

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 697 778**

51 Int. Cl.:

**C12N 9/00** (2006.01)

**C12N 15/52** (2006.01)

**C07K 1/02** (2006.01)

**C12P 21/02** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.10.2014 PCT/US2014/060058**

87 Fecha y número de publicación internacional: **16.04.2015 WO15054587**

96 Fecha de presentación y número de la solicitud europea: **10.10.2014 E 14852519 (9)**

97 Fecha y número de publicación de la concesión europea: **10.10.2018 EP 3055321**

54 Título: **ARNt sintetasas de aminoácidos no naturales para para-metilazido-L-fenilalanina**

30 Prioridad:

**11.10.2013 US 201361890028 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**28.01.2019**

73 Titular/es:

**SUTRO BIOPHARMA, INC. (100.0%)  
310 Utah Avenue, Suite 150  
South San Francisco, California 94080, US**

72 Inventor/es:

**ZIMMERMAN, ERIK S. y  
THANOS, CHRISTOPHER**

74 Agente/Representante:

**PONS ARIÑO, Ángel**

ES 2 697 778 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

ARNt sintetasa de aminoácidos no naturales para para-metilazido-L-fenilalanina

5 **Antecedentes de la invención**

Los conjugados de fármaco proteínico tales como conjugados de fármaco de anticuerpo (ADC) son un agente quimioterapéutico dirigido actualmente en la vanguardia de la medicina oncológica. Los ADC, por ejemplo, consisten en un anticuerpo específico de antígeno tumoral que está acoplado a una citotoxina de molécula pequeña quimioterapéutica. A través del suministro dirigido de citotoxinas potentes, los conjugados de fármaco proteínico muestran índice terapéutico mejorado sobre las quimioterapias tradicionales con eficacia potenciada respecto a los tratamientos con anticuerpo monoclonal convencionales. Sin embargo, los métodos actuales utilizan modos no específicos de conjugación de los fármacos con proteínas, dando lugar de ese modo a productos farmacéuticos heterogéneos con cantidades variadas de fármacos conjugados a través de varios sitios posibles. Los problemas técnicos asociados con la conjugación de fármacos a proteínas usando aminoácidos de origen natural se deben principalmente a los grados heterogéneos y a la ubicación de la carga del fármaco, así como a la inestabilidad del conjugado.

Para reducir la heterogeneidad del producto, varios grupos han presentado estrategias dirigidas al sitio que utilizan cisteínas sustituidas o modificación enzimática de glutamina genomaniplada para la conjugación. Los ADC específicos de sitio tienen potencia comparable con los ADC conjugados de forma aleatoria, mientras muestran al mismo tiempo un índice terapéutico y farmacocinética superiores. Sin embargo, existen limitaciones en la estabilidad del acoplamiento basado en tiol debido a la hidrólisis plasmática del anillo succinimida del conjugado de tiomaleimida, que provoca transferencia del fármaco a seroalbúmina. Además, la reducción parcial y la nueva formación de enlaces disulfuro que facilitan la conjugación a la cisteína sin genomaniplada, puede dar lugar a una estructura cuaternaria aberrante mediada por disulfuro. Una alternativa al uso de restos de cisteína sin introducir es usar incorporación específica de sitio de aminoácidos no naturales con cadenas laterales químicas que son compatibles con la química de conjugación bioortogonal.

Los componentes esenciales de cualquier sistema de incorporación de aminoácido no natural (nnAA) consiste en una aminoacil ARNt sintetasa (aaRS) que carga un ARNt específico con un nnAA. EL par de aaRS-ARNt debe ser ortogonal con respecto a la célula hospedadora o sistema de expresión en que se emplean. Es decir, la sintetasa específica de nnAA no debe reconocer ningún ARNt del hospedador o aminoácido equivalente, y el ARNt no debe aminoacilarse por ningún aaRS del hospedador. Adicionalmente, el anticodón del ARNt ortogonal a menudo se muta para que reconozca un codón de parada o sin sentido. El reposicionamiento de los codones no proteinogénicos, tal como el codón de parada ámbar TAG, posibilita la incorporación de un nnAA en cualquier sitio en una proteína mediante mutagénesis de la secuencia codificante del ARNm en TAG. La supresión ámbar es el modo más ampliamente usado de incorporación de nnAA catalizada por enzima, cotraduccional. Los aminoácidos no naturales con cadenas laterales químicas reactivas bioortogonales pueden usarse como "empuñadura" química para conjugar diversas cargas en sitios concretos en una proteína. Esta estrategia puede generar funcionalidad adicional en las proteínas por conjugación directa de un aducto biológicamente activo, tal como marcadores fluorescentes o radiactivos, marcadores fotoactivables, PEG de modificación farmacocinética o agentes quimioterapéuticos. Desafortunadamente, los métodos actuales para la incorporación de nnAA y la conjugación de cadenas laterales químicas reactivas bioortogonales en proteínas están impedidos por el bajo rendimiento de producto global, la ineficacia de la incorporación de nnAA y la baja eficacia de conjugación.

El documento US 2008/233661 se refiere a pares ortogonales de ARNt y aminoacil-ARNt sintetasa que pueden incorporar aminoácidos no naturales en proteínas producidas en células hospedadoras eubacterianas tales como *E. coli*, o en hospedadores eucariotas tal como una célula de levadura. Bazewicz *et al.*, describen la incorporación genética del aminoácido no natural, 4-azidometil-L-fenilalanina, de una manera específica de sitio, en proteínas utilizando una aminoacil ARNt sintetasa ortogonal genomaniplada (J Phys Chem B., 2013, vol. 117, n.º 30, páginas 8987 - 93).

Existe una necesidad en la técnica de métodos mejorados de incorporación específica de sitio de nnAA y conjugación de aductos biológicamente activos en proteínas para formar agentes terapéuticos de conjugado de fármaco proteínico conjugado homogéneamente. La presente invención satisface estas y otras necesidades.

**Breve resumen de la invención**

En un primer aspecto, la presente invención proporciona una composición que comprende una aminoacil-ARNt sintetasa (RS), en la que la RS:

- i) aminoacila preferentemente hasta un grado de más de un 90 % un ARNt con para-metilazido-L-fenilalanina (pAMF) en comparación con los 20 aminoácidos de origen natural comunes;
- ii) tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y

iii) usando la SEQ ID NO: 1 como secuencia de referencia, tiene:

- 5 a) en la posición del aminoácido 65: A o V;  
 b) en la posición del aminoácido 108: Y o W;  
 c) en la posición de aminoácido 158: A;  
 d) en la posición del aminoácido 32: T o V o A; y  
 e) en la posición del aminoácido 159: S o G o V.

10 La composición de la variante de RS puede tener sustituciones de aminoácido en la posición del aminoácido 65 para A o V (por ejemplo, L65A o L65V), en la posición del aminoácido 108 para Y o W (por ejemplo, F108Y o F108W), en la posición del aminoácido 158 para A (por ejemplo, D158A); en la posición del aminoácido 32 para T o V o A (por ejemplo, Y32T o Y32V o Y32A); y en la posición del aminoácido 159 para S o G o V (por ejemplo, I159S o I159G o I159V). En algunas realizaciones, la composición incluye además una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consisten en L o M o I (por ejemplo, Q109L, Q109M o Q109I).

15 En algunas realizaciones, la RS tiene una identidad de secuencia de más de un 80 %, por ejemplo, de un 81 %, de un 82 %, de un 83 %, de un 84 %, de un 85 %, de un 86 %, de un 87 %, de un 88 %, de un 89 %, de un 90 %, de un 91 %, de un 92 %, de un 92 %, de un 93 %, de un 94 %, de un 95 %, de un 96 %, de un 97 %, de un 98 %, 99 % o 100 % de la SEQ ID NO: 1. En algunas realizaciones, la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y tiene sustituciones de aminoácidos seleccionadas del grupo que consiste en: 1) Y32T, L65A, F108Y, Q109L e I159S; ii) Y32V, L65A, F108W, Q109M e I159G; iii) Y32A, L65V, F108W e I159G; iv) Y32V, L65V, F108Y e I159V; v) Y32T, L65V, F108W e I159S; y vi) Y32V, L65V, F108W, G109I e I159S. La RS puede tener una identidad secuencia de aminoácidos de más de un 80 %, por ejemplo, de un 80 %, de un 81 %, de un 82 %, de un 83 %, de un 84 %, de un 85 %, de un 86 %, de un 87 %, de un 88 %, de un 89 %, de un 90 %, de un 91 %, de un 92 %, de un 93 %, de un 94 %, de un 95 %, de un 96 %, de un 97 %, de un 98 %, 99 % o 100 % de identidad, con la secuencia de la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* de tipo silvestre (SEQ ID NO: 1). La RS puede tener una identidad de secuencia de aminoácidos de al menos un 90 % de la SEQ ID NO: 1, por ejemplo, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, 99 % o 100 % de identidad de secuencia con la SEQ ID NO: 1.

30 En un segundo aspecto, la presente invención proporciona un polinucleótido que codifica una aminoacil-ARNt sintetasa (RS), en el que la RS:

- 35 i) aminoacila preferentemente hasta un grado de más de un 90 % un ARNt con para-metilazido-L-fenilalanina (pAMF) en comparación con los 20 aminoácidos de origen natural comunes;  
 ii) tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y  
 iii) usando la SEQ ID NO: 1 como secuencia de referencia, tiene:

- 40 a) en la posición del aminoácido 65: A o V;  
 b) en la posición del aminoácido 108: Y o W;  
 c) en la posición de aminoácido 158: A;  
 d) en la posición del aminoácido 32: T o V o A; y  
 e) en la posición del aminoácido 159: S o G o V.

45 En algunas realizaciones, el polinucleótido tiene además una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I (por ejemplo, Q109L, Q109M o Q109I).

50 En algunas realizaciones, la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y tiene sustituciones de aminoácidos seleccionadas del grupo que consiste en: 1) Y32T, L65A, F108Y, Q109L e I159S; ii) Y32V, L65A, F108W, Q109M e I159G; iii) Y32A, L65V, F108W e I159G; iv) Y32V, L65V, F108Y e I159V; v) Y32T, L65V, F108W e I159S; y vi) Y32V, L65V, F108W, G109I e I159S.

55 En un tercer aspecto, la presente invención proporciona un sistema de síntesis de proteínas sin células para incorporar selectivamente para-metilazido-L-fenilalanina (pAMF) en una proteína de interés, comprendiendo el sistema:

- 60 a) un extracto sin células de bacterias que tienen ARNt que funciona biológicamente, aminoácidos y ribosomas necesarios para la síntesis de proteína sin células;  
 b) un polinucleótido que tiene una región codificante que codifica la proteína de interés y que incluye un codón de supresión ubicado selectivamente dentro de su región codificante;  
 c) metilazido-L-fenilalanina en una concentración suficiente para permitir la incorporación selectiva de pAMF en la proteína de interés;  
 d) un ARNt que puede cargarse con pAMF y complementario al codón de supresión de la proteína de interés; y  
 e) una aminoacil-ARNt sintetasa (RS), en el que la RS:

65

i) aminoacila preferentemente hasta un grado de más de un 90 % un ARNt con para-metilazido-L-fenilalanina (pAMF) en comparación con los 20 aminoácidos de origen natural comunes;

ii) tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y

5 iii) usando la SEQ ID NO: 1 como secuencia de referencia, tiene: a) en la posición del aminoácido 65: A o V; b) en la posición del aminoácido 108: Y o W; c) en la posición de aminoácido 158: A; d) en la posición del aminoácido 32: T o V o A; y e) en la posición del aminoácido 159: S o G o V.

10 En algunas realizaciones, el sistema de síntesis de proteína sin células tiene además una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I.

15 En algunas realizaciones, la RS tiene una identidad de secuencia de más de un 80 %, por ejemplo, de un 81 %, de un 82 %, de un 83 %, de un 84 %, de un 85 %, de un 86 %, de un 87 %, de un 88 %, de un 89 %, de un 90 %, de un 91 %, de un 92 %, de un 92 %, de un 93 %, de un 94 %, de un 95 %, de un 96 %, de un 97 %, de un 98 %, de un 99 % o 100 % de la SEQ ID NO: 1. En algunas realizaciones, la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y comprende sustituciones de aminoácidos seleccionadas del grupo que consiste en: i) Y32T, L65A, F108Y, Q109L e I159S; ii) Y32V, L65A, F108W, Q109M e I159G; iii) Y32A, L65V, F108W e I159G; iv) Y32V, L65V, F108Y e I159V; v) Y32T, L65V, F108W e I159S; y vi) Y32V, L65V, F108W, G109I e I159S.

20 En algunas realizaciones, el extracto sin células tiene un sistema de fosforilación oxidativa activo. En algunas realizaciones, la proteína de interés es un anticuerpo o fragmento de anticuerpo.

25 En un cuarto aspecto, la presente invención proporciona un método para incorporar selectivamente para-metilazido-L-fenilalanina (pAMF) en una proteína de interés, comprendiendo el método las etapas de:

a) combinar un extracto sin células de bacterias que tienen ARNt que funciona biológicamente, aminoácidos y ribosomas necesarios para la síntesis de proteína sin células con los siguientes reactivos:

30 i) un polinucleótido que tiene una región codificante que codifica la proteína de interés y que incluye un codón de supresión ubicado selectivamente dentro de su región codificante;

ii) metilazido-L-fenilalanina en una concentración suficiente para permitir la incorporación selectiva de pAMF en la proteína de interés;

35 iii) un ARNt que puede cargarse con pAMF y complementario al codón de supresión de la proteína de interés; y

40 iv) una aminoacil-ARNt sintetasa (RS), en el que la RS: aminoacila preferentemente hasta un grado de más de un 90 %, un ARNt con para-metilazido-L-fenilalanina (pAMF) en comparación con los 20 aminoácidos de origen natural comunes; tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1 (WT); y usando la SEQ ID NO: 1 como secuencia de referencia, tiene: 1) en la posición del aminoácido 65: A o V; 2) en la posición del aminoácido 108: Y o W; 3) en la posición del aminoácido 158: A; 4) en la posición del aminoácido 32: T o V o A; y 5) en la posición del aminoácido 159: S o G o V; y

45 b) incubar la combinación de la etapa (a) en condiciones que permiten la incorporación selectiva de pAMF en la proteína de interés.

En algunas realizaciones, el método tiene además una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I.

50 En algunas realizaciones, la RS tiene una identidad de secuencia de más de un 80 %, por ejemplo, de un 81 %, de un 82 %, de un 83 %, de un 84 %, de un 85 %, de un 86 %, de un 87 %, de un 88 %, de un 89 %, de un 90 %, de un 91 %, de un 92 %, de un 92 %, de un 93 %, de un 94 %, de un 95 %, de un 96 %, de un 97 %, de un 98 %, de un 99 % o 100 % de la SEQ ID NO: 1. En algunas realizaciones, la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y tiene sustituciones de aminoácidos seleccionadas del grupo que consiste en: 1) Y32T, L65A, F108Y, Q109L e I159S; ii) Y32V, L65A, F108W, Q109M e I159G; iii) Y32A, L65V, F108W e I159G; iv) Y32V, L65V, F108Y e I159V; v) Y32T, L65V, F108W e I159S; y vi) Y32V, L65V, F108W, G109I e I159S.

En algunas realizaciones, el extracto sin células tiene un sistema de fosforilación oxidativa activo.

60 En algunas realizaciones, la proteína de interés es un anticuerpo o fragmento de anticuerpo.

En algunas realizaciones, el método comprende además la etapa de conjugar un aducto biológicamente activo con la pAMF de la proteína de interés. En algunos casos, la conjugación es mediante 1,3-cicloadiación de una azida con un alquino sobrecargado.

65 Otros objetivos, características y ventajas de la presente invención serán evidentes para los expertos en la materia a partir de la siguiente descripción detallada y las figuras.

### Breve descripción de los dibujos

La fig. 1 muestra el porcentaje de expresión de GFP en comparación con el nivel de proteína RS usado en el ensayo de supresión ámbar de GFP K49TAG para las muteínas A01, A04, B03, C02, C10 y D08. Las sustituciones de aminoácidos respecto a la TyrRS de *M. jannaschii* de tipo silvestre para cada una de las muteínas se presentan en la fig. 3A.

Las fig. 2A-C muestran la cinética de conjugación de click sin cobre. La fig. 2A muestra la estructura química de para-azido-L-fenilalanina (izquierda) y para-azidometil-L-fenilalanina (derecha). La fig. 2B muestra la cinética de la formación del producto conjugado. La fig. 2C muestra la reacción de click sin cobre.

Las fig. 3A-B muestran que se usaron variante de RS específicas de pAMF para producir potentes conjugados de anticuerpo-fármaco. La fig. 3A muestra que ADC de trastuzumab (trastuzumab-HC136-MMAF) generados con variantes de RS específicas de pAMF tienen altos valores DAR. La fig. 3B muestra la actividad citotóxica para diferentes ADC de trastuzumab generados usando una variante de RS particular.

### Descripción detallada de la invención

En este documento se proporcionan composiciones y métodos para la incorporación específica de sitio del aminoácido no natural para-metilazido-L-fenilalanina (pAMF) en proteínas de interés usando un sistema de síntesis sin células. Además, en este documento se proporciona una composición de una variante de la tirosil ARNt sintetasa TyrRS) de *Methanococcus jannaschii* que tiene actividad y especificidad aumentada por pAMF en comparación con los aminoácidos de origen natural comunes.

#### I. Definiciones

Salvo que se definan de otro modo, los términos técnicos y científicos usados en este documento tienen el mismo significado que el comprendido habitualmente por un experto en la materia. Véase, por ejemplo, Lackie, DICTIONARY OF CELL AND MOLECULAR BIOLOGY, Elsevier (4.º ed. 2007); Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL, Cold Springs Harbor Press (Cold Springs Harbor, NY 1989); Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley y Sons (Hoboken, NY 1995).

Como se usa en este documento, las formas singulares "un", "una", y "el", "la", incluyen referencias en plural salvo que el contexto indique claramente lo contrario. Por tanto, por ejemplo, una referencia a "una célula" incluye una pluralidad de dichas células y una referencia a "la proteína" incluye una referencia a una o más proteínas y equivalentes de las mismas conocidas para los expertos en la materia, y así sucesivamente. Todos los términos técnicos y científicos usados en este documento tienen el mismo significado que el comprendido habitualmente por un experto en la materia a la que pertenece esta invención, salvo que se indique claramente lo contrario.

El término "aminoacilación" o "aminoacilar" se refiere al proceso completo en que un ARNt se carga con su aminoácido correcto, que es un resultado de añadir un grupo aminoacilo a un compuesto. En lo que respecta a esta invención, un ARNt que experimenta aminoacilación o se ha aminoacilado es uno que se ha cargado con un aminoácido, y un aminoácido que experimenta aminoacilación o se ha aminoacilado es uno que se ha cargado en una molécula de ARNt.

Las expresiones "aminoacil-ARNt sintetasa" o "ARNt sintetasa" o "sintetasa" o "aaRS" o "RS" se refieren a una enzima que cataliza un enlace covalente entre un aminoácido y una molécula de ARNt. Esto produce una molécula de ARNt aminoacilada, que es una molécula de ARNt que tiene su aminoácido respectivo adherido mediante un enlace éster.

El término "cargado" es el contexto de ARNt se refiere a la aminoacilación de un ARNt con un aminoácido, tanto natural como no natural, donde la aminoacilación permite que un ribosoma incorpore el aminoácido en un polipéptido que se está traduciendo desde el ARNm.

La expresión "aducto biológicamente activo" se refiere a un agente químico, molécula o reactivo que puede realizar una función en una célula o en un organismo. Por ejemplo, la función puede incluir proliferación celular, apoptosis, modificación postraduccional (por ejemplo, fosforilación), activación de la señalización celular, inactivación de la señalización celular, muerte celular, marcaje celular, etc.

La expresión "incorporación selectiva" en el contexto de la traducción de proteínas se refiere a incluir o introducir un aminoácido específico (por ejemplo, un aminoácido no natural específico) en una posición de aminoácido deseada predeterminada, en la secuencia de la proteína sin alterar la función deseada de la proteína.

La expresión "concentración suficiente para permitir la incorporación selectiva" se refiere a una concentración mínima de un componente (por ejemplo, RS, aminoácido no natural, ARNt supresión ámbar) necesario para la incorporación específica de sitio del aminoácido no natural en una proteína de interés.

- La expresión "aminoacilar preferentemente" se refiere a la preferencia de una ARNt sintetasa por aminoacilar (cargar) una molécula de ARNt particular con una molécula de aminoácido predeterminada en comparación con otra molécula de aminoácido. En otras palabras, la ARNt sintetasa puede aminoacilar selectivamente un aminoácido no natural (nnAA) sobre un aminoácido de origen natural. Por ejemplo, la ARNt sintetasa puede aminoacilar un nnAA específico a una frecuencia de más de un 90 %, por ejemplo, de un 91 %, de un 92 %, de un 93 %, de un 94 %, de un 95 %, de un 96 %, de un 97 %, de un 98 %, de un 99 % o de un 100 %, en comparación con todos y cada uno de los demás aminoácidos naturales.
- La expresión "aminoácido de origen natural" se refiere a uno cualquiera de los 20 aminoácidos codificados por el código genético, tales como, (arginina, Arg, R; histidina, His, H; lisina, Lys, K; ácido aspártico, Asp, D; ácido glutámico, Glu, E; serina, S, Ser; treonina, Thr, T; asparagina, Asn, N; glutamina, Gln, Q; cisteína, Cys, G; glicina, Gli, G; prolina, Pro, P; alanina, Ala, A; isoleucina, Ile, I; leucina, Leu, L; metionina, Met, M; fenilalanina, Phe, F; triptófano, Trp, W; tirosina, Tyr, Y y valina, Val, V, que son precursores de las proteínas.
- Los aminoácidos pueden mencionarse en este documento por los símbolos de tres letras habitualmente conocidos o por los símbolos de una letra recomendados por la Comisión de Nomenclatura bioquímica IUPAC-IUB. Los nucleótidos, asimismo, pueden mencionarse por sus códigos de una letra habitualmente aceptados.
- La expresión "ácido nucleico" o "polinucleótido" se refiere a ácidos desoxirribonucleicos (ADN) o ácidos ribonucleicos (ARN) y polímeros de los mismos en forma monocatenaria o bicatenaria. Salvo que se limite específicamente, la expresión abarca ácidos nucleicos que contienen análogos conocidos de nucleótidos naturales que tienen propiedades de unión similares al ácido nucleico de referencia y se metabolizan de una manera similar a los nucleótidos de origen natural. Salvo que se indique de otro modo, una secuencia de ácido nucleico particular también abarca de forma implícita variantes modificadas de forma conservativa de la misma (por ejemplo, sustituciones por codones degenerados), alelos, ortólogos, SNP y secuencias complementarias, así como la secuencia indicada de forma explícita. Específicamente, las sustituciones de codones degenerados pueden conseguirse generando secuencias en que la tercera posición de uno o más codones seleccionados (o todos) está sustituido con restos de base mixta y/o desoxiinosina (Batzet et al., *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka et al., *J. Biol. Chem.* 260:2605-2608 (1985); y Rossolini et al., *Mol. Cell. Probes* 8:91-98 (1994)). La expresión ácido nucleico o polinucleótido se usa indistintamente con gen, ADNc y ARNm codificado por un gen.
- El término "péptido", "proteína" y "polipéptido" se usan indistintamente en este documento y se refieren a un polímero de restos de aminoácidos. Los términos se aplican a los polímeros de aminoácidos en donde uno más restos de aminoácidos es un mimético químico artificial de un aminoácido correspondiente de origen natural, así como a polímeros de aminoácidos de origen natural y a polímeros de aminoácidos de origen no natural. Como se usa en este documento, los términos abarcan cadenas de aminoácidos de cualquier longitud, incluyendo proteínas de longitud completa y proteínas truncadas, en las que los restos de aminoácido están unidos por enlaces peptídicos covalentes.
- El término "muteína" se refiere a una proteína que tiene una sustitución de aminoácidos respecto a la secuencia de aminoácidos de tipo silvestre o de referencia para la proteína.
- La expresión "identidad de secuencia" o "porcentaje de identidad" en el contexto de dos o más ácidos nucleicos o secuencias polipeptídicas, se refieren a dos o más secuencias que son iguales o tienen un porcentaje especificado de restos de aminoácido o nucleótidos que son iguales, cuando se comparan y se alinean para una máxima correspondencia sobre una ventana de comparación, medida usando un algoritmo de comparación de secuencias, por ejemplo, BLASTP. Para los propósitos de este documento, el porcentaje de identidad se determina sobre la secuencia de tipo silvestre de longitud completa, tal como la secuencia de referencia expuesta en la SEQ ID NO: 1. El método para calcular la identidad de secuencia proporcionado en este documento es el programa BLASTP que tiene sus valores por defecto establecidos en un longitud de palabra (W) de 3, una expectativa (E) de 10, y la matriz de valores BLOSUM62 (véase, por ejemplo, Henikoff y Henikoff, 1989, *Proc Natl Acad Sci USA* 89:10915). Una determinación ejemplar de la alineación de secuencias y el % de identidad de secuencia puede emplear los programas BESTFIT o GAP en el paquete del programa informático GCG Wisconsin (Accelrys, Madison WI), usando los parámetros por defecto proporcionados.
- La expresión "sustitución en la posición del aminoácido" se refiere a un cambio de un resto de aminoácido en una posición específica de la secuencia de aminoácidos de una proteína. Por ejemplo, el término "X20Y" se refiere al remplazo del aminoácido X de tipo silvestre (de referencia) en la posición del aminoácido 20 de la proteína con el aminoácido Y.
- La expresión "codón de supresión" se refiere a un triplete de nucleótidos que se introduce en un polinucleótido en una ubicación predeterminada y se reconoce por un ARNt específico que puede reconocer un codón de parada (por ejemplo, un codón de parada ámbar, ocre u ópalo) y permite que la traducción lea más allá del codón para producir la proteína, suprimiendo de ese modo el codón de parada.
- La expresión "proteína biológicamente activa" se refiere a una proteína que retiene al menos algo de la actividad biológica de la proteína de interés. La actividad biológica puede determinarse comparando la actividad, la función y/o

la estructura de la proteína de interés expresada por los métodos descritos en este documento con la actividad de una proteína de interés de referencia. Por ejemplo, si la proteína de interés de referencia es una IgG, la proteína biológicamente activa comprenderá una molécula IgG plegada y ensamblada apropiadamente. En algunas realizaciones, la proteína de referencia puede ser una proteína expresada por un sistema bacteriano de síntesis sin células que no contiene una chaperona de proteínas exógena. La actividad biológica puede determinarse usando un ensayo *in vitro* o *in vivo* que es apropiado para la proteína de interés. La actividad biológica de la proteína de interés puede expresarse como la actividad biológica por unidad de volumen de la mezcla de reacción de síntesis de proteína sin células. En algunas realizaciones, la actividad biológica de una proteína producida por los métodos descritos en este documento es al menos un 30 %, un 35 %, un 40 %, un 45 %, un 50 %, un 55 %, un 60 %, un 65 %, un 70 %, un 70 %, un 80 %, un 85 %, un 90 %, un 91 %, un 92 %, un 93 %, un 94 %, un 95 %, un 96 %, un 97 %, un 98 %, un 99 % o un 100 % de la actividad de una proteína de referencia.

El término "anticuerpo" se refiere a una proteína definida funcionalmente como una proteína de unión y definida estructuralmente como comprendiendo una secuencia de aminoácidos que se reconoce por un experto en la materia como obtenida de la región flanqueante de un gen codificante de inmunoglobulina de un animal que produce anticuerpos. Un anticuerpo puede consistir en uno o más polipéptidos sustancialmente codificados por genes de inmunoglobulina o fragmentos de genes de inmunoglobulina. Los genes de inmunoglobulina reconocidos incluyen los genes de la región constante kappa, lambda, alfa, gamma, delta, épsilon y mu, así como una miríada de genes de la región variable de inmunoglobulina. Las cadenas ligeras se clasifican como kappa o lambda. Las cadenas pesadas se clasifican como gamma, mu, alfa, delta o épsilon, que a su vez definen las clases de la inmunoglobulina, IgG, IgM, IgA, IgD e IgE, respectivamente.

Se sabe que una unidad estructural de inmunoglobulina (anticuerpo) típica comprende un tetrámero. Cada tetrámero está compuesto de dos pares idénticos de cadenas polipeptídicas, teniendo cada par una cadena "ligera" (aproximadamente 25 kD) y una cadena "pesada" (aproximadamente 50-70 kD). El extremo N de cada cadena define una región variable de aproximadamente 100 a 110 o más aminoácidos responsables principalmente del reconocimiento del antígeno. Las expresiones cadena ligera variable (VL) y cadena pesada variable (VH) se refieren a estas cadenas ligeras y pesadas, respectivamente.

Los anticuerpos existen como inmunoglobulinas intactas o como varios fragmentos bien caracterizados producidos por digestión con diversas peptidasas. Por tanto, por ejemplo, la pepsina digiere un anticuerpo por debajo de los enlaces disulfuro en la región de bisagra para producir F(ab)'<sub>2</sub>, un dímero de Fab que es en sí mismo una cadena ligera unida a VH-CH1 por un enlace disulfuro. El F(ab)'<sub>2</sub> puede reducirse en condiciones moderadas para romper el enlace disulfuro en la región de bisagra convirtiendo de ese modo el dímero (Fab)'<sub>2</sub> en un monómero Fab'. El monómero de Fab' es esencialmente un Fab con parte de la región de bisagra (véase, *Fundamental Immunology*, W.E. Paul, ed., Raven Press, N. Y. (1993), para una descripción más detallada de los fragmentos de anticuerpo). Aunque diversos fragmentos de anticuerpos se definen en términos de la digestión de un anticuerpo intacto, un experto apreciará que dichos fragmentos Fab' pueden sintetizarse *de novo* de forma química o utilizando metodología de ADN recombinante. Por tanto, el término "anticuerpo", como se usa en este documento, también incluye fragmentos de anticuerpo producidos por la modificación de anticuerpos completos o sintetizados *de novo* usando metodologías de ADN recombinante. Los anticuerpos también incluyen anticuerpos de cadena sencilla (anticuerpos que existen como una única cadena polipeptídica) y anticuerpos Fv de cadena sencilla (sFv o scFv) en que una cadena pesada variable y una cadena ligera variable están unidas juntas (directamente o mediante un conector peptídico) para formar un polipéptido continuo. El anticuerpo Fv de cadena sencilla es un heterodímero VH-VL unido covalentemente que puede expresarse a partir de un ácido nucleico que incluye secuencias codificantes de VH y VL unidas directamente o unidas mediante un conector codificante del péptido. Huston, et al., (1988) *Proc. Nat. Acad. Sci. USA*, 85: 5879-5883. Aunque la VH y VL están conectadas entre sí como una única cadena polipeptídica, los dominios VH y VL se asocian de forma no covalente. Las primeras moléculas de anticuerpo funcional a expresar sobre la superficie de fagos filamentosos fueron Fv de cadena sencilla (scFv); sin embargo, estrategias de expresión alternativas también han sido satisfactorias. Por ejemplo, pueden presentarse moléculas Fab en fagos si una de las cadenas (pesada o ligera) se fusiona a la proteína de la cápsida g3 y la cadena complementaria se exporta al periplasma como una molécula soluble. Las dos cadenas pueden estar codificadas en el mismo replicón o en replicones diferentes; el punto importante es que las dos cadenas de anticuerpo en cada molécula Fab se ensamblan de forma postraducciona y el dímero se incorpora en la partícula fágica mediante unión de una de las cadenas a g3p (véase, por ejemplo, la patente de Estados Unidos n.º 5733743). Los fragmentos scFv y otras varias estructuras que convierten las cadenas polipeptídicas ligera y pesada agregadas, pero químicamente separadas de la región V del anticuerpo en una molécula que se pliega en una estructura tridimensional sustancialmente similar a la estructura de un sitio de unión a antígeno son conocidos para los expertos en la materia (véanse, por ejemplo, las patentes de los Estados Unidos n.º 5.091.513 5.132.405 y 4.956.778). Los anticuerpos también incluyen todos aquellos que se han presentado en fagos (por ejemplo, scFv, Fv, Fab y Fv unidos por disulfuro (Reiter et al., (1995) *Protein Eng.* 8: 1323-1331). Los anticuerpos también pueden incluir diacuerpos, minicuerpos y fusiones scFv-Fc.

Como se usa en este documento, la expresión "fragmento Fab" es un fragmento de anticuerpo que contiene la parte del anticuerpo de longitud completa que provoca la digestión de una inmunoglobulina de longitud completa con papaína, o un fragmento que tiene la misma estructura que se produce sintéticamente, por ejemplo, de forma recombinante. Un fragmento Fab contiene una cadena ligera (que contiene un dominio de región variable (VL) y

constante ( $C_L$ ) y otra cadena que contiene un dominio variable de una cadena pesada ( $V_H$ ) y una parte del dominio de la región constante de la cadena pesada ( $C_{H1}$ ).

5 Como se usa en este documento, un fragmento  $F(ab')_2$  es un fragmento de anticuerpo producido por digestión de una inmunoglobulina con pepsina a pH 4,0-4,5, o un anticuerpo producido de forma sintética, por ejemplo, de forma recombinante, que tiene la misma estructura. El fragmento  $F(ab')_2$  contiene dos fragmentos Fab, pero donde cada parte de cadena pesada contiene unos pocos aminoácidos adicionales, incluyendo restos de cisteína para formar enlaces disulfuro que unen los dos fragmentos.

10 La expresión "extracto sin células derivado de bacterias" se refiere a la preparación de mezclas de reacción *in vitro* que pueden transcribir ADN en ARNm y/o traducir el ARNm en polipéptidos. Las mezclas incluyen ribosomas, ATP, aminoácidos y ARNt. Pueden obtenerse directamente de bacterias lisadas, de componentes purificados o combinaciones de los mismos.

15 La expresión "sistema de síntesis sin células" o "sistema CFPS" se refiere a la síntesis *in vitro* de polipéptidos en una mezcla de reacción que comprende extractos biológicos y/o reactivos definidos. La mezcla de reacción comprenderá un molde para la producción de la macromolécula, por ejemplo, ADN, ARNm, etc.; monómeros para la macromolécula a sintetizar, por ejemplo, aminoácidos, nucleótidos, etc.; y cofactores, enzimas y otros reactivos que son necesarios para la síntesis, por ejemplo, ribosomas, ARNt no cargados, ARNt cargados con aminoácidos no naturales, polimerasas, factores de transcripción, ARNt sintetasas, etc.

20 La expresión "sistema de fosforilación oxidativa activo" se refiere a un lisado bacteriano que muestra fosforilación oxidativa activa durante la síntesis de proteínas. Por ejemplo, el lisado bacteriano puede generar ATP usando enzimas ATP sintasa y reducción de oxígeno. Se entenderá que otros sistemas de traducción conocidos en la técnica también pueden usar una fosforilación oxidativa activa durante la síntesis de proteínas. La activación de la fosforilación oxidativa puede demostrarse mediante la inhibición de la ruta usando inhibidores específicos, tales como inhibidores de la cadena de transporte de electrones.

## 30 II. Descripción detallada de realizaciones

En este documento se proporcionan composiciones y métodos de producción de una variante de RS que contiene sustituciones de aminoácido en la tirosil ARNt sintetasa de *M. jannaschii* expuesta en la SEQ ID NO: 1. Las variantes de RS tienen alta actividad y especificidad aumentada para el aminoácido no natural pAMF en comparación con los aminoácidos de origen natural comunes. También se proporcionan métodos de introducción de un codón supresor en un polinucleótido que codifica una proteína de interés, de modo que pAMF se incorpore en una ubicación específica en la proteína de interés. Además, se describen métodos para producir dicha proteína de interés en un sistema de síntesis de proteína sin células. Además, se describen métodos para conjugar un aducto biológicamente activo a la proteína de interés mediante el aminoácido no natural. Se encuentran realizaciones ejemplares de producción de: 1) una variante de RS, 2) una proteína diana con pAMF en la ubicación deseada, y 3) una proteína diana que contiene un pAMF específico de sitio conjugado con un aducto biológicamente activo en los ejemplos.

### 40 A. Métodos generales

Salvo que se definan de otro modo, todos los términos técnicos y científicos usados en este documento tienen el significado habitualmente comprendido por un experto en la materia a la que pertenece esta invención. A los facultativos se les remite particularmente a Green, M. R. y Sambrook, J., eds., *Molecular Cloning: A Laboratory Manual*, 4ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y. (2012), y Ausubel, F. M., *et al.*, *Current Protocols in Molecular Biology* (Suplemento 99), John Wiley & Sons, Nueva York (2012) para las definiciones y términos de la técnica. También aparecen métodos convencionales en Bindereif, Schön, y Westhof (2005) "Handbook of RNA Biochemistry", Wiley-VCH, Weinheim, Alemania, que describe métodos detallados para la manipulación y el análisis de ARN. Los ejemplos de técnicas moleculares apropiadas para generar ácidos nucleicos recombinantes, e instrucciones suficientes para dirigir a los expertos en la materia a través de muchos ejercicios de clonación, se encuentran en Green, M. R. y Sambrook, J., (Id.); Ausubel, F. M., *et al.*, (Id.); Berger y Kimmel, "Guide to Molecular Cloning Techniques", *Methods in Enzymology* (Volumen 152 Academic Press, Inc., San Diego, Calif. 1987); y "PCR Protocols: A Guide to Methods and Applications" (Academic Press, San Diego, Calif. 1990).

Los métodos para la purificación de proteínas, cromatografía, electroforesis, centrifugación y cristalización se describen en Coligan *et al.* (2000) "Current Protocols in Protein Science", Vol. 1, John Wiley and Sons, Inc., Nueva York. Los métodos para la síntesis sin células se describen en Spirin y Swartz (2008) "Cell-free Protein Synthesis", Wiley-VCH, Weinheim, Alemania. Los métodos para la incorporación de aminoácidos no naturales en proteínas usando síntesis sin células se describen en Shimizu *et al.*, (2006) *FEBS Journal*, 273, 4133-4140.

Los métodos de amplificación por PCR son bien conocidos en la técnica y se describen, por ejemplo, en Innis *et al.*, "PCR Protocols: A Guide to Methods and Applications", Academic Press Inc. San Diego, Calif., 1990. Una reacción de amplificación típicamente incluye el ADN que tiene que amplificarse, una ADN polimerasa termoestable, dos cebadores oligonucleotídicos, trifosfatos de desoxinucleótido (dNTP), tampón de reacción y magnesio. Típicamente,

una cantidad deseable de ciclos térmicos es entre 1 y 25. Los métodos de diseño de cebadores y optimización de las condiciones de PCR son bien conocidos en la técnica y pueden encontrarse en textos convencionales de biología molecular tales como Ausubel et al., *Short Protocols in Molecular Biology*, 5ª edición, Wiley, 2002, e Innis et al., *PCR Protocols*, Academic Press, 1990. Los programas informáticos son útiles en el diseño de cebadores con la especificidad requerida y las propiedades de amplificación óptimas (por ejemplo, Oligo versión 5.0 (National Biosciences)). En algunas realizaciones, los cebadores de PCR pueden contener además sitios de reconocimiento para endonucleasas de restricción, para facilitar la inserción del fragmento de ADN amplificado en sitios de enzima de restricción específicos en un vector. Si los sitios de restricción tienen que añadirse en el extremo 5' de los cebadores de PCR, es preferible incluir unas pocas (por ejemplo, dos o tres) bases 5' adicionales para permitir una escisión más eficaz por la enzima. En algunas realizaciones, los cebadores de PCR también pueden contener un sitio promotor de ARN polimerasa, tales como T7 o SP6, para permitir la posterior transcripción *in vitro*. Los métodos para la transcripción *in vitro* son bien conocidos para los expertos en la materia (véanse, por ejemplo, Van Gelder et al., *Proc. Natl. Acad. Sci. U.S.A.* 87:1663-1667, 1990; Eberwine et al., *Proc. Natl. Acad. Sci. U.S.A.* 89:3010-3014, 1992).

## 15 B. Introducción de sustituciones de aminoácido en la aminoacil-ARNt sintetasa de tipo silvestre.

Sorprendentemente, se ha determinado que sustituciones de aminoácido específicas en la tirosil ARNt sintetasa de *M. jannaschii* de tipo silvestre permite incorporar de forma preferente o selectiva para-metilazido-L-fenilalanina (pAMF), pero no cualquiera de los aminoácidos de origen natural comunes en proteínas emergentes durante la síntesis de proteína basada en células o sin células. Además, la sintetasa no aminoacila ningún ARNt de *E. coli* endógeno con tirosina, pero puede aminoacilar un ARNt supresor ámbra mutante. Las sustituciones de aminoácido, tales como L65A, L65V, F108Y, F108W, D158A, Y32T, Y32V, Y32A, I159S, I159G, I159V, Q109L, Q109M y Q109I pueden hacerse mutando la secuencia codificante de la tirosil ARNt sintetasa de *M. jannaschii* de tipo silvestre expuesta en la SEQ ID NO: 1. La variante RS se genera a partir de la secuencia de ADN de la tirosil ARNt sintetasa de tipo silvestre o la parte de la misma que contiene el marco abierto de lectura, con cambios realizados según lo necesario en los codones correspondientes a las sustituciones descritas en este documento.

La secuencia de aminoácidos de una variante de RS específica proporciona una descripción de todos los polinucleótidos que pueden codificar la variante de RS a causa de la correspondencia conocida entre los aminoácidos y el código genético. Para la mayoría de los organismos, el código genético es "aminoácido (código de una letra) [codones]": fenilalanina (F) [TTT, TTC]; leucina (L) [TTA, TTG, CTT, CTC, CTA, CTG]; isoleucina (I) [ATT, ATC, ATA]; metionina (M) [ATG]; valina (V) [TGG, GTC, GTA, GTG]; serina(S) [TCT, TCC, TCA, TCG, AGT, AGC]; prolina (P) [CCT, CCC, CCA, CCG]; treonina (T) [ACT, ACC, ACA, ACG]; alanina (A) [GCT, GCC, GCA, GCG]; tirosina (Y) [TAT, TAC]; histidina (H) [CAT, CAC]; glutamina (Q) [CAA, CAG]; asparagina (N) [AAT, AAC]; lisina (K) AAA, AAG; ácido aspártico (D) [GAT, GAC]; ácido glutámico (E) [GAA, GAG]; cisteína (C) [TGT, TGC]; triptófano (W) [TGG]; arginina (R) [CGT, CGC, CGA, CGG, AGA, AGG]; y glicina (G) [GGT, GGC, GGA, GGG].

En algunos casos, la secuencia de aminoácidos puede diferir en uno o más aminoácidos de los de las proteínas RS proporcionadas en este documento como resultado de una o más de las sustituciones de aminoácido conservativas bien conocidas. Las sustituciones conservativas para un aminoácido dentro de la secuencia de la proteína RS proporcionada en este documento pueden seleccionarse de otros miembros de una clase a la que pertenece el aminoácido de origen natural. Los aminoácidos representativos dentro de estas diversas clases incluyen, aunque sin limitación: (1) aminoácidos ácidos (cargados negativamente) tales como ácido aspártico y ácido glutámico; (2) aminoácidos básicos (cargados positivamente) tales como arginina, histidina y lisina; (3) aminoácidos polares neutros tales como glicina, serina, treonina, cisteína, tirosina, asparagina y glutamina; y (4) aminoácidos apolares neutros (hidrófobos) tales como alanina, leucina, isoleucina, valina, prolina, fenilalanina, triptófano y metionina. Los sustitutos conservados para un aminoácido dentro de una secuencia de aminoácidos natural pueden seleccionarse de otros miembros del grupo al que pertenece el aminoácido de origen natural. Por ejemplo, un grupo de aminoácidos que tiene cadenas laterales alifáticas es glicina, alanina, valina, leucina e isoleucina; un grupo de aminoácidos que tiene cadenas laterales de hidroxilo alifáticas es serina y treonina; un grupo de aminoácidos que tiene cadenas laterales que contienen amida es asparagina y glutamina; un grupo de aminoácidos que tiene cadenas laterales aromáticas es fenilalanina, tirosina y triptófano; un grupo de aminoácidos que tiene cadenas laterales básicas es lisina, arginina e histidina; y un grupo de aminoácidos que tiene cadenas laterales que contienen sulfuro es cisteína y metionina. Cada uno de los siguientes ocho grupos contiene otros aminoácidos ejemplares que son sustituciones conservativas entre sí:

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

(véase, por ejemplo, Creighton, *Proteins* (1984)).

Por ejemplo, las variantes de RS proporcionadas en este documento pueden contener sustituciones de aminoácido conservativas, tales como, aunque sin limitación, una variante de RS con las siguientes sustituciones de aminoácido:

- 5 i) M6L, Y32A, L65V, F108W e I159G;
- ii) M6L, Y32V, L65V, F108Y e I159V;
- iii) M6L, Y32T, L65V, F108W e I159S;
- iv) N10Q, Y32A, L65V, F108W e I159G;
- v) N10Q, Y32V, L65V, F108Y e I159V;
- 10 vi) N10Q, Y32T, L65V, F108W e I159S;
- vii) I14L, Y32A, L65V, F108W e I159G;
- viii) I14L, Y32V, L65V, F108Y e I159V;
- ix) I14L, Y32T, L65V, F108W e I159S;
- x) S16T, Y32A, L65V, F108W e I159G;
- xi) S16T, Y32V, L65V, F108Y e I159V;
- 15 xii) S16T, Y32T, L65V, F108W e I159S, y similares.

Las variantes de RS de esta invención se definen además por su capacidad de unirse a anticuerpos policlonales generados contra la TyrRS de tipos silvestre que tiene la secuencia de aminoácidos de la SEQ ID NO: 1. En condiciones de inmunoensayo indicadas, las variantes de RS de esta invención incluyen aquellas definidas por su función, es decir, su capacidad de incorporar selectivamente el aminoácido no natural y por su capacidad de unirse también a los anticuerpos policlonales especificados a una tasa de al menos dos veces por encima del fondo. Típicamente, una reacción específica o selectiva será al menos dos veces la señal o ruido de fondo y, más típicamente, de más de 10 a 100 veces la de fondo. Puede usarse una diversidad de formatos de inmunoensayo para determinar si los anticuerpos policlonales de ensayo reaccionan con una aaRS de esta invención. Véase, por ejemplo, Harlow y Lane, Antibodies, A Laboratory Manual (1988) para una descripción de formatos de inmunoensayo y condiciones que pueden usarse para determinar la inmunoreactividad específica.

Los métodos para generar una secuencia codificante que codifique una variante de RS con sustituciones de aminoácido incluyen técnicas convencionales de clonación molecular tales como, aunque sin limitación, PCR, mutagénesis y clonación con enzimas de restricción. Como alternativa, la secuencia codificante de las variantes de RS puede producirse usando química sintética de acuerdo con métodos convencionales, por ejemplo, el método de triéster de fosforamidita en fase sólida.

Los expertos en la materia reconocen que están disponibles diversas técnicas de mutagénesis bien conocidas para generar las variantes de RS de esta invención. Por ejemplo, pueden introducirse mutaciones puntuales correspondientes a las sustituciones de aminoácido deseadas descritas en este documento en la secuencia codificante para la tirosil ARNt sintetasa de *M. jannaschii* de tipo silvestre mediante PCR solapante. Como alternativa, la secuencia codificante de tipo silvestre puede mutarse usando una técnica de mutagénesis basada en PCR, por ejemplo, mutagénesis dirigida al sitio, para mutar los codones correspondientes a las sustituciones de aminoácido deseadas de las variantes de RS. El polinucleótido deseado que codifica la variante de RS se usa para generar la proteína variante de RS.

Más específicamente, dichas técnicas de mutagénesis incluyen, por ejemplo, mutagénesis dirigida al sitio (por ejemplo, kit de mutagénesis dirigida al sitio Quikchangell™, Agilent Technologies), mutagénesis aleatoria (kit de mutagénesis aleatoria por PCR Diversify™, Clontech), recombinaciones homólogas, mutagénesis dirigida por oligonucleótido (por ejemplo, kit Transformer™, Clontech), mutagénesis de ADN modificado por fosforotioato, etc., que pueden usarse para generar una secuencia codificante correspondiente a sustituciones de aminoácido. Véase, por ejemplo, Ling, et al., "Approaches to DNA mutagenesis: an overview," Anal. Biochem., 254(2):157-78 (1997); Dale, et al., "Oligonucleotide-directed random mutagenesis using the phosphorothioate method," Methods Mol. Biol., 57:369-74 (1996); Smith, "In vitro mutagenesis," Ann. Rev. Genet., 19:423-462 (1985); Botstein, et al., "Strategies and applications of in vitro mutagenesis," Science, 229:1193-1201 (1985); Carter, "Site-directed mutagenesis," Biochem. J., 237:1-7 (1986); Kramer, et al., "Point Mismatch Repair," Cell, 38:879-887 (1984); Wells, et al., "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites," Gene, 34:315-323 (1985); Minshull, et al., "Protein evolution by molecular breeding," Current Opinion in Chemical Biology, 3:284-290 (1999); Christians, et al., "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling," Nature Biotechnology, 17:259-264 (1999); Cramer, et al., "DNA shuffling of a family of genes from diverse species accelerates directed evolution," Nature, 391:288-291; Cramer, et al., "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," Nature Biotechnology, 15:436-438 (1997); Zhang, et al., "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening," Proceedings of the National Academy of Sciences, EE. UU., 94:45-4-4509; Cramer, et al., "Improved green fluorescent protein by molecular evolution using DNA shuffling," Nature Biotechnology, 14:315-319 (1996); Stemmer, "Rapid evolution of a protein in vitro by DNA shuffling," Nature, 370:389-391 (1994); Stemmer, "DNA shuffling by random fragmentation and reassembly: In vitro recombination," Proceedings of the National Academy of Sciences, EE. UU., 91:10747-10751 (1994); documento WO 95/22625; documento WO 97/0078; documento WO 97/35966; documento WO 98/27230; documento WO 00/42651; documento WO 01/75767; y el documento WO 2009/152336.

**C. Expresión de aminoacil-ARNt sintetasas específicas para metilazido-L-fenilalanina.**

Una vez obtenido el polinucleótido variante de RS deseado, puede usarse para generar el polipéptido correspondiente en un sistema de expresión basado en células o sin células.

5 En un sistema de síntesis de proteína basado en células, el polinucleótido que codifica la variante de RS se clona en un vector de expresión tal como un plásmido, fago, fagómido, cósmido, bacteriófago, vector de baculovirus, plásmido bacteriano y similares. El vector de expresión puede incluir el ácido nucleico de la presente invención que está unido de forma funcional a un promotor. Posteriormente, el vector de expresión se introduce en una célula hospedadora para que exprese la variante de RS. Puede usarse cualquier otro vector siempre que sea replicable y viable en la célula hospedadora. La célula hospedadora puede ser una célula bacteriana, una célula arqueobacteriana, una célula fúngica, una célula de levadura, una célula de insecto o una célula de mamífero. La variante de RS puede producirse en una célula y después purificarse del lisado celular.

15 En un sistema de síntesis de proteína sin células (CFPS), el polinucleótido que codifica la variante de RS sirve como molde polinucleotídico para la reacción. Otros componentes de la reacción incluyen uno o más extractos bacterianos y/o reactivos definidos. La mezcla de reacción también puede incluir al menos ATP o una fuente de energía, cofactores, enzimas y otros reactivos que son necesarios para la síntesis de polipéptidos, por ejemplo, ribosomas, ARNt, polimerasas, factores de transcripción, aminoacil sintetasas, factores de elongación, factores de inicio, etc. A  
20 continuación se describe una descripción adicional de sistemas de reacción CFPS ejemplares.

La proteína variante de RS puede purificarse después de la síntesis de la proteína de acuerdo con métodos convencionales conocidos en la técnica incluyendo, aunque sin limitación, cromatografía (por ejemplo, intercambio iónico, afinidad, hidrófobo, cromatofoco y de exclusión molecular), procedimientos electroforéticos (por ejemplo, isoelectroenfoque preparativo), solubilidad diferencial (por ejemplo, precipitación con sulfato de amonio), SDS-PAGE o extracción (véase, por ejemplo, Protein Purification, J.C. Janson y Lars Ryden, editores, VCH Publishers, Nueva York, 1989) para obtener polipéptidos sustancialmente puros.

30 La actividad y selectividad de la variante de RS para cargar el ARNt supresor con el aminoácido no natural pueden determinarse usando un gen indicador que contiene un codón supresor. La actividad de la variante de RS en presencia de pAMF en comparación con los 20 aminoácidos de origen natural comunes puede medirse detectando la presencia o la ausencia de la proteína indicadora. A continuación se proporciona una descripción de métodos para introducir un codón supresor en un polinucleótido que codifica una proteína diana tal como una proteína indicadora.

35 Puede usarse un indicador de selección positiva (por ejemplo, un indicador fluorescente, indicador luminiscente o indicador basado en afinidad) de modo que la supresión ámbar provoque una señal positiva detectable. Por ejemplo, puede introducirse un codón ámbar en el gen indicador (por ejemplo, gen de GFP). Si la variante de RS puede cargar el ARNt supresor en presencia de pAMF, se detecta GFP, indicando de ese modo que la variante de RS es específica para el codón ámbar y pAMF. Si la variante de RS no puede cargar el ARNt supresor en presencia de pAMF, no se traduce proteína GFP funcional debido al codón de parada prematura insertado en el gen indicador de GFP. Además, si la variante de RS puede cargar el ARNt supresor en ausencia de pAMF, la variante no es específica para el nAA. Véase, por ejemplo, el ejemplo 1 y Takimoto et al., Molecular Biosystems, 5:931-934 (2009). En otros casos, el marcador de selección puede ser un gen de resistencia que contiene un codón de parada ámbar en el gen, de modo que en presencia de un agente de selección (por ejemplo, antibiótico, anticuerpo, nutriente y similares), las células  
45 que contienen una variante de RS eficaz en la supresión de codones puede distinguirse de las que no tienen una variante de RS funcional.

**D. Introducción de codones supresores en polinucleótidos que codifican proteínas de interés.**

50 Una vez obtenida la proteína aaRS deseada, es necesario generar un polinucleótido diana que codifique una proteína diana con un sitio deseado para pAMF. En este documento se proporcionan métodos para introducir un codón supresor (por ejemplo, TAG ámbar, TAA ocre, TGA ópalo) que se reconozca por la variante de RS descrita en este documento en el polinucleótido que codifica una proteína de interés. Ejemplos no limitantes de una proteína de interés incluyen un anticuerpo, fragmento de anticuerpo, anticuerpo anti-idiotipo, anticuerpo quimérico, anticuerpo humanizado, proteína de fusión de anticuerpo, proteína secretada (por ejemplo, hormona), proteína transmembranaria (por ejemplo, receptor), enzima, proproteína, fragmento de proteína, proteína farmacéuticamente activa y una proteína que tiene valor industrial o terapéutico potencial. En algunas realizaciones, la proteína de interés es un anticuerpo o un fragmento de anticuerpo.

60 La región codificante del polinucleótido puede mutarse para introducir un codón supresor, por ejemplo, un codón ámbar TAG en el marco abierto de lectura del polinucleótido que codifica la proteína de interés, de modo que la variante de RS descrita en este documento puede suprimir selectivamente el codón introducido durante la traducción y generar la proteína de interés que contiene un aminoácido no natural en una ubicación deseada. La mutación puede introducirse para evitar cambiar el marco abierto de lectura de la proteína. Por ejemplo, el polinucleótido que codifica la proteína de interés se muta para insertar un codón supresor en la región codificante de la proteína en una posición  
65 seleccionada, de modo que la proteína de interés contenga el aminoácido no natural y se produzca con la ayuda del

par de variante de RS/ARNt supresor. La posición del codón supresor puede seleccionarse de acuerdo con la estructura primaria, secundaria, terciaria o cuaternaria de la proteína, la función de la proteína, el aminoácido no natural a incorporar en la proteína y/o el aducto biológicamente activo a conjugar con la proteína.

- 5 Los métodos para introducir un codón supresor en el polinucleótido que codifica la proteína de interés incluyen, aunque sin limitación, técnicas convencionales de biología molecular tales como PCR, clonación y mutagénesis. Como se describe anteriormente, las técnicas de mutagénesis útiles incluyen, aunque sin limitación, mutagénesis dirigida al sitio (por ejemplo, kit de mutagénesis dirigida al sitio QuikchangeII™, Agilent Technologies), mutagénesis aleatoria (kit de mutagénesis aleatoria por PCR Diversify™, Clontech), recombinaciones homólogas, mutagénesis dirigida por oligonucleótido (por ejemplo, kit Transformer™, Clontech), mutagénesis de ADN modificado por fosforotioato, etc., que pueden usarse para generar sustituciones de aminoácido, etc., que pueden usarse para generar sustituciones de aminoácido. Anteriormente se describen otras técnicas de mutagénesis.

- 15 El polinucleótido mutado que contiene el codón supresor puede clonarse en un vector de expresión tal como un plásmido, fago, fagómido, cósmido, bacteriófago, vector de baculovirus, plásmido bacteriano y similares. El vector de expresión puede incluir el ácido nucleico de la presente invención que está unido de forma funcional a un promotor. Posteriormente, el vector de expresión se introduce en una célula hospedadora para que expresa la proteína de interés modificada. Puede usarse cualquier otro vector siempre que sea replicable y viable en la célula hospedadora. La célula hospedadora puede ser una célula bacteriana, una célula arqueobacteriana, una célula fúngica, una célula de levadura, una célula de insecto o una célula de mamífero.

El polinucleótido mutado que codifica la proteína de interés puede usarse en una reacción de síntesis de proteína sin células (CFPS) para producir la proteína de interés con un aminoácido no natural pAMF específico de sitio.

## 25 E. Síntesis de proteína sin células (CFPS) de proteínas que contienen pAMF.

- Después de generar tanto la variante de RS deseada como el polinucleótido diana que codifica una proteína diana con un sitio deseado para pAMF, es posible sintetizar la proteína diana que contiene pAMF. En este documento se proporcionan métodos de síntesis de proteína sin células para incorporar para-metilizado-L-fenilalanina en la proteína de interés. En particular, la reacción de CFPS incluye para-metilizado-L-fenilalanina en una concentración suficiente para permitir la incorporación selectiva de pAMF en la proteína de interés, un polinucleótido que tiene una región codificante que codifica la proteína de interés y que incluye un codón de supresión ubicado selectivamente dentro de su región codificante, un ARNt que puede cargarse con pAMF y complementario al codón de supresión de la proteína de interés y la variante de RS proporcionada en este documento.

- 35 La cantidad de proteína variante de RS de *M. jannaschii* purificada en la reacción de CFPS puede estar en el intervalo de aproximadamente 1  $\mu$ M a aproximadamente 10  $\mu$ M, por ejemplo, 1  $\mu$ M, 2  $\mu$ M, 3  $\mu$ M, 4  $\mu$ M, 5  $\mu$ M, 6  $\mu$ M, 7  $\mu$ M, 8  $\mu$ M, 9  $\mu$ M o 10  $\mu$ M.

- 40 En algunos casos, la cantidad de la proteína variante de RS usada está en exceso de la cantidad necesaria para aminoacilar el ARNt supresor ámbar ortogonal. La cantidad de proteína variante de RS en la reacción de CFPS puede estar en el intervalo de aproximadamente 1  $\mu$ M a aproximadamente 10  $\mu$ M, por ejemplo, 1  $\mu$ M, 2  $\mu$ M, 3  $\mu$ M, 4  $\mu$ M, 5  $\mu$ M, 6  $\mu$ M, 7  $\mu$ M, 8  $\mu$ M, 9  $\mu$ M o 10  $\mu$ M.

- 45 La cantidad de para-metilizado-L-fenilalanina en la reacción de CFPS puede estar en el intervalo de aproximadamente 1 mM a aproximadamente 10 mM, por ejemplo, 1 mM, 2 mM, 3 mM, 4 mM, 5 mM, 6 mM, 7 mM, 8 mM, 9 mM o 10 mM.

- Los sistemas de reacción de síntesis de proteína sin células son bien conocidos en la técnica y se han descrito en, por ejemplo, Zawada et al., 2011, Biotechnol. Bioeng., 108(7):1570-1578; Shimizu et al., Nature Biotechnology, 2001, 19:751-755; y las patentes de Estados Unidos n.º 7.338.789 y 8.492.115.

- Los sistemas de síntesis de proteína sin células (CFPS) utilizan extractos celulares para dar soporte a las síntesis de proteína *in vitro* a partir de transcritos de ARNm purificados o a partir de ARNm transcritos a partir de ADN durante la reacción de síntesis *in vitro*. La mezcla de reacción puede catalizar la síntesis de polipéptidos a partir de un molde de ácido nucleico. La mezcla incluye uno o más extractos bacterianos y/o reactivos definidos. Los extractos bacterianos ejemplares incluyen, aunque sin limitación, extractos de *E. coli* S30, extracto de *E. coli* atenuado RF-1 sensible a OmpT y variantes de los mismos. Las células bacterianas del extracto pueden sobreexpresar un componente del sistema de CFPS tal como enzimas, por ejemplo, DsbA, PpiA, FkpA, DegP y SlyD, y ARNt (por ejemplo, un ARNt que codifica un supresor ortogonal tal como un ARNt que codifica CUA ortogonal).

- 60 Los métodos de preparación de un extracto celular para CFPS se describen en, por ejemplo, Zawada, J. "Preparation and Testing of E.coli S30 In Vitro Transcription Translation Extracts", Douthwaite, J.A. y Jackson, R.H. (eds.), Ribosome Display and Related Technologies: Methods and Protocols, Methods in Molecular Biology, vol. 805, pág. 31-41 (Humana Press, 2012); Jewett et al., Molecular Systems Biology: 4, 1-10 (2008); Shin J. y Norieaux V., J. Biol. Eng., 4:8 (2010). En resumen, se cultiva un cultivo bacteriano y se recoge; se suspende en un tampón apropiado (por ejemplo, tampón S30) y se homogeneiza para lisar las células.

La mezcla de reacción del sistema de CFPS también puede incluir al menos ATP o una fuente de energía; un molde para la producción de la macromolécula, por ejemplo, ADN, ARNm, etc.; aminoácidos y cofactores, enzimas y otros reactivos que son necesarios para la síntesis de la proteína de interés, por ejemplo, ribosomas, ARNt (incluyendo ARNt codificante de supresor ortogonal), polimerasas, factores de la transcripción, aminoacil sintetasas (por ejemplo, variante de RS proporcionada en este documento), chaperonas, factores de elongación, factores de inicio, etc.

Cuando el codón supresor es un codón ámbar, es útil para que la mezcla de reacción incluya un componente que puede atenuar la actividad del factor de liberación (RF) o tiene actividad disminuida, por ejemplo, un mutante RF-1, RF-2 o RF-3. Un experto en la materia reconoce que la eficacia de supresión del ARNt supresor ámbar depende de la competición con el factor de liberación 1 (RF1) para descodificar el codón ámbar por el ribosoma. RF1 actúa terminando la cadena polipeptídica en el codón de parada, mientras que el ARNt supresor ámbar compete con RF1 para incorporar el aminoácido no natural durante la síntesis de la proteína. Por tanto, disminuir la cantidad o la actividad de RF1 en el sistema de CFPS puede aumentar la eficacia de supresión de los ARNt supresores cargados. La actividad RF1 puede manipularse usando, por ejemplo, aptámeros inhibidores de RF1, anticuerpos contra RF1 y lisados celulares con RF-1 reducido. En algunos casos, RF1 se omite del sistema de CFPS.

El sistema de CFPS puede incluir una fuente de energía tal como una fuente de energía homeostática. También puede incluirse una o más enzimas que catalizan la regeneración de ATP a partir de enlaces fosfato de alta energía, por ejemplo, acetato cinasa, creatina cinasa, etc., mediante fosforilación oxidativa. Dichas enzimas pueden estar presentes en los extractos usados para la traducción, o pueden añadirse a la mezcla de reacción. Opcionalmente, puede añadirse un compuesto tal como un dinucleótido de nicotinamida adenina (NADH), NAD<sup>+</sup> o acetil-coenzima A para la activación de la fosforilación oxidativa.

Para sintetizar una proteína de interés *in vitro*, un extracto de CFPS en algún punto comprende una molécula de ARN que codifica la proteína de interés y contiene un codón de supresión ubicado selectivamente (por ejemplo, un codón ámbar). En algunos sistemas de CFPS, se añade ARNm de forma exógena después de purificarse de fuentes naturales o prepararse sintéticamente *in vitro* de ADN clonado usando ARN polimerasas tales como ARN polimerasa II, ARN polimerasa SP6, ARN polimerasa T3, ARN polimerasa T7, ARN polimerasa III y/o ARN polimerasas derivadas de fago. En otros sistemas, el ARNm se produce *in vitro* a partir de un polinucleótido de ADN molde; tanto la transcripción como la traducción se producen en este tipo de reacción de CFPS. El ADN molde contiene un codón de supresión ubicado selectivamente dentro de la región codificante. En otros sistemas más, los sistemas de transcripción y traducción están acoplados o se han desarrollado sistemas de transcripción y traducción complementarios, que realizan la síntesis tanto del ARN como de la proteína de interés en la misma reacción. En dichos sistemas de transcripción y traducción *in vitro*, los extractos de CFPS contienen todos los componentes (exógenos o endógenos) necesarios tanto para transcripción (para producir ARNm) como para la traducción (para sintetizar la proteína de interés) en un único sistema.

En algunas realizaciones, la reacción de CFPS se realiza usando el sistema Cytomim™ que comprende NTP, ARNt de *E. coli* y un ARNt que codifica un supresor ortogonal, aminoácidos y pAMF, acetato Mg<sup>2+</sup>, glutamato Mg<sup>2+</sup>, acetato K<sup>+</sup>, glutamato K<sup>+</sup>, ácido fólico, Tris pH 8,2, DTT, piruvato cinasa, ARN polimerasa T7, disulfuro isomerasa, piruvato de sodio, NAD, CoA, oxalato Na<sup>+</sup>, putrescina, espermidina y extracto bacteriano. En algunas realizaciones, el sustrato de energía para el sistema Cytomim™ es piruvato, ácido glutámico y/o glucosa. En algunas realizaciones del sistema, los trifosfatos de nucleósido (NTP) se reemplazan con monofosfatos de nucleósido (NMP).

En algunas realizaciones, la reacción de CFPS se realiza usando el sistema PANOx-SP que comprende NTP, ARNt y un ARNt que codifica un supresor ortogonal, aminoácidos y pAMF, acetato Mg<sup>2+</sup>, glutamato Mg<sup>2+</sup>, acetato K<sup>+</sup>, glutamato K<sup>+</sup>, ácido fólico, Tris pH 8,2, DTT, piruvato cinasa, ARN polimerasa T7, disulfuro isomerasa, fosfoenol piruvato (PEP), NAD, CoA, oxalato Na<sup>+</sup>, putrescina, espermidina y extracto S30.

Para incorporar una para-metilazido-L-fenilalanina en una proteína de interés, el sistema de CFPS incluye la proteína variante de RS de *M. jannaschii* de la presente invención, un ARNt que puede cargarse con pAMF y es complementario al codón de supresión de la proteína de interés, un polinucleótido que tiene una región codificante que codifica la proteína de interés y que alberga un codón supresor (por ejemplo, codón ámbar) ubicado selectivamente dentro de la región codificante, y para-metilazido-L-fenilalanina. El aminoácido no natural puede ser sintético u obtenerse de otra fuente biológica. Por ejemplo, puede sintetizarse para-metilazido-L-fenilalanina de acuerdo con el método descrito en el ejemplo 2.

La proteína de interés que alberga el aminoácido no natural puede purificarse de acuerdo con métodos convencionales conocidos en la técnica incluyendo, aunque sin limitación, cromatografía (por ejemplo, intercambio iónico, afinidad, hidrófobo, cromatoenfoco y de exclusión molecular), procedimientos electroforéticos (por ejemplo, isoelectroenfoco preparativo), solubilidad diferencial (por ejemplo, precipitación con sulfato de amonio), SDS-PAGE o extracción (véase, por ejemplo, PROTEIN PURIFICATION, J. C. Janson y Lars Ryden, eds., VCH Publishers, Nueva York, 1989) para obtener polipéptidos sustancialmente puros.

La incorporación del resto de para-metilazido-L-fenilalanina puede determinarse por métodos tales como espectrometría de masas, secuenciación de proteínas, marcaje de aminoácidos tal como por fluorescencia,

radioactividad, ELISA u otros ensayos de cribado con anticuerpo, ensayos funcionales u otros métodos conocidos para los expertos en la materia.

- 5 En algunas realizaciones, el método incluye además la etapa de conjugar un aducto biológicamente activo con el pAMF de la proteína de interés. En algunos casos, la conjugación es mediante 1,3-cicloaddición de una azida con un alquino sobrecargado.

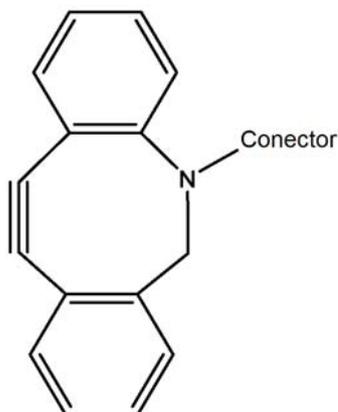
#### F. Química de conjugación para proteínas que contienen pAMF.

- 10 Cuando se tiene la proteína diana que contiene pAMF en la ubicación de aminoácido deseada, puede conjugarse un aducto biológicamente activo con la pAMF usando una reacción química tal como química click. Por ejemplo, la proteína de interés que contiene pAMF se purifica por procedimientos convencionales y después la proteína purificada se somete a una reacción de química click (por ejemplo, reacción de 1,3-cicloaddición de azida-alquino catalizada por cobre(I) o reacción de 1,3-cicloaddición de azida-alquino catalizada sin cobre) para conjugar directamente un aducto biológicamente activo con el resto de pAMF.

15 Los aductos biológicamente activos ejemplares para su uso en la presente invención incluyen, aunque sin limitación, moléculas pequeñas, oligonucleótidos, péptidos, aminoácidos, ácidos nucleicos, azúcares, oligosacáridos, polímeros, polímeros sintéticos, quelantes, fluoróforos, cromóforos, otros agentes detectables, restos farmacéuticos, agentes citotóxicos, agentes detectables y similares.

20 Sen encuentran descripciones detalladas de una reacción de química click para la conjugación de un aducto biológicamente activo en, por ejemplo, Baskin et al., Proc. Natl. Acad. Sci., 2007, 104:16793-16797; Kim et al., Curr. Opin. Chem. Biol., 2013, 14:412-419; y Bundy y Swartz, Bioconjug. Chem., 2010, 21(2):255-263.

- 25 En una reacción de química click, los alquinos se activan para la [3+2] cicloaddición con azidas. El aducto biológicamente activo que contiene (por ejemplo, ligado a) un alquino sobrecargado (por ejemplo, ciclooctino o variante del mismo) puede experimentar cicloaddición de alquino-azida promovida por sobrecarga con el aminoácido no natural para-metilazido-L-fenilalanina en la proteína de interés, conjugando de ese modo el aducto biológicamente activo con la proteína en la posición de aminoácido del aminoácido no natural. Un reactivo de alquino sobrecargado preferido es el reactivo DBCO, mostrado a continuación.



- 35 La proteína conjugada de interés puede purificarse de acuerdo con métodos convencionales conocidos en la técnica incluyendo, aunque sin limitación, cromatografía (por ejemplo, intercambio iónico, afinidad, hidrófobo, cromatoenfoco y de exclusión molecular), procedimientos electroforéticos (por ejemplo, isoelectroenfoco preparativo), solubilidad diferencial (por ejemplo, precipitación con sulfato de amonio), SDS-PAGE o extracción (véase, por ejemplo, Protein Purification, J. - C. Janson y Lars Ryden, editores, VCH Publishers, Nueva York, 1989) para obtener polipéptidos sustancialmente puros.

40 La proteína conjugada de interés puede cuantificarse de acuerdo con métodos convencionales conocidos en la técnica incluyendo, aunque sin limitación, espectrometría de masas (por ejemplo, espectrometría de masas ESI-TOF y espectrometría de masas en tándem), electroforesis de microfluidos, electroforesis en gel, transferencia de Western, inmunoensayos (por ejemplo, ELISA) y otros ensayos para evaluar la actividad de la proteína conjugada.

### III. Ejemplos

- 50 La presente invención se describirá en mayor detalle mediante ejemplos específicos. Los siguientes ejemplos se ofrecen con fines ilustrativos, y no pretenden limitar la invención de ninguna manera. Los expertos en la materia reconocerán fácilmente una diversidad de parámetros no críticos que pueden cambiarse o modificarse para producir

esencialmente los mismos resultados.

Ejemplo 1. Generación de variantes de TyrRS específicas de para-metilazido-L-fenilalanina (pAMF) y determinación de la eficacia de supresión.

5 Este ejemplo ilustra un método de cribado de variantes de RS específicas de pAMF novedosas para su uso en un sistema de síntesis de proteína sin células. Se encuentra una descripción detallada del método en, por ejemplo, Zimmerman et al., *Bioconjugate Chemistry*, 2014, 25, 351-361.

10 Las variantes de RS específicas de pAMF descritas en este documento tales como las que contienen las siguientes sustituciones de aminoácido: A1) Y32T, L65A, F108Y, Q109L e I159S; A4) Y32V, L65A, F108W, Q109M e I159G; B03) Y32A, L65V, F108W e I159G; C10) Y32V, L65V, F108Y e I159V; D08) Y32T, L65V, F108W e I159S; y C02) Y32V, L65V, F108W, G109I e I159S se generaron por PCR y se expresaron en células de *E.coli*. En resumen, se usó PCR de extensión solapante con cebadores específicos para cada mutación diseñada para construir las variantes de aaRS. Los productos de PCR se ligaron en un vector de expresión dirigido por el promotor T7 (por ejemplo, pYD317) y se usaron para transformar células de *E.coli*. La secuencia codificante del aaRS final contenía una marca de hexahistidina en el extremo carboxi en fase, para facilitar la purificación posterior de la proteína y la caracterización bioquímica. El ADN plasmídico purificado de las colonias se usó como moldes de ADN para la expresión del aaRS dirigida por la ADN polimerasa T7 en reacciones de síntesis de proteína en células (CFPS) a pequeña escala (por ejemplo, 100 µl) (véase, por ejemplo, Zawada et al., 2011, *Biotechnol. Bioeng.*, 108(7), 1570-1578). Las reacciones de CFPS se realizaron usando un extracto celular que contenía un ARNt supresor ámbar que se reconoce por TyrRS de *M. jannaschii* y que se ha optimizado para expresión elevada y toxicidad baja en el extracto de la cepa usada. Estas reacciones se usaron como fuente de aaRS para las reacciones del indicador supresión ámbar, como se describe a continuación, para evaluar la actividad y la selectividad de las proteínas variantes de RS específicas de pAMF.

25 Los ensayos de indicador de supresión ámbar se basan en las reacciones de síntesis de proteínas sin células usando el gen GFP indicador de supresor ámbar. En particular, las reacciones de indicador de supresión ámbar expresaban GFP con un codón TAG ámbar en la lisina 49 (K49TAG) del gen indicador TurboGFP (Evrogen, Moscú, Rusia). La lisina 49 está en un bucle expuesto al disolvente, y es un sitio permisivo que permite la fluorescencia de GFP independientemente del aminoácido presente en esta posición. La comparación de la fluorescencia de GFP en la presencia de pAMF frente a la ausencia de pAMF se usó para indicar la especificidad de la aaRS hacia el nnAA and su capacidad de discriminar frente a cualquiera de los 20 aminoácidos naturales comunes.

35 Las reacciones indicadoras de supresión ámbar se realizaron de la siguiente manera: se añadieron 4 µl de la CFPS de producción de sintetasa como se describe anteriormente a una reacción de CFPS indicadora de GFP K49TAG de 20 µl en placas de microvaloración negras de 384 pocillos en ausencia de pAMF o en presencia de pAMF 1 mM. Después de 12 horas de expresión a 30 °C durante 12 horas, se midió la señal de fluorescencia de GFP de punto final a excitación a 476 nm, emisión a 510 nm en un lector de placa de fluorescencia SpectraMax M5. Las variantes de sintetasa mostraron la actividad elevada y selectividad deseadas hacia pAMF, mostrando una eficacia de supresión al menos 3 veces mayor en presencia de pAMF que en ausencia de pAMF. Las variantes mostraron un alto grado de supresión ámbar de GFP únicamente en presencia de pAMF. Las variantes ensayadas tenían alta selectividad por el aminoácido no natural pAMF, discriminando al mismo tiempo frente a los 20 aminoácidos naturales comunes. Las variantes de RS A01, B04, B03, C02, C10 y D08 tuvieron eficacias de supresión elevadas en presencia de pAMF y sustancialmente ninguna supresión de la fluorescencia de GFP en las reacciones que carecen de pAMF (fig. 1).

45 En resumen, este ejemplo muestra que pueden usarse varias proteínas mutantes de RS (por ejemplo, muteínas) para la incorporación específica de sitio de pAMG con alta fidelidad traduccional en proteínas en respuesta al codón ámbar, TAG.

50 Ejemplo 2. Generación de conjugado de anticuerpo-fármaco (conjugados de trastuzumab-MMAF) usando síntesis de proteína sin células de anticuerpos que contienen para-metilazido-L-fenilalanina.

Este ejemplo ilustra el uso de la aminoacil-ARNt sintetasa (RS) descrita en este documento para la producción de conjugados de fármacos de anticuerpo mediante incorporación específica de sitio de para-metilazido-L-fenilalanina en el anticuerpo trastuzumab. El ejemplo también describe la conjugación de un fármaco de DBCO-PEG-monometil auristatina (DBCO-MMAF) con la IgG específica de tumor trastuzumab mediante el aminoácido no natural y química click sin cobre. El conjugado de fármaco de anticuerpo (ADC) específico de sitio resultante presentaba una alta relación de fármaco a anticuerpo (DAR) y un perfil de conjugación homogéneo. De forma importante, Los ADC eran muy potentes y específicos en ensayos de citotoxicidad celular *in vitro*.

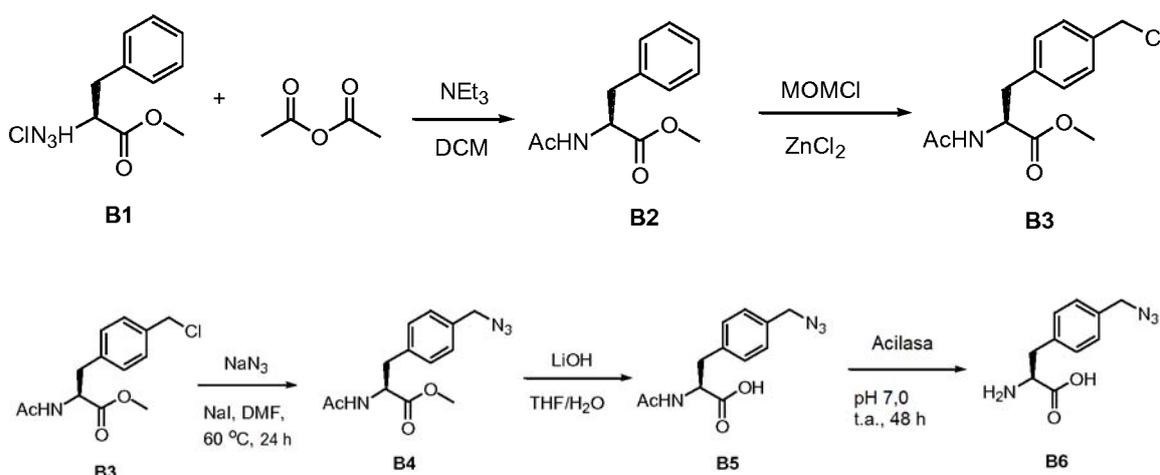
60 A. Síntesis de para-metilazido-L-fenilalanina.

Se ha demostrado que el aminoácido no natural para-azido-L-fenilalanina (pAzF) puede incorporarse en proteínas tales como dihidrofolato reductasa (DHFR) y que la pAzF que contiene DHFR puede conjugarse con un tinte fluorescente en una reacción de (3+2) cicloadición catalizada con cobre(I). En este documento se describe un aminoácido no natural novedoso y el uso del mismo para mejorar 1) el rendimiento global del producto, 2) la eficacia

de incorporación del aminoácido no natural y 3) la eficacia de conjugación de proteínas que contienen el aminoácido no natural conjugadas de forma bioortogonal.

5 Para generar para-azidometil-L-fenilalanina (pAMF), se modificó pAzF mediante la inserción de un espaciador de metileno entre el grupo ácido y el anillo fenilo para aumentar la reactividad del grupo ácido rompiendo la resonancia dentro del sistema cíclico fenilo. Se razonó que la adición de un grupo metileno entre el anillo fenilo y el grupo azida en la posición para movería el grupo azida del anillo fenilo que extrae potencialmente electrones, favoreciendo, por tanto, la reacción de click sin cobre. Además, una vez incorporado en una proteína de interés, el espaciador de metileno podría ubicar el grupo azida más lejos de la superficie de la proteína, disminuyendo de ese modo el potencial de impedancias electrostáticas y estéricas locales para la conjugación química.

El siguiente esquema muestra la reacción de síntesis de para-azidometil-L-fenilalanina (pAMF).



15 A una solución del éster metílico de Phe **B1** (50 g, 232 mmol, 1,0 equiv.) en DCM (300 ml) se le añadió trietilamina (81 ml, 580 mmol, 2,5 equiv.) a 0 °C, se le añadió anhídrido acético (33 ml, 348 mmol, 1,5 equiv.) gota a gota durante 15 min. La mezcla se agitó a 25 °C durante 3 h. La reacción se lavó con NaHCO<sub>3</sub> (2x), se secó sobre Na<sub>2</sub>SO<sub>4</sub> y se purificó mediante una columna de gel de sílice (DCM:MeOH = 9:1) para dar el producto **B2** (50 g, 97 %) como un sólido blanco. Se agitó una mezcla de éster **B2** (52 g, 0,235 mol, 1,0 equiv.), MOM-Cl (136 ml, 1,79 mol, 7,6 equiv.) y ZnCl<sub>2</sub> (128 g, 0,94 mol, 4,0 equiv.) a 6-8 °C durante 8 h. Después de retirar los volátiles en un evaporador rotatorio a 6-8 °C, el residuo se vertió en agua enfriada en hielo y se extrajo con acetato de etilo (3x). Las capas orgánicas combinadas se lavaron con NaHCO<sub>3</sub> y se secaron sobre Na<sub>2</sub>SO<sub>4</sub>, y se concentraron hasta un volumen pequeño. Después se añadió éter (50 ml). La solución de éter se mantuvo en un frigorífico a -20 °C durante una noche. El producto cristalizado se filtró y se secó al vacío para dar el producto **B3** (31 g, 49 %) como un sólido blanco.

20 A una solución de **B3** (20 g, 74,2 mmol, 1,0 equiv.) en DMF (100 ml) se le añadió NaN<sub>3</sub> (9,64 g, 148 mmol, 2,0 equiv.) y NaI (1,11 g, 7,42 mmol, 0,1 equiv.). La mezcla de reacción se calentó a 60 °C en un baño de aceite durante una noche. El disolvente DMF se retiró en un evaporador rotatorio y la mezcla de reacción se disolvió en acetato de etilo, y se lavó con NaHCO<sub>3</sub> (3x), se secó sobre Na<sub>2</sub>SO<sub>4</sub>, y se purificó por columna de gel de sílice (DCM:MeOH = 9:1) para dar el producto **B4** (20,5 g, 100 %) como un aceite amarillo.

35 A una solución de **B4** (20 g, 724 mmol, 1,0 equiv.) en una solución mezclada de THF: MeOH: H<sub>2</sub>O (50 ml: 50 ml: 20 ml) se le añadió LiOH-H<sub>2</sub>O (6,94 g, 144,8 mmol, 2 equiv.). La reacción se agitó a 25 °C durante 2 h. El disolvente se retiró para dar un residuo, que se pretrató con acetato de etilo. La capa orgánica se lavó con HCl 1 N (3x), se secó sobre Na<sub>2</sub>SO<sub>4</sub> y se concentró para dar el producto **B5** (18,3 g, 96 %) como un aceite amarillo.

40 A una solución de la amida **B5** anterior (18 g, 68,7 mmol) en DMSO (20 ml) y tampón NaH<sub>2</sub>PO<sub>4</sub>-Na<sub>2</sub>HPO<sub>4</sub> 50 mM (1,8 l) se le añadió acilasa (1 g). La solución se calentó a 37 °C durante 48 h. Se añadió carbón vegetal (20 g) a la mezcla de reacción y se agitó a 25 °C durante 10 min y después se filtró a través de Celite. El filtrado se lavó con acetato de etilo. La capa acuosa se concentró hasta un volumen pequeño y el producto se precipitó como un sólido blanco. El sólido se filtró y se secó al vacío para dar el producto **B6** (10 g, 66 %) como un sólido blanco.

45 Análisis del producto: CL-EM (IEN): 221 (M+1). RMN de <sup>1</sup>H (300 MHz, DMSO-d<sub>6</sub>) δ 7,27 (s a, 4H), 4,37 (s, 2H), 3,43 (m, 2H), 3,17 (m, 1H), 2,84 (m, 1H).

Se comparó la cinética de conjugación de molécula pequeña de pAzF y pAMF. En particular, pAzF y pAMF se hicieron reaccionar a concentraciones crecientes con un exceso molar de factor 10 de compuesto dibenzociclooctil-

polietilenglicol (DBCO-PEG) (fig. 2A y fig. 2C) en una reacción de click sin cobre. Se determinó la cinética de primer orden a 25 °C, fuerza iónica 0,5 M (KCl) y Tween-20 AL 0,005 % en tampón fosfato en condiciones de pseudo-primer orden por triplicado, usando un lector de placas SpectraMax M5 con placas de poliestireno de 96 pocillos (Molecular Devices, Sunnyvale, CA). La formación de producto conjugado se midió por la disminución en la absorbancia del ciclooctalquino DBCO sobrecargado a 306 nm en presencia de concentraciones crecientes de pAzF o pAMF. Se calcularon las constantes de velocidad de segundo orden de acuerdo con  $k_{cat} (M^{-1} s^{-1}) = k_{obs}/[Azido]$ . La velocidad de reacción para pAMF fue siete veces mayor que para pAzF (fig. 2B). Los datos muestran que pAMF es más reactiva que pAzF en reacciones de conjugación de click sin cobre.

#### 10 B. Sistema de síntesis de proteína sin células para producir anticuerpos que contienen para-metilazido-L-fenilalanina.

Las variantes de RS descritas en el ejemplo 1 se usaron para incorporar el aminoácido no natural pAMF en el anticuerpo monoclonal específico de Her2 trastuzumab. El anticuerpo se expresó en una reacción de síntesis de proteína sin células como se describe en Zawada et al., 2011, *Biotechnol. Bioeng.*, 108(7), 1570-1578 con las siguientes modificaciones. Los extractos sin células se prepararon a partir de una cepa atenuada para RF-1 sensible a OmpT que también se genomanipuló para que expresara DsbC de *E. coli* y una cepa de *E. coli* atenuada para RF-1 similar que se genomanipuló para que produjera un ARNt codificante de CUA ortogonal para la inserción de un aminoácido no natural en un codón de parada ámbar. Los extractos sin células se mezclaron (a una relación de 85:15), se trataron con yodoacetamida 50 µM durante 30 min a TA (20 °C) y después se añadieron a una premezcla que contenía todos los demás componentes de un sistema de síntesis de proteína sin células excepto por el ADN que codifica las cadenas pesada y ligera de IgG. La concentración final en la reacción de síntesis de proteína sin células fue extracto celular al 30 %, para-metilazido-L-fenilalanina (pAMF) 2 mM (RSP Amino Acids, Shirley, MA), variante de RS específica para pAMF 5 µM del ejemplo 1 (por ejemplo, variante A01, B04, B03, C02, C10 o D08), GSSG 2 mM, glutamato de magnesio 8 mM, glutamato de amonio 10 mM, glutamato de potasio 130 mM, piruvato sódico 35 mM, AMP 1,2 mM, 0,86 mM de cada uno de GMP, UMP y CMP, aminoácidos 2 mM (excepto 0,5 mM para tirosina y fenilalanina), oxalato de sodio 4 mM, putrescina 1 mM, espermidina 1,5 mM, fosfato de potasio 15 mM, ARN P T7 100 nM, 2 µg/ml de ADN de cadena ligera de trastuzumab, 8 µg/ml de ADN de cadena pesada de trastuzumab-(His)6 que contiene un codón ámbar en el sitio S136 (S136TAG). Las reacciones de síntesis sin células se iniciaron mediante la adición del ADN plasmídico que codifica la cadena pesada y ligera de IgG. Las reacciones se incubaron a 30 °C durante 12 h en un agitador a 650 rpm en placas Flower de 48 pocillos (m2p-labs n.º MTP-48-B). La reacción se incubó adicionalmente a 4 °C durante 6 h hasta que se procesó para su purificación.

#### C. Purificación de anticuerpos que contienen para-metilazido-L-fenilalanina.

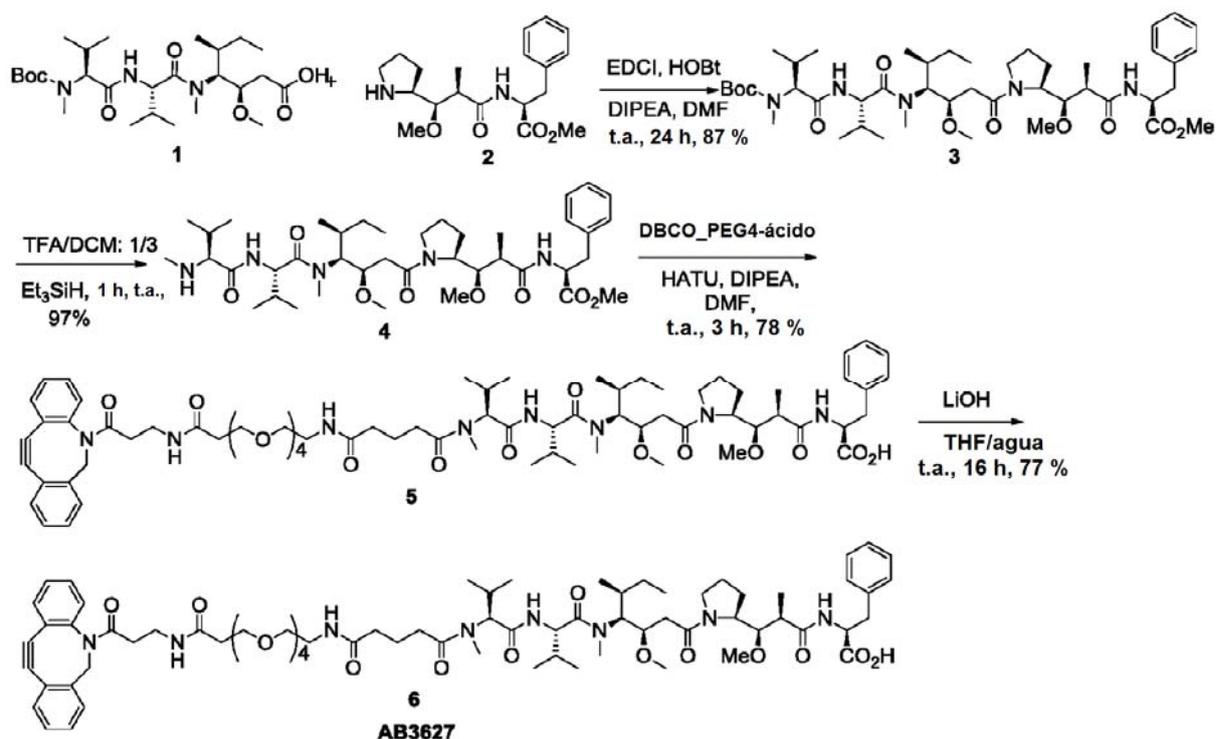
Después de la reacción de síntesis de proteína sin células, la mezcla que contenía anticuerpos que contienen para-metilazido-L-fenilalanina se transfirió a una placa de 96 pocillos (DyNa Block™, 2 ml; Labnet, Edison, NJ) y se centrifugó a 5000 xg durante 15 minutos a 40 °C. La purificación de IgG a partir del sobrenadante sin células se realizó usando IMAC Phytips (Phynexus, San Jose, CA) que contenía 40 µl de resina. El lecho de resina se preequilibró en tampón de equilibrado IMAC (Tris 50 mM pH 8,0, NaCl 300 mM e imidazol 10 mM) y el sobrenadante aclarado se pipeteó arriba y abajo 10 veces a través IMAC Phytips equilibrado a un caudal de 4,2 µl/min. La proteína unida se lavó con tampón de equilibrado IMAC y después se eluyó con 125 µl de tampón de elución IMAC (Tris 50 mM pH 8,0, NaCl 300 mM e imidazol 500 mM).

La cuantificación de IgG purificada y el análisis de la concentración y la pureza de aaRS se realizaron usando electroforesis de microfluidos, por ejemplo, ensayo de chip de expresión de proteínas en el sistema LabChip GXII (Caliper LifeSciences/Perkin Elmer). Los reactivos de ensayo se prepararon según las directrices del fabricante y se procesaron usando el modo de baja sensibilidad. Las muestras (2 µl) se mezclaron con 7 µl de tampón de muestra de expresión de proteínas usando una placa de PCR de 96 pocillos, se desnaturalizaron a 65 °C durante 10 min y se mezclaron con 32 µl de agua Milli-Q. Para la cuantificación, se preparó una curva patrón de Herceptin® o TyrRS de *M. jannaschii* que variaba de 500 a 10 µg/ml de una manera idéntica a las muestras analizadas. La placa de PCR de 96 pocillos se procesó en el LabChip GXII y las muestras se cuantificaron usando el programa informático LabChip GX v3.1 (Caliper LifeSciences/Perkin Elmer).

#### 55 D. Conjugación química de anticuerpos que contienen para-metilazido-L-fenilalanina para formar conjugados de anticuerpo-fármaco.

El fármaco DBCO-PEG-monometil auristatina (DBCO-MMAF) se conjugó con trastuzumab que contiene pAMF usando química click sin cobre. El siguiente esquema muestra la reacción de síntesis química para DBCO-PEG-MMAF (mencionado a continuación como AB3627).

60



5 A una solución de un ácido **1** (410 mg, 0,80 mmol, 1,0 equiv.), una amina **2** (279 mg, 0,80 mmol, 1,0 equiv.) y DIPEA (357  $\mu$ l, 2,2 mmol, 2,7 equiv.) en DMF (4 ml) a 25 °C se le añadió EDCI-HCl (169 mg, 0,88 mmol, 1,1 equiv.) y HOBt (135 mg, 0,88 mmol, 1,1 equiv.). La mezcla resultante se agitó a 25 °C durante 20 h y se diluyó con agua. La mezcla se extrajo con DCM (4 x 30 ml) y se concentró. El producto en bruto se purificó por FCC usando iPrOH/DCM (1/50, 1/25, 1/10) como eluyente para producir un producto **3** casi puro (586 mg, 87 %) como un sólido blanco. Análisis del compuesto **3** por CLEM m/z: 846 (M+1), 848 (M+Na).

10 Se añadió lentamente TFA (3 ml) a una solución de **3** (502 mg, 0,59 mmol) y trietilsilano (300  $\mu$ l) en DCM (9 ml) a 25 °C. La mezcla resultante se agitó a temperatura ambiente durante 1 h y se concentró. El residuo se repartió entre DCM (90 ml) y 1 M K<sub>2</sub>CO<sub>3</sub> (30 ml). La fase acuosa se extrajo con DCM (4 x 40 ml). Las capas orgánicas combinadas se secaron sobre K<sub>2</sub>CO<sub>3</sub> y se filtraron. El filtrado se concentró para proporcionar una amina libre **4** (429 mg, 97 %). Análisis del compuesto **4** por CLEM m/z: 746 (M+1).

15 Una mezcla del DBCO-PEG4-ácido (102 mg, 0,16 mmol, 1 equiv.), HATU (80 mg, 0,20 mmol, 1,3 equiv.) y DIPEA (70  $\mu$ l, 0,42 mmol, 2,6 equiv.) en DMF (2 ml) se agitó durante 1 h. Se añadió una solución de la amina **4** (120 mg, 0,16 mmol, 1 equiv.) en DMF (4 ml) a la solución anterior. La mezcla resultante se agitó durante 2 h y se concentró hasta sequedad. El residuo se disolvió en EA/Hex (80 ml / 20 ml) y se lavó con agua. El producto en bruto se purificó por FCC (NEt<sub>3</sub>/iPrOH/DCM = 1/0/200, 1/15/200) para producir el éster **5** (170 mg, 78 %) como un sólido blanco. Análisis por CLEM m/z: 683,6 [(M+2)/2].

25 El éster metílico **5** (138 mg, 0,10 mmol, 1 equiv.) se trató con LiOH-H<sub>2</sub>O (6,4 mg, 0,15 mmol) en THF/agua (5 ml /1,7 ml) a 25 °C durante 20 h. La mezcla se concentró hasta sequedad. El producto en bruto se purificó por FCC (NEt<sub>3</sub>/iPrOH/DCM = 2/10/200, 2/20/200, 2/40/200) para producir la sal de litio de DBCO-PEG-MMAF (105 mg, 77 %) como un sólido blanco. Análisis de DBCO-PEG-MMAF por CLEM m/z: 677 [(M+2)/2], 1352 (M+1), 1350 (M-1).

30 Las variantes de trastuzumab se conjugaron con un agente citotóxico ejemplar, MMAF, usando un reactivo de ciclooctino restringido. En resumen, se disolvió DBCO-MMAF (ACME Bioscience; Palo Alto, CA) en DMSO hasta una concentración final de 5 mM. El compuesto se diluyó con PBS hasta 1 mM y después se añadió a la muestra de proteína purificada en tampón de elución IMAC hasta una concentración de fármaco final de 100  $\mu$ M. La mezcla se incubó a TA (20 °C) durante 17 horas. La reacción se detuvo añadiendo azida de sodio hasta una concentración final de 100 mM y se intercambiaron el tampón usando placas Zeba™ (Thermo Scientific, Waltham, MA) equilibradas en PBS 1x. El filtrado entonces se pasó a través de una placa MUSTANG® Q (Pall Corp., Port Washington, NY) para retirar las endotoxinas.

35

## E. Análisis de conjugados de anticuerpo-fármaco.

Después de la conjugación, las muestras de conjugado de anticuerpo-fármaco se cuantificaron en un sistema Caliper GXII (Perkin Elmer, Waltham, MA) por comparación con una serie de patrones de masa de HERCEPTIN® procesadas en el mismo Protein Express LabChip® (Caliper Life Sciences n.º 760499). Las muestras se prepararon para el análisis como se especifica en el kit de reactivo Protein Express Reagent (Caliper Life Sciences n.º 760328) con la excepción que las muestras (mezcladas en tampón de muestra + NEM 50 mM) se calentaron a 65 °C durante 10 minutos antes del análisis en el sistema Caliper. La conjugación del fármaco no causaba ninguna disociación o agregación del esqueleto de IgG intacto determinado por cromatografía de exclusión molecular analítica.

Estos conjugados de fármaco de anticuerpo se ensayaron para el grado de conjugación, o la relación de fármaco a anticuerpo (DAR) por análisis de espectrometría de masas de conjugados de proteína intactos. Los valores DAR se calcularon como el promedio ponderado de las intensidades de los picos del espectro de masas desconvolucionado para cada especie de conjugado de conector-fármaco de IgG (es decir, DAR 0, 1 y 2). Se observó un alto grado de conjugación con valores DAR de uno bajo de 1,2 a uno elevado de 1,9, por ejemplo, de 1,48 a 1,70, que se correlacionaban bien con la actividad y especificidad de la variante pAMFRS usada para producir cada muestra ADC respectiva (fig. 2A).

## F. Citotoxicidad de conjugados de anticuerpo-fármaco.

La actividad de eliminación celular (citotoxicidad) del ADC conjugado se midió usando un ensayo de proliferación basado en células. La línea celular de cáncer de mama positiva a Her2 SKBR3 se obtuvo de la ATCC y se mantuvo en DMEM/nutriente F-12 de Ham (50:50), elevado contenido de glucosa (Cellgro-Mediatech; Manassas, VA) complementado con suero bovino fetal inactivado por calor al 10 % (Hyclone; Thermo Scientific; Waltham, MA), glutamax 2 mM (Invitrogen; Carlsbad, CA) y penicilina/estreptomicina 1x (Cellgro-Mediatech; Manassas, VA). Las células adherentes se lavaron dos veces con calcio y solución salina equilibrada con fosfato sin magnesio (PBS) y se recogieron con HYQ®TASE™ (Hyclone; Thermo Scientific; Waltham, MA). Se sembró un total de  $1 \times 10^3$  o  $3 \times 10^3$  células en un volumen de 40 ml en cada pocillo de una placa de poliestireno blanca de fondo plano de semiárea de 96 pocillos. Se permitió que las células se adhirieran durante una noche a 37 °C en una incubadora de CO<sub>2</sub>. Las variantes de ADC se formularon a concentración 2x en medio DMEM/F12 y se filtraron a través de centrífugo de 2 ml filtrado con acetato de celulosa SpinX estéril (Corning Costar, Cat n.º 8160). Se añadieron 40 µl de ADC esterilizados en filtro a pocillos de tratamiento y las placas se cultivaron a 37 °C en una incubadora de CO<sub>2</sub> durante 120 horas. Para la medición de la viabilidad celular, se añadieron 80 µl del reactivo Cell Titer-Glo® (Promega Corp. Madison, WI) a cada pocillo, y las placas se procesaron según las instrucciones del producto. Se midió la luminiscencia relativa en un lector de placa ENVISION® (Perkin-Elmer; Waltham, MA). Las lecturas de luminiscencia relativas se convirtieron al % de viabilidad usando células no tratadas como controles. Los datos se ajustaron con análisis de regresión no lineal, usando la ecuación de ajuste de 4 parámetros log(inhibidor) frente a respuesta-pendiente variable, usando GraphPad Prism® (GraphPad v 5.00, Software; San Diego, CA).

En comparación con el control trastuzumab no conjugado, que es ineficaz para eliminar las células a las concentraciones ensayadas, los conjugados de trastuzumab-MMAF mostraron potente citotoxicidad con valores de CE50 que varían de 0,043 nM a 0,071 nM (fig. 2B). Estos valores están muy de acuerdo con los valores de CE50 publicados para los ADC de trastuzumab-DMI generados por métodos de conjugación aleatoria.

En resumen, este ejemplo demuestra que la incorporación específica de sitio de pAMP facilita una conjugación casi completa de un fármaco de DBCO-PEG-monometil auristatina (DBCO-PEG-MMAF) con la IgG de unión a HER2 específica de tumor (trastuzumab) usando química click sin cobre de cicloadiición de azida-alquino promovida por sobrecarga (SPAAC). Los ADC resultantes son muy potentes para inducir citotoxicidad en ensayos celulares *in vitro*. El método descrito en este documento puede usarse para producir ADC homogéneos con valores de relación de fármaco a anticuerpo que se aproximan al límite teórico de dos fármacos por anticuerpo.

Ejemplo 3. Determinación de la alta fidelidad de las ARNt sintetasas específicas de pAMF en síntesis sin células de conjugados de anticuerpo-fármaco

El alto grado de supresión de codón ámbar que se observó únicamente en presencia de pAMF, el alto valor DAR de ADC y la alta potencia de fármaco mediada por la variante pAMFRS B03 (fig. 4A), todo en conjunto, indicó que esta variante de sintetasa poseía un alto grado de especificidad por pAMF. La diferencia en DAR entre las 6 preparaciones de trastuzumab producidas con las 6 diferentes sintetetasas (fig. 4A y 4B) muy probablemente se puede atribuir a diferencias en la eficacia y especificidad de la incorporación de pAMF. Esto por varias razones: (1) el sitio de incorporación de pAMF era el mismo, de modo que los efectos de posición que pueden dar lugar a una eficacia de conjugación variable no son un factor; (2) las reacciones de conjugación proseguían durante 16 h, pasado el punto temporal en que la reacción de conjugación del ADC ha alcanzado estabilidad; y (3) el método espectrométrico de masas del análisis de DAR ofrece un análisis muy preciso y robusto, haciendo que se muy improbable la imprecisión de la medición, y (4) los valores de DAR y CE<sub>50</sub> observados se correlacionan muy bien con la relación de actividad/especificidad ( $GFP^{1 \text{ mM pAMF}}/GFP^{No \text{ pAMF}}$ ) de la variante de sintetasa usada para producir cada ADC. Por tanto, parece muy probable que las muestras de DAR inferior sean el resultado de eventos de incorporación incorrecta por

sintetasas de menor fidelidad, que dan lugar a la producción de especies HC de longitud completa con aminoácidos naturales en la posición 136 que no pueden reaccionar con el conector del fármaco DBCO. Para hacer un seguimiento de la fidelidad de incorporación de pAMFRS, se usó análisis de espectrometría de masas como método de alta resolución para determinar la fidelidad precisa de la variante pAMFRS B03. El porcentaje de fidelidad de la aminoacil ARNt sintetasa puede definirse como la frecuencia con la que la enzima carga su ARNt afín con su aminoácido afín frente a cualquier otro aminoácido. Para obtener este número, se mide la frecuencia de incorporación de pAMF frente a cualquier otro aminoácido en la posición 136 del codón TAG en la cadena pesada de trastuzumab. Trastuzumab HC S136-TAG se coexpresó con la cadena ligera de trastuzumab en presencia de la variante pAMFRS B03, el ARNt supresor ámbar y pAMF 1 mM en una reacción OCFS de 30 ml OCFS. La IgG ensamblada de longitud completa se purificó hasta >90 % de pureza antes de la digestión con tripsina y el análisis de espectrometría de masas (EM) del péptido HC con pAMF incorporada. Los datos de EM se adquirieron en primer lugar con un método únicamente de EM, seguido por un método de EM/EM dependiente de datos con selección preferente por los péptidos tratados con tripsina diana que contienen pAMF, Phe, Tyr o Gln en la posición 136. Se observó previamente la incorporación incorrecta de Phe, Tyr y Gln cuando se usaban otros sistemas de supresión ámbar y como eventos de supresión ámbar endógenos que implican el acoplamiento incorrecto que descodifica los codones TAG por los ARNt endógenos. La EM en tándem del péptido tratado con tripsina trastuzumab HC-S136pAMF identificó de forma positiva la presencia de pAMF en la posición 136. Además, los péptidos tratados con tripsina que contiene Phe, Tyr o Gln en la posición 136 de HC y todas las demás posiciones de Phe en la proteína no podían detectarse por encima del nivel de ruido espectral. Además, no se observó ningún producto de truncamiento S136TAG no suprimido medible en una reacción OCFS paralela que contenía <sup>14</sup>CLeu, que indica un alto grado de incorporación de pAMF. Como no se observaron péptidos incorporados incorrectamente, la tasa máxima de incorporación incorrecta sería el nivel de ruido, o el límite de detección de nuestro método de EM. El porcentaje de fidelidad se calculó como la relación de la intensidad de señal para el péptido diana con pAMF incorporada a la intensidad del ruido espectral en la posición esperada de un péptido diana que contiene Phe, Tyr o Gln para producir un porcentaje de fidelidad de al menos un 99,8 % para la variante pAMF-RS B03 (tabla 1).

Tabla 1. Péptidos que contienen fenilalanina encontrados en la digestión con tripsina de trastuzumab HC-S136pAMF

Cadena	Posición de Phe	Secuencia de aminoácidos	SEQ ID NO	Fidelidad
HC	136	ST-pAMF-GGTAALGCLVK	SEQ ID NO: 4	99,8
LC	53	LLIYSASFLYSGVPSR	SEQ ID NO: 5	99,8
LC	116, 118	RTVAAPSVFIFPPSDEQLK	SEQ ID NO: 6	99,8
HC	67	FTISADTSK	SEQ ID NO: 7	99,1
HC	27	LSCAASGFNIK	SEQ ID NO: 8	99,7
HC	275	FNWYVDGVEVHNAK	SEQ ID NO: 9	99,6
HC	404, 405	TTPPVLDSDGSFFLYSK	SEQ ID NO: 10	99,7
HC	372	GFYPSDIAVEWESNGQPENNYK	SEQ ID NO: 11	99,7
HC	67	GRFTISADTSK	SEQ ID NO: 12	99,9

Se observó una pequeña cantidad de trastuzumab de longitud completa sintetizado en ausencia de pAMF-RS, supuestamente debido a eventos de supresión ámbar que se pueden atribuir al emparejamiento incorrecto que descodifica el codón TAG por ARNt cargados con tirosina y glutamina endógenos (TAT/TAC, y ARNt que descodifican CAG, respectivamente). Sin embargo, estos eventos de incorporación incorrecta no se observaron en nuestro análisis peptídico basado en CL-EM/EM. Esto sugiere que dichos eventos de supresión ámbar endógenos se producen infrecuentemente, si es que se producen, en presencia del sistema de supresión ámbar basado en pAMF altamente eficaz descrito en este documento.

Aunque la invención anterior se ha descrito en algún detalle mediante ilustración y ejemplo con fines de claridad de comprensión, un experto en la materia apreciará que pueden ponerse en práctica determinados cambios y modificaciones.

**Lista de secuencias informal**

**SEQ ID NO: 1**

Tirosil-ARNt sintetasa (TyrRS) de tipo silvestre de *Methanococcus jannaschii*

EC 6.1.1.1

Secuencia de referencia del NCBI n.º NP 247363

## ES 2 697 778 T3

1 MDEFEMIKRN TSEIISEEEL REVLKKDEKS AYIGFEPGK IHLGHYLQIK KMIDLQNAGF  
61 DIIILLADLH AYLNQGELD EIRKIGDYNK KVFEAMGLKA KYVYGSEFQL DKDYTLNVYR  
121 LALKTTLKRA RRSMEIARE DENPKVAEVI YPIMQVNDIH YLGVDVAVGG MEQRKIHMLA  
181 RELPKKVVC IHNPVLTGLD GEGKMSSSKG NFIAVDDSP EIRAKIKKAY CPAGVVEGNP  
241 IMEIAKYFLE YPLTIKRPEK FGGDLTVNSY EELESLEFKNK ELHPMDLKNA VAEELIKILE  
301 PIRKRL

**REIVINDICACIONES**

1. Una composición que comprende una aminoacil-ARNt sintetasa (RS), en la que la RS:

- 5 i) aminoacila preferentemente hasta un grado de más de un 90 % un ARNt con para-metilazido-L-fenilalanina (pAMF) en comparación con los 20 aminoácidos de origen natural comunes;  
 ii) tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y  
 iii) usando la SEQ ID NO: 1 como secuencia de referencia, tiene:
- 10 a) en la posición del aminoácido L65: A o V;  
 b) en la posición del aminoácido F108: Y o W;  
 c) en la posición del aminoácido D158: A;  
 d) en la posición del aminoácido Y32: T o V o A; y  
 15 e) en la posición del aminoácido I159: S o G o V.

2. La composición de la reivindicación 1, que tiene una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I.

20 3. La composición de la reivindicación 1, en la que la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y tiene las sustituciones de aminoácidos seleccionadas del grupo que consiste en:

- 25 i) Y32T, L65A, F108Y, Q109L e I159S;  
 ii) Y32V, L65A, F108W, Q109M e I159G;  
 iii) Y32A, L65V, F108W e I159G;  
 iv) Y32V, L65V, F108Y e I159V;  
 v) Y32T, L65V, F108W e I159S; y  
 vi) Y32V, L65V, F108W, G109I e I159S.

30 4. La composición de la reivindicación 1, en la que la RS, usando la SEQ ID NO: 1 como secuencia de referencia, tiene

- 35 a) en la posición del aminoácido L65: V;  
 b) en la posición del aminoácido F108: W;  
 c) en la posición del aminoácido D158: A;  
 d) en la posición del aminoácido Y32: A; y  
 e) en la posición del aminoácido I159: G.

40 5. Un polinucleótido que codifica una aminoacil-ARNt sintetasa (RS), en el que la RS:

- 45 i) aminoacila preferentemente hasta un grado de más de un 90 % un ARNt con para-metilazido-L-fenilalanina (pAMF) en comparación con los 20 aminoácidos de origen natural comunes;  
 ii) tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y  
 iii) usando la SEQ ID NO: 1 como secuencia de referencia, tiene:
- 50 a) en la posición del aminoácido L65: A o V;  
 b) en la posición del aminoácido F108: Y o W;  
 c) en la posición del aminoácido D158: A;  
 d) en la posición del aminoácido Y32: T o V o A; y  
 e) en la posición del aminoácido I159: S o G o V.

55 6. El polinucleótido de la reivindicación 5, que tiene una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I.

7. El polinucleótido de la reivindicación 5, en el que la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y tiene las sustituciones de aminoácidos seleccionadas del grupo que consiste en:

- 60 i) Y32T, L65A, F108Y, Q109L e I159S;  
 ii) Y32V, L65A, F108W, Q109M e I159G;  
 iii) Y32A, L65V, F108W e I159G;  
 iv) Y32V, L65V, F108Y e I159V;  
 v) Y32T, L65V, F108W e I159S; y  
 65 vi) Y32V, L65V, F108W, G109I e I159S.

8. El polinucleótido de la reivindicación **5**, en el que la RS, usando la SEQ ID NO: 1 como secuencia de referencia, tiene
- en la posición del aminoácido L65: V;
  - en la posición del aminoácido F108: W;
  - en la posición del aminoácido D158: A;
  - en la posición del aminoácido Y32: A; y
  - en la posición del aminoácido I159: G.
9. Un sistema de síntesis de proteína sin células para incorporar selectivamente para-metilazido-L-fenilalanina (pAMF) en una proteína de interés, comprendiendo el sistema:
- un extracto sin células de bacterias que tienen ARNt que funciona biológicamente, aminoácidos y ribosomas necesarios para la síntesis de proteína sin células;
  - un polinucleótido que tiene una región codificante que codifica la proteína de interés y que incluye un codón de supresión ubicado selectivamente dentro de su región codificante;
  - metilazido-L-fenilalanina en una concentración suficiente para permitir la incorporación selectiva de pAMF en la proteína de interés;
  - un ARNt que puede cargarse con pAMF y complementario al codón de supresión de la proteína de interés; y
  - una aminoacil-ARNt sintetasa (RS), en el que la RS:
    - aminoacila preferentemente hasta un grado de más de un 90 % un ARNt con pAMF en comparación con los 20 aminoácidos de origen natural comunes;
    - tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y
    - usando la SEQ ID NO: 1 como secuencia de referencia, tiene:
      - en la posición del aminoácido L65: A o V;
      - en la posición del aminoácido F108: Y o W;
      - en la posición del aminoácido D158: A;
      - en la posición del aminoácido Y32: T o V o A; y
      - en la posición del aminoácido I159: S o G o V.
10. El sistema de síntesis libre de células de la reivindicación **9**, en el que la RS tiene una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I.
11. El sistema de síntesis libre de células de la reivindicación **9**, en el que la RS tiene una identidad de secuencia de al menos un 90 % de la SEQ ID NO: 1 y tiene las sustituciones de aminoácidos seleccionadas del grupo que consiste en:
- Y32T, L65A, F108Y, Q109L e I159S;
  - Y32V, L65A, F108W, Q109M e I159G;
  - Y32A, L65V, F108W e I159G;
  - Y32V, L65V, F108Y e I159V;
  - Y32T, L65V, F108W e I159S; y
  - Y32V, L65V, F108W, G109I e I159S.
12. El sistema de síntesis libre de células de la reivindicación **9**, en el que la RS, usando la SEQ ID NO: 1 como secuencia de referencia, tiene
- en la posición del aminoácido L65: V;
  - en la posición del aminoácido F108: W;
  - en la posición del aminoácido D158: A;
  - en la posición del aminoácido Y32: A; y
  - en la posición del aminoácido I159: G.
13. El sistema de síntesis libre de células de la reivindicación **9**, en el que el extracto sin células tiene un sistema de fosforilación oxidativa activo.
14. El sistema de síntesis libre de células de la reivindicación **9**, en el que la proteína de interés es un anticuerpo o un fragmento de anticuerpo.
15. Un método para incorporar selectivamente para-metilazido-L-fenilalanina (pAMF) en una proteína de interés, comprendiendo el método las etapas de:
- combinar un extracto sin células de bacterias que tienen ARNt que funciona biológicamente, aminoácidos y

ribosomas necesarios para la síntesis de proteína sin células con los siguientes reactivos:

- 5 i) un polinucleótido que tiene una región codificante que codifica la proteína de interés y que incluye un codón de supresión ubicado selectivamente dentro de su región codificante;
- ii) metilazido-L-fenilalanina en una concentración suficiente para permitir la incorporación selectiva de pAMF en la proteína de interés;
- iii) un ARNt que puede cargarse con pAMF y complementario al codón de supresión de la proteína de interés;
- 10 y
- iv) una aminoacil-ARNt sintetasa (RS), en el que la RS:

10 aminoacila preferentemente hasta un grado de más de un 90 %, un ARNT con pAMF en comparación con los 20 aminoácidos de origen natural comunes;

15 tiene una identidad de secuencia de más de un 80 % con la tirosil ARNt sintetasa (TyrRS) de *Methanococcus jannaschii* que tiene la SEQ ID NO: 1; y usando la SEQ ID NO: 1 como secuencia de referencia, tiene:

- 20 a) en la posición del aminoácido L65: A o V;
- b) en la posición del aminoácido F108: Y o W;
- c) en la posición del aminoácido D158: A;
- 20 d) en la posición del aminoácido Y32: T o V o A; y
- e) en la posición del aminoácido I159: S o G o V; y

25 b) incubar la combinación de la etapa (a) en condiciones que permiten la incorporación selectiva de pAMF en la proteína de interés.

16. El método de la reivindicación 15, en el que la RS además tiene una sustitución de aminoácido en la posición Q109 seleccionada del grupo de aminoácidos que consiste en L o M o I.

17. El método de la reivindicación 15, en el que la RS tiene una identidad de secuencia de al menos 90 % de la SEQ ID NO: 1 y tiene una sustitución de aminoácido seleccionado del grupo que consiste en:

- 35 i) Y32T, L65A, F108Y, Q109L e I159S;
- ii) Y32V, L65A, F108W, Q109M e I159G;
- iii) Y32A, L65V, F108W e I159G;
- iv) Y32V, L65V, F108Y e I159V;
- v) Y32T, L65V, F108W e I159S; y
- vi) Y32V, L65V, F108W, G109I e I159S.

18. El método de la reivindicación 15, en el que la RS, usando la SEQ ID NO: 1 como secuencia de referencia, tiene

- 40 a) en la posición del aminoácido L65: V;
- b) en la posición del aminoácido F108: W;
- c) en la posición del aminoácido D158: A;
- d) en la posición del aminoácido Y32: A; y
- 45 e) en la posición del aminoácido I159: G.

19. El método de la reivindicación 15, en el que el extracto sin células tiene un sistema de fosforilación oxidativa activo.

20. El método de la reivindicación 15, en el que la proteína de interés es un anticuerpo o un fragmento de anticuerpo.

21. El método de la reivindicación 15, que comprende además conjugar un aducto biológicamente activo con la pAMF de la proteína de interés.

22. El método de la reivindicación 21, en el que la conjugación es mediante 1,3-cicloaddición de una azida con un alquino sobrecargado.

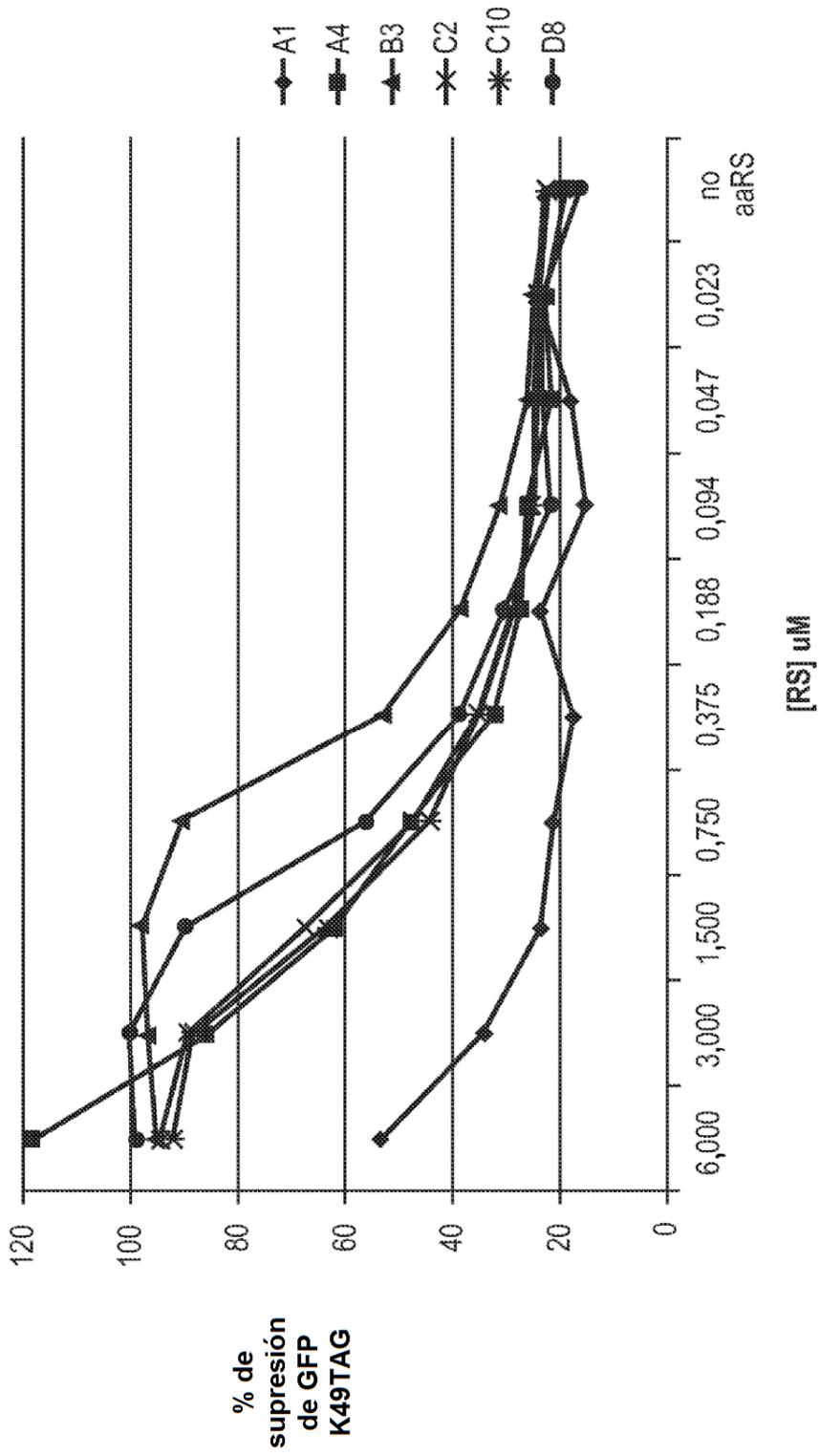


FIG. 1

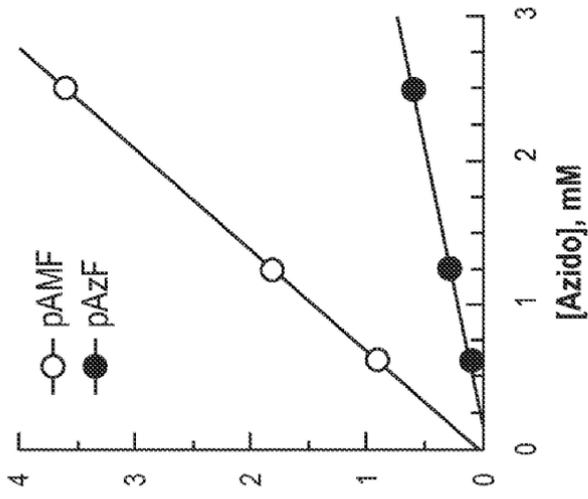


FIG. 2B

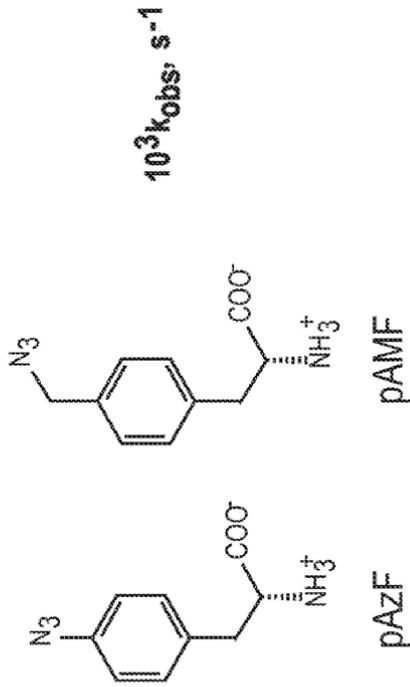


FIG. 2A

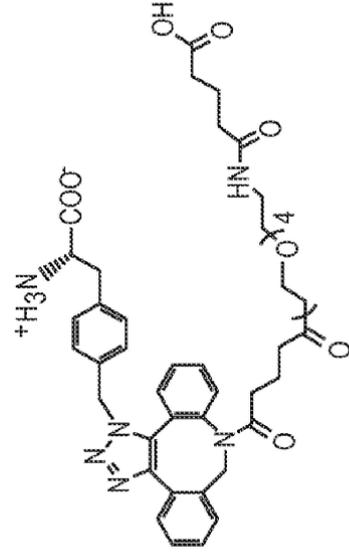
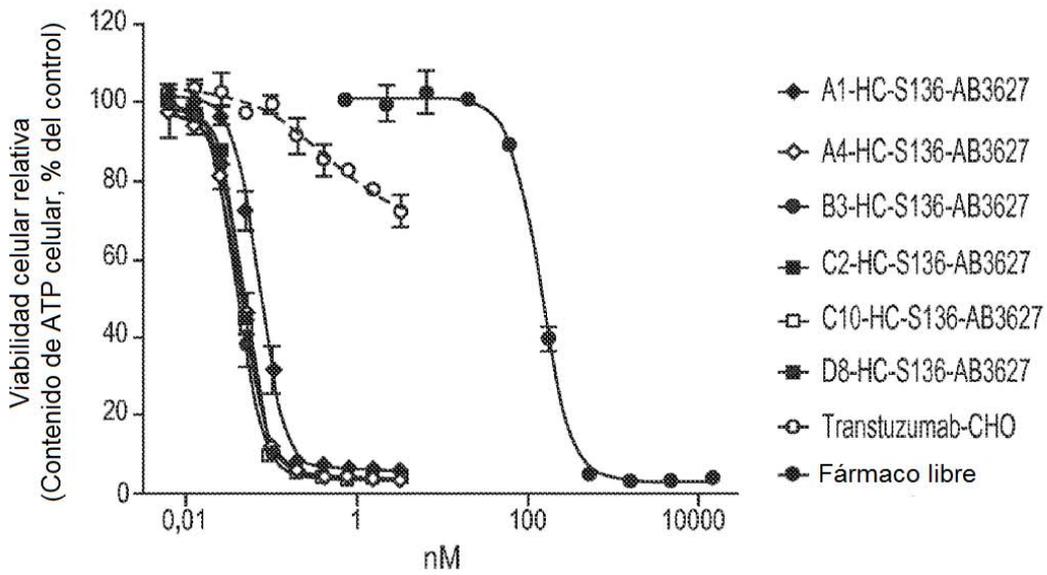


FIG. 2C

Variante pAMFRS	32	65	108	109	158	159	DAR de IgG HC 136-AB3627	CE50 (nM)
A01	T	A	Y	L	A	S	1,48	0,071
A04	V	A	W	M	A	G	1,68	0,047
B03	A	V	W	Q	A	G	1,70	0,043
C10	V	V	Y	Q	A	V	1,68	0,048
D08	T	V	W	Q	A	S	1,68	0,044
C02	V	V	W	I	A	S	1,64	0,044

**FIG. 3A**



**FIG. 3B**