ala Val Ala Asp Ala Val Arg Ala Gly Ala Gly
          755          760          765
Asp Leu Leu Glu Asp Ile Ala Leu Phe Asp Val Phe Thr Gly Pro Gln
          770          775          780
Ile Gly Glu His Arg Lys Ser Leu Thr Phe Ala Leu Arg Phe Arg Ala
785          790          795          800
Pro Asp Arg Thr Leu Thr Glu Asp Asp Ala Ser Ala Ala Arg Asp Ala
          805          810          815
Ala Val Gln Ser Ala Ala Glu Arg Val Gly Ala Val Leu Arg Gly
          820          825          830

```

<210> 90

<211> 764

<212> PRT

<213> Helicobacter pylori

<400> 90

```

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

```

35					40					45					
Ile	Leu	Glu	Lys	Ala	Pro	His	Lys	Asn	Ala	Glu	Lys	Leu	Ser	Val	Cys
50					55					60					
Gln	Val	Asp	Val	Gly	Lys	Glu	Val	Leu	Gln	Ile	Val	Cys	Gly	Ala	Lys
65					70					75					80
Asn	Val	Ala	Pro	Asn	Gln	Phe	Val	Pro	Val	Ala	Leu	Asn	Gly	Ala	Leu
				85					90					95	
Ile	Gly	Ser	Thr	Thr	Ile	Ala	Lys	Thr	Glu	Leu	Arg	Gly	Val	Glu	Ser
			100					105					110		
His	Gly	Met	Ile	Cys	Ser	Ser	Ile	Glu	Leu	Gly	Phe	Pro	Lys	Ile	Asn
		115					120					125			
Asp	Gly	Ile	Leu	Glu	Leu	Asp	Glu	Ser	Val	Gly	Glu	Leu	Val	Leu	Gly
	130					135					140				
Lys	Glu	Leu	Asn	Glu	Tyr	Ala	Pro	Phe	Asn	Thr	His	Val	Leu	Glu	Ile
145					150					155					160
Ser	Leu	Thr	Pro	Asn	Arg	Gly	Asp	Cys	Leu	Ser	Val	Leu	Gly	Ile	Ala
				165					170					175	
Arg	Glu	Ile	Ser	Ala	Phe	Tyr	His	Thr	Pro	Leu	Lys	Pro	Ile	Lys	Ala
			180					185					190		
Leu	Asn	Phe	Thr	Pro	Lys	Ser	Gly	Leu	Ile	Thr	Leu	Ser	Ala	Gly	Glu
		195					200					205			
Asn	Ile	Glu	Ser	His	Leu	Ala	Tyr	Tyr	Leu	Ile	Cys	Asn	His	Ser	Leu
	210					215					220				
Lys	Thr	Pro	Leu	Asn	Ile	Lys	Leu	Ser	Leu	Ala	His	Asn	Asn	Ala	Leu
225					230					235					240
Ser	Glu	Asn	Asp	Leu	Asn	Asn	Phe	Ile	Glu	Phe	Ser	Thr	His	Phe	Ser
			245						250					255	
Gly	Val	Ile	Met	Asn	Ala	Tyr	Ser	Leu	Asn	Thr	Thr	Pro	Met	Asp	Leu
			260					265					270		
Ser	Val	Lys	Asn	Asp	Glu	Asn	Asn	Leu	Glu	Ser	Val	Tyr	Ile	Asn	His
		275					280					285			
Gln	Lys	Arg	Ser	Thr	Ile	Ala	Ile	Lys	His	Gln	Val	Gln	Lys	Asp	Leu
	290					295					300				
Ser	Glu	Cys	Leu	Leu	Leu	Glu	Ala	Ser	Tyr	Thr	Asp	Pro	Ile	Ser	Leu
305					310					315					320
Ser	Leu	Lys	Leu	His	Ala	Leu	Lys	Asp	Lys	Thr	Leu	Gln	Lys	Asp	Asn
			325						330					335	
Ala	Leu	Ile	Tyr	Arg	Ser	Ala	Arg	Gly	Ser	Asn	Pro	Asn	Leu	Ser	Asp
			340					345					350		
Gly	Leu	Asn	Phe	Leu	Ser	Ala	His	Leu	Lys	Ala	Thr	Ile	Leu	Glu	Ser
		355					360					365			
Lys	Gln	Thr	Glu	His	Ser	Leu									

Pro	Leu	Glu	Leu	Gln	Asn	Pro	Ile	Thr	Thr	Glu	Leu	Asn	Thr	Leu	Arg
		515					520					525			
Thr	Ser	Leu	Val	Cys	Gly	Leu	Leu	Asp	Ala	Ser	Leu	Arg	Asn	Lys	Asn
		530				535					540				
Leu	Gly	Phe	Lys	Ser	Ile	Ala	Leu	Tyr	Glu	Lys	Gly	Ser	Val	Tyr	Asn
545					550					555					560
Ser	Lys	Arg	Glu	Glu	Ile	Gln	Lys	Leu	Gly	Phe	Leu	Ile	Ser	Gly	Leu
				565					570					575	
Gln	Lys	Lys	Glu	Ser	Tyr	Pro	Asp	Thr	Lys	Gly	Lys	Ala	Trp	Asp	Phe
			580					585					590		
Tyr	Ser	Phe	Ala	Glu	Cys	Val	Ser	Lys	Val	Ile	Gly	Asp	Phe	Ser	Leu
		595					600					605			
Glu	Lys	Leu	Thr	Thr	Gln	Thr	Pro	Ile	Asn	His	Pro	Tyr	Gln	Ser	Ala
		610				615					620				
Lys	Ile	Ile	Gln	Asn	His	Glu	Ile	Ile	Gly	Val	Ile	Ala	Lys	Ile	His
625				630						635					640
Pro	Lys	Val	Ile	Gln	Glu	Leu	Asp	Leu	Phe	Glu	Ser	Tyr	Tyr	Ala	Glu
				645					650					655	
Ile	Asp	Ala	Phe	Lys	Leu	Lys	Arg	Pro	Ala	Met	Leu	Leu	Lys	Pro	Phe
		660						665					670		
Ser	Ile	Tyr	Pro	Ser	Ser	Val	Arg	Asp	Leu	Thr	Leu	Ile	Ile	Asp	Glu
		675					680					685			
Asn	Thr	Ala	Phe	Ser	Gly	Ile	Lys	Lys	Ala	Leu	Lys	Asp	Ala	Gln	Ile
		690				695					700				
Pro	Asn	Leu	Ser	Glu	Ile	Leu	Pro	Leu	Asp	Ile	Phe	Lys	Glu	Ser	Asn
705				710						715					720
Asn	Ser	Ile	Ala	Leu	Ser	Val	Arg	Cys	Val	Ile	His	Ser	Leu	Glu	Lys
			725						730					735	
Thr	Leu	Asn	Asp	Glu	Glu	Val	Asn	Ser	Ala	Val	Gln	Lys	Ala	Leu	Glu
			740					745					750		
Ile	Leu	Glu	Lys	Glu	Phe	Asn	Ala	Arg	Leu	Lys	Gly				
		755					760								

<210> 91
<211> 792
<212> PRT
<213> Pseudomonas aeruginosa

<400> 91

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

130	135	140
Asp Val Arg Thr Tyr Leu Glu Leu Ala Asp Tyr Thr Ile Glu Val Gly		
145	150	155
Leu Thr Pro Asn Arg Gly Asp Cys Leu Ser Leu Ala Gly Leu Ala Arg		160
	165	170
Glu Val Ser Ala Ile Tyr Asp Val Pro Leu Ala Pro Val Ala Val Asp		175
	180	185
Ala Val Ala Ala Gln His Asp Glu Thr Arg Pro Val Glu Leu Ala Ala		190
	195	200
Pro Ala Ala Cys Pro Arg Tyr Leu Gly Arg Val Ile Arg Asn Val Asp		205
	210	215
Leu Ser Arg Pro Thr Pro Leu Trp Met Val Glu Arg Leu Arg Arg Ser		220
225	230	235
Asp Ile Arg Ser Ile Asp Pro Val Val Asp Val Thr Asn Tyr Val Met		240
	245	250
Ile Glu Leu Gly Gln Pro Met His Ala Phe Asp Leu Ala Glu Ile Asn		255
	260	265
Gly Gly Val Arg Val Arg Met Ala Glu Asp Gly Glu Lys Leu Val Leu		270
	275	280
Leu Asp Gly Gln Glu Ile Thr Leu Arg Ala Asp Thr Leu Val Ile Ala		285
	290	295
Asp His Gln Arg Ala Leu Ala Ile Ala Gly Val Met Gly Gly Glu His		300
305	310	315
Ser Gly Val Ser Asp Ser Thr Arg Asp Leu Phe Leu Glu Ala Ala Phe		320
	325	330
Phe Asp Thr Ile Ala Leu Ala Gly Lys Ala Arg Ser Tyr Gly Leu His		335
	340	345
Thr Asp Ser Ser His Arg Phe Glu Arg Gly Val Asp Ser Gln Leu Ala		350
	355	360
Arg Lys Ala Met Glu Arg Ala Thr Arg Leu Ile Leu Asp Ile Val Gly		365
	370	375
Gly Glu Pro Gly Pro Ile Val Glu Gln Val Ser Glu Ala His Leu Pro		380
385	390	395
Lys Val Ala Pro Ile Thr Leu Arg Ala Glu Arg Val Thr Gln Met Leu		400
	405	410
Gly Met Pro Leu Asp Ala Ala Glu Ile Val Arg Leu Leu Gln Ala Leu		415
	420	425
Glu Leu Thr Val Val Ala Asp Gly Glu Gly Gln Trp Ser Val Gly Val		430
	435	440
Pro Ser His Arg Phe Asp Ile Ser Leu Glu Val Asp Leu Ile Glu Glu		445
	450	455
Leu Ala Arg Leu Tyr Gly Tyr Asn Arg Leu Pro Val Arg Tyr Pro Gln		460
465	470	475
Ala Arg Leu Ala Pro Asn Asn Lys Pro Glu Ala Arg Ala Ala Leu Pro		480
	485	490
Leu Leu Arg Arg Leu Leu Val Ala Arg Gly Tyr Gln Glu Ala Ile Thr		495
	500	505
Phe Ser Phe Ile Asp Pro Ala Leu Phe Glu Leu Phe Asp Pro Gly Thr		510
	515	520
Gln Pro Leu Thr Leu Ala Asn Pro Ile Ser Ala Asp Met Ala Ala Met		525
	530	535
Arg Ser Ser Leu Trp Pro Gly Leu Val Lys Ala Leu Gln His Asn Leu		540
545	550	555
Asn Arg Gln Gln Ser Arg Val Arg Leu Phe Glu Ser Gly Leu Arg Phe		560
	565	570
Val Gly Gln Leu Glu Gly Leu Lys Gln Glu Ala Met Leu Ala Gly Ala		575
	580	585
Ile Cys Gly Lys Arg Leu Pro Glu Gly Trp Ala Asn Gly Arg Asp Gly		590

```

      595              600              605
Val Asp Phe Phe Asp Ala Lys Ala Asp Val Glu Ala Val Leu Ala Ser
  610              615              620
Ala Gly Ala Leu Gly Asp Phe Ser Phe Val Pro Gly Glu His Pro Ala
  625              630              635              640
Leu His Pro Gly Gln Thr Ala Arg Ile Glu Arg Glu Gly Arg Leu Val
      645              650              655
Gly Tyr Leu Gly Ala Leu His Pro Glu Leu Ala Lys Lys Leu Asp Leu
      660              665              670
Glu Gln Pro Val Phe Leu Phe Glu Leu Leu Leu Ala Glu Val Val Asp
      675              680              685
Gly His Leu Pro Lys Phe Arg Glu Leu Ser Arg Phe Pro Glu Val Arg
      690              695              700
Arg Asp Leu Ala Leu Leu Val Asp Gln Asp Val Pro Ala Gln Asp Ile
  705              710              715              720
Leu Thr Gln Ile Arg Ala Ala Ala Gly Glu Trp Leu Thr Asp Leu Arg
      725              730              735
Leu Phe Asp Val Tyr His Gly Lys Gly Ile Asp Pro His Arg Lys Ser
      740              745              750
Leu Ala Val Gly Leu Thr Trp Gln His Pro Ser Arg Thr Leu Asn Asp
      755              760              765
Asp Glu Val Asn Ser Thr Thr Gln Asn Ile Val Thr Ser Leu Glu Glu
  770              775              780
Arg Phe Asn Ala Thr Leu Arg Lys
  785              790

```

<210> 92
 <211> 551
 <212> PRT
 <213> Escherichia coli

<400> 92

```

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

```

		195					200					205					
Gly	Ala	Leu	Gln	Glu	Gln	Val	Ala	Asn	Lys	Met	Gln	Glu	Trp	Phe	Glu		
	210					215					220						
Ser	Gly	Leu	Gln	Gln	Trp	Asp	Ile	Ser	Arg	Asp	Ala	Pro	Tyr	Phe	Gly		
225					230					235					240		
Phe	Glu	Ile	Pro	Asn	Ala	Pro	Gly	Lys	Tyr	Phe	Tyr	Val	Trp	Leu	Asp		
				245					250					255			
Ala	Pro	Ile	Gly	Tyr	Met	Gly	Ser	Phe	Lys	Asn	Leu	Cys	Asp	Lys	Arg		
			260					265					270				
Gly	Asp	Ser	Val	Ser	Phe	Asp	Glu	Tyr	Trp	Lys	Lys	Asp	Ser	Thr	Ala		
			275				280					285					
Glu	Leu	Tyr	His	Phe	Ile	Gly	Lys	Asp	Ile	Val	Tyr	Phe	His	Ser	Leu		
	290					295					300						
Phe	Trp	Pro	Ala	Met	Leu	Glu	Gly	Ser	Asn	Phe	Arg	Lys	Pro	Ser	Asn		
305					310					315					320		
Leu	Phe	Val	His	Gly	Tyr	Val	Thr	Val	Asn	Gly	Ala	Lys	Met	Ser	Lys		
				325					330						335		
Ser	Arg	Gly	Thr	Phe	Ile	Lys	Ala	Ser	Thr	Trp	Leu	Asn	His	Phe	Asp		
			340					345					350				
Ala	Asp	Ser	Leu	Arg	Tyr	Tyr	Tyr	Thr	Ala	Lys	Leu	Ser	Ser	Arg	Ile		
		355					360					365					
Asp	Asp	Ile	Asp	Leu	Asn	Leu	Glu	Asp	Phe	Val	Gln	Arg	Val	Asn	Ala		
	370					375					380						
Asp	Ile	Val	Asn	Lys	Val	Val	Asn	Leu	Ala	Ser	Arg	Asn	Ala	Gly	Phe		
385					390					395					400		
Ile	Asn	Lys	Arg	Phe	Asp	Gly	Val	Leu	Ala	Ser	Glu	Leu	Ala	Asp	Pro		
				405					410					415			
Gln	Leu	Tyr	Lys	Thr	Phe	Thr	Asp	Ala	Ala	Glu	Val	Ile	Gly	Glu	Ala		
			420					425					430				
Trp	Glu	Ser	Arg	Glu	Phe	Gly	Lys	Ala	Val	Arg	Glu	Ile	Met	Ala	Leu		
		435					440					445					
Ala	Asp	Leu	Ala	Asn	Arg	Tyr	Val	Asp	Glu	Gln	Ala	Pro	Trp	Val	Val		
	450					455					460						
Ala	Lys	Gln	Glu	Gly	Arg	Asp	Ala	Asp	Leu	Gln	Ala	Ile	Cys	Ser	Met		
465					470					475					480		
Gly	Ile	Asn	Leu	Phe	Arg	Val	Leu	Met	Thr	Tyr	Leu	Lys	Pro	Val	Leu		
				485					490					495			
Pro	Lys	Leu	Thr	Glu	Arg	Ala	Glu	Ala	Phe	Leu	Asn	Thr	Glu	Leu	Thr		
			500					505					510				

<210> 93
<211> 594
<212> PRT
<213> Homo sapiens

<400> 93

Thr 1	Met	Leu	Arg	Thr 5	Ser	Val	Leu	Arg	Leu 10	Leu	Gly	Arg	Thr	Gly 15	Ala
Ser	Arg	Leu	Ser 20	Leu	Leu	Glu	Asp	Phe 25	Gly	Pro	Arg	Tyr	Tyr 30	Ser	Ser
Gly	Ser	Leu	Ser	Ala	Gly	Asp	Asp	Ala	Cys	Asp	Val	Arg	Ala	Tyr	Phe

		35				40					45				
Thr	Thr	Pro	Ile	Phe	Tyr	Val	Asn	Ala	Ala	Pro	His	Ile	Gly	His	Leu
50						55					60				
Tyr	Ser	Ala	Leu	Leu	Ala	Asp	Ala	Leu	Cys	Arg	His	Arg	Arg	Leu	Arg
65					70					75					80
Gly	Pro	Ser	Thr	Ala	Ala	Thr	Arg	Phe	Ser	Thr	Gly	Thr	Asp	Glu	His
				85					90					95	
Gly	Leu	Lys	Ile	Gln	Gln	Ala	Ala	Ala	Thr	Ala	Gly	Leu	Ala	Pro	Thr
			100					105					110		
Glu	Leu	Cys	Asp	Arg	Val	Ser	Glu	Gln	Phe	Gln	Gln	Leu	Phe	Gln	Glu
		115					120					125			
Ala	Gly	Ile	Ser	Cys	Thr	Asp	Phe	Ile	Arg	Thr	Thr	Glu	Ala	Arg	His
	130					135					140				
Arg	Val	Ala	Val	Gln	His	Phe	Trp	Gly	Val	Leu	Lys	Ser	Arg	Gly	Leu
145					150					155					160
Leu	Tyr	Lys	Gly	Val	Tyr	Glu	Gly	Trp	Tyr	Cys	Ala	Ser	Asp	Glu	Cys
			165						170					175	
Phe	Leu	Pro	Glu	Ala	Lys	Val	Thr	Gln	Gln	Pro	Gly	Pro	Ser	Gly	Asp
			180					185					190		
Ser	Phe	Pro	Val	Ser	Leu	Glu	Ser	Gly	His	Pro	Val	Ser	Trp	Thr	Lys
		195					200					205			
Glu	Glu	Asn	Tyr	Ile	Phe	Arg	Leu	Ser	Gln	Phe	Arg	Lys	Pro	Leu	Gln
	210					215					220				
Arg	Trp	Leu	Arg	Gly	Asn	Pro	Gln	Ala	Ile	Thr	Pro	Glu	Pro	Phe	His
225					230					235					240
His	Val	Val	Leu	Gln	Trp	Leu	Asp	Glu	Glu	Leu	Pro	Asp	Leu	Ser	Val
			245						250					255	
Ser	Arg	Arg	Ser	Ser	His	Leu	His	Trp	Gly	Ile	Pro	Val	Pro	Gly	Asp
			260					265					270		
Asp	Ser	Gln	Thr	Ile	Tyr	Val	Trp	Leu	Asp	Ala	Leu	Val	Asn	Tyr	Leu
		275					280					285			
Thr	Val	Ile	Gly	Tyr	Pro	Asn	Ala	Glu	Phe	Lys	Ser	Trp	Trp	Pro	Ala
	290					295					300				
Thr	Ser	His	Ile	Ile	Gly	Lys	Asp	Ile	Leu	Lys	Phe	His	Ala	Ile	Tyr
305					310					315					320
Trp	Pro	Ala	Phe	Leu	Gly	Ala	Gly	Met	Ser	Pro	Pro	Gln	Arg	Ile	
			325					330					335		
Cys	Val	His	Ser	His	Trp	Thr	Val	Cys	Gly	Gln	Lys	Met	Ser	Lys	Ser
			340					345					350		
Leu	Gly	Asn	Val	Val	Asp	Pro	Arg	Thr	Cys	Leu	Asn	Arg	Tyr	Thr	Val
		355					360					365			
Asp	Gly	Phe	Arg	Tyr	Phe	Leu	Leu	Arg	Gln	Gly	Val	Pro	Asn	Trp	Asp
	370					375					380				
Cys	Asp	Tyr	Tyr	Asp	Glu	L									

```

          500          505          510
Val Ala Leu Glu Cys Leu Arg Val Phe Gly Thr Leu Leu Gln Pro Val
      515          520          525
Thr Pro Ser Leu Ala Asp Lys Leu Leu Ser Arg Leu Gly Val Ser Ala
      530          535          540
Ser Glu Arg Ser Leu Gly Glu Leu Tyr Phe Leu Pro Arg Phe Tyr Gly
545          550          555          560
His Pro Cys Pro Phe Glu Gly Arg Arg Leu Gly Pro Glu Thr Gly Leu
      565          570          575
Leu Phe Pro Arg Leu Asp Gln Ser Arg Thr Trp Leu Val Lys Ala His
      580          585          590
Arg Thr

```

<210> 94
 <211> 562
 <212> PRT
 <213> *Xenopus laevis*
 <400> 94

```

Met Leu Arg Ser Leu Ala Leu Arg Thr Phe Ala Asn Ile Leu Gly Leu
 1      5      10      15
Ser Ser Arg Ser Gly Ser Thr Ala Ala Met Ser Arg Asn Ala Val Ala
      20      25      30
Ser Arg Pro His Leu Tyr Thr Thr Pro Ile Phe Tyr Val Asn Ala Ala
      35      40      45
Pro His Leu Gly His Val Tyr Ser Ala Leu Leu Ala Asp Val Gln His
 50      55      60
Arg Tyr Ser Ala Met Cys Gly Ile Glu Ser Lys Leu Ser Thr Gly Thr
65      70      75      80
Asp Glu His Gly Met Lys Val Gln Gln Ala Ala Ser Ala Leu Gly Leu
      85      90      95
Asp Pro Gln Thr Phe Cys Ser Thr Val Ser Leu Gln Phe Arg Thr Ile
      100      105      110
Phe Asp Ala Leu Asp Ile Ser Tyr Thr Asp Phe Val Arg Thr Thr Glu
      115      120      125
Pro Arg His Ile Glu Ala Val Ser Arg Phe Trp Met Thr Leu Glu Glu
      130      135      140
Gln Gly Tyr Ile Tyr Lys Gly Thr Tyr Glu Gly Trp Tyr Cys Thr Ser
145      150      155      160
Asp Glu Ala Phe Leu Ser Glu Gly Gln Thr Ala Glu His Thr Asp Phe
      165      170      175
Glu Gly Asn Lys Ile Arg Val Ser Leu Glu Ser Gly His Gln Val His
      180      185      190
Trp Val Ser Glu Glu Asn Tyr Met Phe Arg Leu Ser Ser Leu Arg Pro
      195      200      205
Ala Leu Leu Asn Trp Leu Gln Thr Glu Pro Val His Pro Ala Pro Phe
      210      215      220
Leu Lys Leu Val His His Trp Leu Glu Glu Glu Leu Pro Asp Leu Ser
225      230      235      240
Val Ser Arg Gln Arg Ser Arg Leu Ser Trp Gly Ile Pro Val Pro Ser
      245      250      255
Asp Ser Ser His Val Ile Tyr Val Trp Leu Asp Ala Leu Val Asn Tyr
      260      265      270
Leu Thr Ala Ala Gly Tyr Pro Asn Pro Gln Leu Ala Pro Trp Gly Pro
      275      280      285
Ser Thr His Leu Leu Gly Lys Asp Ile Leu Arg Phe His Ala Ile Tyr

```

```

      290      295      300
Trp Pro Ala Phe Leu Ile Ala Ala Gly Leu Pro Pro Gln Lys Leu
305      310      315      320
Leu Val His Ser His Trp Thr Ser Glu Gly Thr Lys Met Ser Lys Ser
      325      330      335
Leu Lys Asn Val Val Asp Pro Ser Asp Cys Ile Arg Arg Tyr Thr Thr
      340      345      350
Asp Gly Leu Arg Tyr Tyr Leu Leu Arg His Gly Ala Pro Glu Arg Asp
      355      360      365
Cys Asp Phe Thr His Arg Thr Ala Arg Met Leu Leu Asn Ser Glu Leu
      370      375      380
Ala Asp Ala Leu Gly Gly Leu Leu Asn Arg Cys Thr Ala Pro Ala Ile
385      390      395      400
Asn Pro Met Gln His Phe Pro Lys Phe Gln Tyr Glu Asn Phe Pro Val
      405      410      415
Ala Ser Arg Asp Gln Val His Asp Leu Leu Gly Ala Leu Gln Glu Leu
      420      425      430
Pro Val Glu Val Asp Gln Trp Ile Lys Lys Phe Gln Val His Lys Ala
      435      440      445
Leu Glu Cys Ile Asp Ala Cys Val Arg Arg Ser Asn Ala Phe Phe Gln
      450      455      460
Ser Gln Ala Pro Trp Lys Leu Gln Arg Gly Val Glu Lys Glu Ala Ala
465      470      475      480
Leu Arg Asp Ser Val Ile Tyr Leu Thr Leu Glu Ala Leu Arg Leu Tyr
      485      490      495
Ala Thr Leu Leu His Pro Ala Val Pro Gly Leu Ala Thr Val Val Leu
      500      505      510
Asp Arg Leu Gly Val Pro His Lys Met Arg Thr Leu Lys Lys Asn Thr
      515      520      525
Phe Leu Ala Ala Thr Arg Gly Glu Ile Cys Tyr Phe Gln Ala Gln Thr
      530      535      540
Leu Gly Pro Asp Lys Gly Leu Leu Phe Pro Arg Leu Glu Lys Ser Glu
545      550      555      560
Ala Phe

```

<210> 95

<211> 539

<212> PRT

<213> Schizosaccharomyces pombe

<400> 95

```

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

```

115				120				125							
Gln	Lys	Ala	Gly	Met	Ile	Ser	Phe	Glu	Arg	His	Glu	Gly	Trp	Tyr	Cys
130				135				140							
Val	Ser	Asp	Glu	Thr	Phe	Tyr	Pro	Glu	Ser	Ala	Ile	Gln	Lys	Val	Val
145				150				155							
Asp	Pro	Ala	Thr	Lys	Gln	Glu	Lys	Arg	Val	Ser	Met	Glu	Thr	Gly	Lys
165				170				175							
Glu	Val	Gln	Trp	Ser	Ser	Glu	Met	Asn	Tyr	His	Phe	Leu	Leu	Ser	Lys
180				185				190							
Phe	Gln	Ser	Arg	Leu	Ile	Glu	His	Tyr	Asn	Lys	Asn	Pro	Asn	Phe	Val
195				200				205							
Gln	Pro	Ser	Ile	Phe	His	Thr	Gln	Val	Leu	Glu	Glu	Leu	Lys	Thr	Gly
210				215				220							
Ile	Ser	Asp	Leu	Ser	Ile	Ser	Arg	Pro	Lys	Gln	Arg	Leu	Ser	Trp	Gly
225				230				235							
Ile	Pro	Val	Pro	Gly	Asn	Ser	Gln	Gln	Thr	Ile	Tyr	Val	Trp	Leu	Asp
245				250				255							
Ala	Leu	Ile	Asn	Tyr	Ile	Ser	Val	Ile	Gly	Tyr	Pro	Trp	Leu	Asn	Glu
260				265				270							
Lys	Ser	Ser	Leu	Ser	Ala	Gly	Trp	Pro	Ala	Asn	Met	His	Val	Ile	Gly
275				280				285							
Lys	Asp	Ile	Ile	Arg	Phe	His	Cys	Ile	Tyr	Trp	Pro	Ala	Phe	Leu	Met
290				295				300							
Ala	Ala	Gly	Leu	Pro	Leu	Pro	Glu	Lys	Ile	Leu	Val	His	Ser	His	Trp
305				310				315							
Thr	Met	Asn	Lys	Val	Lys	Met	Ser	Lys	Ser	Leu	Gly	Asn	Val	Val	Asp
325				330				335							
Pro	Phe	Trp	Leu	Ile	Glu	Lys	Tyr	Gly	Val	Asp	Thr	Ile	Arg	Tyr	Tyr
340				345				350							
Leu	Leu	Lys	Arg	Gly	Arg	Leu	Thr	Ser	Asp	Ser	Asn	Phe	Asp	Ile	Glu
355				360				365							
Glu	Leu	Glu	Lys	Asp	Glu	Glu	His	Asp	Leu	Arg	Arg	Ser	Leu	Gly	Val
370				375				380							
Leu	Leu	Ser	Arg	Leu	Gln	Ser	Lys	Lys	Leu	Phe	Ile	Ser	Asn	Glu	Ile
385				390				395							
Gln	Lys	Gln	Trp	His	Lys	Lys	Asp	Asp	Phe	Thr	Glu	Tyr	Glu	Asp	Ile
405				410				415							
Val	His	Glu	Leu	Ile	Glu	Leu	Pro	Val	Val	Cys	Ala	Gln	Ser	Ile	Asp
420				425				430							
Gly	Gly	Cys	Val	Tyr	Glu	Val	Ile	Asn	Leu	Val	Gln	Ser	Val	Leu	Arg
435				440				445							
Arg	Val	Thr	Lys	Leu	Phe	Gln	Leu	Lys	Glu	Pro	Trp	Lys	Leu	Ser	Asp
450				455				460							
Asp	Ser	Gln	Glu	Lys	Ile	Asp	Thr	Leu	Met	Leu	Val	Ala	His	Ser	Leu
465				470				475							
Arg	Ile	Ser	Gly	Ile	Leu	Leu	Gln	Pro	Ile	Met	Pro	Thr	Lys	Ser	Thr
485				490				495							
Glu	Leu	Leu	Asp	Gln	Leu	Gly	Ile	Pro	Lys	Asn	Gln	Arg	Ser	Leu	Gln
500				505				510							
Asn	Ala	Thr	Asn	Val	Phe	Glu	Pro	Thr	Glu	Phe	Thr	Phe	His	Ser	Gly
515				520				525							
Asn	Asn	Ser	His	Leu	Phe	Asp	Lys	Arg	Thr	Gln					
530				535											

<210> 96
 <211> 913
 <212> PRT
 <213> Danio rerio

<400> 96

Met	Lys	Leu	Phe	Ile	Gly	Glu	Gly	Asn	Pro	His	Cys	Leu	Lys	Val	Leu
1				5				10						15	
Ala	Ala	Leu	Glu	Leu	Ser	Gly	Val	Arg	Cys	Glu	Thr	Gln	Leu	Val	Lys
		20						25					30		
His	Glu	Glu	Lys	Val	Val	Pro	Tyr	Leu	Thr	His	Pro	Val	Leu	Pro	Ile
		35					40					45			
Leu	Gln	Leu	Pro	Ser	Gly	Gln	His	Leu	Phe	Ser	Pro	Asn	Ser	Ile	Cys
	50				55						60				
Gln	Tyr	Leu	Phe	Asp	Ile	Ser	Gly	Gln	Lys	Ala	Thr	Asp	Ala	Thr	Asn
65				70						75					80
Gln	Trp	Leu	Glu	Trp	Glu	Ala	Thr	Asn	Leu	Gln	Pro	Ala	Val	Leu	Gln
			85					90						95	
Ser	Leu	Gln	Leu	Val	Ala	Leu	Gln	Gly	Lys	Arg	Val	Glu	Ser	Ala	Ala
		100						105					110		
Val	Met	Lys	Glu	Pro	Leu	Ser	Trp	Leu	Glu	Gln	Ser	Leu	Ser	Lys	Arg
		115					120					125			
Lys	Thr	Ser	Phe	Leu	Thr	Asp	Glu	Val	Val	Ser	Val	Ala	Asp	Val	Val
	130					135					140				
Leu	Trp	Ala	Ala	Leu	Tyr	Pro	Leu	Leu	Ser	Asp	Ser	Ala	Phe	Glu	Pro
145				150						155					160
Gly	Asp	Leu	Gln	Ala	Val	Arg	Gly	Trp	Phe	Glu	Arg	Val	Cys	Ser	Val
			165					170						175	
Ser	Ala	Cys	Gln	Ser	Ala	Ala	Leu	Arg	Val	Leu	Gln	Gly	Lys	Gly	Ala
		180						185					190		
Glu	Ala	Leu	Lys	Ser	Phe	Leu	Gln	Lys	Gln	Pro	Val	Thr	His	Thr	Pro
		195					200					205			
Arg	Arg	Asp	Ser	Pro	Ser	Asn	Ser	Thr	Glu	Ala	Glu	Asp	Ser	Glu	Arg
	210					215					220				
Glu	Leu	Ser	Pro	Glu	Glu	Ile	Glu	Glu	Ala	Ala	Gln	Val	Tyr	Ser	Glu
225				230						235					240
Gly	Leu	Lys	Asp	Phe	Thr	Val	Val	Thr	Glu	Arg	Lys	His	Pro	Val	Leu
			245					250						255	
Pro	Gln	Glu	Gly	Lys	Arg	Asn	Val	Leu	Ile	Thr	Ser	Ala	Leu	Pro	Tyr
		260						265					270		
Val	Asn	Asn	Val	Pro	His	Leu	Gly	Asn	Ile	Ile	Gly	Cys	Val	Leu	Ser
	275						280					285			
Ala	Asp	Val	Phe	Ala	Arg	Tyr	Gly	Arg	Leu	Arg	Gly	Trp	Asn	Leu	Leu
	290					295					300				
Tyr	Ile	Cys	Gly	Thr	Asp	Glu	Tyr	Gly	Thr	Ala	Thr	Glu	Asn	Lys	Ala
305				310						315					320
Arg	Glu	Glu	Gly	Leu	Thr	Pro	Gln	Gln	Ile	Cys	Asp	Lys	Tyr	His	Cys
			325					330						335	
Ile	His	Ala	Ser	Ile	Tyr	Gln	Trp	Phe	Gln	Ile	Asp	Phe	Asp	Phe	Phe
		340						345					350		
Gly	Arg	Thr	Thr	Gln	His	Gln	Thr	Glu	Ile	Ala	Gln	Asp	Ile	Phe	
	355					360					365				
Trp	Arg	Leu	His	Glu	Arg	Gly	Phe	Leu	Leu	Glu	Asp	Thr	Val	Glu	Gln
	370					375					380				
Leu	Arg	Cys	Glu	Gly	Cys	Gln	Arg	Phe	Leu	Ala	Asp	Arg	Phe	Val	Glu
385				390						395					400
Gly	Glu	Cys	Pro	His	Cys	Arg	Tyr	Pro	Glu	Ala	Arg	Gly	Asp	Gln	Cys
			405					410						415	
Asp	Lys	Cys	Gly	Arg	Leu	Ile	Asn	Ala	Val	Glu	Leu	Lys	Asn	Pro	Gln
			420					425					430		
Cys	Lys	Val	Cys	Lys	Glu	Thr	Pro	Val	Ile	Arg	Ser	Ser	Lys	His	Leu

			435					440					445		
Phe	Leu	Asn	Leu	Pro	Lys	Leu	Glu	Gln	Asp	Leu	Glu	Gln	Trp	Leu	Gln
	450					455					460				
Thr	Ser	Thr	Ala	Ala	Gly	Asp	Trp	Thr	Thr	Asn	Ala	Arg	His	Ile	Thr
465					470					475					480
Arg	Ser	Trp	Leu	Arg	Asp	Gly	Leu	Lys	Pro	Arg	Cys	Ile	Thr	Arg	Asp
				485					490						495
Leu	Lys	Trp	Gly	Thr	Pro	Val	Pro	His	Pro	Asp	Tyr	Lys	Glu	Lys	Val
			500					505					510		
Phe	Tyr	Val	Trp	Phe	Asp	Ala	Pro	Ile	Gly	Tyr	Leu	Ser	Ile	Thr	Ala
		515					520					525			
Asn	Tyr	Thr	Asp	Gln	Trp	Glu	Arg	Trp	Trp	Lys	Asn	Pro	Gln	Gln	Val
	530					535					540				
Glu	Leu	Tyr	Asn	Phe	Met	Ala	Lys	Asp	Asn	Val	Pro	Phe	His	Ser	Val
545					550					555					560
Val	Phe	Pro	Cys	Ser	Leu	Leu	Gly	Ala	Gln	Asp	Asn	Tyr	Thr	Leu	Val
				565					570						575
Asn	Asn	Leu	Ile	Ala	Thr	Glu	Tyr	Leu	Asn	Tyr	Glu	Asp	Thr	Lys	Phe
			580					585					590		
Ser	Lys	Ser	Arg	Gly	Val	Gly	Val	Phe	Gly	Asp	Met	Ala	Lys	Asp	Thr
		595					600					605			
Gly	Ile	Pro	Ser	Asp	Val	Trp	Arg	Phe	Tyr	Leu	Leu	Tyr	Leu	Arg	Pro
	610					615					620				
Glu	Gly	Gln	Asp	Ser	Ala	Phe	Ser	Trp	Thr	Asp	Met	Ala	Leu	Lys	Asn
625					630					635					640
Asn	Ser	Glu	Leu	Leu	Asn	Asn	Leu	Gly	Asn	Phe	Ile	Asn	Arg	Ala	Gly
				645					650						655
Met	Phe	Val	Ser	Lys	Phe	Phe	Glu	Gly	Cys	Val	Pro	Glu	Met	Leu	Leu
			660					665					670		
Asn	Asp	Asp	Asp	Lys	Arg	Leu	Ile	Ala	Gln	Val	Cys	Trp	Glu	Leu	Lys
		675					680					685			
Gln	Tyr	Ile	Gln	Leu	Leu	Asp	Lys	Val	Arg	Ile	Arg	Asp	Ala	Leu	Lys
	690					695					700				
Cys	Ile	Leu	Asn	Met	Ser	Arg	His	Gly	Asn	Gln	Tyr	Ile	Gln	Leu	Asn
705					710					715					720
Glu	Pro	Trp	Lys	Lys	Ile	Lys	Gly	Gly	Ala	Glu	Asp	Arg	Cys	Arg	Ala
				725					730						735
Gly	Thr	Val	Thr	Gly	Val	Ser	Val	Asn	Val	Ala	Cys	Leu	Leu	Ala	Ala
			740					745					750		
Val	Leu	Glu	Pro	Phe	Met	Pro	Thr	Val	Ser	Leu	Ser	Ile	Arg	Ser	Gln
		755					760					765			
Leu	Gln	Ala	Pro	Glu	Ser	Ser	Ser	Arg	Ala	Met	Leu	Ser	Gly	Pro	Gly
	770					775					780				
Ala	Phe	Ile	Cys	Thr	Leu	Pro	Ala	Gly							

	900	905	910
Lys			

<210> 97
 <211> 650
 <212> PRT
 <213> Helicobacter pylori

<400> 97

```

Met Gln Lys Ser Leu Ile Thr Thr Pro Ile Tyr Tyr Val Asn Asp Ile
 1          5          10          15
Pro His Ile Gly His Ala Tyr Thr Thr Leu Ile Ala Asp Thr Leu Lys
 20          25          30
Lys Tyr Tyr Thr Leu Gln Gly Glu Val Phe Phe Leu Thr Gly Thr
 35          40          45
Asp Glu His Gly Gln Lys Ile Glu Gln Ser Ala Arg Leu Arg Asn Gln
 50          55          60
Ser Pro Lys Ala Tyr Ala Asp Ser Ile Ser Thr Ile Phe Lys Asp Gln
 65          70          75          80
Trp Asp Phe Phe Asn Leu Asp Tyr Asp Gly Phe Ile Arg Thr Thr Asp
 85          90          95
Ser Glu His Gln Lys Cys Val Gln Asn Ala Phe Glu Ile Met Phe Glu
100          105          110
Lys Gly Asp Ile Tyr Lys Gly Ala Tyr Ser Gly Tyr Tyr Cys Val Ser
115          120          125
Cys Glu Ser Tyr Cys Ala Ile Ser Lys Ala Asp Asn Thr Asn Asp Lys
130          135          140
Val Leu Cys Pro Asp Cys Leu Arg Glu Thr Thr Leu Leu Glu Glu Glu
145          150          155          160
Ser Tyr Phe Phe Arg Leu Ser Ala Tyr Glu Lys Pro Leu Leu Asp Phe
165          170          175
Tyr Ala Lys Asn Pro Glu Ala Ile Leu Pro Val Tyr Arg Lys Asn Glu
180          185          190
Val Thr Ser Phe Ile Glu Gln Gly Leu Leu Asp Leu Ser Ile Thr Arg
195          200          205
Thr Ser Phe Glu Trp Gly Ile Pro Leu Pro Lys Lys Met Asn Asp Pro
210          215          220
Lys His Val Val Tyr Val Trp Leu Asp Ala Leu Leu Asn Tyr Ala Ser
225          230          235          240
Ala Leu Gly Tyr Leu Asn Asp Leu Asp Asn Lys Met Ala His Phe Glu
245          250          255
Cys Ala Arg His Ile Val Gly Lys Asp Ile Leu Arg Phe His Ala Ile
260          265          270
Tyr Trp Pro Ala Phe Leu Met Ser Leu Asn Leu Pro Leu Phe Lys Gln
275          280          285
Leu Cys Val His Gly Trp Trp Thr Ile Glu Gly Val Lys Met Ser Lys
290          295          300
Ser Leu Gly Asn Val Leu Asp Ala Gln Lys Ile Ala Met Glu Tyr Gly
305          310          315          320
Ile Glu Glu Leu Arg Tyr Phe Leu Leu Arg Glu Val Pro Phe Gly Gln
325          330          335
Asp Gly Asp Phe Ser Lys Lys Ala Leu Ile Glu Arg Ile Asn Ala Asn
340          345          350
Leu Asn Asn Asp Leu Gly Asn Leu Leu Asn Arg Leu Leu Gly Met Ala
355          360          365
Lys Lys Tyr Phe Asn His Ser Leu Lys Ser Thr Lys Ile Thr Ala Tyr
    
```

```

      370      375      380
Tyr Ser Lys Glu Leu Glu Lys Val His Gln Ile Leu Asp Asn Ala Asn
385      390      395      400
Ser Phe Val Pro Lys Met Gln Leu His Lys Ala Leu Glu Glu Leu Phe
      405      410      415
Asn Val Tyr Asp Phe Leu Asn Lys Leu Ile Ala Lys Glu Glu Pro Trp
      420      425      430
Val Leu His Lys Asn Asn Glu Ser Glu Lys Leu Glu Ala Leu Leu Ser
      435      440      445
Leu Ile Ala Asn Ala Leu Leu Gln Ser Ser Phe Leu Leu Tyr Ala Phe
      450      455      460
Met Pro Lys Ser Ala Val Lys Leu Ala Asn Ala Phe Asn Thr Glu Ile
465      470      475      480
Thr Pro Asp Asn Tyr Glu Arg Phe Phe Lys Ala Lys Lys Leu Gln Asp
      485      490      495
Met Ile Leu Gln Asp Thr Glu Pro Leu Phe Ser Lys Met Glu Lys Ile
      500      505      510
Glu Lys Thr Glu Lys Ala Gly Glu Ala Ser Pro Glu Lys Asn Glu Lys
      515      520      525
Glu Lys Lys Asp Ala Lys Glu Lys Ala Pro Leu Lys Gln Glu Asn Tyr
530      535      540
Ile Gly Ile Glu Asp Phe Lys Lys Val Glu Ile Lys Val Gly Leu Ile
545      550      555      560
Lys Glu Ala Gln Arg Ile Glu Lys Ser Asn Lys Leu Leu Arg Leu Lys
      565      570      575
Val Asp Leu Gly Glu Gly Arg Leu Arg Gln Ile Ile Ser Gly Ile Ala
      580      585      590
Leu Asp Tyr Glu Pro Glu Ser Leu Val Gly Gln Met Val Cys Val Val
      595      600      605
Ala Asn Leu Lys Pro Ala Lys Leu Met Gly Glu Met Ser Glu Gly Met
610      615      620
Ile Leu Ala Val Arg Asp Ser Asp Asn Leu Ala Leu Ile Ser Pro Thr
625      630      635      640
Arg Glu Lys Ile Ala Gly Ser Leu Ile Ser
      645      650

```

<210> 98
 <211> 582
 <212> PRT
 <213> *Drosophila melanogaster*

 <400> 98

```

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

```

115	120	125
His Thr Arg Gly His Ile Tyr Ser Ala Ala Tyr Ser Gly Trp Tyr Cys		
130	135	140
Val Ser Asp Glu Thr Phe Leu Thr Asp Ser Gln Leu Arg Leu Asp Glu		
145	150	155
Ala Thr Gly Thr Arg Tyr Ser Leu Glu Ser Gly His Pro Val Glu Trp		
165	170	175
Thr Glu Glu Thr Asn Tyr Met Phe Arg Leu Ser Gln Phe Gln Asp Asp		
180	185	190
Val Arg His Trp Val Lys Thr Glu Ala Arg Val Arg Pro Ala Lys Phe		
195	200	205
Glu Lys Ile Leu Leu Asp Thr Leu Ser Glu Pro Leu Pro Asp Val Ser		
210	215	220
Val Ser Arg Pro Ser Asn Arg Val His Trp Ala Ile Pro Val Pro Asp		
225	230	235
Asp Asp Ser Gln Thr Val Tyr Val Trp Leu Asp Ala Leu Val Asn Tyr		
245	250	255
Leu Ser Ser Val Gly Tyr Pro Asp Glu Lys Phe Ser Ala His Trp Pro		
260	265	270
Pro Ala Gln Gln Val Ile Gly Lys Asp Ile Leu Lys Phe His Gly Ile		
275	280	285
Tyr Trp Thr Ala Phe Leu Leu Ala Ala Gly Leu Glu Pro Pro Gly Gln		
290	295	300
Leu Tyr Val His Ser His Trp Thr Val Asp Gly Gln Lys Met Ser Lys		
305	310	315
Ser Lys His Asn Val Val Asp Pro Leu Gln Ala Ala Gln Gln Tyr Thr		
325	330	335
Met Glu Gly Leu Arg Tyr Phe Leu Leu Arg Glu Gly Val Ala His Ser		
340	345	350
Asp Gly Asn Tyr Ser His Val Lys Ala Gln Arg Ile Leu Asn Ser Glu		
355	360	365
Leu Ala Asp Thr Leu Gly Asn Leu Leu Ser Arg Ala Ser Ala Lys Ser		
370	375	380
Leu Asn Pro Gly Gln Ile Tyr Pro Ser Pro Ser Ala Glu His Leu Ala		
385	390	395
Asp Leu Leu Arg Ser Leu Asp Val Ala Lys Arg Leu Gln Asp Ser Leu		
405	410	415
Leu Gln Leu Ser Glu Arg Cys Glu Ser His Tyr Glu Cys Asn His Phe		
420	425	430
His Leu Val Ala Asp Thr Thr Met Ala Ala Leu His Ala Ala Asn Asn		
435	440	445
Phe Phe Glu Ser Ser Lys Pro Trp Thr Leu Lys Ala Gly Ala Pro Asp		
450	455	460
Gly Asn Gln Ala Arg Leu Glu Thr Ile Ile Ala Met Thr Met Asp Ala		
465	470	475
Leu Arg Leu Ser Gly Ile Val Leu Gln Pro Ile Ile Pro Gln Leu Ala		
485	490	495
Asn Arg Leu Leu Asp Lys Leu Ser Val Pro Thr Ala Gln Arg Gly Trp		
500	505	510
Asn Tyr Leu Ala Glu Ser Phe Ala Thr Ser Pro Asn Ser Ser Asn Ser		
515	520	525
Pro Ala Gly Leu Gly Glu Ser Arg Gln Leu Asp Gly Gln Thr Ser Ala		
530	535	540
Leu Leu Phe Gln Arg Ile Leu Glu Glu Thr Ser Ala Lys Glu Val Lys		
545	550	555
Glu Gln Lys Pro Gln Pro Ala Lys Arg Ser Lys Ser Lys Lys Lys Glu		
565	570	575
Arg Arg Glu Thr Met Ser		

<211> 519

<212> PRT

<213> Mycobacterium tuberculosis

<400> 99

```

Met Lys Pro Tyr Tyr Val Thr Thr Ala Ile Ala Tyr Pro Asn Ala Ala
 1          5          10          15
Pro His Val Gly His Ala Tyr Glu Tyr Ile Ala Thr Asp Ala Ile Ala
          20          25          30
Arg Phe Lys Arg Leu Asp Arg Tyr Asp Val Arg Phe Leu Thr Gly Thr
          35          40          45
Asp Glu His Gly Leu Lys Val Ala Gln Ala Ala Ala Ala Gly Val
          50          55          60
Pro Thr Ala Ala Leu Ala Arg Arg Asn Ser Asp Val Phe Gln Arg Met
65          70          75          80
Gln Glu Ala Leu Asn Ile Ser Phe Asp Arg Phe Ile Arg Thr Thr Asp
          85          90          95
Ala Asp His His Glu Ala Ser Lys Glu Leu Trp Arg Arg Met Ser Ala
          100          105          110
Ala Gly Asp Ile Tyr Leu Asp Asn Tyr Ser Gly Trp Tyr Ser Val Arg
          115          120          125
Asp Glu Arg Phe Phe Val Glu Ser Glu Thr Gln Leu Val Asp Gly Thr
          130          135          140
Arg Leu Thr Val Glu Thr Gly Thr Pro Val Thr Trp Thr Glu Glu Gln
145          150          155          160
Thr Tyr Phe Phe Arg Leu Ser Ala Tyr Thr Asp Lys Leu Leu Ala His
          165          170          175
Tyr His Ala Asn Pro Asp Phe Ile Ala Pro Glu Thr Arg Arg Asn Glu
          180          185          190
Val Ile Ser Phe Val Ser Gly Gly Leu Asp Asp Leu Ser Ile Ser Arg
          195          200          205
Thr Ser Phe Asp Trp Gly Val Gln Val Pro Glu His Pro Asp His Val
210          215          220
Met Tyr Val Trp Val Asp Ala Leu Thr Asn Tyr Leu Thr Gly Ala Gly
225          230          235          240
Phe Pro Asp Thr Asp Ser Glu Leu Phe Arg Arg Tyr Trp Pro Ala Asp
          245          250          255
Leu His Met Ile Gly Lys Asp Ile Ile Arg Phe His Ala Val Tyr Trp
          260          265          270
Pro Ala Phe Leu Met Ser Ala Gly Ile Glu Leu Pro Arg Arg Ile Phe
          275          280          285
Ala His Gly Phe Leu His Asn Arg Gly Glu Lys Met Ser Lys Ser Val
          290          295          300
Gly Asn Ile Val Asp Pro Val Ala Leu Ala Glu Ala Leu Gly Val Asp
305          310          315          320
Gln Val Arg Tyr Phe Leu Leu Arg Glu Val Pro Phe Gly Gln Asp Gly
          325          330          335
Ser Tyr Ser Asp Glu Ala Ile Val Thr Arg Ile Asn Thr Asp Leu Ala
          340          345          350
Asn Glu Leu Gly Asn Leu Ala Gln Arg Ser Leu Ser Met Val Ala Lys
          355          360          365
Asn Leu Asp Gly Arg Val Pro Asn Pro Gly Glu Phe Ala Asp Ala Asp
          370          375          380
Ala Ala Leu Leu Ala Thr Ala Asp Gly Leu Leu Glu Arg Val Arg Gly

```

```

385          390          395          400
His Phe Asp Ala Gln Ala Met His Leu Ala Leu Glu Ala Ile Trp Leu
          405          410          415
Met Leu Gly Asp Ala Asn Lys Tyr Phe Ser Val Gln Gln Pro Trp Val
          420          425          430
Leu Arg Lys Ser Glu Ser Glu Ala Asp Gln Ala Arg Phe Arg Thr Thr
          435          440          445
Leu Tyr Val Thr Cys Glu Val Val Arg Ile Ala Ala Leu Leu Ile Gln
          450          455          460
Pro Val Met Pro Glu Ser Ala Gly Lys Ile Leu Asp Leu Leu Gly Gln
          465          470          475          480
Ala Pro Asn Gln Arg Ser Phe Ala Ala Val Gly Val Arg Leu Thr Pro
          485          490          495
Gly Thr Ala Leu Pro Pro Pro Thr Gly Val Phe Pro Arg Tyr Gln Pro
          500          505          510
Pro Gln Pro Pro Glu Gly Lys
          515

```

<210> 100

<211> 575

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 100

```

Met Gln Cys Arg Ser Ile Val His Arg Leu Tyr Ser Lys Val Ser His
1          5          10          15
Val Thr Thr Pro Ile Phe Tyr Pro Asn Ala Lys Pro His Leu Gly His
          20          25          30
Leu Tyr Ser Ser Leu Leu Ser Asp Val Tyr His Arg Trp Gln Leu Phe
          35          40          45
Lys Gly Asn Leu Ser Phe Phe Thr Thr Gly Thr Asp Glu His Gly Leu
          50          55          60
Lys Ile Gln Cys Ala Ser Glu Ser Asn Gly Phe Asp Gln Pro Lys Lys
65          70          75          80
Phe Val Asp Lys Leu Tyr Pro Glu Phe Val Gln Leu Asp Lys Ile Tyr
          85          90          95
Gly Ile Asn Tyr Thr Arg Phe Ile Arg Thr Thr Asp Pro Asp His Ile
          100          105          110
Glu Asn Val Met Lys Leu Trp Glu Leu Cys Leu Lys Asn Gly Tyr Ile
          115          120          125
Tyr Met Gly Glu His Lys Gly Trp Tyr Ser Ile Ser Asp Glu Thr Phe
          130          135          140
Tyr Pro Glu Ser Lys Val Ile Lys Asp Pro Lys Asn Asp Gly Lys Tyr
145          150          155          160
Leu Asn Thr Glu Ser Lys Asn Glu Val Val Tyr Gln Ser Glu Thr Asn
          165          170          175
Tyr Phe Phe Arg Leu Ser Leu Phe Asn Lys Lys Ile Val Asp His Ile
          180          185          190
Arg Lys Asn Pro Asp Phe Ile Phe Pro Ala Ser Lys Arg Asp Gln Ile
          195          200          205
Leu Lys Glu Leu Glu Thr Gly Gly Thr Leu Pro Asp Leu Ser Ile Ser
210          215          220
Arg Pro Ser Ala Arg Leu Lys Trp Gly Ile Pro Thr Pro Asn Asp Pro
225          230          235          240
Ser Gln Lys Val Tyr Val Trp Phe Asp Ala Leu Cys Asn Tyr Leu Ser
          245          250          255
Ser Ile Gly Gly Ile Pro Ser Ile Leu Ser Asn Ala Thr Glu Val Val

```

Ser	Arg	His	Tyr	Ser	Asp	Lys	Ser	Asn	Val	Lys	Gly	Gln	Leu	Leu	Ile
		260						265							270
Pro	Tyr	Pro	Lys	Glu	Val	Gln	Arg	Asn	Thr	Ile	His	Val	Ile	Gly	Lys
		275						280				285			
Asp	Ile	Ala	Lys	Phe	His	Thr	Val	Tyr	Trp	Pro	Ser	Phe	Leu	Leu	Ala
305					310					315					320
Ala	Gly	Leu	Pro	Leu	Pro	Arg	Gln	Ile	Val	Val	His	Gly	His	Trp	Leu
				325					330					335	
Cys	Asn	Gly	Met	Lys	Met	Ser	Lys	Ser	Leu	Gly	Asn	Val	Val	Asp	Pro
			340					345					350		
Ile	Asp	Met	Ala	Arg	Tyr	Tyr	Gly	Ala	Asp	Ile	Val	Arg	Trp	Phe	Leu
		355					360					365			
Leu	Glu	Asn	Ser	Lys	Leu	Glu	Glu	Asp	Gly	Asp	Phe	Gln	Glu	Ala	Lys
		370				375					380				
Leu	Tyr	Glu	Thr	Arg	Glu	Leu	Leu	Val	Ser	Lys	Trp	Gly	Asn	Leu	Ile
385					390					395					400
Asn	Arg	Cys	Cys	Gly	Ser	Lys	Phe	Asn	Ile	Glu	Arg	Ala	Val	Met	Lys
				405					410					415	
Phe	Ser	Asp	Lys	Ala	Asn	Phe	Gln	Phe	Gln	Glu	Ile	Phe	Gln	Asn	Glu
			420					425					430		
Pro	Ile	Val	Ser	Glu	Arg	Ile	Glu	Asn	Leu	Ala	Lys	Leu	Leu	Asn	Lys
		435					440					445			
Ser	Gln	Glu	Val	Phe	Asp	Glu	Lys	Ile	Ala	Ile	Phe	Gln	Tyr	Pro	Gln
		450				455					460				
Leu	Leu	Arg	His	Val	Trp	Ser	Ile	Ile	Asn	Asp	Ala	Asn	Thr	Leu	Val
465					470					475					480
Gln	Asn	Ser	Lys	Pro	Trp	Glu	Arg	Glu	Leu	Asp	Gln	Gln	Asp	Asn	Ile
				485					490					495	
Ile	Phe	Leu	Ala	Met	Glu	Thr	Ser	Arg	Ile	Leu	Ser	Ile	Leu	Cys	Gln
			500					505				510			
Ser	Ile	Ile	Pro	Ser	Leu	Ser	Gln	Ser	Phe	Leu	Asp	Arg	Ile	Asp	Val
		515					520					525			
Ser	Lys	Glu	Lys	Arg	Thr	Ile	Asn	Tyr	Ala	Arg	Leu	Gly	Ser	Asp	Lys
		530				535					540				
Thr	Tyr	Gly	Lys	Gln	Ser	Asn	Lys	Lys	Gly	Arg	Glu	Val	Pro	Leu	Lys
545					550					555					560
Lys	Ile	Pro	Phe	Arg	Leu	Gln	Glu	Glu	Gln	Thr	Asn	Met	Arg	Ser	
				565					570					575	

<210> 101
<211> 917
<212> PRT
<213> *Caenorhabditis elegans*

<400> 101

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

										85								90								95		
Gly	Cys	Thr	Pro	Arg	Glu	Leu	Cys	Asp	Lys	Tyr	His	Ala	Ile	His	Lys													
										100			105					110										
Gly	Ile	Tyr	Glu	Trp	Phe	Gly	Ile	Asp	Phe	Ser	His	Phe	Gly	Arg	Thr													
										115			120					125										
Thr	Thr	Asp	His	Gln	Thr	Glu	Ile	Cys	Gln	Asp	Met	Phe	Leu	Lys	Leu													
										130			135					140										
His	Lys	Asn	Gly	Tyr	Thr	Ser	Ser	Gln	Ser	Val	Asp	Gln	Leu	Tyr	Cys													
										145			150					155			160							
Asn	Gln	Cys	Glu	Lys	Phe	Leu	Ala	Asp	Arg	Phe	Val	Thr	Gly	Thr	Cys													
										165			170					175										
Pro	Met	Cys	Ala	Tyr	Asp	Asp	Ala	Arg	Gly	Asp	Gln	Cys	Asp	Gly	Cys													
										180			185					190										
Gly	Lys	Leu	Ile	Asn	Ala	Val	Asp	Leu	Lys	Asp	Ala	Lys	Cys	His	Met													
										195			200					205										
Cys	Lys	Ala	Thr	Pro	Glu	Val	Lys	Gln	Ser	Thr	His	Ile	Phe	Leu	Ser													
										210			215					220										
Leu	Asp	Lys	Leu	Gln	Gln	Lys	Thr	Thr	Glu	His	Leu	Asp	Arg	Glu	Leu													
										225			230					235			240							
Ala	Lys	Glu	Asp	Asn	Arg	Trp	Ser	Ser	Asn	Ala	Val	Gly	Ile	Thr	Lys													
										245			250					255										
Ala	Trp	Met	Lys	Leu	Gly	Leu	Asp	Pro	Arg	Cys	Ile	Thr	Arg	Asp	Leu													
										260			265					270										
Lys	Trp	Gly	Thr	Ala	Val	Pro	Leu	Asp	Gly	Phe	Glu	Lys	Lys	Val	Phe													
										275			280					285										
Tyr	Val	Trp	Phe	Asp	Ala	Pro	Ile	Gly	Tyr	Leu	Ser	Ile	Thr	Lys	Cys													
										290			295					300										
Val	Leu	Gly	Asp	Asn	Trp	Thr	Lys	Trp	Trp	Lys	Asn	Pro	Glu	Asn	Val													
										305			310					315			320							
Glu	Leu	Phe	Asn	Phe	Val	Gly	Lys	Asp	Asn	Val	Ala	Phe	His	Ala	Val													
										325			330					335										
Met	Phe	Pro	Cys	Ser	Gln	Leu	Gly	Ala	Asn	Asp	Asn	Tyr	Thr	Val	Val													
										340			345					350										
Asn	Asn	Leu	Cys	Ala	Thr	Glu	Tyr	Leu	Asn	Tyr	Glu	Asp	Thr	Lys	Phe													
										355			360					365										
Ser	Lys	Ser	Arg	Gly	Thr	Gly	Ile	Phe	Gly	Asp	Ala	Ala	Gln	Gly	Thr													
										370			375					380										
Glu	Ile	Pro	Ala	Asp	Ile	Trp	Arg	Phe	Tyr	Leu	Leu	Tyr	Met	Arg	Pro													
										385			390					395			400							
Glu	Ser	Gln	Asp	Thr	Ala	Phe	Ser	Trp	Asp	Asp	Phe	Val	Leu	Lys	Val													
										405			410					415										
Asn	Ser	Glu	Leu	Leu	Asn	Asn	Leu	Gly	Asn	Phe	Ile	Asn	Arg	Ala	Leu													
										420			425					430										
Ser	Phe	Val	Ala	Asn	Ser	Phe	Gly	Gly	Val	Val	Pro	Glu	Met	Asn	Leu													
										435			440					445										
Thr	Asn	Asp	Asp	Ala	Glu	Val	Leu	Ser	Glu	Ile	His	Asn	Glu	Cys	Met													
										450			455					460										
Gln	Trp	Asp	Lys	Gln	Phe	Asp	Gly	Val	His	Leu	Lys	Asp	Ala	Val	Lys													
										465			470					475			480							
Thr	Ile	Leu	Asn	Val	Ser	Arg	Leu	Gly	Asn	Gln	Tyr	Met	Gln	Ala	Gln													
										485			490					495										
Thr	Pro	Trp	Val	Leu	Met	Lys	Lys	Asp	Glu	Glu	Gly	Lys	Lys	Arg	Ala													
										500			505					510										
Gly	Thr	Ile	Ile	Gly	Val	Ala	Ala	Asn	Ile	Ala	Tyr	His	Val	Ser	Val													
										515			520					525										
Leu	Leu	Tyr	Pro	Ile	Met	Pro	Thr	Ile	Ser	Ala	Thr	Ile	Arg	Glu	Gln													
										530			535					540										
Cys	Gly	Leu	Pro	Ala	Leu	Pro	Leu	Phe	Thr	Pro	Phe	Pro	Ile	Cys	Tyr													

```

545          550          555          560
Leu Lys Ala Gly His Lys Ile Gly Gln Pro Ser Pro Leu Phe Gln Lys
565          570          575
Leu Asp Pro Ala Gln Ile Ala Glu Phe Lys Ala Lys Phe Gly Gly Ser
580          585          590
Gln Asp Ala Gln Ser Ser Ala Pro Lys Thr Ala Glu Lys Pro Lys Gln
595          600          605
Gln Lys Lys Gln Ala Pro Thr Lys Asp Lys Lys Gly Asp Lys Lys Met
610          615          620
Ala Ser Thr Ala Ala Phe Val Glu Leu Glu Gln Gly Ala Lys Val Ile
625          630          635
Ser Gln Leu Ile Ala Gln Asn Leu Lys Lys Phe Asp Gln Ala Lys Ala
645          650          655
Leu Phe Thr Arg Asn Gln Leu Gln Arg Leu Asp Gly Glu Asn Lys Gln
660          665          670
Leu Thr Ile Asp Val Lys Thr Leu Gln His Gln Leu Ile Glu Leu Glu
675          680          685
Thr Ala Ala Gly Ile Lys Gln Val Pro Lys Pro Val Val Ser Cys Thr
690          695          700
Pro Thr Pro Thr Ser Thr Pro Ala Ser Gly Ile Ile Thr Glu Ala Pro
705          710          715
Lys Lys Glu Ala Pro Ser Thr Pro Ala Pro Ser Glu Pro Lys Lys Ala
725          730          735
Lys Glu Gln Lys Lys Gly Lys Gly Gly Ala Ala Ala Ala Pro Val Asp
740          745          750
Asp Thr Ile Asp Val Gly Arg Leu Asp Met Arg Val Gly Arg Ile Ile
755          760          765
Lys Cys Glu Lys His Pro Asp Ala Asp Ala Leu Tyr Val Glu Gln Ile
770          775          780
Asp Val Gly Glu Ser Ala Pro Arg Thr Val Val Ser Gly Leu Val Arg
785          790          795
His Val Pro Leu Asp Gln Met Gln Asn Arg Leu Val Val Val Leu Cys
805          810          815
Asn Leu Lys Pro Ala Lys Met Arg Gly Val Glu Ser Arg Ala Met Val
820          825          830
Met Cys Ala Ser Ser Pro Asp Lys Val Glu Ile Met Glu Val Pro Ala
835          840          845
Asp Ser Lys Pro Gly Thr Pro Val Val Cys Pro Pro Tyr Thr His Arg
850          855          860
Pro Asp Glu Gln Leu Asn Pro Lys Lys Lys Ile Trp Glu Thr Val Ala
865          870          875
Glu Asp Leu Lys Val Ser Ala Glu Gly Phe Ala Glu Trp Lys Gly Gln
885          890          895
Pro Leu Leu Ile Gly Ser Glu Ser Lys Met Thr Ala Pro Thr Leu Arg
900          905          910
Gly Val His Val Lys
915

```

<210> 102

<211> 545

<212> PRT

<213> Streptococcus pneumoniae

<400> 102

```

Met Ser Ile Phe Ile Gly Gly Ala Trp Pro Tyr Ala Asn Gly Ser Leu
1          5          10          15
His Ile Gly His Ala Ala Ala Leu Leu Pro Gly Asp Ile Leu Ala Arg

```

20										25					30				
Tyr	Tyr	Arg	Gln	Lys	Gly	Glu	Glu	Val	Leu	Tyr	Val	Ser	Gly	Ser	Asp				
		35					40					45							
Cys	Asn	Gly	Thr	Pro	Ile	Ser	Ile	Arg	Ala	Lys	Lys	Glu	Asn	Lys	Ser				
	50					55					60								
Val	Lys	Glu	Ile	Ala	Asp	Phe	Tyr	His	Lys	Glu	Phe	Lys	Glu	Thr	Phe				
65					70						75				80				
Glu	Lys	Leu	Gly	Phe	Thr	Tyr	Asp	Leu	Tyr	Ser	Arg	Thr	Asp	Ser	Pro				
				85					90					95					
Leu	His	His	Glu	Ile	Val	Gln	Glu	Leu	Phe	Leu	Gln	Leu	Tyr	Glu	Lys				
			100					105						110					
Lys	Phe	Leu	Tyr	Thr	Lys	Lys	Ile	Lys	Gln	Leu	Tyr	Cys	Thr	Phe	Asp				
		115					120					125							
Asn	Gln	Phe	Leu	Pro	Asp	Arg	Phe	Val	Glu	Gly	Lys	Cys	Pro	Asn	Cys				
	130					135					140								
Gly	Thr	His	Ser	Arg	Gly	Asp	Gln	Cys	Asp	Asn	Cys	Ser	Ala	Ile	Leu				
145					150						155				160				
Asp	Pro	Ile	Asp	Leu	Val	Asp	Lys	Arg	Cys	Ser	Ile	Cys	Ser	Asn	Glu				
				165					170					175					
Pro	Glu	Val	Arg	Glu	Thr	Glu	His	Phe	Tyr	Tyr	Val	Phe	Ser	Glu	Phe				
			180					185						190					
Gln	Asn	Leu	Glu	Glu	Thr	Tyr	Leu	Asn	Asp	Ala	Glu	Glu	Thr	Val	Arg				
		195					200					205							
Trp	Arg	Lys	Asn	Ala	Ile	Asn	Leu	Thr	Lys	Arg	Tyr	Leu	Arg	Glu	Gly				
	210					215					220								
Leu	Pro	Asp	Arg	Ala	Val	Thr	Arg	Asp	Leu	Pro	Asn	Gly	Ile	Pro	Val				
225					230						235				240				
Pro	Ile	Asp	Gly	Phe	Arg	Asp	Lys	Lys	Ile	Tyr	Val	Trp	Phe	Glu	Ala				
				245					250					255					
Val	Ala	Gly	Tyr	Tyr	Thr	Ala	Ser	Val	Asp	Trp	Ala	Gln	Lys	Leu	Gln				
			260					265					270						
Asn	Asn	Ile	Thr	Asp	Phe	Trp	Asn	Asn	Arg	Thr	Lys	Ser	Tyr	Tyr	Val				
		275					280					285							
His	Gly	Lys	Asp	Asn	Ile	Pro	Phe	His	Thr	Ile	Ile	Trp	Pro	Ala	Ile				
	290					295					300								
Leu	Ser	Gly	Leu	Glu	Ile	Glu	Pro	Leu	Pro	Glu	Tyr	Ile	Ile	Ser	Ser				
				310							315				320				
Glu	Tyr	Leu	Thr	Leu	Glu	Asn	Lys	Lys	Ile	Ser	Thr	Ser	Asn	Asn	Trp				
				325					330					335					
Ala	Ile	Trp	Leu	Asn	Asp	Ile	Ile	Lys	Lys	Tyr	Asp	Ala	Asp	Ser	Ile				
			340					345				350							
Arg	Tyr	Phe	Leu	Thr	Ile	Asn	Ala	Pro	Glu	Met	Lys	Asp	Ala	Asn	Phe				
		355					360												

ES 2 504 521 T3

```

              485              490              495
Glu Asp Met Phe Lys Thr Lys Leu Asn Thr Trp Asn Tyr Ile Ser Asn
              500              505              510
Leu Pro Asn Lys Leu Ser Asp Val Ser Met Leu Phe Asp Arg Ile Asp
              515              520              525
Leu Lys Lys Ile Asp Glu Glu Val Leu Glu Leu Gln Gln Thr Ser Ser
              530              535              540
Arg
545

```

<210> 103
 <211> 677
 <212> PRT
 <213> Pseudomonas aeruginosa

<400> 103

```

Met Ser Glu Pro Arg Lys Ile Leu Val Thr Ser Ala Leu Pro Tyr Ala
 1              5              10              15
Asn Gly Ser Ile His Leu Gly His Met Leu Glu Tyr Ile Gln Thr Asp
              20              25              30
Met Trp Val Arg Phe Gln Lys Met Arg Gly Asn Gln Ala Val Tyr Val
              35              40              45
Cys Ala Asp Asp Ala His Gly Ser Ala Ile Met Leu Arg Ala Glu Arg
              50              55              60
Glu Gly Ile Thr Ser Glu Gln Leu Ile Asp Ala Val Arg Ala Glu His
65              70              75              80
Met Gly Asp Phe Ala Asp Phe Leu Val Asp Phe Asp Asn Tyr His Ser
              85              90              95
Thr His Ser Glu Asn Arg Glu Leu Ser Ser Ala Ile Tyr Leu Lys
              100              105              110
Leu Arg Asp Ala Gly His Ile Asp Thr Arg Pro Val Thr Gln Tyr Phe
              115              120              125
Asp Pro Glu Lys Gln Met Phe Leu Ala Asp Arg Phe Ile Lys Gly Thr
              130              135              140
Cys Pro Lys Cys Gly Thr Ala Asp Gln Tyr Gly Asp Asn Cys Glu Ala
145              150              155              160
Cys Gly Ala Thr Tyr Ala Pro Thr Glu Leu Lys Asp Pro Lys Ser Ala
              165              170              175
Ile Ser Gly Ala Thr Pro Val Leu Lys Glu Ser Leu His Tyr Phe Phe
              180              185              190
Lys Leu Pro Asp Phe Glu Ala Met Leu Lys Gln Trp Thr Arg Ser Gly
              195              200              205
Ala Leu Gln Glu Ser Val Ala Asn Lys Leu Ala Glu Trp Leu Asp Ser
              210              215              220
Gly Leu Gln Gln Trp Asp Ile Ser Arg Asp Ala Pro Tyr Phe Gly Phe
225              230              235              240
Glu Ile Pro Asp Ala Pro Gly Lys Tyr Phe Tyr Val Trp Leu Asp Ala
              245              250              255
Pro Ile Gly Tyr Met Ala Ser Phe Lys Asn Leu Cys Ala Arg Arg Pro
              260              265              270
Glu Leu Asp Phe Asp Ala Phe Trp Gly Lys Asp Ser Gly Ala Glu Leu
              275              280              285
Tyr His Phe Ile Gly Lys Asp Ile Val Asn Phe His Ala Leu Phe Trp
              290              295              300
Pro Ala Met Leu Glu Gly Ala Gly Tyr Arg Lys Pro Thr Ala Leu Asn
305              310              315              320
Val His Gly Tyr Leu Thr Val Asn Gly Gln Lys Met Ser Lys Ser Arg

```

```

          325          330          335
Gly Thr Phe Val Lys Ala Arg Thr Tyr Leu Asp His Leu Asp Pro Glu
          340          345          350
Tyr Leu Arg Tyr Tyr Tyr Ala Ser Lys Leu Gly Arg Gly Val Glu Asp
          355          360          365
Leu Asp Leu Asn Leu Glu Asp Phe Val Gln Lys Val Asn Ser Asp Leu
          370          375          380
Val Gly Lys Val Val Asn Ile Ala Ser Arg Cys Ala Gly Phe Ile His
          385          390          395          400
Lys Gly Asn Ala Gly Val Leu Val Gly Ala Asp Pro Ala Pro Glu Leu
          405          410          415
Leu Ala Ala Phe Arg Glu Ala Ala Pro Gly Ile Ala Glu Ala Tyr Glu
          420          425          430
Ala Arg Asp Phe Asn Arg Ala Met Arg Glu Ile Met Ala Leu Ala Asp
          435          440          445
Arg Ala Asn Ala Trp Ile Ala Glu Gln Ala Pro Trp Ala Leu Ala Lys
          450          455          460
Gln Glu Gly Gln Gln Asp Lys Val Gln Ala Val Cys Gly Leu Gly Ile
          465          470          475          480
Asn Leu Phe Arg Gln Leu Val Ile Phe Leu Lys Pro Val Leu Pro Lys
          485          490          495
Leu Ala Ala Ala Glu Ala Phe Leu Asn Val Ala Pro Leu Thr Trp
          500          505          510
Ala Asp His Gln Thr Leu Leu Ala Asn His Gln Leu Asn Pro Phe Gln
          515          520          525
Pro Leu Met Thr Arg Ile Glu Pro Ala Lys Val Glu Ala Met Ile Glu
          530          535          540
Ala Ser Lys Glu Asp Leu Ala Ala Ala Ser Gln Pro Ala Gly Asn Gly
          545          550          555          560
Glu Leu Val Lys Glu Pro Ile Ala Ala Glu Ile Asp Phe Asp Ala Phe
          565          570          575
Ala Ala Val Asp Leu Arg Ile Ala Leu Ile Glu Lys Cys Glu Phe Val
          580          585          590
Glu Gly Ala Asp Lys Leu Leu Arg Leu Ser Leu Asp Ile Gly Asp Ala
          595          600          605
Lys Arg Asn Val Phe Ser Gly Ile Lys Ser Ala Tyr Pro Asp Pro Ser
          610          615          620
Ala Leu Glu Gly Arg Leu Thr Leu Tyr Val Ala Asn Leu Ala Pro Arg
          625          630          635          640
Lys Met Lys Phe Gly Val Ser Glu Gly Met Val Leu Ala Ala Gly Pro
          645          650          655
Gly Gly Glu Glu Ile Tyr Leu Leu Ser Pro Asp Ser Gly Ala Lys Pro
          660          665          670
Gly Gln Arg Val Lys
          675

```

<210> 104
 <211> 618
 <212> PRT
 <213> Thermus thermophilus HB8

<400> 104

```

Met Glu Lys Val Phe Tyr Val Thr Thr Pro Ile Tyr Tyr Val Asn Ala
  1          5          10          15
Glu Pro His Leu Gly His Ala Tyr Thr Thr Val Val Ala Asp Phe Leu
  20          25          30
Ala Arg Trp His Arg Leu Asp Gly Tyr Arg Thr Phe Phe Leu Thr Gly

```

35					40					45					
Thr	Asp	Glu	His	Gly	Glu	Thr	Val	Tyr	Arg	Ala	Ala	Gln	Ala	Ala	Gly
50					55					60					
Glu	Asp	Pro	Lys	Ala	Phe	Val	Asp	Arg	Val	Ser	Gly	Arg	Phe	Lys	Arg
65					70					75					80
Ala	Trp	Asp	Leu	Leu	Gly	Ile	Ala	Tyr	Asp	Phe	Ile	Arg	Thr	Thr	
				85					90				95		
Glu	Glu	Arg	His	Lys	Lys	Val	Val	Gln	Leu	Val	Leu	Lys	Lys	Val	Tyr
			100					105					110		
Glu	Ala	Gly	Asp	Ile	Tyr	Tyr	Gly	Glu	Tyr	Glu	Gly	Leu	Tyr	Cys	Val
			115				120					125			
Ser	Cys	Glu	Arg	Phe	Tyr	Thr	Glu	Lys	Glu	Leu	Val	Glu	Gly	Leu	Cys
130					135					140					
Pro	Ile	His	Gly	Arg	Pro	Val	Glu	Arg	Arg	Lys	Glu	Gly	Asn	Tyr	Phe
145					150					155					160
Phe	Arg	Met	Glu	Lys	Tyr	Arg	Pro	Trp	Leu	Gln	Glu	Tyr	Ile	Gln	Glu
				165					170					175	
Asn	Pro	Asp	Leu	Ile	Arg	Pro	Glu	Gly	Tyr	Arg	Asn	Glu	Val	Leu	Ala
			180					185					190		
Met	Leu	Ala	Glu	Pro	Ile	Gly	Asp	Leu	Ser	Ile	Ser	Arg	Pro	Lys	Ser
			195				200					205			
Arg	Val	Pro	Trp	Gly	Ile	Pro	Leu	Pro	Trp	Asp	Glu	Asn	His	Val	Thr
			210				215					220			
Tyr	Val	Trp	Phe	Asp	Ala	Leu	Leu	Asn	Tyr	Val	Ser	Ala	Leu	Asp	Tyr
225					230					235					240
Pro	Glu	Gly	Glu	Ala	Tyr	Arg	Thr	Phe	Trp	Pro	His	Ala	Trp	His	Leu
				245					250					255	
Ile	Gly	Lys	Asp	Ile	Leu	Lys	Pro	His	Ala	Val	Phe	Trp	Pro	Thr	Met
			260					265					270		
Leu	Lys	Ala	Ala	Gly	Ile	Pro	Met	Tyr	Arg	His	Leu	Asn	Val	Gly	Gly
			275				280					285			
Phe	Leu	Leu	Gly	Pro	Asp	Gly	Arg	Lys	Met	Ser	Lys	Thr	Leu	Gly	Asn
			290			295				300					
Val	Val	Asp	Pro	Phe	Ala	Leu	Leu	Glu	Lys	Tyr	Gly	Arg	Asp	Ala	Leu
305					310					315					320
Arg	Tyr	Tyr	Leu	Leu	Arg	Glu	Ile	Pro	Tyr	Gly	Gln	Asp	Thr	Pro	Val
				325					330					335	
Ser	Glu	Glu	Ala	Leu	Arg	Thr	Arg	Tyr	Glu	Ala	Asp	Leu	Ala	Asp	Asp
			340					345					350		
Leu	Gly	Asn	Leu	Val	Gln	Arg	Thr	Arg	Ala	Met	Leu	Phe	Arg	Phe	Ala
			355				360					365			
Glu	Gly	Arg	Ile	Pro	Glu	Pro	Val	Ala	Gly	Glu	Glu	Leu	Ala	Glu	Gly
</															

		500					505			510			
Trp	Ile	Gly	Ile	Glu	Asp	Phe	Ala	Lys	Val	Glu	Leu	Arg	Val
		515					520			525			
Val	Leu	Ala	Ala	Glu	Lys	His	Pro	Asn	Ala	Asp	Arg	Leu	Leu
		530				535				540			
Arg	Leu	Ser	Leu	Gly	Asn	Glu	Glu	Arg	Thr	Val	Val	Ser	Gly
					550					555			560
Lys	Trp	Tyr	Arg	Pro	Glu	Glu	Leu	Val	Gly	Lys	Lys	Val	Val
				565					570				575
Ala	Asn	Leu	Lys	Pro	Ala	Lys	Leu	Arg	Gly	Ile	Glu	Ser	Gln
		580						585				590	
Ile	Leu	Ala	Ala	Gln	Glu	Gly	Glu	Ala	Leu	Ala	Leu	Val	Thr
		595					600					605	
Gly	Glu	Val	Pro	Pro	Gly	Ala	Val	Val	Lys				
		610				615							

REIVINDICACIONES

- 1.- Una composición, que comprende un primer vector que contiene un polinucleótido que codifica una fenilalanina ARNt sintetasa (PheRS) de levadura modificada, en donde dicho polinucleótido está mutado en los codones que codifican el aminoácido en las posiciones de la secuencia seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3, y en donde dicha sintetasa modificada es capaz de cargar una molécula de ARNt con un aminoácido no natural.
- 2.- La composición de la reivindicación 1, que comprende, además, un segundo vector que contiene un polinucleótido que codifica una molécula de ARNt modificada.
- 3.- La composición de la reivindicación 2, en donde dichos primer y segundo vectores son el mismo vector.
- 4.- La composición de la reivindicación 2, en donde dichos primer y segundo vectores son vectores diferentes.
- 5.- La composición de una cualquiera de las reivindicaciones 2 a 4, en donde dicho ARNt está modificado de manera que contiene un anticodón mutado que forma pares de bases con un correspondiente codón degenerado por bamboleo con una afinidad mayor que la afinidad del ARNt natural.
- 6.- Un polipéptido que comprende una fenilalanina ARNt sintetasa (PheRS) de levadura modificada, en donde dicha sintetasa modificada está mutada en las posiciones de las secuencias de aminoácidos seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3, y en donde dicha sintetasa modificada es capaz de cargar una molécula de ARNt con un aminoácido no natural.
- 7.- Un sistema de traducción que comprende un polinucleótido que codifica una fenilalanina ARNt sintetasa (PheRS) de levadura modificada, en donde dicho polinucleótido de sintetasa modificada está mutado en uno o más codones que codifican el aminoácido en las posiciones de las secuencias seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3, y en donde dicha sintetasa modificada es capaz de cargar una molécula de ARNt con un aminoácido no natural.
- 8.- El sistema de traducción de la reivindicación 7, en donde dicho sistema comprende una célula huésped.
- 9.- El sistema de traducción de la reivindicación 7 o la reivindicación 8, que comprende, además, un polinucleótido que codifica una molécula de ARNt modificada.
- 10.- El sistema de traducción de una cualquiera de las reivindicaciones 7 a 9, en donde dicho ARNt está modificado de manera que contiene un anticodón mutado que forma pares de bases con un correspondiente codón degenerado por bamboleo con una afinidad mayor que la afinidad del ARNt natural.

11.- El sistema de traducción de la reivindicación 10, en donde el ARNt modificado es ARNt^{Phe} equipado con el anticodón AAA.

5 12.- El sistema de traducción de la reivindicación 8, en donde dicha PheRS modificada y/o dicha molécula de ARNt modificada se deriva de un organismo diferente de la célula huésped.

13.- El sistema de traducción de la reivindicación 12, en donde dicha molécula de ARNt modificada se deriva de una célula eucariótica, y la célula huésped es una célula procariótica.

10

14.- El sistema de traducción de la reivindicación 8, en donde la célula huésped es un auxótrofo.

15.- El sistema de traducción de una cualquiera de las reivindicaciones 7 a 14, que comprende un medio de cultivo que contiene uno o más aminoácidos no naturales.

15

16.- Un método para incorporar al menos un aminoácido no natural en un polipéptido diana en una o más ubicaciones especificadas, comprendiendo el método proporcionar un sistema de traducción que contiene al menos un aminoácido no natural; proporcionar al sistema de traducción una o más fenilalanina ARNt sintetasa (PheRS) de levadura modificada, en donde dicha sintetasa modificada está mutada en las posiciones de la secuencia de aminoácidos seleccionadas del grupo que consiste en los números de posición de la secuencia de aminoácidos (i) 412 y 415; (ii) 415 y 418; (iii) 415 y 437; (iv) 412, 415 y 437; (v) 415, 418 y 437; (vi) 412, 415 y 418; y (vii) 412, 415, 418 y 437 del polipéptido codificado por SEQ ID NO: 3, y en donde dicha sintetasa modificada es capaz de cargar una molécula de ARNt con un aminoácido no natural, proporcionar al sistema de traducción un polinucleótido que codifica un polipéptido diana de interés y permitir la traducción del polipéptido diana de interés, incorporando con ello al menos un aminoácido no natural en el polipéptido diana.

20

25

17.- Un método de acuerdo con la reivindicación 16, en el que dicha molécula de ARNt se modifica de manera que contiene un anticodón mutado que forma pares de bases con un correspondiente codón degenerado por bamboleo con una afinidad mayor que la afinidad del ARNt natural.

30

18.- Un método de acuerdo con la reivindicación 16 o la reivindicación 17, en el que el sistema de traducción comprende una célula huésped, la molécula de ARNt se modifica, y en el que dicha PheRS modificada y/o dicha molécula de ARNt modificada se deriva de un organismo diferente a la célula huésped.

35 19.- Una composición, polipéptido, sistema de traducción o método de acuerdo con una cualquiera de las reivindicaciones 1 a 18, en donde la sintetasa modificada está mutada de manera que los codones que codifican los aminoácidos en las posiciones 412, 415 y 437 están mutados.

20.- Una composición, polipéptido, sistema de traducción o método de acuerdo con una cualquiera de las reivindicaciones 1 a 18, en donde la sintetasa modificada está mutada de manera que los codones que codifican los aminoácidos en las posiciones 415, 418 y 437 están mutados.

40

21.- Una composición, polipéptido, sistema de traducción o método de acuerdo con una cualquiera de las reivindicaciones 1 a 18, en donde la sintetasa modificada está mutada de manera que los codones que codifican los aminoácidos en las posiciones 412, 415, 418 y 437 están mutados.

5

22.- Una composición, polipéptido, sistema de traducción o método de acuerdo con una cualquiera de las reivindicaciones 1, 6, 7 ó 16, en donde dicha PheRS modificada utiliza un codón de terminación ámbar para la incorporación de un aminoácido no natural en una ubicación particular en el polipéptido.

	261	314
T. thermophilus:	RFQPVYFPFVEP...	GFAFGLGVERLAMLR
E. coli:	RFRPSTFPFTEP...	GFAFGMGMERLTMLR
S. cerevisiae:	RFKPTYNPYTEP...	VLGMGLSLERPTMIKY
	415	460

FIG. 1A

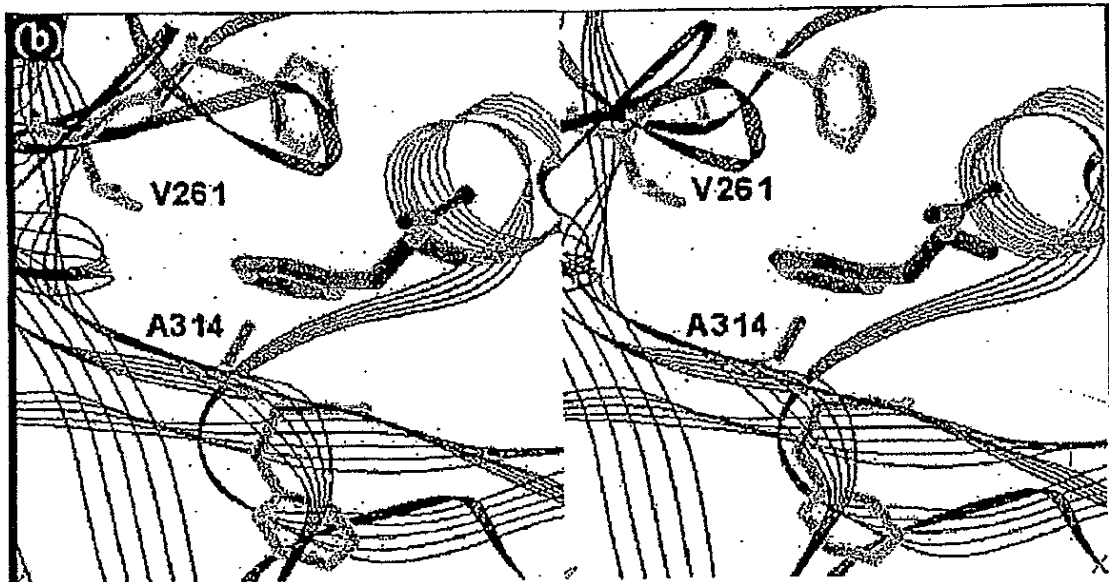


FIG. 1B

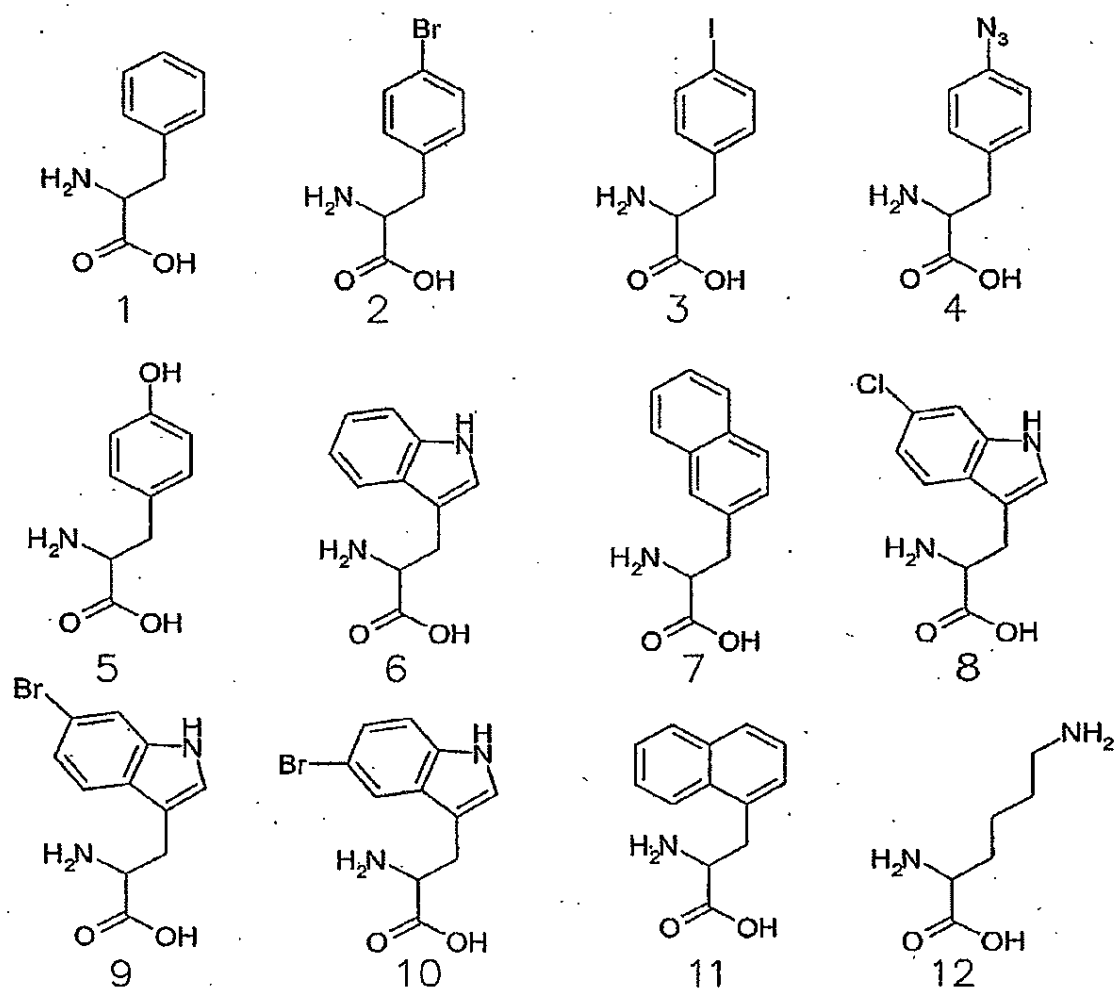


FIG. 2

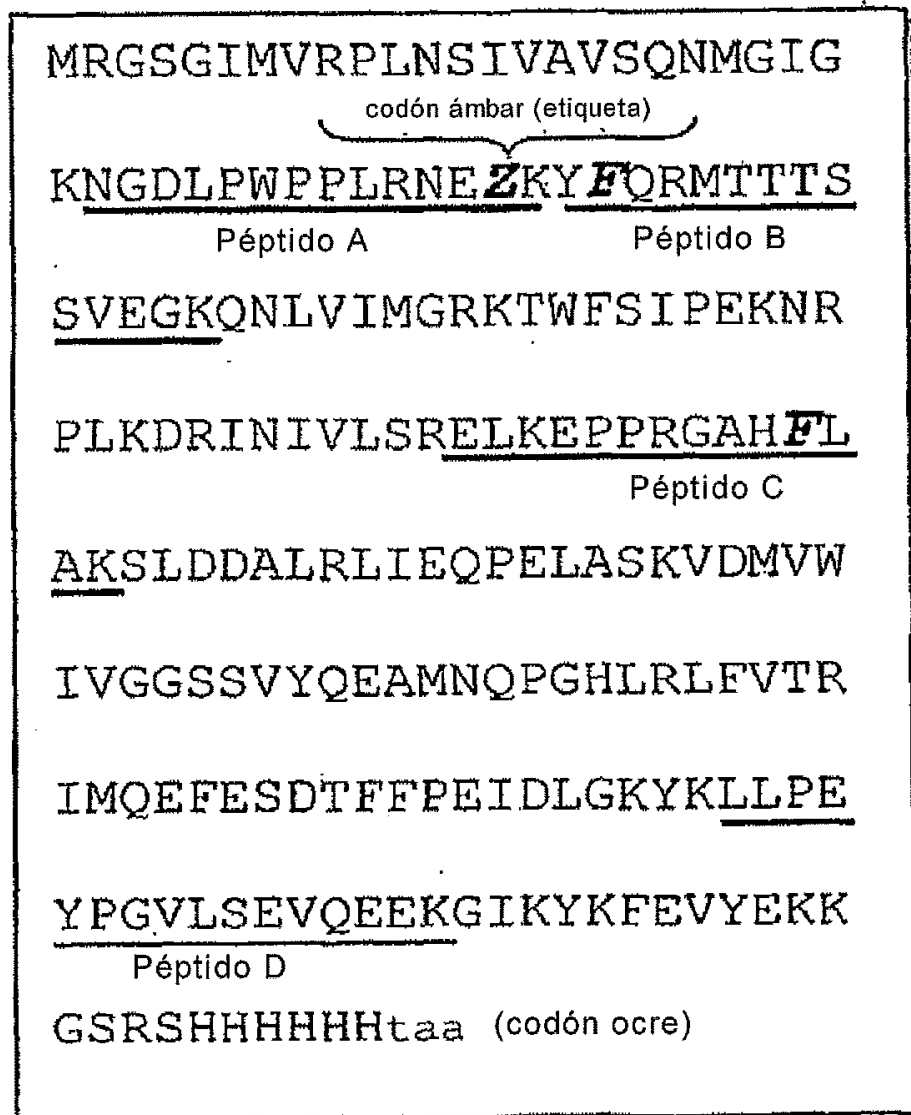


FIG. 3

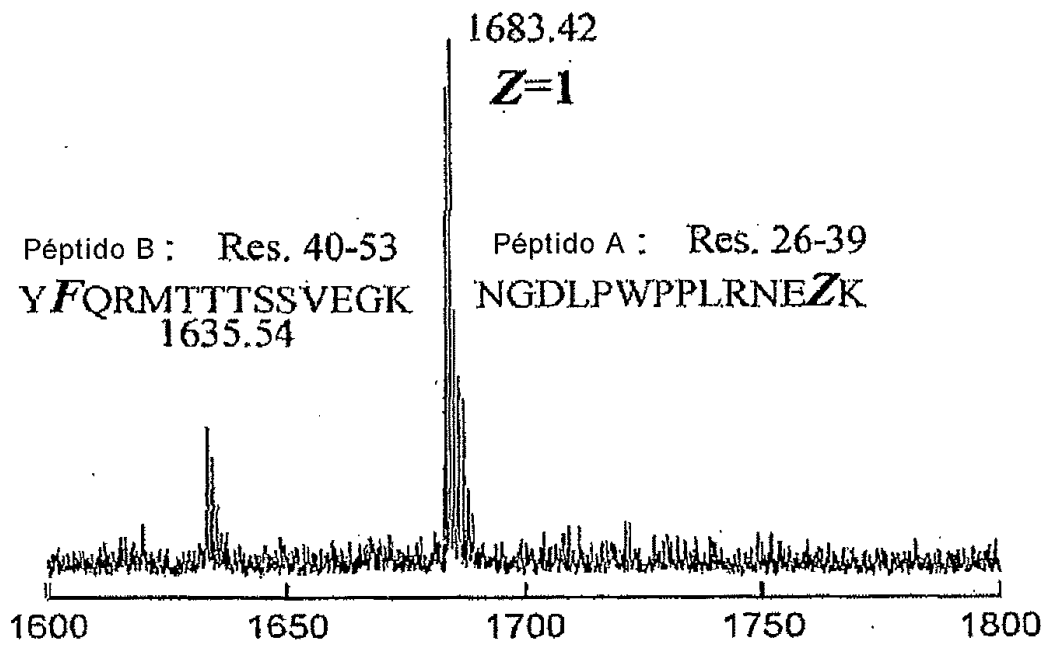


FIG. 4A

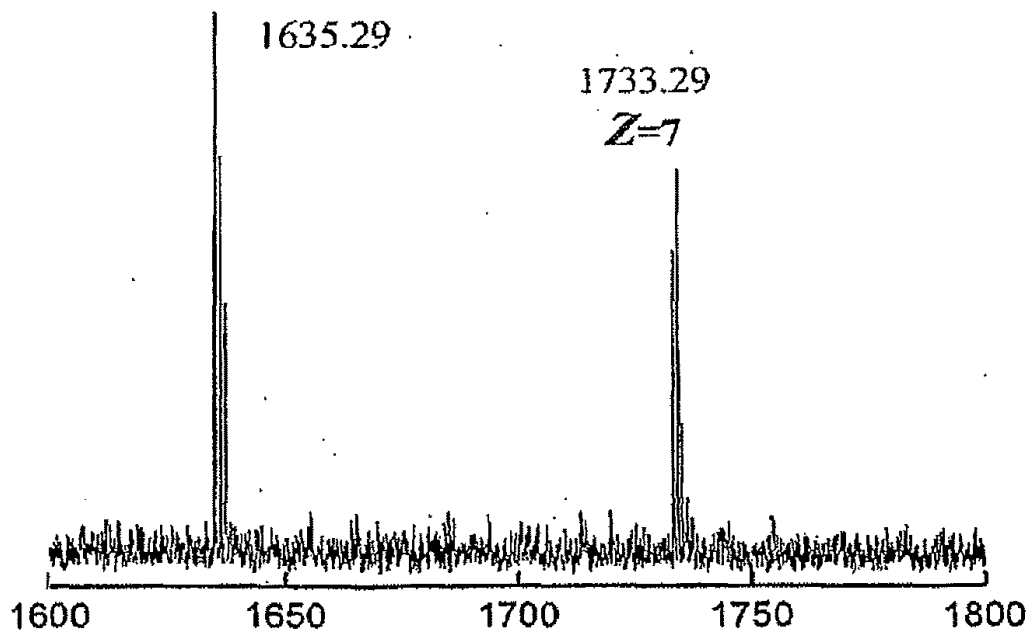


FIG. 4B

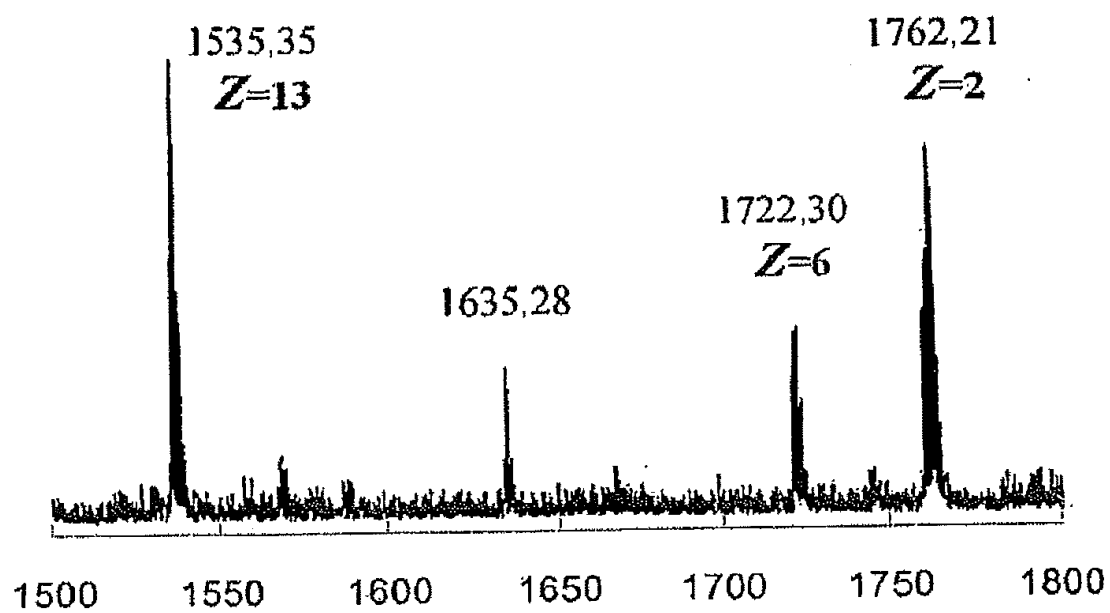


FIG. 4C

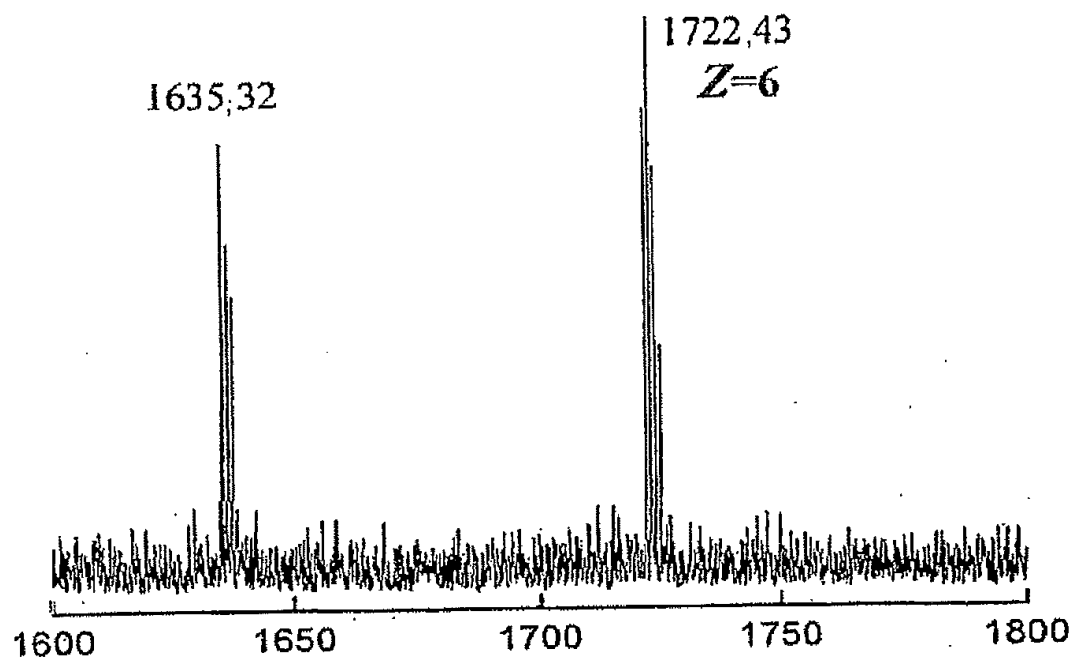


FIG. 4D

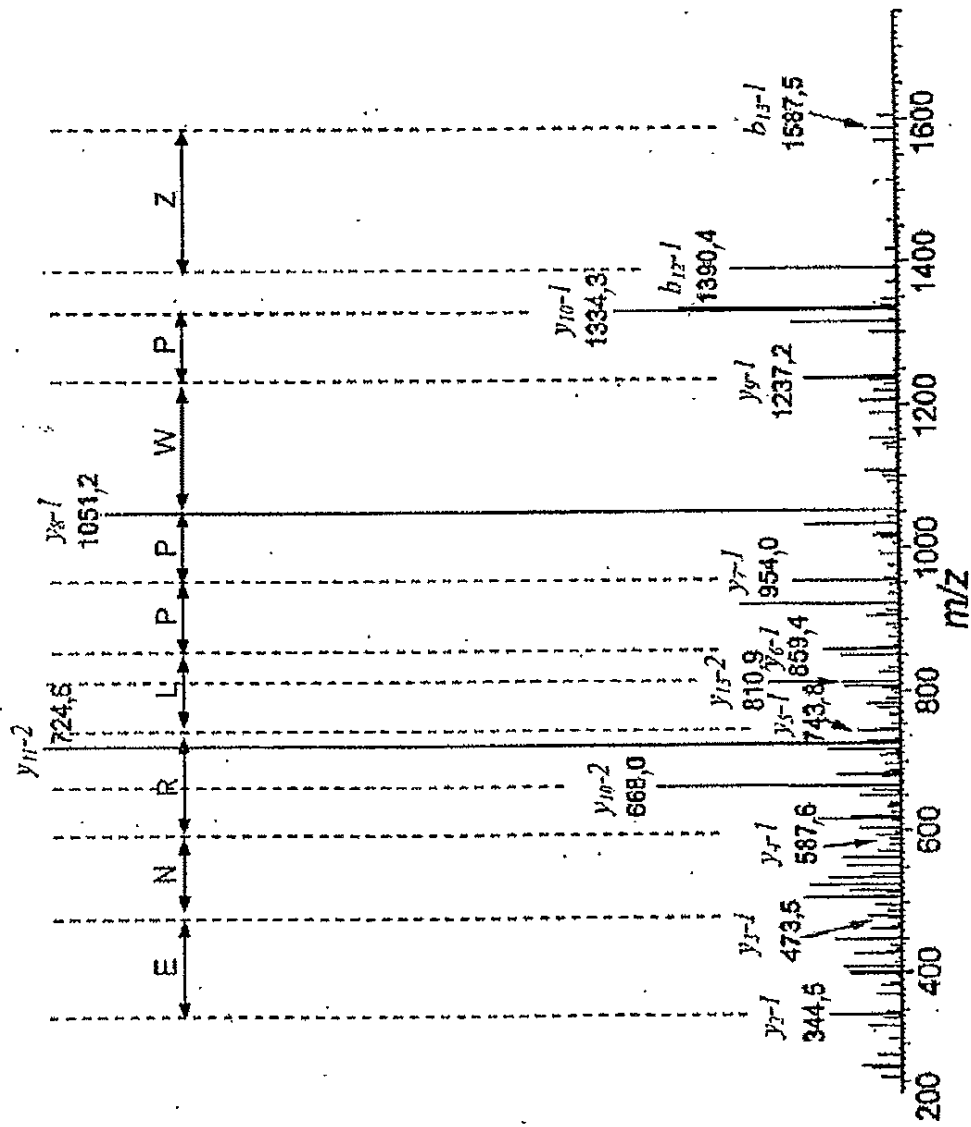


FIG. 5

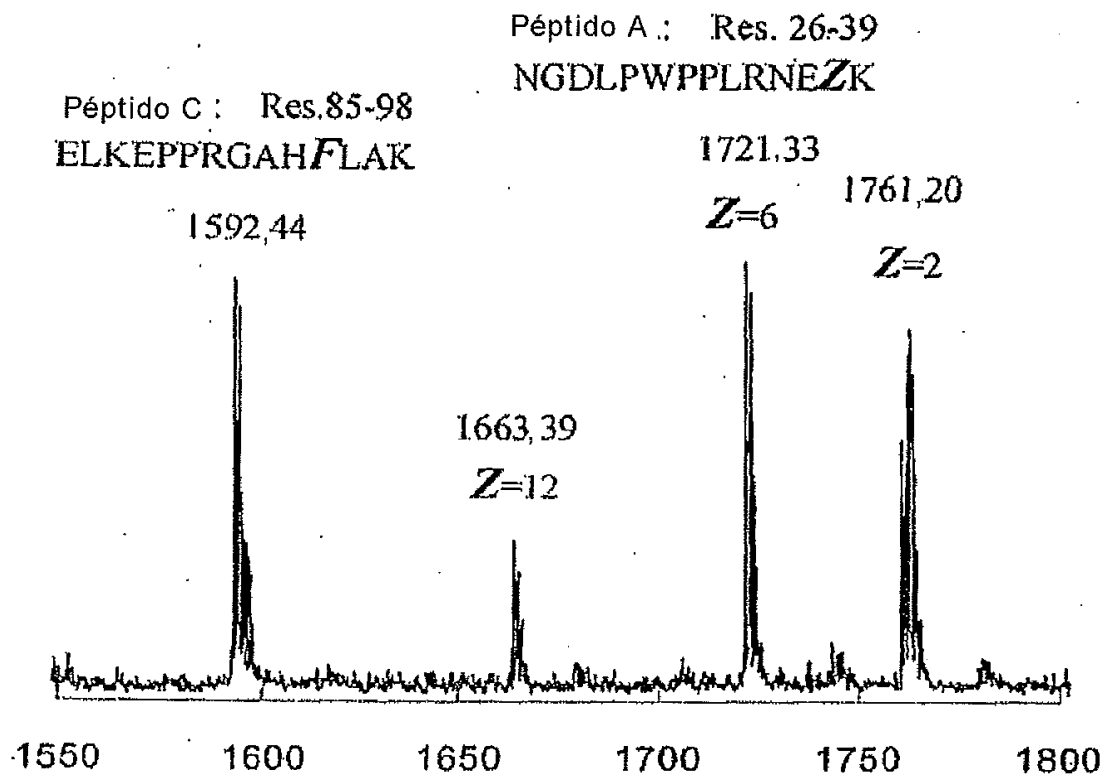


FIG. 6A

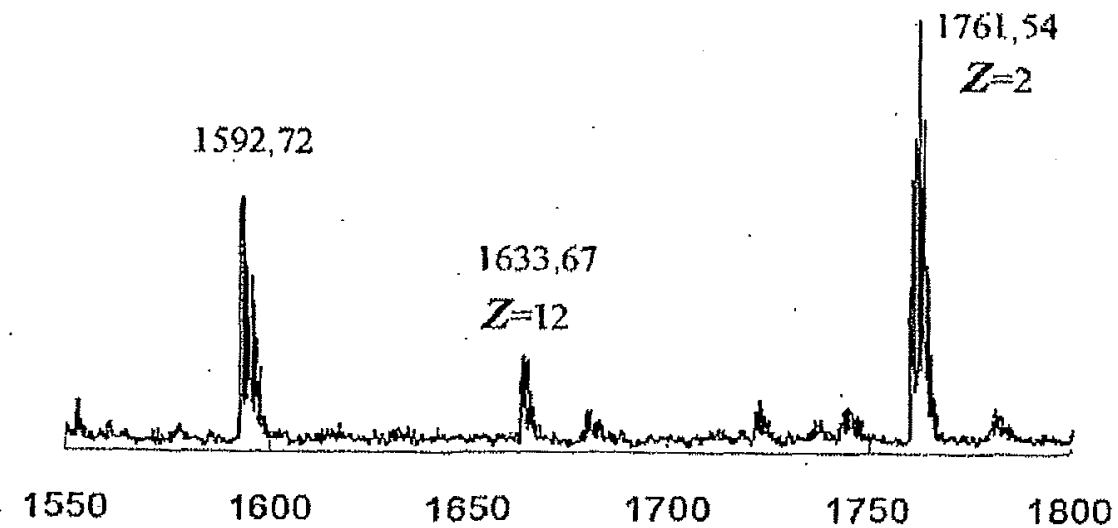


FIG. 6B

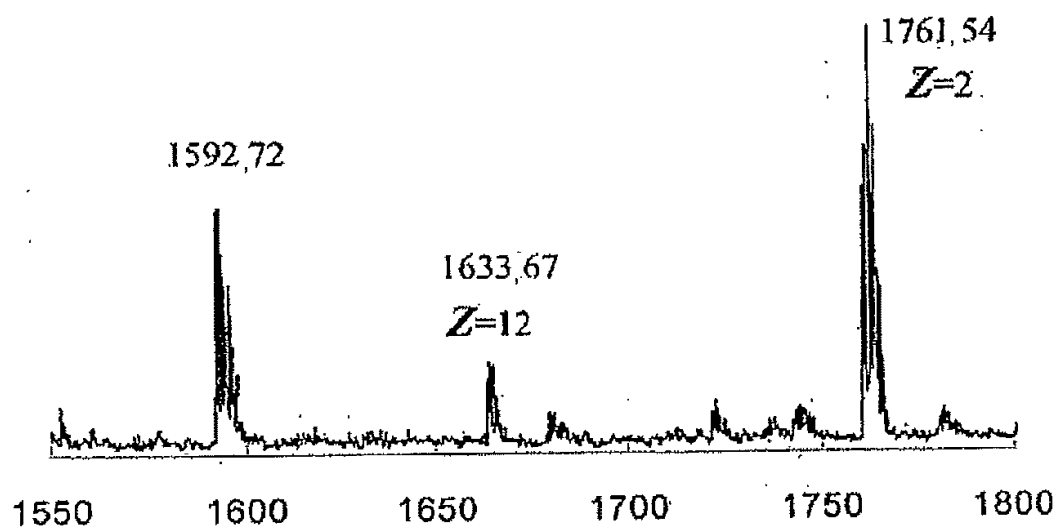


FIG. 6C

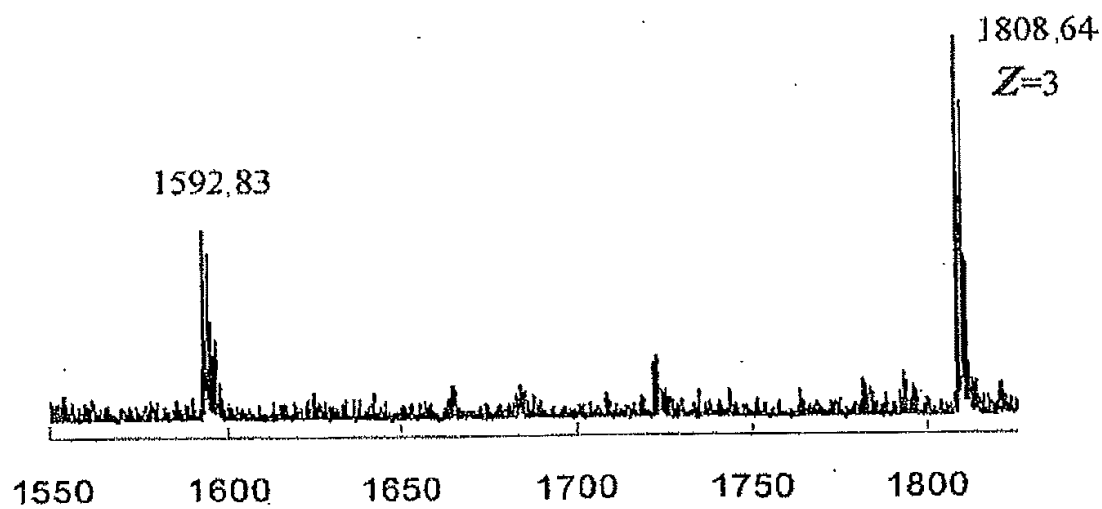


FIG. 6D

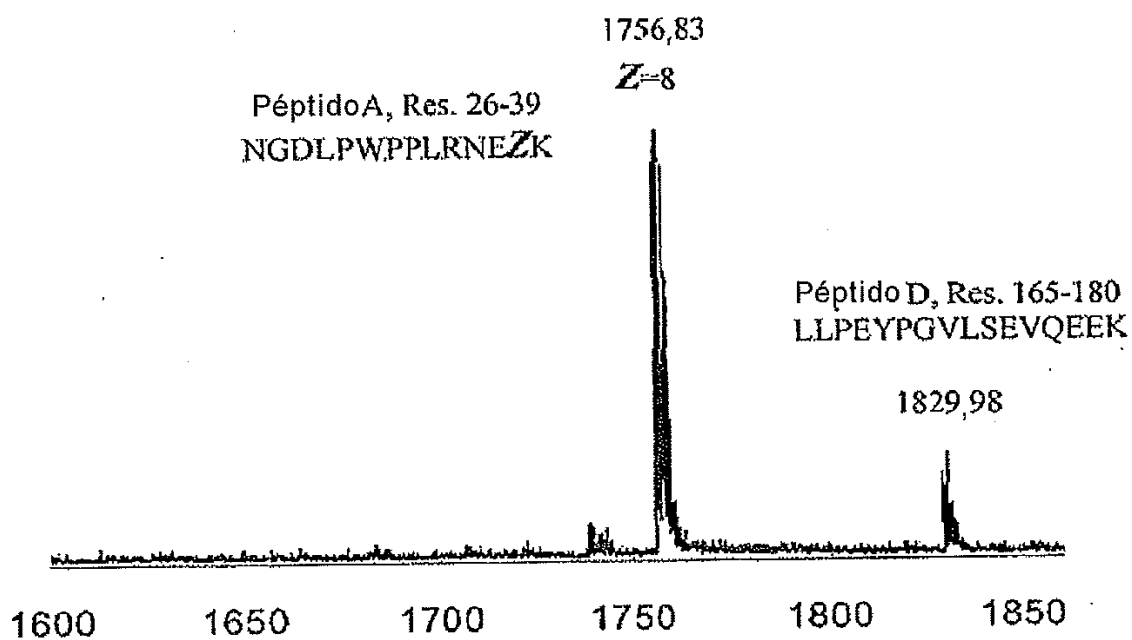


FIG. 7A

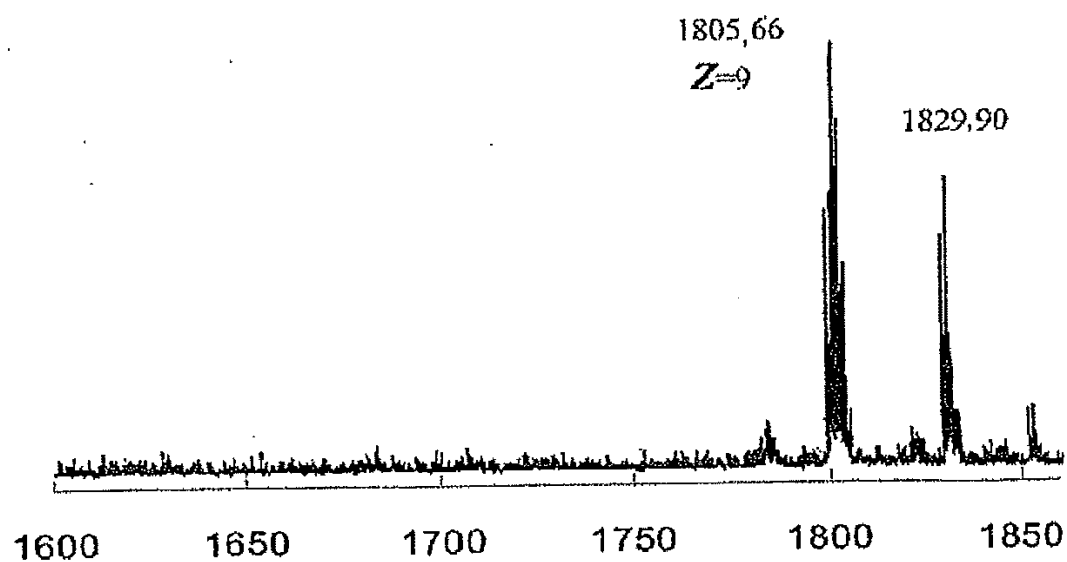


FIG. 7B

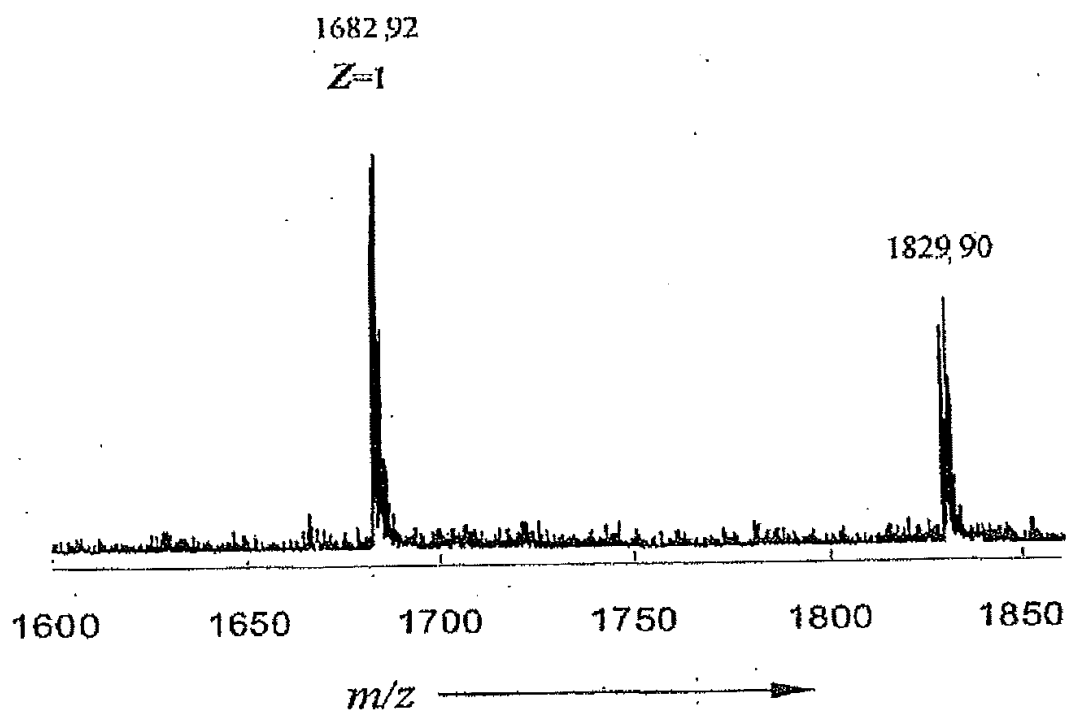


FIG. 7C

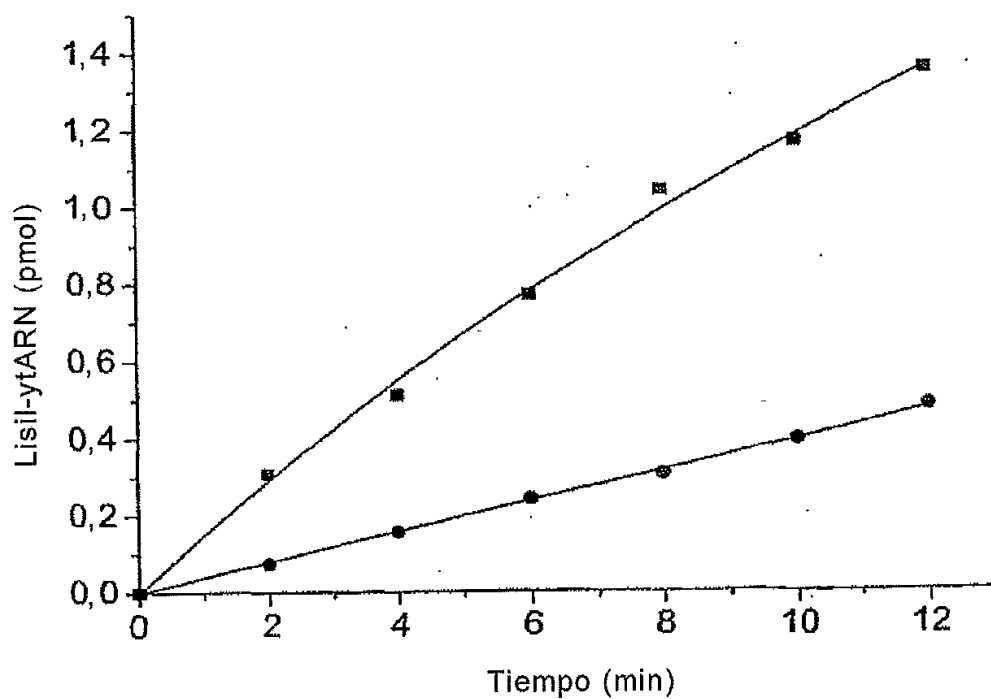
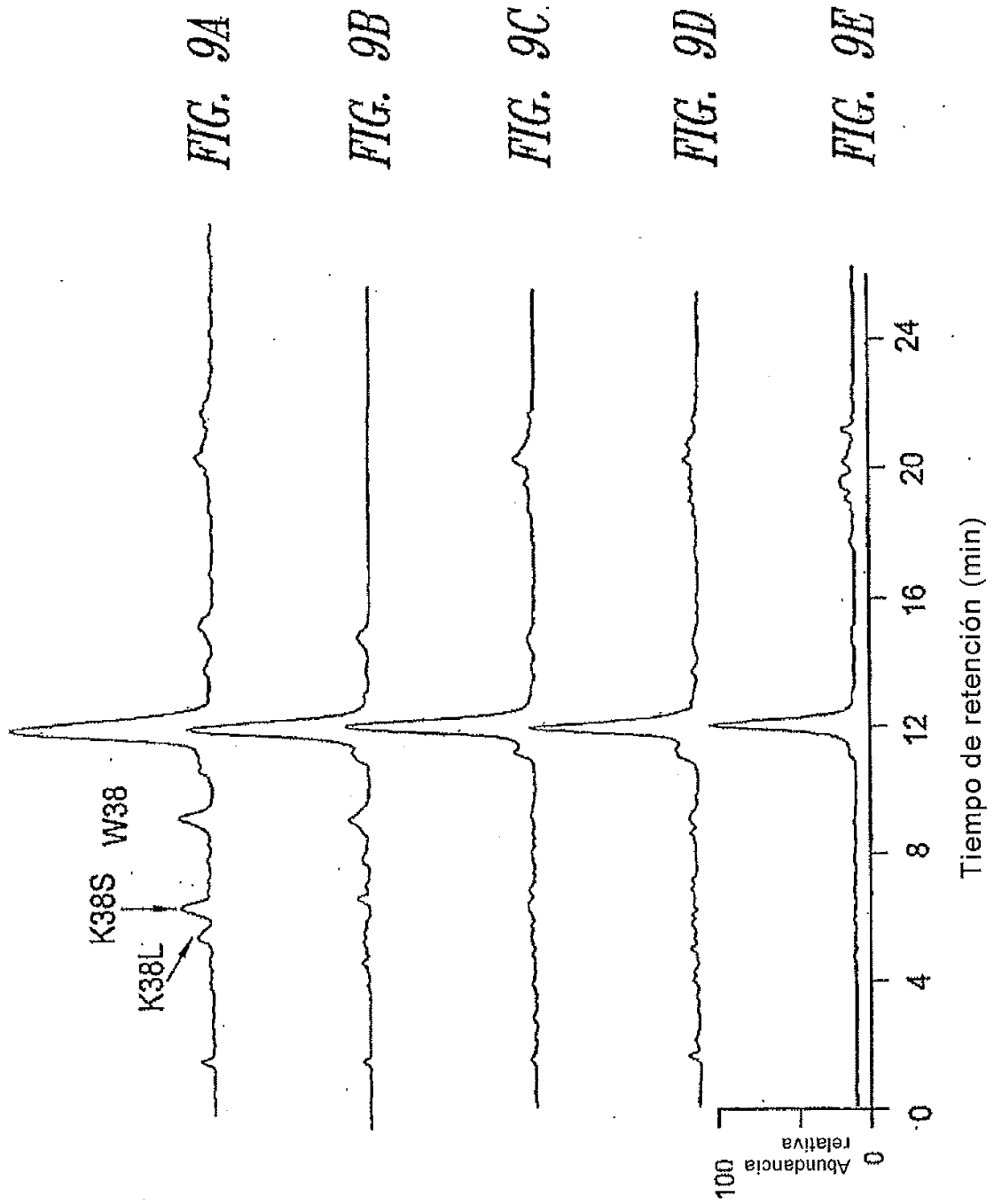
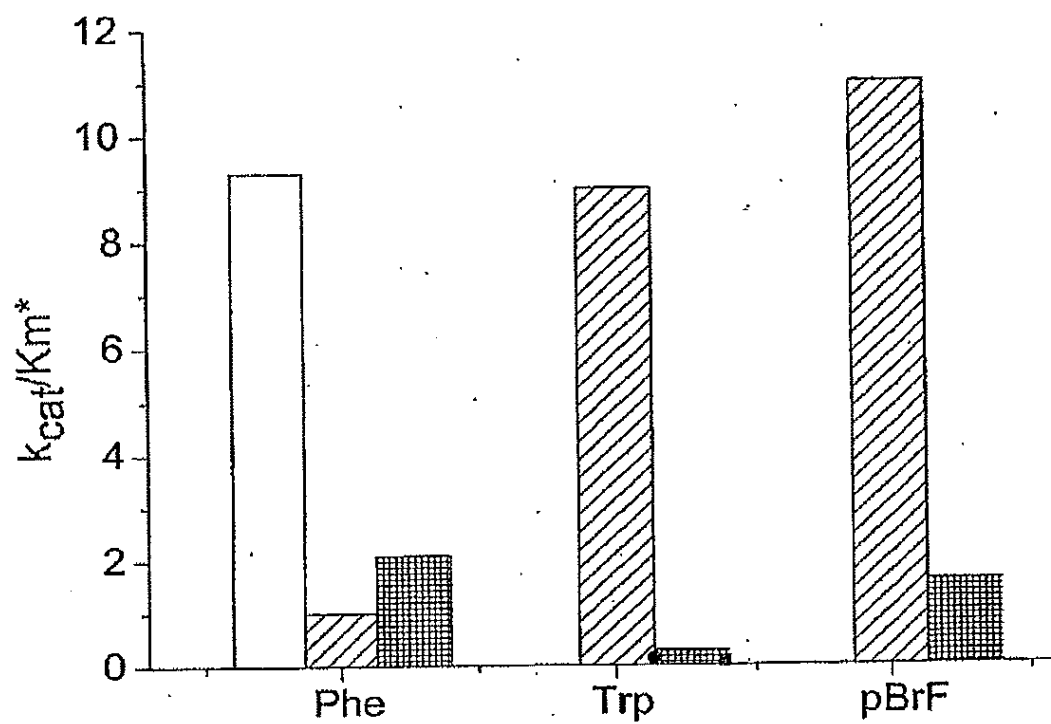
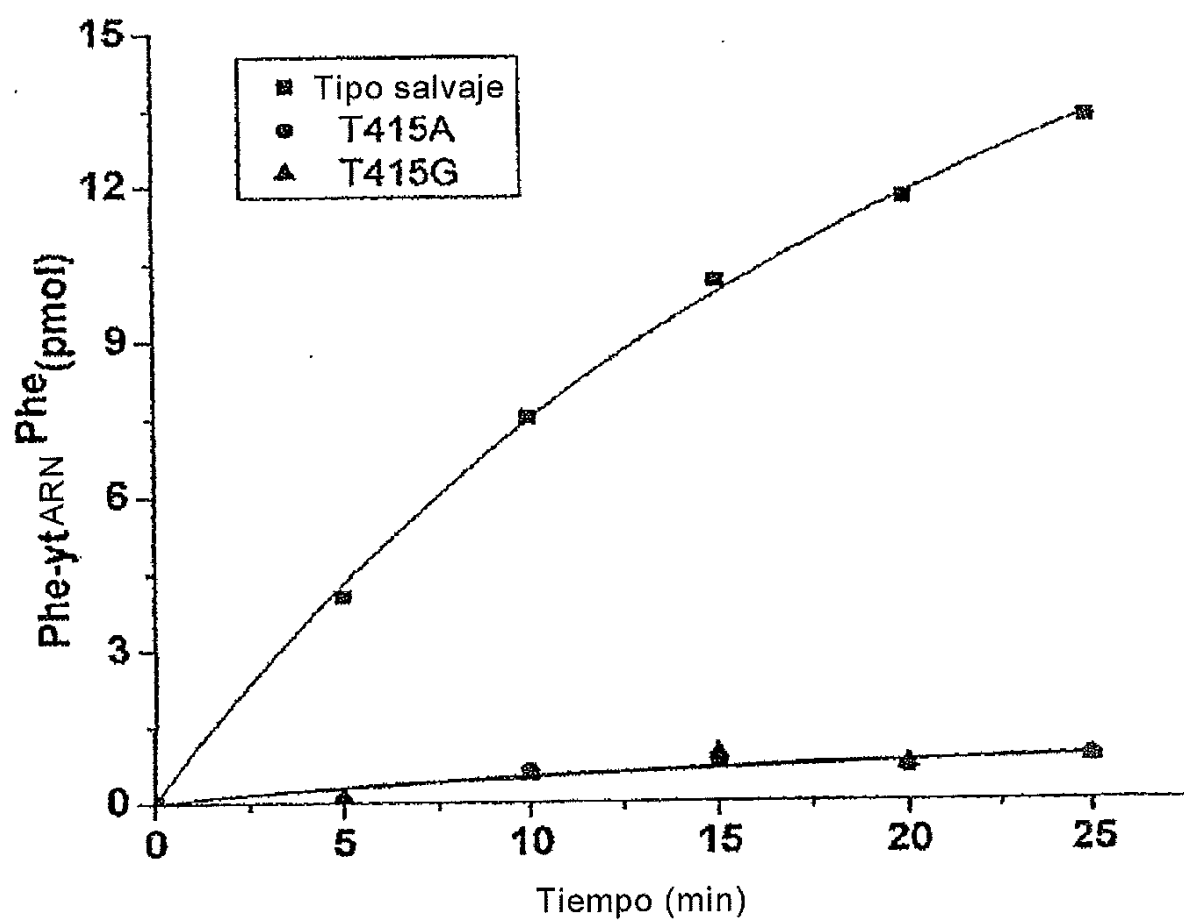


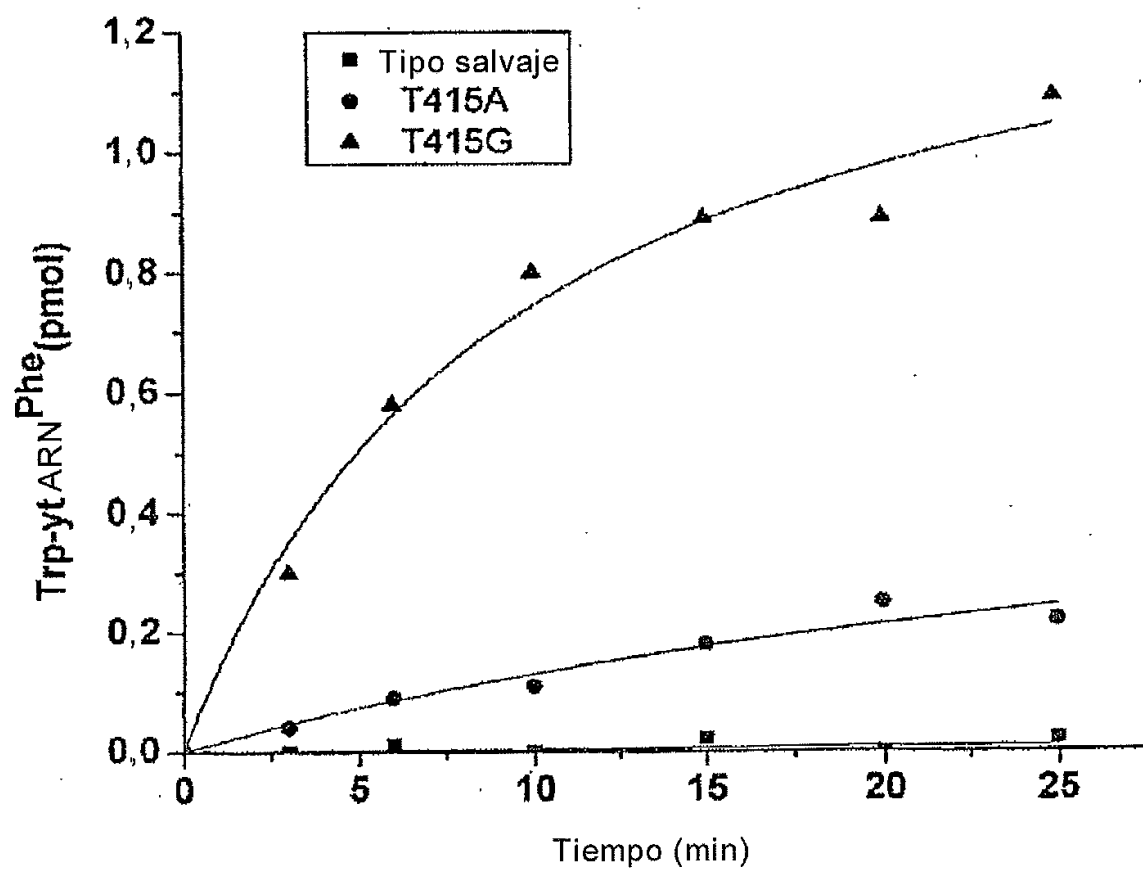
FIG. 8

Z38



*FIG. 10*

*FIG. 11A*

*FIG. 11B*

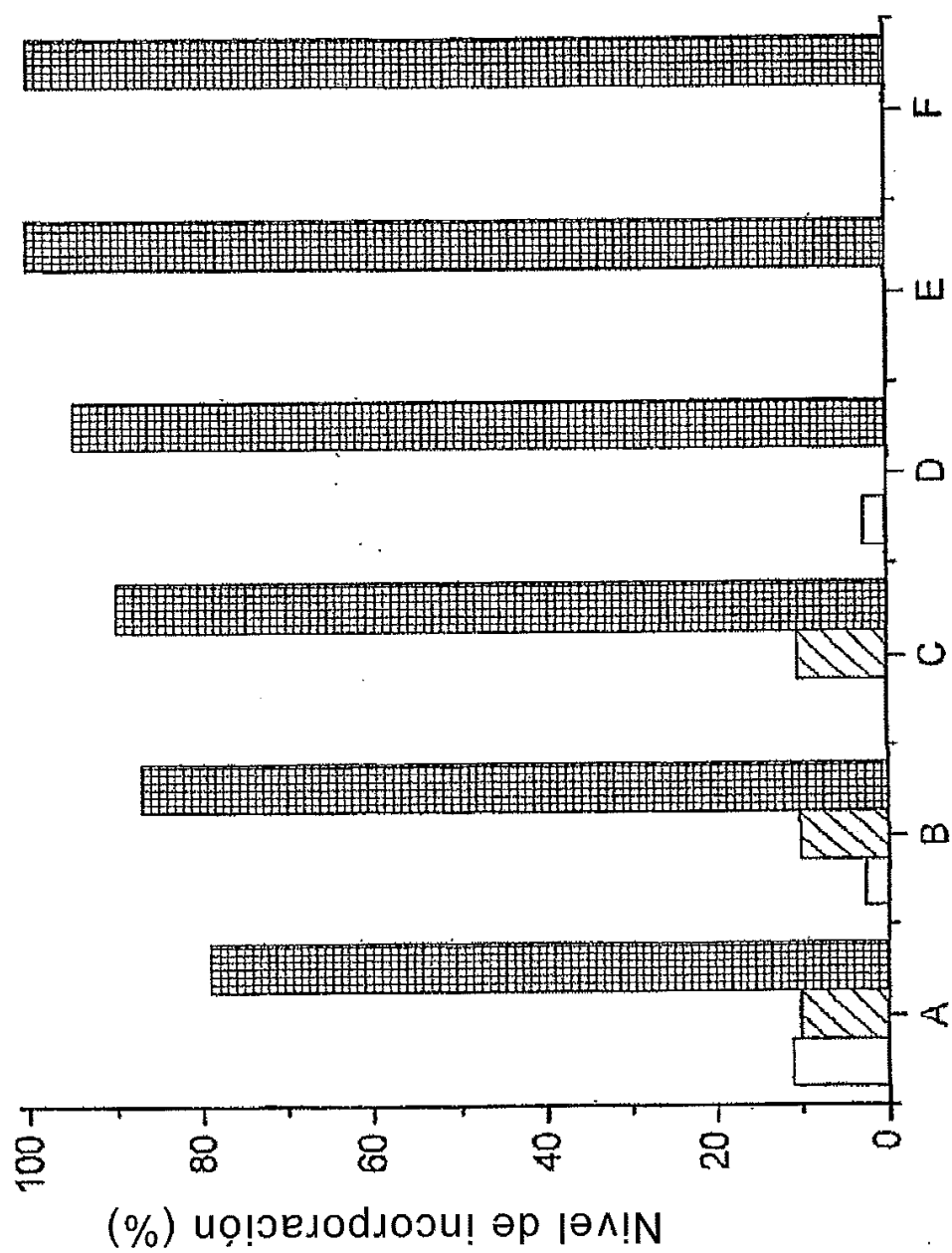


FIG. 12

Utilizando la variante yPheRS (T415G) y el supresor de ámbar yARN^{Phe}, PEtF se puede insertar en mDHFR en respuesta a un codón ámbar en el sistema de *E. coli*.

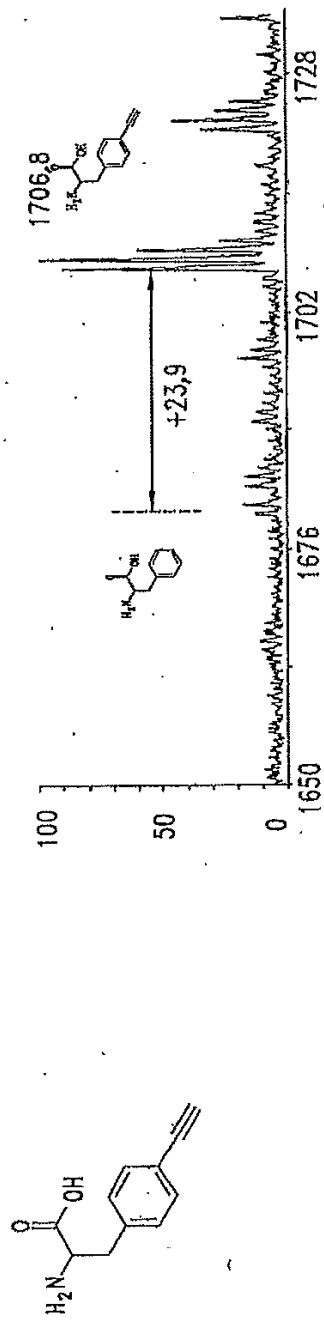
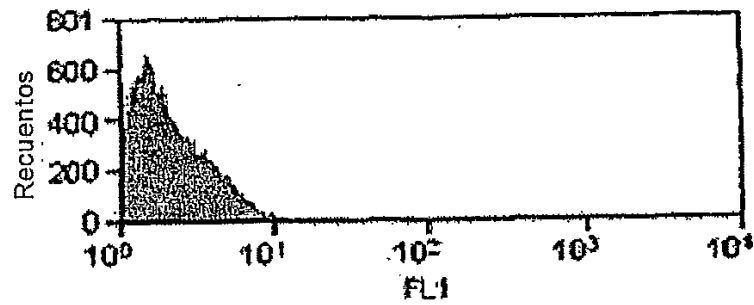


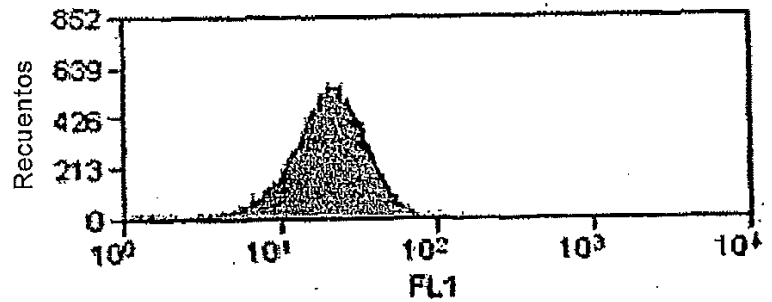
FIG. 13

Supresión ámbar GFP_081406 GFP_158 (L64, Am158)

Antes de la inducción



Phe 0,01 mM
Trp 0,01 mM



Phe 0,05 mM
Trp 0,05 mM
2NaI 3,0 mM

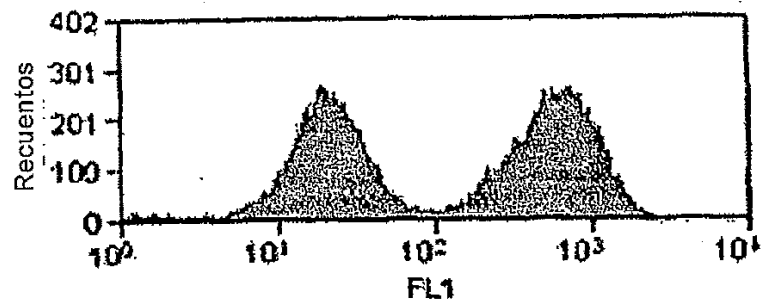
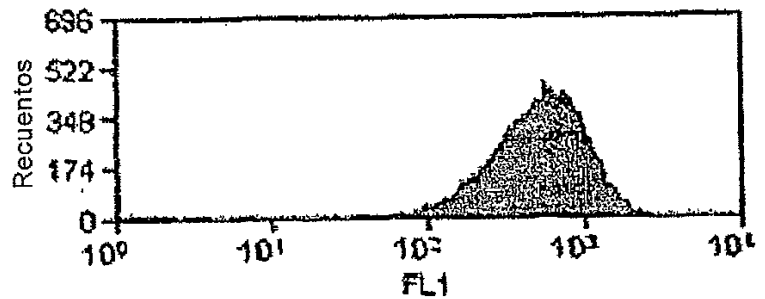


FIG. 14A

Supresión ámbar de GFP_081406 GFP_158 (L64, Am158)

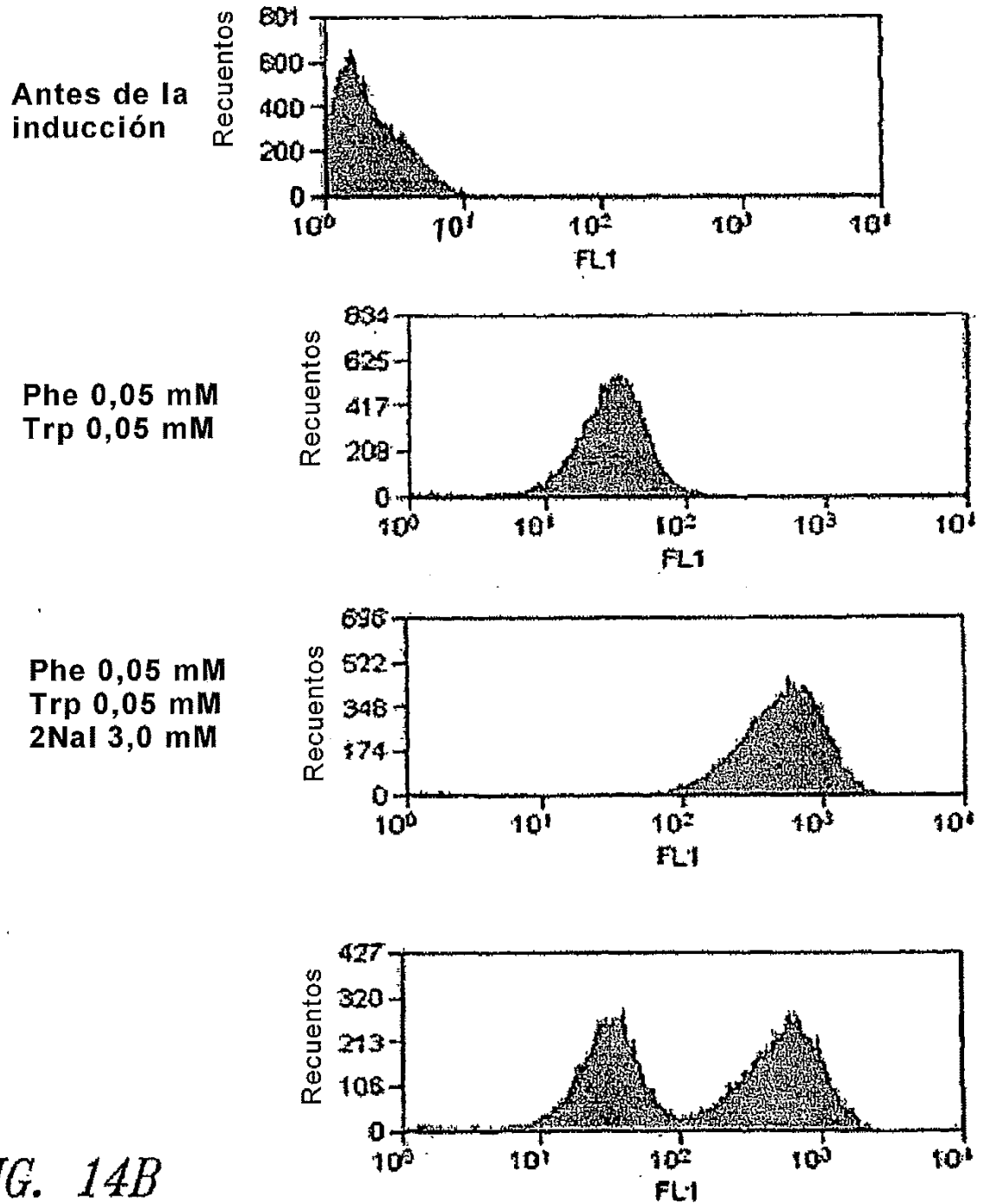


FIG. 14B

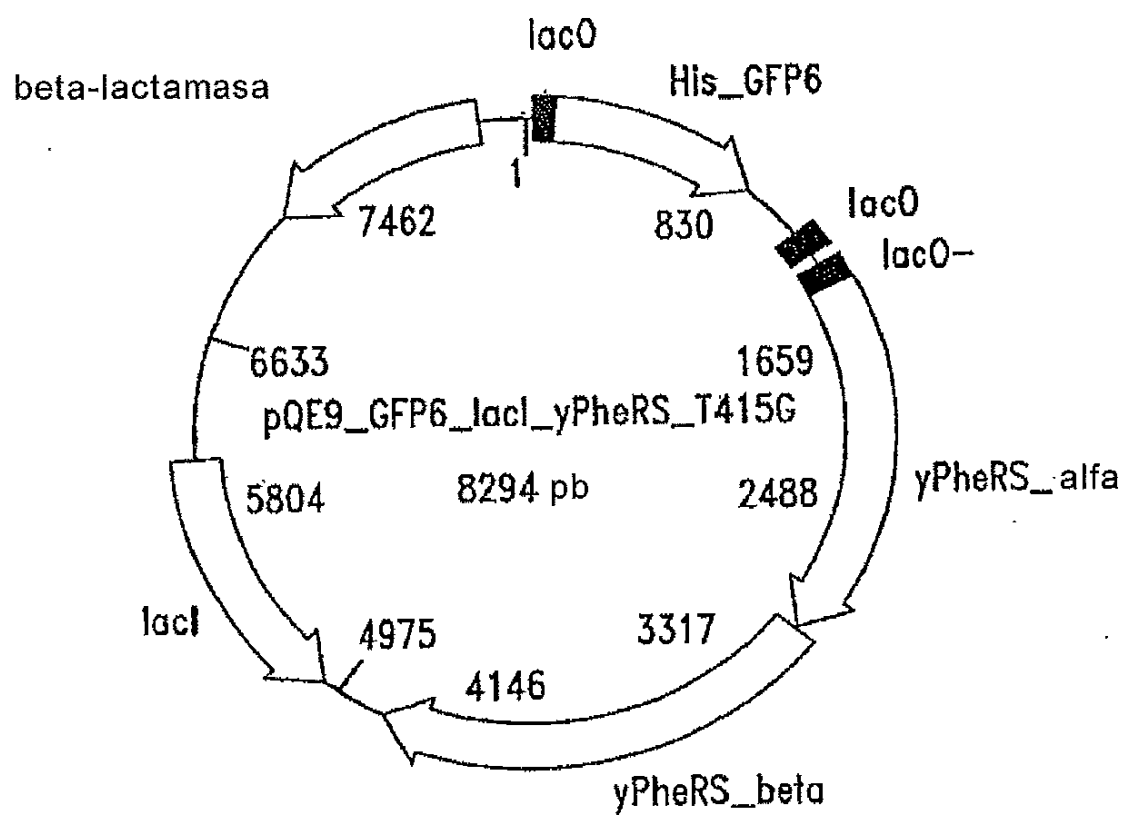


FIG. 15

>los_sitios_de_mutación_alfa_de_PheRS_de_levadura_están_marcados_con_NNN

atgtctgact tccaattaga aattctaaag aaactagatg aattggatga gatcaagtcc
 acactggcaa ctttccctca gcacggctct caagatgttc ttccgcttt gaactctttg
 aaagcccaca acaagtiaga gttttccaag gtcgacacgg ttacgtatga cttgaccaa
 gaagggtctc aaattttgaa tgaagggttcg tacgaaatta aactagtcaa gtcacatcaa
 gagttgggtc aacttcaa atcaagatgtg atgtccaaac taggcccctc agttggtaag
 gtcggtcagg ctagagcttt caagaacggc tggatcgcca aaaacgcctc aaacgagctt
 gaactctccg caaaattgca aaataccgat ttaaatgagc ttactgatga aacgcaatct
 attctagcgc aaatcaagaa caactcgc atggatagca ttgacgcaa gattttgaac
 gacttgaaga aaagaaagt aattgtctca ggtaaatca cagatttcag tgcaccaa
 gggccagagt tctcgaccga cctcaccaaa ttggaaccg atcttacctc cgacatggct
 tccaccaatg catacaagga cttgaagttc aagccttaca atttcaattc tcaagggtg
 caaatatctt caggtgctct tcacctta aacaaagtca gagaggaatt tagacaaatt
 ttcttttcca tgggattcac agagatgcc tcgaaccaat acgtcgagac aggtttctg
 aacttcgatg ccttttacgt ccacaacag catctgtc gtgacctga agacacttc
 tacatcaagg acccactaac cgtgagttg cccgatgaca agacatacat ggacaatatc
 aaagccgttc acgaacagg gagattcggg tccatcggt atcggtaca ctggaagcca
 gaagaatgtc aaaaattgg cttgagaact cactccacag ccatctctgc cagaatgtg
 cagcatttgg ccaaagatcc aaagcccacc agattgttt ctatcgaccg tgtttccgt
 aacgaagcag ttgacgccac ccatttggcc gaattccacc aggtggaagg tgttcttgc
 gactacaaca ttactctggg tgacctgatc aagttcatgg aagagtttt cgaagaatg
 ggtgtcaccg gtttgagatt caagcctacc tacNNNcctt acNNNgagcc aNNNatggaa
 atcttttctt ggcacgaagg ttgcaaaaa tgggtcgaaa tcggtaacNNNggtatgttc
 agaccagaaa tgctcgagtc catgggtcta ccaaaggatc taagagtcct tggttgggg
 ttatccttgg aaagacctac catgatcaaa tataagggtc aaaacatcag agaactgtta
 ggtcataaag tctctttgga cttatcgaa accaatcctg ctgctagatt ggacgaagac
 ttgtacgaat aa

	Nombre	412	415	418	437
	T415G	AAT	GGC	TCA	TCT
1	2Na1	GGG	GGG	TGT	TTT
2	412_415	GGG	GGC		
3	415_418		GGC	TGT	
4	415_437		GGC		TTT
5	412_415_437	GGG	GGC		TTT
6	415_418_437		GGC	TGT	TTT
7	412_415_418	GGG	GGC	TGT	

FIG. 16