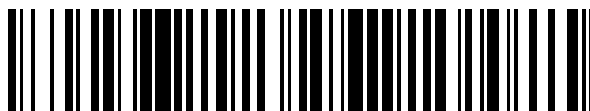


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 478 218**

51 Int. Cl.:

**C07K 1/00** (2006.01)

**C07K 1/13** (2006.01)

**C07K 1/107** (2006.01)

**C12P 21/00** (2006.01)

**C07K 16/46** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.09.2007 E 07838549 (9)**

97 Fecha y número de publicación de la concesión europea: **11.06.2014 EP 2064334**

54 Título: **Etiquetas de aldehído, usos de estas en la modificación sitio específica de proteínas**

30 Prioridad:

**21.09.2006 US 846644 P**

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

**21.07.2014**

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF CALIFORNIA (100.0%)  
1111 FRANKLIN STREET, 12TH FLOOR  
OAKLAND, CA 94607-5200, US**

72 Inventor/es:

**CARRICO, ISAAC S.;  
CARLSON, BRIAN L.;  
BERTOZZI, CAROLYN y  
WU, PENG**

74 Agente/Representante:

**LAZCANO GAINZA, Jesús**

**ES 2 478 218 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

Etiquetas de aldehído, usos de estas en la modificación sitio específica de proteínas

## 5 Antecedentes

10 El marcaje sitio específico de las proteínas es una herramienta importante para la disección de las redes celulares y bioquímicas. Se ha desarrollado una variedad de tecnologías para enfrentar esta necesidad. Una de tales tecnologías usadas para la localización de proteínas y su seguimiento es el marcaje con proteínas fluorescentes, tales como la proteína verde fluorescente (GFP). Sin embargo, el tamaño de estas proteínas fluorescentes puede interferir con el tráfico, la localización y las interacciones proteína-proteína del objetivo (Lisenbee y otros Traffic 2003, 4, (7), 491-501).

15 Como resultado, muchos grupos han centrado su atención en el uso de fusiones más pequeñas para dirigir a reactivos de marcaje secundario específicos. FIAsh, desarrollado por Roger Tsien y colegas, utiliza la interacción entre motivos de tetracisteína dispuestos específicamente y fluoróforos de biarsenilo (Chen y otros Science 1998, 281, (5374), 269-272). A pesar de la afinidad picomolar entre la secuencia mínima de 8 aminoácidos y las sondas biarsenicales (Adams y otros J. Amer.Chem. Soc. 2002, 124, (21), 6063-6076), el fondo debido a los motivos de cisteína nativos sigue siendo un problema (Stroffekova y otros Pflugers Archiv-Eur.J. Physiol. 2001, 442, (6), 859-866).

20 Para aumentar la especificidad, se han explorado los motivos dirigidos a péptido que dependen del marcaje secundario mediante enzimas. Tal estrategia depende de la fusión con O<sup>6</sup>-alquilguanina ADN transferasa (hAGT), que puede ligar una amplia variedad de moléculas pequeñas a una cisteína interna. Si bien las fusiones con hAGT permiten la unión covalente muy específica, de una amplia variedad de sondas de molécula pequeña estas se basan en una fusión de 207 aminoácidos (George y otros J. Amer.Chem. Soc. 2004, 126, (29), 8896-8897; Guignet y otros Nature Biotechnol 2004, 22, (4), 440-444). En un enfoque distinto, las fusiones de las proteínas con la proteína portadora de acilo de aproximadamente 80 aminoácidos pueden marcarse específicamente con sondas derivadas de CoA con el uso de la enzima fosfopanteteína transferasa. Alternativamente, se ha usado la biotina ligasa para transferir biotina o un isómero de biotina que contiene cetona a un péptido aceptor de 15 aminoácidos. El apéndice del isómero de cetona permite la formación de hidrazonas y conjugados oxima.

30 Prescher y otros (2005) Nature Chemical Biology Vol. 1(1) págs. 13-21 se refieren a cómo la enzima generadora de formilglicina (FGE) puede usarse para convertir un residuo de cisteína a formilglicina la que subsecuentemente puede marcarse químicamente con reactivos aminooxi o hidrazida. La modificación con FGE del residuo de cisteína debe producirse en una secuencia consenso objetivo conservada de 13 residuos.

35 Dierks y otros (2003) Cell vol 113, págs. 435 - 444 describe un ensayo in vitro de FGE. La FGE se purifica a partir de testículo bovino, la FGE humana se clona y se secuencia. La actividad de la FGE en células transfectadas se detecta mediante inmunofluorescencia indirecta y transferencia Western. Un péptido arilsulfatasa A (ASA) de 23 aminoácidos de origen natural se usó como sustrato para un ensayo de FGE. Otro péptido de 16 aminoácidos se unió a una matriz de cromatografía de afinidad y se usó en la purificación por afinidad de la FGE.

40 Dierks y otros (1999) EMBO J. col 18 págs. 2081 - 2091 describen los experimentos y un sistema de ensayo para establecer los determinantes basados en la secuencia responsables de la conversión de cisteína a FGly en las sulfatasas eucariotas. Se modificaron y probaron secuencias de arilsulfatasa (ASA) de origen natural y se encontró que tiene que estar presente el "motivo auxiliar" de los residuos de aminoácidos 74 - 80 de la ASA.

Hay una necesidad de nuevos enfoques para la modificación sitio específica de proteínas.

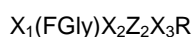
## 50 Literatura

Adams y otros 2002 J. Amer. Chem. Soc. 2002,124, (21), 6063-6076; Banghart y otros 2004 Nat. Neurosci. 7(12):1381-6. Epub 21 de nov. 2004; Berteau y otros 2006 J Biol Chem. 281 (32):22464-70 (Epub 9 de jun. 2006); Chen y otros 2005 Nature Methods 2005, 2, (2), 99-104; Cosma y otros 2003 Cell 113, (4), 445-56; Dierks y otros 1997 Proc Natl Acad Sci Estados Unidos 94, (22),11963-8; Dierks y otros 2003 Cell 113, (4), 435-44; Dierks y otros 2005 Cell 121, (4), 541-52; George y otros 2004 J. Amer. Chem. Soc. 126, (29), 8896-8897; Griffin y otros 1998 Science 281, (5374), 269-272; Guignet y otros 2004 Nature Biotechnol. 22, (4), 440-444; Landgrebe y otros 2003 Gene 316:47-56; Lemieux (1998) Trends Biotechnol 16, 506-13; Lisenbee y otros 2003 Traffic 4, (7), 491-501; Mariappan y otros 2005 J. Biol. Chem. 280(15):15173-9 (Epub 11 de feb. 2005); Mougous y otros 2004 Nat. Struc. Mol. Biol. 11, 721- 729; Preusser y otros 2005 J. Biol. Chem. 280(15): 14900-10 (Epub 18 de ene. 2005); Roeser y otros 2006 Proc Natl Acad Sci Estados Unidos 103(1):81-6 (Epub 20 de dic. 2005); Rush y otros (5 de ene. 2006) Org Lett. 8(1):131-4; Sardiello y otros 2005 Human Mol. Genet. 14, 3203-3217; Schirmer y otros 1998 Chemistry & Biology 5, R181-R186; Schmidt y otros 1995 Cell 82, (2), 271-8; Stroffekova y otros 2001 Archiv-Europ. J. Physiol. 442, (6), 859-866; Szameit y otros 1999 J Biol Chem 274, (22), 15375-81; Yin, J. y otros 2005 Proc. Natl. Acad. Sci. Estados Unidos 102, 15815- 15820 (2005); US20050026234; US20030186229; y US 6,900,304.

## Resumen

5 La invención presenta métodos para la modificación sitio específica de proteínas mediante la incorporación de una etiqueta de aldehído. La modificación enzimática en un motivo sulfatasa de la etiqueta de aldehído a través de la acción de una enzima generadora de formilglicina (FGE) genera un residuo de formilglicina (FGly). La porción aldehído del residuo de FGly puede explotarse como un asa química para la unión sitio específica de una porción de interés a un polipéptido.

10 En consecuencia, la presente invención proporciona un método para modificar un polipéptido, que comprende poner en contacto un polipéptido que comprende un motivo sulfatasa heterólogo convertido con una pareja reactiva que comprende una porción de interés, en donde el motivo sulfatasa convertido comprende:



15

en donde

20 FGly es un residuo de formilglicina;

Z<sub>2</sub> es un residuo de prolina o alanina;

X<sub>1</sub> está presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando el motivo sulfatasa heterólogo está en un N-terminal del polipéptido, X<sub>1</sub> está presente; y

X<sub>2</sub> y X<sub>3</sub> son cada uno independientemente cualquier aminoácido;

25 en donde dicho contacto es bajo condiciones suficientes para la conjugación de la porción de interés de la pareja reactiva a la FGly del polipéptido, que produce de esta manera un polipéptido modificado.

30 El residuo de FGly puede estar posicionado en una secuencia interna del polipéptido, y/o posicionado en un bucle terminal, un C-terminal, o un N-terminal del polipéptido. De particular interés son las situaciones en las que el residuo de FGly puede estar presente en una región accesible al disolvente del polipéptido cuando se pliega. Adicionalmente son de interés las situaciones en las que el residuo de FGly puede estar presente en un sitio de modificación postraducciona del polipéptido, tal como un sitio de glicosilación. Estos sitios de modificación postraducciona pueden ser nativos para el polipéptido progenitor, o el polipéptido puede diseñarse para incluir uno o más sitios no nativos de modificación postraducciona, y el motivo sulfatasa heterólogo está posicionado en dicho uno o más sitios no nativos de modificación postraducciona.

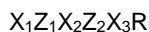
35

40 De particular interés son los motivos sulfatasa donde X<sub>1</sub>, cuando está presente, es L, M, V, S o T. Los motivos sulfatasa adicionales de particular interés son aquellos donde X<sub>2</sub> y X<sub>3</sub> son cada uno independientemente un aminoácido alifático, un aminoácido polar, sin carga, o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas modalidades son cada uno independientemente S, T, A, V, G o C.

40

La presente invención proporciona además una modificación de un polipéptido que comprende poner en contacto un polipéptido que comprende un motivo sulfatasa heterólogo con una enzima generadora de formilglicina (FGE), en donde el motivo sulfatasa heterólogo comprende:

45



en donde

50 Z<sub>1</sub> es cisteína o serina;

Z<sub>2</sub> es un residuo de prolina o alanina;

X<sub>1</sub> puede estar presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando el motivo sulfatasa heterólogo está en un N-terminal del polipéptido, X<sub>1</sub> está presente;

X<sub>2</sub> y X<sub>3</sub> son independientemente cualquier aminoácido,

55

en donde dicho contacto es bajo condiciones suficientes para la conversión de Z<sub>1</sub> a un residuo de formilglicina (FGly) en el polipéptido y produce un polipéptido convertido etiquetado con aldehído.

60 El polipéptido usado en este método puede tener al menos una de las siguientes propiedades: el motivo sulfatasa heterólogo es menor de 16 residuos de aminoácidos de longitud, los motivos sulfatasa heterólogos posicionados en un N-terminal del polipéptido, el motivo sulfatasa heterólogo está posicionado en un sitio interno de una secuencia de aminoácidos nativa para el polipéptido, el motivo sulfatasa heterólogo está posicionado en un bucle terminal del polipéptido, el motivo sulfatasa heterólogo posicionado en un sitio de modificación postraducciona del polipéptido; el polipéptido es un polipéptido de longitud completa, el polipéptido es distinto del polipéptido de preprolactina, un polipéptido de prolactina, o un polipéptido de glutatión-S-transferasa.

65

El motivo sulfatasa heterólogo puede ser menor de 16 residuos de aminoácidos de longitud y puede estar posicionado en un extremo C-terminal del polipéptido. El motivo sulfatasa heterólogo puede estar presente en un sitio interno en un bucle terminal del polipéptido y/o presente en un sitio interno dentro de un bucle extracelular o un bucle intracelular. El motivo sulfatasa heterólogo puede estar presente en un sitio interno o en el N-terminal, y/o puede ser accesible al disolvente cuando el polipéptido se pliega. El motivo sulfatasa heterólogo puede estar presente en un sitio de modificación postraduccional, tal como un sitio de glicosilación. El sitio de modificación postraduccional puede ser nativo para el polipéptido objetivo progenitor o el polipéptido objetivo puede diseñarse para incluir uno o más sitios no nativos de modificación postraduccional, y en donde el motivo sulfatasa heterólogo está posicionado en dicho uno o más sitios no nativos de modificación postraduccional.

De particular interés son los motivos sulfatasa donde  $X_1$ , cuando está presente, es un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas modalidades puede ser L, M, V, S o T. Los motivos sulfatasa adicionales de particular interés son aquellos donde  $X_2$  y  $X_3$  son cada uno independientemente un aminoácido alifático, un aminoácido polar, sin carga, o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas modalidades son cada uno independientemente S, T, A, V, G o C. En una modalidad de interés, el polipéptido se expresa en una célula que contiene la FGE.

En modalidades adicionales, el método comprende además poner en contacto el polipéptido convertido etiquetado con aldehído con una pareja reactiva que comprende una porción de interés; en donde dicho contacto es bajo condiciones que proporcionan la producción de un producto de reacción de un polipéptido modificado etiquetado con aldehído que tiene la porción de interés unida covalentemente al residuo de FGly del motivo sulfatasa heterólogo. La porción de interés puede ser, por ejemplo, un polímero soluble en agua, un marcador detectable, un fármaco, o una porción para la inmovilización del polipéptido en una membrana o sobre una superficie.

En la presente invención se describen polipéptidos convertidos etiquetados con aldehído producidos mediante los métodos de la invención así como también polipéptidos modificados etiquetados con aldehído producidos mediante los métodos de la invención.

En la presente invención se describen polipéptidos que comprenden un motivo sulfatasa heterólogo que tienen una enzima generadora de formilglicina (FGE), en donde el motivo sulfatasa heterólogo comprende



donde

FGly es un residuo de formilglicina;  
 $Z_2$  es un residuo de prolina o alanina;  
 $X_1$  puede estar presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando el motivo de sulfatasa heterólogo está en un N-terminal del polipéptido etiquetado con aldehído,  $X_1$  está presente; y  
 $X_2$  es cualquier aminoácido;

Un polipéptido usado en un método de la invención puede tener al menos una de las siguientes propiedades: el motivo sulfatasa heterólogo es menor de 16 residuos de aminoácidos de longitud, el motivo sulfatasa heterólogo está posicionado en un N-terminal del polipéptido, el motivo sulfatasa heterólogo está posicionado en un sitio interno de una secuencia de aminoácidos nativa para el polipéptido, el motivo sulfatasa heterólogo está posicionado en un bucle terminal del polipéptido, el motivo sulfatasa heterólogo está posicionado en un sitio de modificación postraduccional del polipéptido; el polipéptido es un polipéptido de longitud completa, el polipéptido es distinto de un polipéptido de preprolactina, un polipéptido de prolactina, o un polipéptido de glutatión-S-transferasa.

El motivo sulfatasa heterólogo de tales polipéptidos puede ser menor de 16 residuos de aminoácidos de longitud y puede posicionarse en un extremo C-terminal del polipéptido. El motivo sulfatasa heterólogo puede estar presente en un bucle terminal del polipéptido. El polipéptido puede ser una proteína transmembrana con el motivo sulfatasa heterólogo presente en un sitio interno dentro de un bucle extracelular o un bucle intracelular. El motivo sulfatasa heterólogo del polipéptido puede estar presente en un sitio interno o en el N-terminal del polipéptido, y es accesible al disolvente cuando el polipéptido se pliega. Adicionalmente, el motivo sulfatasa heterólogo puede estar presente en un sitio de modificación postraduccional, tal como un sitio de glicosilación. El sitio de modificación postraduccional puede ser nativo para el polipéptido objetivo progenitor o el polipéptido objetivo puede diseñarse para incluir uno o más sitios no nativos de modificación postraduccional, y en donde el motivo sulfatasa heterólogo está posicionado en dicho uno o más sitios no nativos de modificación postraduccional. De particular interés son los motivos sulfatasa donde  $X_1$ , cuando está presente, es usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga,

(es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas modalidades es L, M, V, S o T. Los motivos sulfatasa adicionales de particular interés son aquellos donde  $X_2$  y  $X_3$  son cada uno independientemente un aminoácido alifático, un aminoácido polar, sin carga, o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas modalidades son cada uno independientemente S, T, A, V, G o C.

Además se describen moléculas de ácido nucleico que comprenden una secuencia de nucleótidos que codifica tales polipéptidos, así como también vectores y células huésped recombinantes que contienen tales moléculas de ácidos nucleicos.

En la presente invención se describen polipéptidos modificados que comprenden un residuo de formilglicina unido covalentemente a una porción de interés, en donde el polipéptido comprende un motivo sulfatasa modificado de fórmula:



donde

FGly' es el residuo de formilglicina que tiene una porción heteróloga, unida covalentemente;

$Z_2$  es un residuo de prolina o alanina;

$X_1$  puede estar presente o ausente y, cuando está presente, es cualquier aminoácido, usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), con la condición de que cuando el motivo sulfatasa heterólogo está en un N-terminal del polipéptido,  $X_1$  está presente; y

$X_2$  es cualquier aminoácido.

La porción de tales polipéptidos modificados puede ser un polímero soluble en agua, un marcador detectable, un fármaco, o una porción para la inmovilización del polipéptido en una membrana o sobre una superficie. El motivo sulfatasa modificado de tales polipéptidos modificados puede posicionarse en el polipéptido modificado en un sitio de modificación postraducciona de un progenitor del polipéptido modificado. El sitio de modificación postraducciona puede ser, por ejemplo, un sitio de glicosilación. El sitio de modificación postraducciona puede ser nativo para el polipéptido objetivo progenitor o el polipéptido objetivo puede diseñarse para incluir uno o más sitios no nativos de modificación postraducciona, y en donde el motivo sulfatasa heterólogo está posicionado en dicho uno o más sitios no nativos de modificación postraducciona.

Además se describen ácidos nucleicos recombinantes que comprenden un casete de expresión que comprende un primer ácido nucleico que comprende una secuencia que codifica una etiqueta de aldehído; y un sitio de restricción posicionado en 5' o 3' de la secuencia que codifica la etiqueta de aldehído, cuyo sitio de restricción proporciona la inserción de un segundo ácido nucleico que codifica un polipéptido de interés; y un promotor operativamente enlazado al casete de expresión para proporcionar la expresión de un polipéptido etiquetado con aldehído producido mediante la inserción del segundo ácido nucleico que codifica el polipéptido de interés en el sitio de restricción.

En la presente invención se describe además la enzima generadora de formilglicina (FGE) de *Mycobacterium tuberculosis* (FGE de Mtb), y los métodos de su uso y las mezclas de reacción que comprenden una enzima generadora de formilglicina (FGE) aislada de *Mycobacterium tuberculosis*; y un polipéptido que comprende un motivo sulfatasa heterólogo de fórmula:



en donde

$Z_1$  es cisteína o serina;

$Z_2$  es un residuo de prolina o alanina;

$X_1$  está presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando el motivo sulfatasa heterólogo está en un N-terminal del polipéptido,  $X_1$  está presente;

$X_2$  y  $X_3$  son cada uno independientemente cualquier aminoácido; en donde el motivo sulfatasa heterólogo de la fórmula.

Además se describen mezclas de reacción de un polipéptido que tiene un motivo sulfatasa heterólogo como se describe en la presente invención y una FGE, que pueden incluir además un polipéptido convertido etiquetado con aldehído en el que un motivo sulfatasa heterólogo del polipéptido contiene un residuo de FGly. La descripción proporciona además composiciones que comprenden una FGE y un polipéptido convertido etiquetado con aldehído en el que un motivo sulfatasa heterólogo del polipéptido contiene un residuo de FGly. En modalidades relacionadas, tales mezclas de reacción pueden incluir además un reactivo para facilitar la unión de una porción de interés a un residuo de FGly de un polipéptido.

Otros elementos de la invención y su descripción relacionada se proporcionan más abajo, y serán fácilmente evidentes para el técnico generalmente con experiencia tras la lectura de la presente descripción.

## 5 Breve descripción de las figuras

La invención se entenderá mejor a partir de la siguiente descripción detallada cuando se lea junto con los dibujos acompañantes. Se destaca que, de acuerdo con la práctica común, los diversos elementos de los dibujos no están a escala. Por el contrario, las dimensiones de los diversos elementos se amplían o reducen arbitrariamente para mayor claridad. En los dibujos se incluyen las siguientes figuras:

La Figura 1A es un esquema que muestra un esbozo ilustrativo de los métodos y composiciones de la invención. En este ejemplo, un motivo sulfatasa ilustrativo ("LCTPSR") está posicionado en una construcción que contiene un ácido nucleico que codifica una proteína de interés.

La Figura 1B es un esquema que muestra un alineamiento de secuencias del motivo sulfatasa de una variedad de sulfatasas que se encuentran en diversos organismos. La secuencia consenso contiene la secuencia de las dos etiquetas de aldehído usadas en este estudio. Los residuos conservados están resaltados.

La Figura 2 es un conjunto de gráficos que muestran el análisis de espectro de masa que confirma la presencia de FGly en un péptido tríptico de sulfotransferasa micobacteriana con un motivo sulfatasa consenso de 13 residuos (ald<sub>13</sub>-Stf0). Los espectros de masa confirman la presencia de FGly en un péptido tríptico de ald<sub>13</sub>-Stf0. (Panel a) Espectro de masa de fragmentos trípticos que incorporan FGly (M+2/2). Teórico: 378.2066 m/z Observado: 378.2065. (Panel b) Espectro FT-ICR M+2/2 de fragmento tríptico que incorpora FGly luego del tratamiento con metoxilamina. Teórico: 392.7198 m/z Observado: 392.7201. (Panel c) Secuenciación MALDI-TOF/TOF del fragmento tríptico que incorpora FGly.

La Figura 3 ilustra los resultados de la cuantificación de la producción de FGly y el marcaje fluorescente selectivo de construcciones etiquetadas con aldehído. (Panel a) Adición de estándar de PL(FGly)TPSR sintético a la digestión tríptica de ald<sub>6</sub>-Stf0. (Panel b) Adición de estándar de PLCTPSR sintético a la digestión tríptica de ald<sub>6</sub>-Stf0. (Panel c) Marcaje fluorescente selectivo de ald<sub>13</sub>-Stf0, ald<sub>6</sub>-Stf0 y ald<sub>6</sub>-MBP con aminooxi-AlexaFluor 647 reflejado directamente en un escáner de gel fluorescente. (Panel d) La carga proteica se evaluó mediante tinción con Sypro Ruby.

La Figura 4 es un conjunto de imágenes que ilustran la modificación selectiva de las proteínas etiquetadas con aldehído. (Panel a) Ensayo de intercambio de ald<sub>6</sub>-MBP. Carril 1: Proteína incubada con biotina hidrazida. Carril 2: Proteína incubada con biotina hidrazida y subsecuentemente modificada con metoxilamina. Carril 3: Proteína incubada con biotina hidrazida y subsecuentemente modificada con aminooxiFLAG. La carga proteica (cuadro inferior) se evaluó mediante tinción con Ponceau. (Panel b) Pegilación de ald<sub>6</sub>-Stf0 con aminooxiPEG de 5,000 Da (carril 1), aminooxiPEG de 10,000 Da (carril 2) y aminooxiPEG de 20,000 Da (carril 3). Debido a la carencia de carga de las cadenas de PEG, las proteínas pegiladas migran más lentamente que las proteínas no pegiladas de peso molecular similar.

La Figura 5 es un conjunto de imágenes de geles que muestran la producción de una proteína canal de K<sup>+</sup> sintética regulada por azobenceno fotoisomerizable (SPARK) etiquetada con aldehído en células CHO (ovario de hámster chino) y HEK (riñón embrionario humano). (Panel a) Gel de SDS-PAGE teñido para proteína con el uso de Ponceau S. (Panel b) Detección de la unión del anticuerpo anti-myc. (Panel c) Detección de la unión del anticuerpo anti-FLAG. V indica que la muestra es un control negativo de vector solo; P, C, e I representan tres estrategias para la inserción de una etiqueta ald ilustrativa de 6 residuos, concretamente la adición de la etiqueta ald dentro de uno de los bucles extracelulares (I), la delección de 6 residuos del bucle y el reemplazo con la etiqueta ald (C), o la delección de 3 residuos del bucle y después la adición de la etiqueta de 6 residuos (P). (+) se refiere a una muestra de control positivo, que es un lisado de células CHO que contiene una proteína de 17 kDa etiquetada con myc.

La Figura 6 ilustra la generación de enlaces hidrazona, oxima y semicarbazona. R y R' se refieren a sustituyentes apropiados que aparecen en el polipéptido etiquetado con aldehído como se describe en la presente. R" se refiere a un sustituyente del reactivo que se transfiere al polipéptido etiquetado con aldehído en el producto de reacción.

La Figura 7 ilustra la activación de sulfatasas por la enzima generadora de formilglicina (FGE) y el mecanismo de la sulfatasa propuesto. (Panel a) La FGE activa las sulfatasas mediante la oxidación de una cisteína del sitio activo a un residuo de 2-formilglicina (FGly). Las estructuras cristalinas de las sulfatasas determinadas previamente indican que la FGly del sitio activo está hidratada, lo que sugiere que el clivaje del éster sulfato está mediado por un mecanismo<sup>6</sup> de transesterificación-eliminación. (Panel b) El motivo sulfatasa está localizado hacia el N-terminal de las sulfatasas y se dirige a la cisteína adecuada (\*) para la modificación por la FGE. Los residuos enmarcados indican un ajuste exacto de residuos; los residuos subrayados indican residuos conservados; los residuos con un punto (\*) indican residuos similares.

La Figura 8 proporciona los resultados que muestran la función de la FGE de *Mtb* (Rv0712) in vitro e in vivo. (Panel a) Un péptido sintético parecido a un motivo sulfatasa se trató con la FGE recombinante de *Mtb* y la oxidación resultante de la cisteína a FGly se monitoreó mediante espectrometría de masa. El mutante de la FGE Cys263Ser es inactivo sobre el sustrato peptídico. Los iones en m/z 1427 y 1445 son aductos de sodio del péptido modificado y no modificado, respectivamente. (Panel b) Tras el tratamiento con biotina hidrazida, el péptido que contiene FGly-forma un aducto de hidrazona con la biotina, lo que resulta en un desplazamiento de masa de +240 Da. (Panel c) Se probaron lisados de cepas silvestres (WT),  $\Delta fge$ , y complementadas ( $\Delta fge + fge$ ) de *Mtb* H37Rv para la actividad

sulfatasa con el uso del sustrato fluorogénico 4-metilumbeliferil sulfato (4MUS) con y sin inhibidores de sulfatasa/fosfatasa. La sulfatasa de lapa se usó como control positivo. (Panel d) Los lisados de cepas WT,  $\Delta fge$ , y  $\Delta fge + FGE$  de *Mtb* H37Rv se probaron para la actividad fosfatasa con el uso del sustrato fluorogénico 4-metilumbeliferil fosfato con y sin inhibidores de sulfatasa/fosfatasa. La fosfatasa PtpA recombinante de *Mtb* se usó como control positivo.

La Figura 9 proporciona los resultados del análisis de transferencia Southern del mutante  $\Delta fge$  de *Mtb*. El ADN genómico se digirió con FspI o NcoI, se separó mediante electroforesis en gel de agarosa, y se transfirió a una membrana de nilón. La transferencia se sondeó con un fragmento de ADN de 474 pb marcado con digoxigenina, que identifica un fragmento FspI de 4.8 kb y un fragmento NcoI de 5.7 kb para la cepa silvestre y un fragmento FspI de 5.5 kb y un fragmento NcoI de 5.1 kb para el mutante.

La Figura 10 muestra los datos que ilustran que Rv2407, Rv3406 y Rv3762c recombinantes no exhiben actividad en el ensayo 4MUS. La sulfatasa de lapa se usó como control positivo.

La Figura 11 es un esquema que muestra la estructura de la FGE de *Strep*. (Panel a) Estereosuperposición de la FGE de *Strep* y la FGE humana. Se indican los elementos de la estructura secundaria de la FGE de *Strep*. Los iones  $Ca^{2+}$  se representan como esferas. La desviación total de la raíz cuadrada media es de 0.65 Å. (Panel b) Comparación de los residuos que rodean el sitio potencial de unión a  $Ca^{2+}$  de la FGE de *Strep* y el segundo sitio de unión  $Ca^{2+}$  de la FGE humana. La geometría de coordinación adecuada se pierde con una mutación Glu66Ala en la FGE de *Strep*. (Panel c) Representación de la superficie del supuesto exosito de la FGE de *Strep*. La superficie, representada en escala de grises, se coloreó de acuerdo con la conservación de residuos entre todas las FGE conocidas y supuestas (basada en el alineamiento de secuencias de aminoácidos); el blanco representa los residuos no conservados, el azul claro representa los residuos débilmente conservados, el azul medio representa los residuos conservados, y el azul oscuro representa los residuos idénticos. El sustrato del péptido de 6 residuos se modela a partir de la estructura 24 del complejo de la FGE humana con el péptido (PDB entrada 2AIK). Un sustrato peptídico hipotético extendido se representa como una cinta. (Paneles d, e) Las cisteínas 272 y 277 del sitio activo parecen existir en un enlace disulfuro parcial. La Cys272 se muestra en dos conformaciones alternativas. La densidad de electrones entre Cys272 y Ser269 puede modelarse como una molécula de agua (d) o un hidróperóxido con una ocupación parcial (Panel e). Se muestra el monómero D. La densidad electrónica OMIT se contornea en  $1 \sigma$ .

La Figura 12 es una tabla que muestra el análisis ICP-AES de FGE de *Strep* y FGE de *Mtb*.

La Figura 13 proporciona los datos que ilustran que la actividad de la FGE de *Mtb* y de *Strep* depende del oxígeno molecular pero es independiente de cofactores metálicos. (a) Un péptido sintético parecido al motivo sulfatasa (masa = 1421.7 Da) se usó como sustrato para la FGE. La conversión de cisteína a FGly resultó en una pérdida de 18 Da, que se detectó mediante espectrometría de masa. (b,g) El quelante de metales EDTA no tuvo ningún efecto sobre la actividad. (c,h,i) La pérdida de cisteínas del sitio activo en la FGE de *Mtb* y de *Strep* suprimió la actividad. (d) La pérdida de la Ser260 del sitio activo en la FGE de *Mtb* redujo significativamente la actividad. (e) La FGE de *Mtb* es inactiva en ausencia de oxígeno molecular. (f) La FGE de *Strep* WT fue capaz de oxidar el péptido sintético. (j) El Trp234 del sitio activo de la FGE de *Strep* no es esencial para la actividad catalítica. El ion en m/z 1427 es el aducto de sodio del producto peptídico que contiene FGly.

La Figura 14 proporciona datos que ilustran que las cisteínas 272 y 277 del sitio activo de la FGE de *Strep* están comprometidas en un enlace disulfuro parcial. (a) La FGE de *Strep* se marcó con NBD y los aductos se detectaron mediante espectrometría de masa. Se observaron dos poblaciones de la FGE de *Strep* modificada con NBD. Una población tenía un único aducto de NBD, que corresponde a las Cys272 y Cys277 conectadas por disulfuro (ver b). La otra población tenía tres aductos de NBD correspondientes a tioles de Cys 272 y Cys277 (ver c). Se muestran múltiples estados de carga para ambas poblaciones. Los números encima de cada pico de ion indican el número de aductos de NBD atribuidos a cada población. (b-d) Los aductos de NBD se correlacionaron a cada tiol expuesto a la superficie con el uso de espectrometría de masa luego de la tripsinólisis. Se muestran los iones con estado de carga +2 correspondientes a Cys272 y Cys277 conectadas por disulfuro (b, masa calculada = 1,471.58; masa observada = 1,471.58), aductos de NBD en ambas Cys272 y Cys277 (c, masa calculada = 1,799.58; masa observada = 1,799.61), y un aducto de NBD en Cys301 (d, masa calculada = 1.486,52; masa observada = 1486.54). (e) La Cys301 no se observó como un tiol libre (posición esperada del ion con estado de carga +2 mostrada). La Cys301 es un tiol adicional expuesto a la superficie que sirvió como un control interno para evaluar la eficiencia del marcaje.

La Figura 15 proporciona datos que ilustran que una subpoblación de Cys277Ser de la FGE de *Strep* no contiene un residuo Cys272 reducido. (a) La FGE de *Strep* Cys277Ser se marcó con NBD y los aductos se detectaron mediante espectrometría de masa. Se observaron dos poblaciones de la FGE de *Strep* Cys277Ser modificadas con NBD. Una población corresponde a Cys272 como un tiol libre con un único aducto de NBD en Cys301 (ver b y d). La otra población tenía aductos de NBD en ambas Cys272 y Cys301 (ver c y d). Se muestran múltiples estados de carga para ambas poblaciones. Los números encima de cada pico de ion indican el número de aductos de NBD atribuidos a cada población. (b-d) Los aductos de NBD se correlacionaron a cada tiol expuesto a la superficie con el uso de espectrometría de masa luego de la tripsinólisis. Se muestran los iones con estado de carga +2 correspondientes a Cys272 como un tiol libre (b, masa calculada = 1,457.61; masa observada = 1,457.60), Cys272 con un aducto de NBD (c, masa calculada = 1,620.62; masa observada = 1,620.60), y Cys301 con un aducto de NBD (d, masa calculada = 1,486.54; masa observada = 1,486.54). (e) La Cys301 no se observó como un tiol libre (posición esperada del ion con estado de carga +2 mostrada). La Cys301 es un tiol adicional expuesto a la superficie que sirvió como un control interno para evaluar la eficiencia del marcaje.

La Figura 16 proporciona datos que ilustran que la FGE de *Strep* no exhibe una anomalía en la masa, indicativa de una modificación covalente estable de Cys272. La etiqueta His6 se clivó de la FGE de *Strep* y la masa de la enzima

se determinó mediante espectrometría de masa. Los números encima de cada pico de ion representan los estados de carga.

La Figura 17 proporciona un esquema de marcaje sitio específico del Fc de IgG recombinante, que incluye un esquema de un anticuerpo y los resultados de la modificación de un fragmento Fc que está etiquetado en N (N-Ald<sub>13</sub>-Fc) o etiquetado en C (C-Ald<sub>13</sub>-Fc) con una etiqueta de aldehído de 13mer (LCTPSRAALLTGR) o modificado en N o C con una etiqueta control (LATPSRAALLTGR).

La Figura 18 proporciona un esquema del marcaje sitio específico de un fragmento Fc de IgG con el uso de una etiqueta de aldehído de 6mer, e incluye los resultados de la modificación de un fragmento Fc que está etiquetado en N (N-Ald<sub>6</sub>-Fc) o etiquetado en C (C-Ald<sub>6</sub>-Fc) con una etiqueta de aldehído de 6mer (LCTPSR) o modificado en N o C con una etiqueta control (LATPSR).

La Figura 19 proporciona los resultados de la identificación de péptidos que contienen formilglicina (FGly) a partir de Fc de IgG etiquetado en N, e incluye un conjunto de gráficos que muestran el análisis de espectro de masa que confirma la presencia de FGly en unos fragmentos trípticos del fragmento Fc etiquetado en N. (Panel a) Espectro de masa de fragmentos trípticos que incorporan FGly. Teórico: 429.7268 m/z; Observado: 429.7321 m/z. (Panel b) Espectro de masa de fragmento tríptico del fragmento Fc etiquetado en N que incorpora FGly luego del tratamiento con 2-yodoacetamida. Teórico: 467.2375 m/z Observado: 467.2410 m/z.

La Figura 20 proporciona los resultados de la identificación de péptidos que contienen formilglicina (FGly) a partir de Fc de IgG etiquetado en C, y un conjunto de gráficos que muestran el análisis de espectro de masa que confirma la presencia de FGly en unos fragmentos trípticos del fragmento Fc etiquetado en C. (Panel a) Espectro de masa de fragmentos trípticos que incorporan FGly. Teórico: 508.7613 m/z; Observado: 508.7755m/z. (Panel b) Espectro de masa de fragmento tríptico del fragmento Fc etiquetado en C que incorpora FGly luego del tratamiento con 2-yodoacetamida. Teórico: 546.2721 m/z; Observado: 546.2811m/z.

La Figura 21 se refiere al marcaje sitio específico de una proteína de superficie celular, y proporciona un esquema del vector pDisplay™ usado para la construcción de la proteína de superficie celular etiquetada con aldehído (con el uso de una etiqueta de aldehído de 13mer de LCTPSRAALLTGR) y un gráfico que muestra el aumento de la fluorescencia media para la proteína de superficie etiquetada con la de 13mer (Ald13-TM) en comparación con el control. (LATPSRAALLTGR; referido como C->A-TM).

La Figura 22 se refiere al marcaje sitio específico de una proteína citosólica, ilustrada por His<sub>6</sub>-Ald<sub>13</sub>-AcGFP, y proporciona los resultados de la modificación de la proteína de fusión proteína verde fluorescente (GFP) que contiene una etiqueta de His y una etiqueta de aldehído de 13mer (referida como His<sub>6</sub>-Ald<sub>13</sub>-Ac-GFP o Ald-AcGFP) o una proteína de fusión GFP que contiene una etiqueta control (LATPSRAALLTGR) (referida como C->A-AcGFP).

La Figura 23 proporciona un esquema de la glicosilación sitio específica del interferón beta (IFN-beta) con el uso de la metodología de etiqueta de aldehído.

### 35 Descripción detallada de las modalidades de la invención

Antes de describir la presente invención en detalles, se debe entender que esta invención no se limita a las modalidades particulares descritas, las cuales, por supuesto, pueden variar. Se deberá entender, además, que la terminología que se usa en la presente invención es para el propósito de describir modalidades particulares solamente, y no pretende ser limitante, ya que el alcance de la presente invención se limitará solamente por las reivindicaciones anexas.

Cuando se proporciona un intervalo de valores, se entiende que cada valor intermedio se describe además específicamente, hasta la décima de la unidad del límite inferior a menos que el contexto dicte claramente lo contrario, entre los límites superiores e inferiores de ese intervalo. Cada intervalo más pequeño entre cualquier valor indicado o valor intermedio en un intervalo indicado y cualquier otro valor indicado o intermedio en ese intervalo indicado se abarca dentro de la invención. Los límites superiores e inferiores de estos intervalos más pequeños pueden incluirse o excluirse independientemente en el intervalo, y cada intervalo donde se incluyen cualquiera, ninguno o ambos límites en los intervalos más pequeños se abarca además dentro de la invención, sometido a cualquier límite excluido específicamente en el intervalo indicado. Donde el intervalo indicado incluya uno o los dos límites, los intervalos que excluyen uno o ambos límites incluidos se incluyen también en la invención.

A menos que se especifique lo contrario, todos los términos técnicos y científicos utilizados en la presente tienen el mismo significado que el conocido comúnmente por aquellos con experiencia en la técnica a la que pertenece esta invención. Si bien los métodos y materiales similares o equivalentes a los descritos en la presente invención pueden usarse en la práctica o prueba de la presente invención, algunos materiales y métodos potenciales e ilustrativos se describen ahora. Todas las publicaciones mencionadas en la presente se incorporan como referencia para revelar y describir los métodos y/o materiales en relación con las publicaciones que se citan. Se entiende que la presente descripción reemplaza a cualquier descripción de una publicación incorporada en la medida que haya una contradicción.

Se debe destacar que, como se usa en la presente y en las reivindicaciones anexas, las formas del singular "un" "una" y "el/la" incluyen referencias del plural a menos que el contexto claramente establezca lo contrario. Así, por ejemplo, la referencia a "una etiqueta de aldehído" incluye una pluralidad de tales etiquetas y la referencia a "el polipéptido" incluye la referencia a uno o más polipéptidos y equivalentes de estos conocidos para los expertos en la técnica, etcétera.



5 Debe destacarse, además, que las reivindicaciones pueden redactarse para excluir cualquier elemento que pueda ser opcional. Como tal, esta declaración pretende servir como base antecedente para el uso de tal terminología exclusiva como "solamente", "solo" y similares en relación con la narración de los elementos de la reivindicación, o el uso de una limitación "negativa".

10 Las publicaciones discutidas en la presente se proporcionan únicamente para su divulgación antes de la fecha de presentación de la presente solicitud. Nada en este documento debe interpretarse como una admisión de que la presente invención no tiene derecho a preceder tal publicación en virtud de una invención previa. Adicionalmente, las fechas de publicación proporcionadas pueden ser diferentes de las fechas de publicación presentes que pueden necesitar ser confirmadas independientemente.

#### Definiciones

15 Los términos "polipéptido", "péptido", y "proteína" se usan indistintamente en la presente para referirse a una forma polimérica de aminoácidos de cualquier longitud. A menos que se indique específicamente lo contrario, "polipéptido", "péptido" y "proteína" puede incluir aminoácidos codificados y no codificados genéticamente, aminoácidos químicamente o bioquímicamente modificados o derivatizados, y polipéptidos que tienen cadenas principales peptídicas modificadas. El término incluye proteínas de fusión, que incluyen, pero sin limitarse a, proteínas de fusión con una  
20 secuencia de aminoácidos heteróloga, fusiones con secuencias líder heterólogas y homólogas, proteínas que contienen al menos un residuo de metionina N-terminal (por ejemplo, para facilitar la producción en una célula huésped bacteriana recombinante); proteínas etiquetadas inmunológicamente; y similares.

25 "Polipéptido objetivo" se usa en la presente para referirse a un polipéptido que debe modificarse mediante el uso de una etiqueta de aldehído como se describe en la presente.

30 "Secuencia de aminoácidos nativa" o "secuencia de aminoácidos progenitora" se usan indistintamente en la presente en el contexto de un polipéptido objetivo para referirse a la secuencia de aminoácidos del polipéptido objetivo antes de la modificación para incluir una etiqueta de aldehído heteróloga.

35 Por "etiqueta de aldehído" o "etiqueta ald" se entiende una secuencia de aminoácidos que contiene una secuencia de aminoácidos derivada de un motivo sulfatasa que es capaz de convertirse, o que se ha convertido, por la acción de una enzima generadora de formilglicina (FGE) para contener un residuo de 2-formilglicina (referido en la presente como "FGly"). Aunque esto es técnicamente incorrecto, el residuo de FGly generado por una FGE se refiere frecuentemente en la literatura como una "formilglicina". Dicho de otra manera, el término "etiqueta de aldehído" se usa en la presente para referirse a una secuencia de aminoácidos que comprende un motivo sulfatasa "no convertido" (es decir, un motivo sulfatasa en el que los residuos de cisteína o serina no se han convertido a FGly por una FGE, pero que son capaces de convertirse) así como también a una secuencia de aminoácidos que comprende un motivo sulfatasa "convertido" (es decir, un motivo sulfatasa en el que los residuos de cisteína o serina se han convertido a FGly por acción de una FGE).  
40

Por "conversión" como se usa en el contexto de la acción de una enzima generadora de formilglicina (FGE) sobre un motivo sulfatasa se refiere a la modificación bioquímica de un residuo de cisteína o serina en un motivo de sulfatasa a un residuo de formilglicina (FGly) (por ejemplo, Cys a FGly, o Ser a FGly).

45 "Modificación" abarca la adición, la eliminación o la alteración de una porción. Como se usa en el contexto de un polipéptido que tiene un motivo sulfatasa convertido, "modificación" pretende referirse a la modificación química o bioquímica de un residuo de FGly de una etiqueta de aldehído de un polipéptido a través de la reacción de la porción aldehído de la FGly con una pareja reactiva. Como se discutió anteriormente, el término "conversión" se refiere a un tipo de modificación bioquímica de un residuo de FGly de una etiqueta de aldehído mediada por una FGE.  
50

55 Por "genéticamente codificable" como se usa en referencia a una secuencia de aminoácidos de un polipéptido, péptido o proteína se entiende que la secuencia de aminoácidos está compuesta de residuos de aminoácidos que son capaces de producir un ácido nucleico que codifica la secuencia de aminoácidos mediante la transcripción y la traducción, donde la transcripción y/o la traducción pueden producirse en una célula o en un sistema in vitro de transcripción/traducción libre de células.

60 El término "secuencias control" se refiere a secuencias de ADN para facilitar la expresión de una secuencia codificante operativamente enlazada en un sistema de expresión particular, por ejemplo células de mamífero, células bacterianas, síntesis libre de células, etc. Las secuencias control que son apropiadas para los sistemas procariotas, por ejemplo, incluyen un promotor, opcionalmente una secuencia operadora, y un sitio de unión al ribosoma. Los sistemas de células eucariotas pueden utilizar promotores, señales de poliadenilación, y potenciadores.

65 Un ácido nucleico está "operativamente enlazado" cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, el ADN para una presecuencia o líder secretora está operativamente enlazado al ADN de un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o

5 potenciador está operativamente enlazado a una secuencia codificante si afecta la transcripción de la secuencia; o un sitio de unión al ribosoma está operativamente enlazado a una secuencia codificante si se posiciona para facilitar la iniciación de la traducción. Generalmente, "operativamente enlazado" significa que las secuencias de ADN que se enlazan son contiguas, y, en el caso de una líder secretora, contigua y en marco de lectura. El enlace se logra mediante ligación o a través de reacciones de amplificación. Los adaptadores sintéticos de oligonucleótidos o enlazadores pueden usarse para enlazar secuencias de acuerdo con la práctica convencional.

10 El término "casete de expresión" tal como se usa en la presente se refiere a un segmento de ácido nucleico, usualmente ADN, que puede insertarse en un ácido nucleico (por ejemplo, mediante el uso de sitios de restricción compatibles con la ligación en una construcción de interés o mediante recombinación homóloga en una construcción de interés o en un genoma de célula huésped). Generalmente, el segmento de ácido nucleico comprende un polinucleótido que codifica un polipéptido de interés (por ejemplo, una etiqueta de aldehído, que puede estar operativamente enlazada a un polinucleótido que codifica un polipéptido objetivo de interés), y el casete y los sitios de restricción están diseñados para facilitar la inserción del casete en el marco de lectura apropiado para la transcripción y la traducción. Los casetes de expresión pueden comprender además elementos que facilitan la expresión de un polinucleótido que codifica un polipéptido de interés en una célula huésped. Estos elementos pueden incluir, pero sin limitarse a: un promotor, un promotor mínimo, un potenciador, un elemento de respuesta, una secuencia de terminación, una secuencia de poliadenilación, y similares.

20 Como se usa en la presente el término "aislado" pretende describir un compuesto de interés que está en un medio ambiente diferente de aquel en el que el compuesto se produce de forma natural. "Aislado" pretende incluir los compuestos que están dentro de muestras que están prácticamente enriquecidas en el compuesto de interés y/o en las que el compuesto de interés está parcialmente o prácticamente purificado.

25 Como se usa en la presente, el término "prácticamente purificado" se refiere a un compuesto que se elimina de su medio ambiente natural y es al menos 60% libre, usualmente 75% libre, y lo más usual 90% libre de otros componentes con el que está asociado de forma natural.

30 El término "condiciones fisiológicas" pretende abarcar aquellas condiciones compatibles con las células vivas, por ejemplo, condiciones predominantemente acuosas de una temperatura, pH, salinidad, etc. que son compatibles con las células vivas.

35 Por "heterólogo" se entiende que una primera entidad y una segunda entidad se proporcionan en una asociación que no se encuentra normalmente en la naturaleza. Por ejemplo, una proteína que contiene un motivo sulfatasa "heterólogo" o una etiqueta ald "heteróloga" es una proteína que normalmente no contiene un motivo sulfatasa en esa posición dentro de su secuencia de aminoácidos (por ejemplo, las proteínas que tienen un solo motivo sulfatasa nativo, pueden contener un segundo motivo sulfatasa que es "heterólogo"; las proteínas adicionales que contienen un motivo sulfatasa pueden modificarse para cambiar la posición del motivo sulfatasa, dando lugar a un motivo sulfatasa reposicionado "heterólogo" para la proteína) En algunas modalidades, está presente un motivo sulfatasa heterólogo en un polipéptido que no contiene ningún motivo de sulfatasa nativo.

40 Por "pareja reactiva" se entiende una molécula o porción molecular que reacciona específicamente con otra pareja reactiva para producir un producto de reacción. Las parejas reactivas ilustrativas incluyen una cisteína o serina del motivo sulfatasa y una enzima generadora de formilglicina (FGE), que reaccionan para formar un producto de reacción de una etiqueta de aldehído convertido que contiene una FGly en lugar de cisteína o serina en el motivo. Otras parejas reactivas ilustrativas incluyen un aldehído de un residuo de formilglicina (FGly) de una etiqueta de aldehído convertido y un reactivo de pareja reactiva que comprende una porción de interés, que reaccionan para formar un producto de reacción de un polipéptido modificado etiquetado con aldehído que tiene la porción de interés conjugada con el polipéptido etiquetado con aldehído en el residuo de FGly.

50 "N-terminal" se refiere al residuo de aminoácido terminal de un polipéptido que tiene un grupo amina libre, cuyo grupo amina en residuos de aminoácidos que no son N-terminales normalmente forma parte de la cadena principal covalente del polipéptido.

55 "C-terminal" se refiere al residuo de aminoácido terminal de un polipéptido que tiene un grupo carboxilo libre, cuyo grupo carboxilo en los residuos de aminoácidos que no son C-terminales normalmente forma parte de la cadena principal covalente del polipéptido.

60 Por "N-terminal" se entiende la región de un polipéptido que está más cercana al N-terminal que al C-terminal.

Por "C-terminal" se entiende la región de un polipéptido que está más cercana al C-terminal que al N-terminal.

65 Por "sitio interno" como se usa en referencia a un polipéptido o una secuencia de aminoácidos de un polipéptido se entiende una región del polipéptido que no está en el N-terminal o en el C-terminal, e incluye ambas regiones N-terminal y C-terminal del polipéptido.

## Introducción

5 La presente invención explota un motivo sulfatasa genéticamente codificable de origen natural, para usar como una etiqueta de péptido, que se refiere en la presente como un "etiqueta de aldehído" o "etiqueta ald", para dirigir la modificación sitio específica de un polipéptido. El motivo sulfatasa de la etiqueta de aldehído, que se basa en un motivo que se encuentra en los sitios activos de las sulfatasas, contiene un residuo de serina o cisteína que es capaz de convertirse (oxidarse) a una formilglicina (FGly) por la acción de una enzima generadora de formilglicina (FGE) ya sea in vivo (por ejemplo, en el momento de la traducción de una proteína que contiene la etiqueta de aldehído en una célula) o in vitro (por ejemplo, al poner en contacto una proteína que contiene la etiqueta de aldehído con una FGE en un sistema libre de células). La porción aldehído del residuo de FGly resultante puede usarse como un "asa química" para facilitar la modificación química sitio específica de la proteína.

15 La Figura 1A es un esquema que muestra métodos ilustrativos y composiciones de la invención. En este ejemplo, un motivo sulfatasa ilustrativo ("LCTPSR") está posicionado en una construcción que contiene un ácido nucleico que codifica una proteína de interés. En este ejemplo, el motivo sulfatasa está posicionado en el N-terminal de la proteína codificada después de la expresión; sin embargo, como se describe con más detalle más abajo, los motivos sulfatasa pueden insertarse en uno o más sitios preferidos del polipéptido (por ejemplo, para proporcionar el motivo en el N-terminal, el C-terminal y/o en el sitio interno del polipéptido codificado). El motivo sulfatasa que se ilustra en la Figura 1A está dentro de un género de motivos sulfatasa como se describe más abajo con más detalle. La Figura 1B es un esquema de un alineamiento de secuencia del motivo sulfatasa de una variedad de sulfatasas que se encuentran en diversos organismos. La secuencia consenso contiene la secuencia de las dos etiquetas de aldehído usadas en este estudio. Se destacan los residuos conservados.

25 Tras la expresión en una célula y/o la exposición a la enzima adecuada (por ejemplo, la FGE tipo AtsB o tipo SUMF1), la cisteína codificada del motivo sulfatasa se convierte en una formilglicina (FGly). El aldehído del residuo de FGly puede usarse como un "asa química" para una variedad de aplicaciones, por ejemplo, para la ligación covalente con una porción de interés o para aplicaciones tales como la inmovilización de proteínas. En la Figura 1A, la porción ilustrativa es un marcador detectable que está unido al residuo de cisteína modificado de la porción de sulfatasa.

30 La ubicación de la etiqueta de aldehído dentro de la proteína objetivo a modificar y la modificación mediada por etiqueta de aldehído como se describen en la presente son generalizables con respecto a una amplia variedad de proteínas. La capacidad de la FGE de facilitar la conversión del motivo sulfatasa para generar un residuo de FGly es independiente de la posición del motivo dentro de la proteína. Debido a que la FGE puede convertir la cisteína/serina del motivo sulfatasa de manera que es independiente del contexto de la secuencia e independiente del contexto estructural, las etiquetas de aldehído pueden posicionarse en cualquier sitio preferido dentro de un polipéptido objetivo a modificar, con la condición de que el motivo sulfatasa es accesible a la FGE en el momento de la conversión enzimática. Además, la reactividad única del aldehído permite la modificación biortogonal y quimioselectiva de proteínas recombinantes, lo que proporciona así un medio sitio específico para la modificación química de proteínas que puede realizarse bajo condiciones fisiológicas y en una manera altamente selectiva.

35 Como se apreciará a partir de la presente descripción, las aplicaciones de las etiquetas de aldehído son numerosas y pueden proporcionar un número de ventajas. Por ejemplo, la etiqueta de aldehído es más pequeña que la mayoría si no todas las etiquetas peptídicas convencionales que permiten la modificación covalente de las proteínas, de esta manera se requieren cambios mínimos en la secuencia de aminoácidos de un polipéptido objetivo. En segundo lugar, la etiqueta de aldehído toma ventaja de las químicas de marcaje secundario bien caracterizadas. En tercer lugar, la etiqueta de aldehído demuestra reversibilidad, y a través de la selección de parejas reactivas que proporcionan la conjugación de porciones a través de uniones covalentes de diferente estabilidad, permite la modificación secuencial y el reemplazo de una porción unida a una etiqueta de aldehído. Adicionalmente, debido a que la etiqueta de aldehído se forma con el uso de la maquinaria biosintética ya presente en la mayoría de los sistemas celulares, y es independiente de la naturaleza del objetivo o la ubicación dentro de la secuencia de aminoácidos progenitora, la etiqueta de aldehído puede usarse para facilitar la modificación de un gran número de polipéptidos con el uso de un sistema de expresión disponible fácilmente.

55 La porción aldehído de una etiqueta de aldehído convertido puede usarse para una variedad de aplicaciones que incluyen, pero sin limitarse a, la visualización con el uso de marcaje de fluorescencia o de epítipo (por ejemplo, microscopía electrónica con el uso de partículas de oro equipadas con grupos aldehído reactivos), inmovilización de proteínas (por ejemplo, producción de micromatrices de proteínas), estudios de dinámica y localización de proteínas y aplicaciones, y la conjugación de proteínas con una porción de interés (por ejemplo, porciones que mejoran el índice terapéutico de una proteína progenitora (por ejemplo, PEG), porciones dirigidas (por ejemplo, para mejorar la biodisponibilidad a un sitio de acción), y porciones biológicamente activas (por ejemplo, una porción terapéutica).

60 Es de particular interés el uso de etiquetas de aldehído para facilitar la unión sitio específica de un polímero soluble en agua, tal como PEG. A pesar de los avances en la química de conjugación de proteínas, la modificación sitio específica controlada de proteínas, sigue siendo un desafío. Muchos métodos convencionales de pegilación unen porciones de

PEG a través de la reacción con, por ejemplo, una lisina o cisteína como un residuo objetivo. Debido a la presencia de múltiples residuos objetivo en una proteína, tales sistemas convencionales pueden resultar en la pegilación en múltiples sitios, lo que crea una colección de distintos conjugados de proteína-PEG con diferentes farmacocinéticas. En contraste, el uso de un residuo de FGly de una etiqueta de aldehído como un residuo objetivo proporciona un sitio único para la unión covalente del polímero, y así aumenta tanto la especificidad y la homogeneidad del producto modificado resultante. Estos y otros elementos y ventajas serán fácilmente evidentes para el técnico generalmente con experiencia tras la lectura de la presente descripción.

Los métodos y composiciones para la práctica de la invención se describirán ahora con más detalle.

Etiquetas de aldehído

Generalmente, una etiqueta de aldehído puede basarse en cualquier secuencia de aminoácidos derivada de un motivo sulfatasa (referido además como un "dominio sulfatasa") que es capaz de convertirse por la acción de una enzima generadora formilglicina (FGE) para contener una formilglicina (FGly). La acción de la FGE se dirige de una manera específica de secuencia en que la FGE actúa en un motivo sulfatasa, pero este motivo sulfatasa puede posicionarse dentro de cualquier región de un polipéptido objetivo. Así, la conversión de un motivo sulfatasa mediada por la FGE es sitio específica (es decir, en que la FGE actúa en la secuencia de aminoácidos de un motivo sulfatasa) pero la capacidad de la FGE de actuar sobre el motivo sulfatasa es independiente del contexto de la secuencia (es decir, la capacidad de la FGE de convertir una cisteína/serina de un motivo sulfatasa es independiente del contexto de la secuencia en la que el motivo sulfatasa se presenta en el polipéptido objetivo).

Etiquetas de aldehído ilustrativas

Un motivo sulfatasa mínimo de una etiqueta de aldehído es usualmente de aproximadamente 5 o 6 residuos de aminoácidos de longitud, usualmente no más de 6 residuos de aminoácidos de longitud. Generalmente, se prefiere normalmente minimizar el grado de modificación de la secuencia de aminoácidos nativa del polipéptido objetivo, para minimizar el número de residuos de aminoácidos que se insertan, se delecionan, se sustituyen (reemplazan), o adicionan (por ejemplo, al N o C-terminal). Usualmente se prefiere minimizar el grado de modificación de la secuencia de aminoácidos del polipéptido objetivo para minimizar el impacto que tales modificaciones puedan tener sobre la función y/o la estructura del polipéptido objetivo. Así, las etiquetas de aldehído de interés particular incluyen aquellas que requieren la modificación (inserción, adición, delección, sustitución/reemplazo) de menos de 16, 15, 14, 13, 12, 11, 10, 9, 8, o 7 residuos de aminoácidos de la secuencia de aminoácidos del polipéptido objetivo.

Debe destacarse que mientras que las etiquetas de aldehído de particular interés son aquellas basadas en un motivo sulfatasa mínimo, se apreciará fácilmente que las etiquetas de aldehído más largas se contemplan y abarcan por la presente descripción y pueden usarse en la composiciones y métodos de la invención. Las etiquetas de aldehído pueden comprender así un motivo sulfatasa mínimo de 5 o 6 residuos, o pueden ser más largas y comprender un motivo sulfatasa mínimo que puede flanquearse en los lados N y/o C-terminal del motivo por residuos de aminoácidos adicionales. Se contemplan las etiquetas de aldehído de, por ejemplo, 5 o 6 residuos de aminoácidos, así como también las secuencias de aminoácidos más largas de más de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más residuos de aminoácidos.

Generalmente, los motivos sulfatasa útiles en etiquetas de aldehído como se describe en la presente son de fórmula:



donde

$Z_1$  es cisteína o serina (las cuales pueden representarse además por (C/S));

$Z_2$  es un residuo de prolina o alanina (las cuales pueden representarse además por (P/A));

$X_1$  está presente o ausente y, cuando está presente, puede ser cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente L, M, V, S o T, más usualmente de L, M, S o V, con la condición de que cuando el motivo sulfatasa está en el N-terminal del polipéptido objetivo,  $X_1$  está presente; y

$X_2$  y  $X_3$  pueden ser independientemente cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido polar, sin carga, o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente S, T, A, V, G o C, más usualmente S, T, A, V o G.

Debe destacarse que, después de la acción de una FGE sobre el motivo sulfatasa,  $Z_1$  se oxida para generar un residuo de formilglicina (FGly). Además, después de la conversión mediada por FGE y la reacción con una pareja reactiva que comprende una porción de interés, la FGly en la posición  $Z_1$  en la fórmula anterior está unida covalentemente a la porción de interés (por ejemplo, marcador detectable, polímero soluble en agua, etc).

5 Cuando la etiqueta de aldehído está presente en un lugar distinto del N-terminal del polipéptido objetivo,  $X_1$  de la fórmula anterior puede proporcionarse por un residuo de aminoácido de la secuencia de aminoácidos nativa del polipéptido objetivo. Por lo tanto, en algunas modalidades, y cuando están presentes en un lugar distinto del N-terminal de un polipéptido objetivo, los motivos sulfatasa son de fórmula:



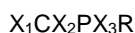
10 donde  $X_1$  y  $X_2$  pueden ser cualquier aminoácido independientemente, aunque usualmente un aminoácido alifático, un aminoácido polar, sin carga, o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente S, T, A, V, C o, más usualmente S, T, A o V.

15 Como se destacó anteriormente, el motivo sulfatasa puede contener residuos adicionales en uno o ambos de los N y C-terminales de la secuencia, por ejemplo, de manera que la etiqueta de aldehído incluya un motivo sulfatasa y un "motivo auxiliar". En una modalidad, el motivo sulfatasa incluye un motivo auxiliar en el C-terminal (es decir, después del residuo de arginina en la fórmula anterior) 1, 2, 3, 4, 5, 6, o los 7 residuos contiguos de una secuencia de aminoácidos de AALLTGR, SLLTGR, AAFMTGR, AAFLTGR, SAFLTGR, ASILTGR, VSFLTGR, ASLLTGL, ASILITG, VSFLTGR, SAIMTGR, SAIVTGR, TNLWRG, TNLWRGQ, TNLCAAS, VSLWTGK, SMLLTG, SMLLTGN, SMLLTGT, ASFMAGQ, o ASLLTGL, (ver, por ejemplo, Dierks y otros (1999) EMHO J 18(8): 2084-2091), o de GSLFTGR. Sin embargo, como se  
20 expone en los ejemplos más abajo, los presentes inventores han encontrado que tales residuos de aminoácidos adicionales en el C-terminal no se requieren para la conversión mediada por la FGE del motivo sulfatasa de la etiqueta de aldehído, y así sólo son opcionales y pueden excluirse específicamente de las etiquetas de aldehído que se describen en la presente. En algunas modalidades la etiqueta de aldehído no contiene una secuencia de aminoácidos CGPSR(M/A)S o CGPSR(M/A), que puede estar presente como una secuencia de aminoácidos nativa en las hidrolasas de monoéster fosfonato.  
25

El motivo sulfatasa de la etiqueta de aldehído se selecciona generalmente para ser capaz de convertirse por una FGE seleccionada, por ejemplo, una FGE presente en una célula huésped en la que se expresa el polipéptido etiquetado con aldehído o una FGE que debe ponerse en contacto con el polipéptido etiquetado con aldehído en un método *in vitro* libre de células.  
30

35 La selección de las etiquetas de aldehído y una FGE que proporcionen parejas reactivas apropiadas para proporcionar la generación de una FGly en el polipéptido objetivo etiquetado con aldehído puede lograrse fácilmente a la luz de la información disponible en la técnica. Generalmente, los motivos sulfatasa susceptibles a la conversión por una FGE procariota contienen ya sea una cisteína o una prolina (es decir, una cisteína y prolina en  $Z_1$  y  $Z_2$ , respectivamente, en la fórmula I anterior (por ejemplo,  $X_1CX_2PX_3R$ );  $CX_1PX_2R$  en la fórmula II anterior) y se modifican por la FGE "tipo SUMF1" (Cosma y otros Cell 2003, 113, (4), 445-56; Dierks y otros Cell 2003, 113, (4), 435-44). Los motivos sulfatasa susceptibles a la conversión por una FGE procariota contienen ya sea una cisteína o una serina en el motivo sulfatasa (es decir, una cisteína o serina en  $Z_1$ , y una prolina en  $Z_2$ , respectivamente, en la fórmula I anterior (por ejemplo,  $X_1(C/S)X_2PX_3R$ );  $(C/S)X_1PX_2R$  en la fórmula II anterior) se modifican por FGE "tipo SUMF1" o el FGE "tipo AtsB", respectivamente (Szameit y otros J Biol Chem 1999, 274, (22), 15375-81). Otros motivos sulfatasa susceptibles a la conversión por una FGE procariota contienen ya sea una cisteína o una serina, y, o bien una prolina o una alanina en el motivo sulfatasa (es decir, una cisteína o serina en  $Z_1$ , y una prolina o alanina en  $Z_2$ , respectivamente, en la fórmula I anterior (por ejemplo,  $X_1CX_2PX_3R$ ;  $X_1SX_2PX_2R$ ;  $X_1CX_2AX_3R$ ;  $X_1S7C_2AX_3R$ );  $CX_1PX_2R$ ;  $SX_1PX_2R$ ;  $CX_1AX_2R$ ;  $SX_1AX_2R$  en la fórmula II anterior), y son susceptibles a la modificación, por ejemplo, puede ser modificada por una FGE de un Firmicutes (por ejemplo, Clostridium perfringens) (ver Berteau y otros J. Biol. Chem..2006; 281:22464-22470).  
45

Por lo tanto, por ejemplo, donde la FGE es una FGE eucariota (por ejemplo, una FGE de mamífero, que incluye una FGE humana), el motivo sulfatasa es usualmente de fórmula:  
50



donde

55  $X_1$  puede estar presente o ausente y, cuando está presente, puede ser cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente L, M, S o V, con la condición de que cuando el motivo sulfatasa está en el N-terminal del polipéptido objetivo,  $X_1$  está presente; y  
60  $X_2$  y  $X_3$  pueden ser independientemente cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente S, T, A, V, G, o C, más usualmente S, T, A, V o G.

Los ejemplos específicos de motivos sulfatasa incluyen LCTPSR, MCTPSR, VCTPSR, LCSPSR, LCAPSR LCVPSR y LCGPSR. Otros motivos sulfatasa específicos son fácilmente evidentes a partir de la descripción proporcionada en la presente.  
65

5 Como se describe con más detalle más abajo, un polipéptido convertido etiquetado con aldehído se hace reaccionar con una pareja reactiva que contiene una porción de interés para proporcionar la conjugación de la porción de interés al residuo de FGly del polipéptido convertido etiquetado con aldehído, y la producción de un polipéptido modificado. Los polipéptidos modificados que tienen una etiqueta de aldehído modificada se describen generalmente que comprenden un motivo sulfatasa modificado de fórmula:



10 donde

FGly' es el residuo de formilglicina que tiene una porción unida covalentemente;

Z<sub>2</sub> es un residuo de prolina o alanina (que pueden representarse además por (P/A));

15 X<sub>1</sub> puede estar presente o ausente y, cuando está presente, puede ser cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente L, M, V, S o T, más usualmente de L, M o V, con la condición de que cuando el motivo sulfatasa está en el N-terminal del polipéptido objetivo, X<sub>1</sub> está presente; y

20 X<sub>2</sub> y X<sub>3</sub> pueden ser independientemente cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, sin carga, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente S, T, A, V, G o C, más usualmente S, T, A, V o G.

25 Ejemplos específicos de motivos sulfatasa convertidos incluyen L(FGly)TPSR, M(FGly)TPSR, V(FGly)TPSR, L(FGly)SPSR, L(FGly)APSR L(FGly)VPSR, y L(FGly)GPSR.

30 Como se describe con más detalle más abajo, la porción de interés puede ser cualquiera de una variedad de porciones tales como un polímero soluble en agua, un marcador detectable, un fármaco, o una porción para la inmovilización del polipéptido en una membrana o sobre una superficie. Como es evidente a partir de la discusión anterior de los polipéptidos etiquetados con aldehído, el motivo sulfatasa modificado del polipéptido modificado puede posicionarse en cualquier sitio preferido del polipéptido. Así, la presente descripción proporciona, por ejemplo, un polipéptido modificado que tiene un motivo sulfatasa modificado posicionado en un sitio de modificación postraducciona de un progenitor del polipéptido modificado (es decir, si el polipéptido objetivo se modifica para proporcionar una etiqueta de aldehído en un sitio de modificación postraducciona, el polipéptido modificado producido más tarde contendrá una porción en una posición correspondiente a este sitio de modificación postraducciona en el polipéptido progenitor). Por ejemplo, después, puede producirse un polipéptido modificado para tener un polímero soluble en agua unido covalentemente, en un sitio correspondiente a un sitio en el que normalmente se produciría la glicosilación en el polipéptido progenitor objetivo. Así, por ejemplo, puede producirse un polipéptido pegilado que tiene la porción PEG posicionada en el mismo o casi el mismo lugar en que se posicionarían los residuos de azúcar en el polipéptido progenitor de origen natural. De manera similar, cuando el polipéptido objetivo progenitor está diseñado para incluir uno o más sitios no nativos de modificación postraducciona, el polipéptido modificado puede contener polímeros solubles en agua unidos covalentemente en uno o más sitios del polipéptido modificado que corresponden a estos sitios no nativos de modificación postraducciona en el polipéptido progenitor.

#### 45 Modificación de un polipéptido objetivo para incluir una etiqueta de aldehído

50 Las etiquetas de aldehído pueden posicionarse en cualquier lugar dentro de un polipéptido objetivo al que se desea proporcionar la conversión y/o modificación del polipéptido objetivo, con la condición de que el sitio de la etiqueta de aldehído sea accesible para la conversión por una FGE y la modificación subsecuente en la FGly, o pueda volverse accesible (por ejemplo, mediante la desnaturalización de la proteína). Los polipéptidos objetivo pueden modificarse para incluir una o más etiquetas de aldehído. El número de etiquetas de aldehído que pueden estar presentes en un polipéptido objetivo variará con el polipéptido objetivo seleccionado, y puede incluir 1, 2, 3, 4, 5, o más etiquetas de aldehído.

55 En algunas modalidades se prefiere posicionar la(s) etiqueta(s) de aldehído en el polipéptido objetivo teniendo en cuenta su estructura cuando se pliega (por ejemplo, en un medio ambiente libre de células, usualmente un medio ambiente fisiológico libre de células) y/o se presenta en o sobre una membrana celular (por ejemplo, para polipéptidos asociados a la membrana celular, tales como proteínas de transmembrana). Por ejemplo, una etiqueta de aldehído puede posicionarse en un sitio accesible al disolvente en el polipéptido objetivo plegado. La etiqueta aldehído accesible al disolvente en un polipéptido no convertido plegado etiquetado con aldehído es accesible así a una FGE para la conversión de la serina o cisteína a una FGly. Similarmente, una etiqueta de aldehído accesible al disolvente de un polipéptido convertido etiquetado con aldehído es accesible a un reactivo de pareja reactiva para la conjugación a una porción de interés para proporcionar un polipéptido modificado etiquetado con aldehído. Cuando una etiqueta de aldehído está posicionada en un sitio accesible al disolvente, la conversión in vitro mediada por la FGE y la modificación con una porción por la reacción con una pareja reactiva puede realizarse sin la necesidad de desnaturalizar la proteína.

Los sitios accesibles al disolvente pueden incluir además regiones del polipéptido objetivo que están expuestas a una superficie celular extracelular o intracelular cuando se expresan en una célula huésped (por ejemplo, distintas de una región transmembrana del polipéptido objetivo).

5 En consecuencia, o más etiquetas de aldehído pueden proporcionarse en los sitios independientemente seleccionados de, por ejemplo, un N-terminal accesible al disolvente, una región N-terminal accesible al disolvente, un C-terminal accesible al disolvente, una región C-terminal accesible al disolvente, y/o una estructura de bucle (por ejemplo, una estructura de bucle extracelular y/o una estructura de bucle intracelular). En algunas modalidades, la etiqueta de aldehído está posicionada en un sitio distinto del C-terminal del polipéptido. En otras modalidades, el polipéptido en el que está posicionada la etiqueta de aldehído es un polipéptido de longitud completa.

10 En otras modalidades, un sitio de etiqueta de aldehído está posicionado en un sitio que está modificado postraduccionalmente en el polipéptido objetivo nativo. Por ejemplo, una etiqueta de aldehído puede introducirse en un sitio de glicosilación (por ejemplo, N-glicosilación, O-glicosilación), fosforilación, sulfatación, ubiquitinación, acilación, metilación, prenilación, hidroxilación, carboxilación, y similares en el polipéptido objetivo nativo. Las secuencias consenso de una variedad de sitios modificados postraduccionalmente, y los métodos para la identificación de un sitio modificado postraduccionalmente en un polipéptido, se conocen bien en la técnica. Se entiende que el sitio de modificación postraduccional puede ser de origen natural o un sitio de un polipéptido que se ha diseñado (por ejemplo, a través de técnicas recombinantes) para incluir un sitio de modificación postraduccional que no es nativo para el polipéptido (por ejemplo, como en un sitio de glicosilación de una variante hiperglicosilada de la EPO). En la última modalidad, son de particular interés los polipéptidos que tienen un sitio de modificación postraduccional no nativo y que se ha demostrado que exhiben una actividad biológica de interés.

25 La descripción proporciona además en la presente los métodos para identificar sitios apropiados para la modificación de un polipéptido objetivo para incluir una etiqueta de aldehído. Por ejemplo, pueden producirse una o más construcciones de polipéptidos objetivo etiquetados con aldehído, y las construcciones expresarse en una célula que expresa una FGE, o exponerse a la FGE después del aislamiento a partir de la célula (como se describe con más detalle más abajo). El polipéptido etiquetado con aldehído después puede ponerse en contacto con una pareja reactiva que, si la etiqueta de aldehído está accesible, proporciona la unión de una porción detectable a la FGly de la etiqueta de aldehído. Después se determina la presencia o ausencia de la porción detectable. Si se detecta la porción detectable, entonces fue exitoso el posicionamiento de la etiqueta de aldehído en el polipéptido. De esta manera, una genoteca de construcciones que tienen una etiqueta de aldehído posicionada en sitios diferentes en la secuencia codificante del polipéptido objetivo puede producirse y tamizarse para facilitar la identificación de una posición óptima de una etiqueta de aldehído. Adicionalmente o alternativamente, el polipéptido etiquetado con aldehído puede probarse para una actividad biológica que se asocia normalmente con el polipéptido objetivo, y/o la estructura del polipéptido etiquetado con aldehído evaluado (por ejemplo, para evaluar si un epítipo presente normalmente en una superficie celular extracelular en el polipéptido objetivo nativo está presente además en el polipéptido etiquetado con aldehído).

40 Puede proporcionarse una etiqueta de aldehído en un polipéptido objetivo mediante inserción (por ejemplo, para proporcionar la inserción de un residuo de 5 o 6 aminoácidos dentro de la secuencia de aminoácidos nativa) o mediante adición (por ejemplo, en un N o C-terminal del polipéptido objetivo). Puede proporcionarse además una etiqueta de aldehído mediante la sustitución completa o parcial de residuos de aminoácidos nativos con la secuencia de aminoácidos contigua de una etiqueta de aldehído. Por ejemplo, puede proporcionarse una etiqueta de aldehído heteróloga de 5 (o 6) residuos de aminoácidos en un polipéptido objetivo mediante el reemplazo de 1, 2, 3, 4, o 5 (o 1, 2, 3, 4, 5, o 6) residuos de aminoácidos de la secuencia de aminoácidos nativa con los residuos de aminoácidos correspondientes de la etiqueta de aldehído. Aunque generalmente pueden ser de menor interés en muchas aplicaciones, los polipéptidos objetivo que tienen más de una etiqueta de aldehído pueden modificarse para proporcionar la unión de la misma porción o de diferentes porciones en la FGly de la etiqueta.

50 La modificación de un polipéptido objetivo para incluir una o más etiquetas de aldehído puede lograrse con el uso de técnicas recombinantes de genética molecular, para producir ácido nucleico que codifica el polipéptido objetivo preferido etiquetado con aldehído. Tales métodos se conocen bien en la técnica, e incluyen métodos de clonación, métodos de mutación sitio específica, y similares (ver, por ejemplo, Sambrook y otros, en "Molecular Cloning: A Laboratory Manual" (Cold Spring Harbor Laboratory Press 1989); "Current Protocols in Molecular Biology" (eds., Ausubel y otros; Greene Publishing Associates, Inc., y John Wiley & Sons, Inc. 1990 y suplementos). Alternativamente, una etiqueta aldehído se puede añadir usando técnicas no recombinantes, por ejemplo, el uso de ligación química nativa o ligación química pseudo-nativa, por ejemplo, para añadir una etiqueta aldehído a un C-terminal del polipéptido objetivo (ver, por ejemplo, US 6,184,344; US 6,307,018; US 6,451,543; US 6,570,040; US 2006/0173159; US 2006/0149039). Ver además Rush y otros (5 de ene., 2006) Org Lett. 8(1):131-4.

#### 60 Polipéptidos objetivos

65 Cualquiera de una amplia variedad de polipéptidos puede modificarse para incluir una etiqueta de aldehído para facilitar la modificación del polipéptido. Los polipéptidos apropiados para la modificación basada en etiquetas de aldehído incluyen proteínas que tienen una secuencia de aminoácidos de origen natural, una secuencia de aminoácidos nativa

que tiene una metionina N-terminal, fragmentos de polipéptidos de origen natural, y polipéptidos que no son de origen natural y fragmentos de estos. En algunas modalidades, el polipéptido objetivo es un polipéptido distinto de una sulfatasa o fragmento de esta, distinto de una proteína reportera, o distinto de preprolactina o prolactina.

- 5 Los siguientes son clases ilustrativas y tipos de polipéptidos que son de interés para la modificación con el uso de los métodos basados en etiquetas de aldehído descritos en la presente.

#### Polipéptidos terapéuticos

- 10 En una modalidad, los métodos de modificación de proteínas basados en etiquetas de aldehído se aplican a la modificación de polipéptidos que pueden proporcionar un beneficio terapéutico, particularmente los polipéptidos para los que la unión a una porción que puede proporcionar uno o más de, por ejemplo, un aumento en la vida media en suero, una disminución en una respuesta inmune adversa, la actividad o funcionalidad biológica adicional o alternativa, y similares, o reducción de un efecto secundario adverso u otro beneficio. Cuando el polipéptido terapéutico es un
- 15 antígeno para una vacuna, la modificación puede proporcionar una inmunogenicidad mejorada del polipéptido.

- Los ejemplos de clases de proteínas terapéuticas incluyen aquellas que son citocinas, quimocinas, factores de crecimiento, hormonas, anticuerpos, y antígenos. Los ejemplos adicionales incluyen eritropoyetina (EPO, por ejemplo, EPO nativa, EPO sintética (ver, por ejemplo, Estados Unidos 2003/0191291), hormona del crecimiento humana (hGH), hormona del crecimiento bovina (bGH), hormona estimulante del folículo (FSH), interferón (por ejemplo, IFN-gamma, IFN-beta, IFN-alfa, IFN-omega, interferón consenso, y similares), insulina, factor de crecimiento similar a la insulina (por ejemplo, IGF-I, IGF-II), factores de la sangre (por ejemplo, Factor VIII, Factor IX, Factor X, activador del plasminógeno tisular (TPA), y similares), factores estimulantes de colonias (por ejemplo, CSF de granulocitos (G-CSF), CSF de macrófagos (M-CSF), CSF de granulocitos y macrófagos (GM-CSF), y similares), factores de crecimiento transformantes (por ejemplo, TGF-beta, TGF-alfa), interleucinas (por ejemplo, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-12, y similares), factor de crecimiento epidérmico (EGF), factor de crecimiento derivado de plaquetas (PDGF), factores de crecimiento de fibroblastos (FGF, por ejemplo, aFGF, bFGF), factor de crecimiento derivado de línea de células gliales (GDNF), factor de crecimiento nervioso (NGF), RANTES, y similares.
- 20
- 25

- 30 Los ejemplos adicionales incluyen anticuerpos, por ejemplo, anticuerpos policlonales, anticuerpos monoclonales, anticuerpos humanizados, fragmentos de unión a antígeno (por ejemplo, F(ab)', Fab, Fv), anticuerpos de cadena única, y similares. De particular interés son los anticuerpos que se unen específicamente a un antígeno tumoral, un antígeno de células inmunológicas (por ejemplo, CD4, CD8, y similares), un antígeno de un microorganismo, particularmente un microorganismo patógeno (por ejemplo, un antígeno bacteriano, viral, fúngico, o parasitario), y similares.
- 35

- Los métodos y composiciones descritos en la presente se pueden aplicar para proporcionar una porción (por ejemplo, un polímero soluble en agua) en un sitio de glicosilación natural o modificado, tal como los que se encuentran en formas hiperglicosiladas de una proteína terapéutica, tales como, por ejemplo: un **interferón** (por ejemplo, IFN- $\gamma$ , IFN- $\alpha$ , IFN- $\beta$ , IFN- $\omega$ ; IFN- $\tau$ ); una insulina (por ejemplo, Novolin, Humulin, Humalog, Lantus, Ultralente, etc.); una **eritropoyetina** (por ejemplo, PROCRIT®, EPREX®, o EPOGEN® (epoetina- $\alpha$ ); ARANESP® (darbepoyetina- $\alpha$ ); NEORECORMON®, EPOGIN® (epoetina- $\beta$ ); y similares); un **anticuerpo (por ejemplo, un anticuerpo monoclonal)** (por ejemplo, RITUXAN® (rituximab); REMICADE® (infliximab); HERCEPTIN® (trastuzumab); HUMIRA™ (adalimumab); XOLAIR® (omalizumab); BEXXAR® (tositumomab); RAPTIVA™ (efalizumab); ERBITUX™ (cetuximab); y similares), que incluyen un fragmento de unión al antígeno de un anticuerpo monoclonal; un **factor de la sangre** (por ejemplo, ACTIVASE® (alteplase) activador del plasminógeno tisular; NOVOSEVEN® (factor VIIa recombinante humano); Factor VIIa; Factor VIII (por ejemplo, KOGENATE®); Factor IX;  $\beta$ -globina; hemoglobina; y similares); un **factor de estimulación de colonias** (por ejemplo, NEUPOGEN® (filgrastim; G-CSF); Neulasta (pegfilgrastim); factor estimulante de colonias de granulocitos (G-CSF), factor estimulante de colonias de granulocitos-monocitos, factor estimulante de colonias de macrófago, factor estimulante de colonias de megacariocito; y similares); a **hormona del crecimiento** (por ejemplo, una somatotropina, por ejemplo, GENOTROPIN®, NUTROPIN®, NORDITROPIN®, SAIZEN®, SEROSTIM®, HUMATROPE®, etc.; una hormona del crecimiento humano; y similares); una **interleucina** (por ejemplo, IL-1; IL-2, que incluyen, por ejemplo, Proleukin®; IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9; etc.); un **factor de crecimiento** (por ejemplo, REGRANEX® (beclaplerin; PDGF); FIBLAST® (trafermin; bFGF); STEMGEN® (ancestim; factor de célula madre); factor de crecimiento de queratinocitos; un factor de crecimiento de fibroblasto ácido, un factor de células madre, un factor de célula madre de fibroblastos básico, un básico de hepatocitos; y similares); un receptor **soluble** (por ejemplo, un receptor soluble de unión a TNF- $\alpha$  tal como ENBREL® (etanercept); un receptor de VEGF soluble; un receptor de interleucina soluble; un receptor celular de  $\gamma/\delta$  T soluble; y similares); una **enzima** (por ejemplo,  $\alpha$ -glucosidasa; CERAZYME® (imiglucarasa;  $\beta$ -glucocerebrosidasa, CEREDASE® (alglucerasa;); un **activador enzimático** (por ejemplo, activador del plasminógeno tisular); una **quimocina** (por ejemplo, IP-10; Mig; Gro $\alpha$ /IL-8, RANTES; MIP-1 $\alpha$ ; MIP-1 $\beta$ ; MCP-1; PF-4; y similares); un **agente angiogénico** (por ejemplo, factor de crecimiento endotelial vascular (VEGF) ; un **agente anti-angiogénico** (por ejemplo, un receptor de VEGF soluble); una vacuna de proteína ; un **péptido neuroactivo** tal como bradiquinina, colecistoquinina, gastina, secretina, oxitocina, hormona liberadora de gonadotropina, beta-endorfina, encefalina, sustancia P, somatostatina, galanina, hormona liberadora de la hormona del crecimiento, bombesina, warfarina, dinorfina, neurotensina, motilina, tiotropina, neuropéptido Y, hormona luteinizante, calcitonina, insulina, glucagon, vasopresina, angiotensina II, hormona liberadora de tiotropina, péptido intestinal
- 40
- 45
- 50
- 55
- 60
- 65



vasoactivo, un péptido del sueño , etc.; **otras proteínas** tales como un agente trombolítico, un péptido natriurético auricular, proteína morfogénica ósea, trombopoyetina, relaxina, proteína ácida glial fibrilar, hormona folículo estimulante, una alfa-1 antitripsina humana, un factor inhibidor de leucemia, un factor de crecimiento transformante, un factor tisular, un factor de crecimiento similar a insulina, una hormona luteinizante, una hormona estimulante del folículo, un factor activador de macrófagos, factor de necrosis tumoral, un factor quimiotáctico de neutrófilos, un factor de crecimiento nervioso, un inhibidor de tejido de metaloproteinasas; un péptido intestinal vasoactivo, angiogenina, angiotropina, fibrina; hirudina; un factor inhibidor de la leucemia; un antagonista del receptor IL-1 (por ejemplo, Kineret® (anakinra)); y similares. Se apreciará fácilmente que las formas nativas de las proteínas terapéuticas anteriores son además de interés como polipéptidos objetivo en la presente invención.

La actividad biológica de un polipéptido objetivo modificado puede ensayarse de acuerdo con los métodos conocidos en la técnica. Son de interés los polipéptidos modificados etiquetados con aldehído que retienen al menos una actividad farmacológica preferida de la proteína progenitora correspondiente. Los ejemplos de ensayos útiles para proteínas terapéuticas particulares incluyen, pero sin limitarse a, GMCSF (Eaves, CA y Eaves CJ, Erythropoiesis in culture. En: McCulloch E A (ed) Cell culture techniques--Clinics in hematology. W B Saunders, Eastbourne, págs. 371-91 (1984); Metcalf, D., International Journal of Cell Cloning 10: 116-25 (1992); Testa, N. G., y otros, Assays for hematopoietic growth factors. En: Balkwill F R (ed) Cytokines A practical Approach, págs. 229-44; IRL Press Oxford 1991) EPO (bioassay: Kitamura y otros, J. Cell. Physiol. 140 p323 (1989)); Hirudin (platelet aggregation assay: Blood Coagul Fibrinolysis 7(2):259-61 (1996)); IFN $\alpha$  (anti-viral assay: Rubinstein y otros, J. Virol. 37(2):755-8 (1981)); ensayo anti-proliferativo : Gao Y, y otros Mol Cell Biol. 19(11):7305-13 (1999); y bioensayo: Czarniecki y otros, J. Virol. 49 p490 (1984)); GCSF (bioensayo: Shirafuji y otros, Exp. Hematol. 17 p116 (1989); proliferación de células NFS-60 murinas (Weinstein y otros, Proc Natl Acad Sci 83:5010-4 (1986)); insulina (<sup>3</sup>ensayo de captación de H-glucosa: Steppan y otros, Nature 409(6818):307-12 (2001)); ensayo de proliferación de hGH (Ba/F3-hGHR: J Clin Endocrinol Metab 85(11):4274-9 (2000); estándar internacional para la hormona del crecimiento: Horm Res, 51 Suppl 1:7-12 (1999)); factor X (factor X activity assay: Van Wijk y otros Thromb Res 22:681-686 (1981)); factor VII (coagulation assay using prothrombin clotting time: Belaouaj y otros, J. Biol. Chem. 275:27123-8(2000); Diaz-Collier y otros, Thromb Haemost 71:339-46 (1994)).

#### *Composiciones inmunogénicas*

La tecnología basada en etiquetas de aldehído descrita en la presente encuentra además aplicación en la producción de componentes de composiciones inmunogénicas (por ejemplo, vacunas terapéuticas). Por ejemplo, puede usarse una etiqueta de aldehído para facilitar la unión de porciones que aumentan la vida media en suero de un antígeno polipeptídico, que aumenta la inmunogenicidad del polipéptido, o que enlaza un antígeno no aminoacídico a un polipéptido portador. Con respecto a esto, las etiquetas de aldehído pueden usarse para facilitar la modificación de antígenos microbianos (por ejemplo, un antígeno bacteriano, viral, fúngico, o parasitario), antígenos tumorales, y otros antígenos que son de interés para la administración a un sujeto para obtener una respuesta inmune en el sujeto. Además es de interés la modificación de los antígenos que son útiles en la obtención de anticuerpos que pueden ser útiles como herramientas de investigación.

Los polipéptidos ilustrativos adicionales de interés para la modificación con el uso de etiqueta(s) de aldehído incluyen aquellos que son de interés para la detección o monitoreo funcional en un ensayo (por ejemplo, como una herramienta de investigación, en un ensayo de tamizaje de fármacos, y similares). Los polipéptidos ilustrativos de este tipo incluyen receptores (por ejemplo, receptores acoplados a proteína G (GPCR, que incluyen los GPCR huérfanos)), ligandos de receptor (que incluyen los de origen natural y sintético), canales proteicos (por ejemplo, canales de iones (por ejemplo, los canales de potasio, canales de calcio, canales de sodio, y similares), y otros polipéptidos. En una modalidad, la modificación de polipéptidos asociados a la superficie celular, tales como polipéptidos transmembrana) es de particular interés, particularmente donde tal modificación se logra mientras que el polipéptido está presente en una membrana. Los métodos para la modificación de polipéptido etiquetado con aldehído bajo condiciones fisiológicas se describen más abajo.

#### Enzimas generadoras de formilglicina (FGEs)

La enzima que oxida cisteína o serina en un motivo sulfatasa a FGly se refiere en la presente como una enzima generadora de formilglicina (FGE). Como se discutió anteriormente, "FGE" se usa en la presente para referirse a las enzimas generadoras de FGly que median la conversión de una cisteína (C) de un motivo sulfatasa a FGly así como también las enzimas generadoras de FGly que median la conversión de serina (S) de un motivo sulfatasa a FGly. Debe destacarse que generalmente, la literatura se refiere a las enzimas generadoras de FGly que convierten una C a FGly en un motivo sulfatasa como FGE, y refiere a las enzimas que convierten S a FGly en un motivo sulfatasa como similar a Ats-B. Sin embargo, para los propósitos de la presente descripción "FGE" se usa genéricamente para referirse a ambos tipos de enzimas generadoras de FGly, con la comprensión de que una FGE adecuada será seleccionada de acuerdo con la pareja reactiva objetivo que contiene el motivo sulfatasa adecuado (es decir, que contiene C o que contiene S).

Como se evidencia por la presencia ubicua de sulfatasas que tienen una FGly en el sitio activo, las FGE se encuentran en una amplia variedad de tipos celulares, que incluyen tanto eucariotas como procariotas. Hay al menos dos formas de

FGE. Las sulfatasas eucariotas contienen una cisteína en su motivo sulfatasa y se modifican por la FGE "tipo SUMF1" (Cosma y otros Cell 2003, 113, (4), 445-56, Dierks y otros Cell 2003, 113, (4), 435-44). La enzima generadora de FGly (FGE) es codificada por el gen *SUMF1*. Las sulfatasas procariotas pueden contener ya sea una cisteína o una serina en su motivo sulfatasa y se modifican ya sea por la FGE "tipo SUMF1" o la FGE "tipo AtsB", respectivamente (Szameit y otros, J. Biol. Chem. 1999, 274, (22), 15375-81). En eucariotas, se cree que esta modificación ocurre cotraduccionalmente o poco después de la traducción en el retículo endoplásmico (ER) (Dierks y otros Proc Natl Acad Sci Estados Unidos 1997, 94 (22):11963-8). Sin restringirse a la teoría, se cree que en procariotas la FGE tipo SUMF1 funciona en el citosol y la FGE tipo AtsB funciona cerca de o en la membrana celular. Se ha descrito además una FGE tipo SUMF2 en Deuterostomia, que incluye vertebrados y equinodermos (ver, por ejemplo, Pepe y otros (2003) Cell 113, 445-456, Dierks y otros (2003) Cell 113, 435-444; Cosma y otros (2004) Hum. Mutat. 23, 576-581).

Generalmente, la FGE que se usa para facilitar la conversión de cisteína o serina en FGly en un motivo sulfatasa de una etiqueta de aldehído de un polipéptido objetivo se selecciona de acuerdo con el motivo sulfatasa presente en la etiqueta de aldehído. La FGE puede ser nativa para la célula huésped en la que se expresa el polipéptido etiquetado con aldehído, o la célula huésped puede estar genéticamente modificada para expresar una FGE adecuada. En algunas modalidades puede preferirse usar un motivo sulfatasa compatible con una FGE humana (por ejemplo, la FGE tipo SUMF1, ver, por ejemplo, Cosma y otros Cell 113, 445-56 (2003); Dierks y otros Cell 113, 435-44 (2003)), y expresar la proteína etiquetada con aldehído en una célula humana que expresa la FGE o en una célula huésped, usualmente una célula de mamífero, genéticamente modificada para expresar una FGE humana.

Generalmente, una FGE para el uso en los métodos descritos en la presente puede obtenerse a partir de fuentes de origen natural o producirse sintéticamente. Por ejemplo, una FGE adecuada puede derivarse de fuentes biológicas que producen naturalmente una FGE o que están modificadas genéticamente para expresar un gen recombinante que codifica una FGE. Los ácidos nucleicos que codifican un número de FGE se conocen en la técnica y están fácilmente disponibles (ver, por ejemplo, Preusser y otros 2005 J. Biol.Chem. 280(15):14900-10 (Epub 18 de ene. 2005); Fang y otros 2004 J Biol Chem. 79(15):14570-8 (Epub28 de ene. 2004); Landgrebe y otros Gene. 16 de oct. 2003:316:47-56; Dierks y otros 1998 FEBS Lett. 423(1):61-5; Dierks y otros Cell. 16 de mayo de 2003;113(4):435-44; Cosma y otros (16 de may. 2003) Cell 113(4):445-56; Baenziger (16 de may. 2003) Cell 113(4):421-2 (revisión); Dierks y otros Cell. 20 de may. 2003;121(4):541-52; Roeser y otros (3 de ene. 2006) Proc Natl Acad Sci Estados Unidos 103(1):81-6; Sardiello y otros (2005 Nov 1) Hum Mol Genet. 14 (21):3203-17; WO 2004/072275 ; y número de acceso a GenBank M\_182760. En consecuencia, la descripción proporciona células huésped recombinantes modificadas genéticamente para expresar una FGE que es compatible para el uso con una etiqueta de aldehído de un polipéptido objetivo etiquetado.

En una modalidad, una FGE obtenida a partir de *Mycobacterium tuberculosis (Mtb)* se usa en los métodos descritos en la presente. Una FGE de *Mtb* ilustrativa se describe con detalle en los Ejemplos más abajo. Una FGE de *Mtb* ilustrativa es una que tiene la secuencia de aminoácidos proporcionada en GenBank Acc.núm. NP\_215226 (gi:15607852):

```
mltelvdlpg gsfrmgstrf ypeeapihtv tvrafaverh pvtnaqfaef vsatgyvtva
egpldpglyp gvdaadlcpq amvfcptagp vdlrdwrqww dwvpgacwrh pfgdrsdiaa
raghpvqvva ypdavayarw agrrlpteae weyaarggtt atyawgdqek pggmlmantw
qgrfpyrndg algwvgtspv grfpangfgl ldmignvwew tttefyphhr idppstacca
pvklataadp tisqtlkqgs hlcapeychr yrpaarspqs qdtatthigf rcvadpvsq
```

Así, la FGE de *Mtb* y el ácido nucleico que codifica la FGE de *Mtb*, se contemplan para el uso en los presentes métodos. Adicionalmente, los métodos usados para identificar y caracterizar la FGE de *Mtb* son aplicables a la identificación y caracterización de otras FGE útiles en los métodos descritos en la presente.

Provisto con la extensa información sobre la secuencia de aminoácidos y la caracterización de FGE proporcionada en la presente así como también en la técnica, será fácilmente evidente para el técnico generalmente con experiencia que las FGE incluyen FGE de origen natural así como también FGE modificadas que comparten identidad de secuencia con una FGE conocida (por ejemplo, una FGE de origen natural) y que retiene la función de modificación específica de una serina o cisteína de un motivo sulfatasa.

Generalmente, las FGE de interés incluyen aquellas que tienen al menos 60%, usualmente 75%, usualmente 80%, más usualmente 90% -95% de identidad de nucleótidos o de residuos de aminoácidos, cuando se comparan y alinean para la máxima correspondencia con una secuencia de nucleótidos o una secuencia de aminoácidos de una FGE progenitora, como se mide con el uso de un algoritmo de comparación de secuencias disponibles en la técnica o por inspección visual. Usualmente existe una identidad de secuencia indicada sobre una región de las secuencias que es al menos aproximadamente 50 residuos de longitud, más usualmente sobre una región de al menos aproximadamente 100 residuos, y más usualmente sobre al menos aproximadamente 150 residuos de la longitud completa de la región codificante o de la proteína, con la condición de que la región de comparación incluya un sitio activo de la FGE requerido para la actividad enzimática.

Para la comparación de secuencias, típicamente una secuencia actúa como secuencia de referencia, con la que se

comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, si es necesario se designan coordenadas posteriores, y se designan los parámetros del programa del algoritmo de secuencia. Después, el algoritmo de comparación de secuencias calcula el porcentaje de identidad de secuencia para la(s) secuencia(s) de prueba(s) con relación a la secuencia de referencia, basado en los parámetros de programa designados.

El alineamiento óptimo de las secuencias para la comparación puede realizarse, por ejemplo, por el algoritmo de homología local de Smith & Waterman, *Adv. PCT. Math.* 2:482 (1981), por el algoritmo de alineamiento por homología de Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), por la búsqueda para el método de similitud de Pearson & Lipman, *Proc. Nat'l. Acad. Sci. Estados Unidos* 85:2444 (1988), mediante implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o inspección visual (ver generalmente, *Current Protocols in Molecular Biology*, Ausubel FM y otros, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., (Suplemento 1995) (Ausubel)).

Ejemplos de algoritmos que son apropiados para determinar el porcentaje de identidad de secuencia y la similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul y otros (1990) *J. Mol. Biol.* 215: 403-410 y Altschuel y otros (1977) *Nucleic Acids Res.* 25: 3389-3402, respectivamente. El software para ejecutar los análisis BLAST está disponible públicamente a través del Centro Nacional para la Información Biotecnológica (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo incluye primero identificar los pares de secuencia con puntuación alta (HSP) mediante la identificación de palabras cortas de longitud W en la secuencia de búsqueda, que coincide o satisface algunas puntuaciones T umbrales de valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de la base de datos. El parámetro T se refiere como el umbral de puntuación de la palabra vecina (Altschul y otros, supra).

Estos aciertos de palabras vecinas iniciales actúan como semillas para iniciar las búsquedas para encontrar HSP más largos que los contengan. Las coincidencias de palabras se extienden entonces en ambas direcciones a lo largo de cada secuencia, hasta alcanzar un punto en el cual la puntuación de alineamiento acumulado no aumente más. Los registros acumulativos se calculan por medio del uso de, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos coincidentes; siempre > 0) y N (puntuación de penalización para residuos no coincidentes; siempre < 0). Para las secuencias de aminoácidos, se usa una matriz de puntuación para calcular el registro acumulativo. Las extensiones de los aciertos de las palabras vecinas en cada dirección se detienen cuando: la puntuación de alineamiento acumulativa disminuye en una cantidad X desde su máximo valor alcanzado; la puntuación acumulativa va a cero o por debajo de cero debido a la acumulación de uno o más alineamientos de residuos con puntuación negativa; o se alcanza el extremo de cualquiera de las secuencias. Los parámetros W, T y X del algoritmo BLAST determinan la sensibilidad y la velocidad del alineamiento. El programa BLASTN (para las secuencias nucleotídicas) usa como defecto una longitud de palabra (W) de 11, una expectativa (E) de 10, M=5, N=-4, y una comparación de ambas cadenas. Para secuencias de aminoácidos, el programa BLASTP usa por defecto una longitud de palabra (W) de 3, una expectativa (E) de 10, y la matriz de puntuación BLOSUM62 (ver Henikoff & Henikoff, *Proc. Natl. Acad. Sci. Estados Unidos* 89:10915 (1989)).

Además de calcular el porcentaje de identidad de la secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (ver, por ejemplo, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. Estados Unidos* 90:5873-5787 (1993)). Una medida de la similitud proporcionada por el algoritmo BLAST es la probabilidad de la suma más pequeña (P(N)), que proporciona una indicación de la probabilidad con que dos secuencias de nucleótidos o de aminoácidos coincidan al azar. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de la suma más pequeña en una comparación entre el ácido nucleico del ensayo y uno de referencia es menor de aproximadamente 0.1, más usualmente menor de aproximadamente 0.01 y lo más usual menor de aproximadamente 0.001.

Las posiciones de los residuos que no son idénticos pueden diferir en sustituciones de aminoácidos conservadoras, que serán fácilmente evidentes a partir del análisis de los alineamientos como se discutió anteriormente. Sustituciones de aminoácidos conservadoras se refieren a la intercambiabilidad de residuos que tienen cadenas laterales similares. Por ejemplo, los grupos de aminoácidos que definen los residuos que pueden intercambiarse por otro residuo dentro del grupo y constituyen una sustitución de aminoácidos conservadora incluyen un grupo de aminoácidos que tienen cadenas laterales alifáticas es glicina, alanina, valina, leucina, prolina, e isoleucina ("aminoácido alifático"); un grupo de aminoácidos que tienen cadenas laterales alifáticas hidroxiladas es serina, y treonina ("aminoácido alifático, hidroxilado", que además están abarcados dentro de "aminoácido polar, sin carga"); un grupo de aminoácidos que tienen cadenas laterales que contienen amida es asparagina y glutamina ("aminoácido que contiene amida", que además están abarcados dentro de "aminoácido polar, sin carga"); un grupo de aminoácidos que tienen cadenas laterales aromáticas es fenilalanina, tirosina, y triptófano ("aminoácido aromático"); un grupo de aminoácidos que tienen cadenas laterales básicas (a pH fisiológico) es lisina, arginina, e histidina ("aminoácido básico"); un grupo de aminoácidos que tienen cadenas laterales que contienen azufre es cisteína y metionina ("aminoácido que contiene azufre"); un grupo de aminoácidos que son polares y sin carga (a pH fisiológico) incluyen serina, treonina, asparagina, y glutamina ("aminoácido polar, sin carga"); y un grupo de aminoácidos que tienen cadenas laterales cargadas (a pH fisiológico) es

ácido aspártico, ácido glutámico, histidina, lisina, y arginina ("aminoácido cargado"). Los grupos de sustitución conservadora de aminoácidos se ilustran por: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina, y asparagina-glutamina.

5 Cuando se usa un método libre de células para convertir un polipéptido que contiene motivos sulfatasa, puede usarse una FGE aislada. Cualquiera de los procedimientos convenientes de purificación de proteínas puede usarse para aislar una FGE, ver, por ejemplo, Guide to Protein Purification, (Deuthser ed). (Academic Press, 1990). Por ejemplo, puede prepararse un lisado a partir de una célula que produce una FGE preferida, y se purifica con el uso de HPLC, cromatografía de exclusión, electroforesis en gel, cromatografía de afinidad, y similares.

10 Vectores de expresión y células huésped para la producción de polipéptidos etiquetados con aldehído

15 La descripción proporciona un ácido nucleico que codifica las etiquetas de aldehído y los polipéptidos etiquetados con aldehído, así como también las construcciones y las células huésped que contienen el ácido nucleico. Tales ácidos nucleicos comprenden una secuencia de ADN que tiene un marco de lectura abierto que codifica una etiqueta de aldehído o un polipéptido etiquetado con aldehído y, en la mayoría de las modalidades, es capaz, bajo condiciones adecuadas, de expresarse. "Ácido nucleico" abarca ADN, ADNc, ARNm, y vectores que comprenden tales ácidos nucleicos.

20 Los ácidos nucleicos que codifican etiquetas de aldehído, así como también polipéptidos etiquetados con aldehído, se proporcionan en la presente. Tales ácidos nucleicos incluyen ADN genómicos modificados mediante la inserción de una secuencia que codifica una etiqueta de aldehído y ADNc de polipéptidos etiquetados con aldehído. El término "ADNc" como se usa en la presente pretende incluir todos los ácidos nucleicos que comparten la disposición de elementos de secuencia que se encuentran en una especie nativa de ARNm maduro (que incluye variantes de empalme), donde los  
25 elementos de secuencia son exones y regiones 3' y 5' no codificantes. Normalmente las especies de ARNm tienen exones contiguos, con los intrones intermedios, cuando están presentes, que se eliminan por corte y empalme del ARN nuclear, para crear un marco de lectura abierto continuo que codifica una proteína de acuerdo con la presente invención.

30 El término "gen" pretende un ácido nucleico que tiene un marco de lectura abierto que codifica un polipéptido (por ejemplo, un polipéptido etiquetado con aldehído), y, opcionalmente, cualquier intron, y puede incluir además secuencias de nucleótidos no codificantes adyacentes a 5' y 3' incluidas en la regulación de la expresión (por ejemplo, reguladores de la transcripción y/o traducción, por ejemplo, promotores, potenciadores, señales reguladoras de la traducción, y similares), hasta aproximadamente 20 kb más allá de la región codificante, pero posiblemente más allá en cualquier  
35 dirección, cuyas secuencias de nucleótidos no codificantes adyacentes a 5' y 3' pueden ser endógenas o heterólogas a la secuencia codificante. Pueden incluirse secuencias reguladoras de la transcripción y de la traducción, tales como, promotores, potenciadores, etc. que incluyen aproximadamente 1 kb, pero posiblemente más, de ADN genómico flanqueante ya sea en el extremo 5' o 3' de la región transcrita.

40 Los ácidos nucleicos contemplados en la presente pueden proporcionarse como parte de un vector (referido además como una construcción), una amplia variedad de los cuales se conoce en la técnica y no necesita elaborarse en la presente. Los vectores ilustrativos incluyen, pero sin limitarse a, plásmidos; cósmidos; vectores virales (por ejemplo, vectores retrovirales); vectores no virales; cromosomas artificiales (YAC, BAC, etc.); minicromosomas; y similares.

45 La elección del vector dependerá de una variedad de factores tales como el tipo de célula en la que se prefiere la propagación y el propósito de la propagación. Ciertos vectores son útiles para amplificar y producir grandes cantidades de la secuencia de ADN preferida. Otros vectores son apropiados para la expresión en células en cultivo. Aún otros vectores son apropiados para la transferencia y la expresión en células en un animal entero. La elección del vector adecuado está bien dentro de la experiencia de la técnica. Muchos de tales vectores están disponibles comercialmente.

50 Para preparar las construcciones, se inserta un polinucleótido en un vector, típicamente por medio de unión con ADN ligasa a un sitio de enzima de restricción clivado en el vector. Alternativamente, la secuencia de nucleótidos preferida puede insertarse mediante recombinación homóloga o recombinación específica del sitio. La recombinación homóloga se logra típicamente mediante la unión de las regiones de homología al vector en los flancos de la secuencia de  
55 nucleótidos preferida, mientras que la recombinación específica del sitio puede lograrse a través del uso de secuencias que facilitan la recombinación específica del sitio (por ejemplo, cre-lox, sitios att, etc.). El ácido nucleico que contiene tales secuencias puede añadirse mediante, por ejemplo, la ligación de oligonucleótidos, o la reacción en cadena de la polimerasa con el uso de iniciadores que comprenden la región de homología y una porción de la secuencia de nucleótidos preferida.

60 Los vectores pueden proporcionar el mantenimiento extracromosómico en una célula huésped o pueden proporcionar la integración en el genoma de la célula huésped. Los vectores se describen ampliamente en numerosas publicaciones bien conocidas por los expertos en la técnica, que incluyen, por ejemplo, Short Protocols in Molecular Biology, (1999) F. Ausubel, y otros, eds., Wiley & Sons. Los vectores pueden proporcionar la expresión de los ácidos nucleicos que

codifican un polipéptido de interés (por ejemplo, un polipéptido etiquetado con aldehído, una FGE, etc.), pueden proporcionar la propagación de los ácidos nucleicos en cuestión, o ambas.

5 Los vectores ilustrativos que pueden usarse incluyen pero sin limitarse a los derivados de ADN recombinante de bacteriófago, ADN de plásmido o ADN de cósmido. Por ejemplo, pueden usarse los vectores plasmídicos tales como pBR322, pUC 19/18, pUC 118, 119 y la serie de vectores M13 mp. Los vectores bacteriófagos pueden incluir *lgt10*, *lgt11*, *lgt18-23*, *λZAP/R* y la serie EMBL de vectores bacteriófagos. Los vectores cósmidos que pueden utilizarse incluyen, pero sin limitarse a, pJB8, PCV 103, 107 PCV, PCV 108, pTM, PMCs, PNNL, pHSG274, COS202, COS203, pWE15, pWE16 y la serie de vectores charomid 9. Alternativamente, pueden diseñarse vectores virales recombinantes, 10 que incluyen pero sin limitarse a los derivados de virus tales como el virus del herpes, retrovirus, virus de vaccinia, poxvirus, adenovirus, virus adenoasociados o virus del papiloma bovino.

15 Para la expresión de un polipéptido de interés, puede emplearse un casete de expresión. Así, la presente invención proporciona un vector de expresión recombinante que comprende un ácido nucleico en cuestión. El vector de expresión proporciona unas secuencias reguladoras de la transcripción y de la traducción, y puede proporcionar la expresión inducible o constitutiva, donde la región codificante está operativamente enlazada bajo el control transcripcional de la región de iniciación transcripcional, y una región de terminación transcripcional y traduccional. Estas regiones de control pueden ser nativas para el gen que codifica el polipéptido (por ejemplo, el polipéptido objetivo o la FGE), o pueden derivarse de fuentes exógenas. Generalmente, las secuencias reguladoras de la transcripción y de la traducción pueden 20 incluir, pero sin limitarse a, secuencias promotoras, sitios de unión ribosomal, secuencias de inicio y de terminación de la transcripción, secuencias de inicio y de terminación de la traducción, y secuencias potenciadoras o activadoras. Adicionalmente a los promotores constitutivos e inducibles, los promotores fuertes (por ejemplo, T7, CMV, y similares) se usan en las construcciones descritas en la presente invención, particularmente donde se prefieren niveles superiores de expresión en un sistema *in vivo* (basado en células) o *in vitro*. Los ejemplos adicionales de promotores incluyen 25 promotores de virus de tumor mamario de ratón (MMTV), promotores del virus del sarcoma de Rous (RSV), promotores de adenovirus, el promotor del gen temprano inmediato de CMV humano (Boshart y otros, Cell 41:521-530, 1985), y el promotor de la repetición terminal larga (LTR) del RSV (Gorman y otros, Proc.Natl. Acad. Sci. Estados Unidos 79:6777-6781, 1982). El promotor puede proporcionarse además, por ejemplo, por una 5'UTR de un retrovirus.

30 Los vectores de expresión generalmente tienen sitios de restricción convenientes localizados cerca de la secuencia promotora para proporcionar la inserción de secuencias de ácidos nucleicos que codifican proteínas de interés. Un marcador seleccionable operativo en el huésped de expresión puede estar presente para facilitar la selección de las células que contienen el vector. Adicionalmente, la construcción de expresión puede incluir elementos adicionales. Por ejemplo, el vector de expresión puede tener uno o dos sistemas de replicación, lo que permite así que se mantenga en 35 los organismos, por ejemplo en células de mamíferos o de insectos para la expresión y en un huésped procarionta para la clonación y la amplificación. Adicionalmente, la construcción de expresión puede contener un gen marcador seleccionable para permitir la selección de las células huésped transformadas. Los genes de selección se conocen bien en la técnica y variarán con la célula huésped usada.

40 Un casete de etiqueta de aldehído se proporciona además en la presente, el que incluye un ácido nucleico que codifica una etiqueta de aldehído, y sitios de restricción apropiados que flanquean la secuencia codificadora de la etiqueta para la inserción en marco de un ácido nucleico que codifica un polipéptido objetivo. Tal construcción de expresión puede proporcionar la adición de una etiqueta de aldehído en el N-terminal o C-terminal de un polipéptido objetivo. El casete de etiqueta de aldehído puede enlazarse operativamente a una secuencia promotora para proporcionar la expresión del 45 polipéptido resultante etiquetado con aldehído, y puede incluir además uno o más marcadores seleccionables.

La presente descripción proporciona además casetes de expresión para la producción de polipéptidos etiquetados con aldehído (por ejemplo, que tienen una etiqueta de aldehído posicionada en un N-terminal, en un C-terminal). Tales casetes de expresión incluyen generalmente un primer ácido nucleico que comprende una secuencia que codifica la 50 etiqueta de aldehído, y al menos un sitio de restricción para la inserción de un segundo ácido nucleico que codifica un polipéptido de interés. Los sitios de restricción pueden estar posicionados en el 5' y/o 3' de la secuencia que codifica la etiqueta de aldehído. La inserción de la secuencia que codifica el polipéptido en marco con la secuencia que codifica la etiqueta de aldehído proporciona la producción de un ácido nucleico recombinante que codifica una proteína de fusión que es un polipéptido etiquetado con aldehído como se describe en la presente. Las construcciones que contienen un 55 casete de expresión tal generalmente incluyen además un promotor operativamente enlazado al casete de expresión para proporcionar la expresión del polipéptido etiquetado con aldehído producido. Otros componentes de la construcción de expresión pueden incluir marcadores seleccionables y otros elementos apropiados.

60 Las construcciones de expresión que codifican polipéptidos etiquetados con aldehído pueden generarse además con el uso de métodos de amplificación (por ejemplo, reacción en cadena de la polimerasa (PCR)), donde al menos un iniciador de amplificación (es decir, al menos uno de un iniciador directo o reverso) incluye una secuencia de ácido nucleico que codifica una etiqueta de aldehído. Por ejemplo, un iniciador de amplificación que tiene una secuencia que codifica una etiqueta de aldehído está diseñado para proporcionar la amplificación de un ácido nucleico que codifica un polipéptido objetivo de interés. El producto de la extensión que resulta de la síntesis mediada por la polimerasa a partir 65 del iniciador directo que contiene la etiqueta de aldehído produce un producto de amplificación de ácido nucleico que

codifica una proteína de fusión compuesta de un polipéptido objetivo etiquetado con aldehído. El producto de amplificación se inserta después en una construcción de expresión de elección para proporcionar una construcción de expresión del polipéptido etiquetado con aldehído.

## 5 Células huésped

Cualquiera de un número de células huésped apropiadas puede usarse en la producción de un polipéptido etiquetado con aldehído. La célula huésped usada para la producción de un polipéptido etiquetado con aldehído puede proporcionar opcionalmente la conversión mediada por la FGE, de manera que el polipéptido producido contiene una etiqueta de aldehído que contiene FGly después de la expresión y la modificación postraduccional por la FGE. Alternativamente la célula huésped puede proporcionar la producción de un polipéptido no convertido etiquetado con aldehído (por ejemplo, debido a la carencia de expresión de una FGE que facilita la conversión de la etiqueta de aldehído).

Generalmente, los polipéptidos descritos en la presente invención pueden expresarse en procariotas o eucariotas de acuerdo con las vías convencionales, que dependen del propósito de la expresión. Así, la presente invención proporciona además una célula huésped, por ejemplo, una célula huésped modificada genéticamente, que comprende un ácido nucleico que codifica un polipéptido etiquetado con aldehído. La célula huésped además puede comprender opcionalmente una FGE recombinante, que puede ser endógena o heteróloga a la célula huésped.

Las células huésped para la producción (que incluye la producción a gran escala) de un polipéptido no convertido o convertido (cuando la célula huésped expresa una FGE apropiada) etiquetado con aldehído, o para la producción de una FGE (por ejemplo, para el uso en un método libre de célula) pueden seleccionarse de cualquiera de una variedad de células huésped disponibles. Células huésped ilustrativas incluyen las de un organismo unicelular procariota o eucariota, tales como bacterias (por ejemplo, cepas de *Escherichia coli*, *Bacillus* spp. (por ejemplo, *B. subtilis*), y similares) levadura u hongos (por ejemplo, *S. cerevisiae*, *Pichia* spp., y similares), y otros tales células huésped se pueden usar. Las células huésped ilustrativas derivadas originalmente de un organismo superior tal como insectos, vertebrados, particularmente mamíferos, (por ejemplo, CHO, HEK, y similares), pueden usarse como células huésped de expresión.

Los sistemas de expresión específicos de interés incluyen bacterias, levaduras, células de insecto y sistemas de expresión derivados de células de mamíferos. Los sistemas representativos de cada una de estas categorías se proporcionan más abajo.

Bacteria. Los sistemas de expresión en las bacterias incluyen aquellos descritos en Chang y otros, Nature (1978) 275:615; Goeddel y otros, Nature (1979) 281:544; Goeddel y otros, Nucleic Acids Res. (1980) 8:4057; EP 0 036,776; la patente de Estados Unidos núm. 4,551,433; DeBoer y otros, Proc. Natl. Acad. Sci. (Estados Unidos) (1983) 80:21-25; y Siebenlist y otros, Cell (1980) 20:269.

Levadura. Los sistemas de expresión en las levaduras incluyen aquellos descritos en Hinnen y otros, Proc. Natl. Acad. Sci. (Estados Unidos) (1978) 75:1929; Ito y otros, J. Bacteriol. (1983) 153:163; Kurtz y otros, Mol. Cell. Biol. (1986) 6:142; Kunze y otros, J. Basic Microbiol. (1985) 25:141; Gleeson y otros, J. Gen. Microbiol. (1986) 132:3459; Roggenkamp y otros, Mol. Gen. Genet. (1986) 202:302; Das y otros, J. Bacteriol. (1984) 158:1165; De Louvencourt y otros, J. Bacteriol. (1983) 154:737; Van den Berg y otros, Bio/Technology (1990) 8:135; Kunze y otros, J. Basic Microbiol. (1985) 25:141; Cregg y otros, Mol. Cell. Biol. (1985) 5:3376; las patentes de Estados Unidos núms. 4,837,148 y 4,929,555; Beach y Nurse, Nature (1981) 300:706; Davidow y otros, Curr. Genet. (1985) 10:380; Gaillardin y otros, Curr. Genet. (1985) 10:49; Ballance y otros, Biochem. Biophys. Res. Commun. (1983) 112:284-289; Tilburn y otros, Gene (1983) 26:205-221; Yelton y otros, Proc. Natl. Acad. Sci. (Estados Unidos) (1984) 81:1470-1474; Kelly y Hynes, EMBO J. (1985) 4:475479; EP 0 244,234; y WO 91/00357.

Células de insecto. La expresión de genes heterólogos en los insectos se logra como se describe en la patente de los Estados Unidos núm. 4,745,051; Friesen y otros, "The Regulation of Baculovirus Gene Expression", en: The Molecular Biology Of Baculoviruses (1986) (W. Doerfler, ed.); EP 0 127,839; EP 0 155,476; y Vlak y otros, J. Gen. Virol. (1988) 69:765-776; Miller y otros, Ann. Rev. Microbiol. (1988) 42:177; Carbonell y otros, Gene (1988) 73:409; Maeda y otros, Nature (1985) 315:592-594; Lebacqz-Verheyden y otros, Mol. Cell. Biol. (1988) 8:3129; Smith y otros, Proc. Natl. Acad. Sci. (Estados Unidos) (1985) 82:8844; Miyajima y otros, Gene (1987) 58:273; y Martin y otros, DNA (1988) 7:99. Numerosas cepas y variantes de baculovirus y las células huésped de insecto permisivas correspondientes de los huéspedes se describen en Luckow y otros, Bio/Technology (1988) 6:47-55, Miller y otros, Generic Engineering (1986) 8:277-279, y Maeda y otros, Nature (1985) 315:592-594.

Células de mamíferos. La expresión en mamíferos se logra como se describe en Dijkema y otros; EMBO J. (1985) 4:761, Gorman y otros, Proc. Natl. Acad. Sci. (Estados Unidos) (1982) 79:6777, Boshart y otros, Cell (1985) 41:521 y la patente de Estados Unidos núm. 4,399,216. Otras características de la expresión en mamífero se facilitan como se describe en Ham y Wallace, Meth. Enz. (1979) 58:44, Bames y Sato, Anal. Biochem. (1980) 102:255, las patentes de Estados Unidos núms.. 4,767,704, 4,657,866,4,927,762,4,560,655, WO 90/103430, WO 87/00195, y U.S. RE 30,985.

5 Cuando cualquiera de las células huésped anteriores, u otras células huésped u organismos adecuados, se usan para replicar y/o expresar los polinucleótidos o ácidos nucleicos de la invención, el ácido nucleico replicado resultante, el ARN, la proteína o el polipéptido expresado, están dentro del alcance de la invención como un producto de la célula huésped u organismo.

10 El producto puede recuperarse por cualquier medio adecuado conocido en la técnica. Adicionalmente, puede emplearse cualquiera de los procedimientos convenientes de purificación de proteínas, donde las metodologías apropiadas de purificación de proteínas se describen en Guide to Protein Purification, (Deuthser ed). (Academic Press, 1990). Por ejemplo, un lisado puede prepararse a partir de una célula que comprende el vector de expresión que expresa un polipéptido de interés, y purificarse con el uso de HPLC, cromatografía de exclusión, electroforesis en gel, cromatografía de afinidad, y similares.

#### 15 Porciones para la modificación de polipéptidos

Los polipéptidos que contienen FGly, etiquetados con aldehído pueden someterse a modificación para proporcionar la unión de una amplia variedad de porciones. Las moléculas ilustrativas de interés incluyen, pero no se limitan necesariamente a, un marcador detectable, una molécula pequeña, un péptido, y similares.

20 La porción de interés se proporciona como componente de una pareja reactiva para la reacción con un aldehído del residuo de FGly de una etiqueta de aldehído convertida del polipéptido etiquetado. Ya que los métodos de modificación del polipéptido etiquetado son compatibles con los procesos químicos convencionales, los métodos de la invención pueden explotar una amplia variedad de reactivos disponibles comercialmente para lograr la unión de una porción de interés a un residuo de FGly de un polipéptido etiquetado con aldehído. Por ejemplo, los derivados aminooxi, hidrazida, hidrazina, o tiosemicarbazida de un número de porciones de interés son parejas reactivas apropiadas, y están fácilmente disponibles o pueden generarse con el uso de métodos químicos estándares.

30 Por ejemplo, un aminooxi-PEG puede generarse a partir de los monoamino-PEG y la aminooxiglicina con el uso de protocolos estándares. El aminooxi-PEG puede reaccionar después con un polipéptido convertido etiquetado con aldehído para proporcionar la unión de la porción de PEG. El suministro de una porción de biotina a un polipéptido convertido etiquetado con aldehído puede lograrse con el uso de aminooxibiotina, biotina hidrazida o 2,4 dinitrofenilhidrazina.

35 Según se proporciona en la presente descripción, el técnico con experiencia ordinaria puede adaptar fácilmente cualquiera de una variedad de porciones para proporcionar una pareja reactiva para la conjugación a un polipéptido etiquetado con aldehído como se contempla en la presente. El el técnico con experiencia ordinaria apreciará que los factores tales como el pH y el impedimento estérico (es decir, la accesibilidad de la etiqueta de aldehído a la reacción con una pareja reactiva de interés) son de importancia. La modificación de las condiciones de reacción para proporcionar las condiciones de conjugación óptimas se conoce bien dentro de la experiencia general del técnico, y es de rutina en la técnica. Generalmente, se prefiere realizar las reacciones de conjugación normalmente a un pH por debajo de 7, con un pH de aproximadamente 5.5, aproximadamente 6, aproximadamente 6.5, usualmente aproximadamente 5.5 es el óptimo. Cuando se realiza la conjugación con un polipéptido etiquetado con aldehído presente en o sobre una célula viva, se seleccionan las condiciones para que sean fisiológicamente compatibles. Por ejemplo, el pH puede disminuirse temporalmente por un tiempo suficiente para permitir que se produzca la reacción pero dentro de un período tolerado por la célula que tiene una etiqueta de aldehído (por ejemplo, de aproximadamente 45 30 min a 1 hora). Las condiciones fisiológicas para la realización de la modificación de polipéptidos etiquetados con aldehído en una superficie celular pueden ser similares a aquellas usadas en una reacción cetona azida en la modificación de células que portan azidas en la superficie celular (ver, por ejemplo, Estados Unidos 6,570,040).

50 Generalmente, la porción o porciones pueden proporcionar una o más de una amplia variedad de funciones o elementos. Las porciones ilustrativas incluyen marcadores detectables (por ejemplo, marcadores colorantes (por ejemplo, cromóforos, fluoróforos), sondas biofísicas (marcadores espín, sondas de NMR), marcadores de tipo FRET (por ejemplo, al menos un miembro de un par FRET, que incluye al menos un miembro de un par fluoróforo/apagador), marcadores de tipo BRET (por ejemplo, al menos un miembro de un par BRET), etiquetas inmunodetectables (por ejemplo, FLAG, His(6), y similares), etiquetas de localización (por ejemplo, para identificar la asociación de un polipéptido etiquetado a nivel de tejido o celular molecular (por ejemplo, asociación con un tipo de tejido, o membrana celular particular)), y similares); las porciones dinámicas activadas por la luz (por ejemplo, cierre de poro mediado por azobenceno, cambios estructurales mediados por azobenceno, motivos de reconocimiento en fotodescomposición); polímeros solubles en agua (por ejemplo, pegilación); etiquetas de purificación (por ejemplo, para facilitar el aislamiento por cromatografía de afinidad (*por ejemplo*, la unión de un epítipo FLAG)); dominios de localización de membrana (por ejemplo, lípidos o anclajes de tipo GPI); etiquetas de inmovilización (por ejemplo, para facilitar la unión del polipéptido a una superficie, que incluye la unión selectiva); fármacos (por ejemplo, para facilitar el direccionamiento del fármaco, por ejemplo, a través de la unión del fármaco a un anticuerpo); porciones de suministro dirigidas, (por ejemplo, ligandos para la unión a un receptor objetivo (*por ejemplo*, para facilitar la unión viral, la unión de una proteína dirigida presente en un liposoma, etc.)), y similares.

Los ejemplos específicos, no limitantes, se proporcionan más abajo.

#### Marcadores detectables.

5 Las composiciones y métodos de la invención pueden usarse para suministrar un marcador detectable a un polipéptido etiquetado con aldehído. Los marcadores detectables ilustrativos incluyen, pero no se limitan necesariamente a, moléculas fluorescentes (*por ejemplo*, moléculas autofluorescentes, moléculas que fluorescen tras ponerse en contacto con un reactivo, etc.), marcadores radiactivos (*por ejemplo*, <sup>111</sup>In, <sup>125</sup>I, <sup>131</sup>I, <sup>212</sup>B, <sup>90</sup>Y, <sup>186</sup>Rh, y similares); biotina (*por ejemplo*, para detectarse a través de la reacción de biotina y avidina); etiquetas fluorescentes; reactivos de imagen, y similares. Los marcadores detectables incluyen además péptidos o polipéptidos que pueden detectarse mediante la unión de anticuerpos, *por ejemplo*, mediante la unión de un anticuerpo marcado de forma detectable o mediante la detección de un anticuerpo unido a través de un ensayo de tipo sándwich.

#### Unión de moléculas objetivo a un soporte.

15 Los métodos pueden proporcionar la conjugación de un polipéptido etiquetado con aldehído a una porción para facilitar la unión del polipéptido a un sustrato sólido (*por ejemplo*, para facilitar los ensayos), o a una porción para facilitar la fácil separación (por ejemplo, un hapteno reconocido por un anticuerpo unido a una perla magnética). En una modalidad, los métodos de la invención se usan para proporcionar la unión de una proteína a una matriz (por ejemplo, chip) en una orientación definida. Por ejemplo, puede generarse un polipéptido que tiene una etiqueta de aldehído en un sitio seleccionado (*por ejemplo*, en o cerca del N-terminal), y los métodos y composiciones de la invención usarse para suministrar una porción a las etiquetas de aldehído convertidas. La porción puede usarse después como el sitio de unión para fijar el polipéptido a un soporte (*por ejemplo*, soporte sólido o semisólido, particularmente un soporte apropiado para usar como un microchip en ensayos de gran productividad).

#### Unión de moléculas para suministrar a un sitio objetivo.

20 La pareja reactiva para el polipéptido etiquetado con aldehído puede comprender un fármaco de molécula pequeña, toxina, u otra molécula para el suministro a la célula y el que puede proporcionar una actividad farmacológica o puede servir como un objetivo para el suministro de otras moléculas.

25 Además, se contempla el uso de una pareja reactiva que comprende una de un par de parejas de unión (*por ejemplo*, un ligando, una porción de unión a ligando de un receptor, una porción de unión al receptor de un ligando, etc.). Por ejemplo, la pareja reactiva puede comprender un polipéptido que sirve como un receptor viral y, tras la unión con una proteína de la envoltura viral o proteína de la cápsida viral, facilita la unión del virus a la superficie celular sobre la que se expresa la proteína modificada etiquetada con aldehído. Alternativamente, la pareja reactiva comprende un antígeno que se une específicamente por un anticuerpo (*por ejemplo*, anticuerpo monoclonal), para facilitar la detección y/o separación de las células huésped que expresan el polipéptido modificado etiquetado con aldehído.

#### Polímeros solubles en agua

30 Una porción de particular interés es un polímero soluble en agua. Un "polímero soluble en agua" se refiere a un polímero que es soluble en agua y usualmente es prácticamente no inmunogénico, y usualmente tiene un peso molecular atómico mayor que aproximadamente 1,000 Daltons. Los métodos y composiciones descritos en la presente invención pueden usarse para unir uno o más polímeros solubles en agua a un polipéptido etiquetado con aldehído. La unión de un polímero soluble en agua (por ejemplo, PEG) de un polipéptido, particularmente un polipéptido farmacéuticamente activo (terapéutico) puede ser preferible ya que tal modificación puede aumentar el índice terapéutico mediante el aumento de la vida media en suero como resultado de la estabilidad proteolítica aumentada y/o el aclaramiento renal disminuido. Además, la unión de uno o más polímeros (por ejemplo, pegilación) puede reducir la inmunogenicidad de los productos farmacéuticos proteicos.

35 En algunas modalidades, el polímero soluble en agua tiene un peso molecular hidrodinámico eficaz mayor que aproximadamente 10,000 Da, mayor que aproximadamente 20,000 a 500,000 Da, mayor que aproximadamente 40,000 Da a 300,000 Da, mayor que aproximadamente 50,000 Da a 70,000 Da, usualmente mayor que aproximadamente 60,000 Da. Por "peso molecular hidrodinámico eficaz" se entiende el tamaño de agua solvatada eficaz de una cadena polimérica que se determina mediante cromatografía de exclusión de tamaño en base acuosa (SEC). Cuando el polímero soluble en agua contiene cadenas de polímero que tienen unidades repetidas de óxido de polialquileno, tales como unidades repetidas de óxido de etileno, cada cadena puede tener un peso molecular atómico de entre aproximadamente 200 Da y aproximadamente 80,000 Da, o entre aproximadamente 1,500 Da y aproximadamente 42,000 Da, siendo de particular interés 2,000 a aproximadamente 20,000 Da. A menos que se refiera específicamente, peso molecular pretende referirse al peso molecular atómico. Los polímeros solubles en agua lineales, ramificados, y cargados en el terminal (por ejemplo, PEG) son de particular interés.

60 Los polímeros útiles como porciones a unirse a un polipéptido etiquetado con aldehído pueden tener una amplia



variedad de pesos moleculares, y subunidades de polímero. Estas subunidades pueden incluir un polímero biológico, un polímero sintético, o una combinación de estos. Los ejemplos de tales polímeros solubles en agua incluyen: dextrano y derivados de dextrano, que incluyen sulfato de dextrano, dextrina entrecruzada con P-amino, y carboximetildextrina, celulosa y derivados de celulosa, que incluyen metilcelulosa y carboximetilcelulosa, almidón y dextrinas, y derivados e hidroxilados de almidón, glicopolialquileno y derivados de este, que incluyen polietilenglicol, metoxipolietilenglicol, homopolímeros de polietilenglicol, homopolímeros de polipropilenglicol, copolímeros de etilenglicol con propilenglicol, en donde dichos homopolímeros y copolímeros están no sustituidos o sustituidos en un extremo con un grupo alquilo, heparina y fragmentos de heparina, alcohol polivinílico y éteres de poliviniletilo, polivinilpirrolidona, aspartamida, y polioles polioxietilados, con el dextrano y derivados de dextrano, dextrina y derivados de dextrina. Se apreciará que se contemplan además diversos derivados de los polímeros solubles en agua indicados específicamente.

Los polímeros solubles en agua tales como aquellos descritos anteriormente se conocen bien, particularmente los polímeros a base de óxido de polialquileno tales como polietilenglicol "PEG" (ver, por ejemplo, "Poly(ethylene glycol) Chemistry: Biotechnical and Biomedical Applications", J. M. Harris, Ed., Plenum Press, Nueva York, N.Y.(1992); y "Poly(ethylene glycol) Chemistry and Biological Applications", J. M. Harris y S. Zalipsky, Eds., ACS (1997); y las solicitudes internacionales de patente: WO 90/13540, WO 92/00748, WO 92/16555, WO 94/04193, WO 94/14758, WO 94/17039, WO 94/18247, WO 94/28937, WO 95/11924, WO 96/00080, WO 96/23794, WO 98/07713, WO 98/41562, WO 98/48837, WO 99/30727, WO 99/32134, WO 99/33483, WO 99/53951, WO 01/26692, WO 95/13312, WO 96/21469, WO 97/03106, WO 99/45964, y las patentes de Estados Unidos núms. 4,179,337; 5,075,046; 5,089,261; 5,100,992; 5,134,192; 5,166,309; 5,171,264; 5,213,891; 5,219,564; 5,275,838; 5,281,698; 5,298,643; 5,312,808; 5,321,095; 5,324,844; 5,349,001; 5,352,756; 5,405,877; 5,455,027; 5,446,090; 5,470,829; 5,478,805; 5,567,422; 5,605,976; 5,612,460; 5,614,549; 5,618,528; 5,672,662; 5,637,749; 5,643,575; 5,650,388; 5,681,567; 5,686,110; 5,730,990; 5,739,208; 5,756,593; 5,808,096; 5,824,778; 5,824,784; 5,840,900; 5,874,500; 5,880,131; 5,900,461; 5,902,588; 5,919,442; 5,919,455; 5,932,462; 5,965,119; 5,965,566; 5,985,263; 5,990,237; 6,011,042; 6,013,283; 6,077,939; 6,113,906; 6,127,355; 6,177,087; 6,180,095; 6,194,580; 6,214,966).

Los polímeros ilustrativos de interés incluyen aquellos que contienen un óxido de polialquileno, óxido de alquileno poliamida, o derivados de ellos, que incluyen óxido de polialquileno y óxido de alquileno poliamida que comprende una unidad repetida de óxido de etileno de fórmula  $-(CH_2-CH_2-O)-$ . Los polímeros ilustrativos adicionales de interés incluyen una poliamida que tiene un peso molecular mayor que aproximadamente 1,000 Daltons de fórmula  $-[C(O)-XC(O)-NH-Y-NH]_n-$  o  $-[NH-Y-NH-C(O)-XC(O)]_n-$ , donde X y Y son radicales divalentes que pueden ser iguales o diferentes y pueden ser ramificadas lineales, y n es un número entero distinto de 2-100, usualmente de 2 a 50, y donde cualquiera o ambos de X y Y comprenden una unidad repetida soluble en agua prácticamente no antigénica, biocompatible que puede ser lineal o ramificada. Las unidades repetidas ilustrativas adicionales solubles en agua comprenden un óxido de etileno de fórmula  $-(CH_2-CH_2-O)-$  o  $-(CH_2-CH_2-O)-$ . El número de tales unidades repetidas solubles en agua puede variar significativamente, con el número usual de tales unidades siendo de 2 a 500, 2 a 400, 2 a 300, 2 a 200, 2 a 100, y lo más usual de 2 a 50. Una modalidad ilustrativa es una en la cual uno o ambos de X y Y se selecciona de:  $-((CH_2)_{n1}-(CH_2-CH_2-O)_{n2}-(CH_2)-$  o  $-((CH_2)_{n1}-(O-CH_2-CH_2)_{n2}-(CH_2)_{n1}-)$ , donde n1 es 1 a 6, 1 a 5, 1 a 4 y lo más usual 1 a 3, y donde n2 es 2 a 50, 2 a 25, 2 a 15, 2 a 10, 2 a 8, y lo más usual 2 a 5. Una modalidad ilustrativa adicional es una en la cual X es  $-(CH_2-CH_2)-$ , y donde Y es  $-(CH_2-(CH_2-CH_2-O)_3-CH_2-CH_2-CH_2)-$  o  $-(CH_2-CH_2-CH_2-(O-CH_2-CH_2)_3-CH_2)-$ .

El polímero puede incluir uno o más espaciadores o enlazadores. Los espaciadores o enlazadores ilustrativos incluyen porciones lineales o ramificadas que comprenden una o más unidades repetidas empleadas en un polímero soluble en agua, unidades diamino o diácida, aminoácidos naturales o no naturales o derivados de estos, así como también porciones alifáticas, que incluyen alquilo, arilo, heteroalquilo, heteroarilo, alcoxi, y similares, que pueden contener, por ejemplo, hasta 18 átomos de carbono o aun una cadena adicional de polímero.

La porción de polímero, o uno o más de los espaciadores o enlazadores de la porción de polímero cuando está presente, puede incluir cadenas de polímero o unidades que son bioestables o biodegradables. Por ejemplo, los polímeros con enlaces repetidos tienen diferentes grados de estabilidad bajo condiciones fisiológicas, que dependen de la labilidad del enlace. Los polímeros con tales enlaces pueden clasificarse por sus velocidades relativas de hidrólisis bajo condiciones fisiológicas en base a las velocidades de hidrólisis conocidas de análogos de bajo peso molecular, por ejemplo, de menos estable a más estable, por ejemplo, poliuretanos  $(-NH-C(O)-O-)$  > polioésteres  $(-O-C((OR)(R'))-O-)$  > poliamidas  $(-C(O)-NH-)$ . Igualmente, los sistemas de enlace que unen un polímero soluble en agua a una molécula objetivo pueden ser bioestables o biodegradables, por ejemplo, de menos estable a más estable: carbonato  $(-OC(O)-O-)$  > éster  $(-C(O)-O-)$  > uretano  $(-NH-C(O)-O-)$  > ortoéster  $(-OC((OR)(R'))-O-)$  > amida  $(-C(O)-NH-)$ . Generalmente, puede preferirse evitar el uso de un polisacárido sulfatado, que depende de la labilidad del grupo sulfato. Adicionalmente, puede preferirse menos usar policarbonatos y poliésteres. Estos enlaces se proporcionan a modo de ejemplo, y no pretenden limitar los tipos de enlaces que pueden emplearse en las cadenas de polímero o sistemas de enlace de los polímeros solubles en agua útiles en los polipéptidos modificados etiquetados con aldehído descritos, en la presente.

#### Métodos para la conversión y modificación de una etiqueta de aldehído

Conversión de una etiqueta de aldehído presente en un polipéptido etiquetado con aldehído logrado por métodos basados en células (in vivo) o libres de células (in vitro). Igualmente, la modificación de una etiqueta de aldehído

convertida de un polipéptido etiquetado con aldehído puede lograrse mediante métodos basados en células (in vivo) o métodos libres de células (in vitro). Estos se describen con más detalle más abajo.

#### Conversión y modificación en células huésped "in vivo"

5 La conversión de una etiqueta de aldehído de un polipéptido etiquetado con aldehído puede lograrse mediante la expresión del polipéptido etiquetado con aldehído en una célula que contiene una FGE apropiada. En esta modalidad, la conversión de la cisteína o serina de la etiqueta de aldehído se produce durante o después de la traducción en la célula huésped. En esta modalidad, la FGE de la célula huésped puede ser endógena para la célula huésped, o la célula huésped puede ser recombinante para una FGE apropiada que es heteróloga para la célula huésped. La expresión de la FGE puede proporcionarse por un sistema de expresión endógeno para el gen de la FGE (por ejemplo, la expresión se proporciona por un promotor y otros elementos de control presentes en el gen de la FGE nativa de la célula huésped), o puede proporcionarse por un sistema de expresión recombinante en el que la secuencia codificante de la FGE está operativamente enlazada a un promotor heterólogo para proporcionar la expresión constitutiva o inducible. El uso de un promotor fuerte para proporcionar niveles superiores de expresión de la FGE puede ser de particular interés en alguna modalidad.

20 En dependencia de la naturaleza del polipéptido objetivo que contiene la etiqueta de aldehído, después de la conversión el polipéptido convertido etiquetado con aldehído, se retiene intracelularmente en la célula huésped, se secreta, o se asocia con la membrana extracelular de la célula huésped. Cuando está presente la etiqueta de aldehído del polipéptido etiquetado con aldehído en la superficie celular, la modificación de la etiqueta de aldehído convertida puede lograrse mediante el uso de una pareja reactiva para unir una porción de la pareja reactiva a un residuo de FGly de una etiqueta de aldehído accesible en la superficie bajo condiciones fisiológicas. Las condiciones apropiadas para el uso para lograr la conjugación de una porción de una pareja reactiva a un polipéptido etiquetado con aldehído son similares a aquellas descritas en Mahal y otros (1997 16 de mayo) Science 276(5315):1125-8.

#### Conversión y modificación "in vitro" (libre de células)

30 La conversión in vitro (libre de células) de una etiqueta de aldehído de un polipéptido etiquetado con aldehído puede lograrse al poner en contacto un polipéptido etiquetado con aldehído con una FGE bajo condiciones apropiadas para la conversión de una cisteína o serina de un motivo sulfatasa de la etiqueta de aldehído a un FGly. Por ejemplo, el ácido nucleico que codifica un polipéptido etiquetado con aldehído puede expresarse in vitro en un sistema de transcripción/traducción en presencia de una FGE apropiada para proporcionar la producción de polipéptidos convertidos etiquetados con aldehído.

35 Alternativamente, un polipéptido no convertido aislado, etiquetado con aldehído puede aislarse después de la producción recombinante en una célula huésped que carece de una FGE apropiada o mediante producción sintética. El polipéptido aislado etiquetado con aldehído después se pone en contacto con una FGE apropiada bajo condiciones para proporcionar la conversión de etiquetas de aldehído. En esta modalidad, si la etiqueta de aldehído puede no ser fácilmente accesible al disolvente en el polipéptido aislado, el polipéptido etiquetado con aldehído puede desplegarse por métodos conocidos en la técnica (por ejemplo, con el uso de calor, ajuste del pH, agentes caotrópicos, (por ejemplo, urea, y similares), disolventes orgánicos (por ejemplo, hidrocarburos: octano, benceno, cloroformo), etc.) y la proteína desnaturalizada ponerse en contacto con una FGE apropiada. El polipéptido convertido etiquetado con aldehído puede replegarse después bajo condiciones apropiadas.

40 Con respecto a la modificación del aldehído etiquetado convertido, normalmente la modificación se lleva a cabo in vitro. El polipéptido convertido etiquetado con aldehído se aísla a partir de una fuente de producción (por ejemplo, producción por la célula huésped recombinante, producción sintética), y se pone en contacto con una pareja reactiva bajo condiciones apropiadas para proporcionar la conjugación de una porción de la pareja reactiva a la FGly de la etiqueta de aldehído. Si la etiqueta de aldehído no está accesible al disolvente, el polipéptido etiquetado con aldehído puede desplegarse por métodos conocidos en la técnica previa a la reacción con una pareja reactiva.

#### Porciones intercambiables unidas a la etiqueta de aldehído

55 En algunas modalidades, los polipéptidos etiquetados con aldehído pueden modificarse de una manera para facilitar la eliminación de una porción conjugada en el residuo de FGly de la etiqueta de aldehído y el reemplazo con una porción diferente. Este aspecto de la invención explota la estabilidad termodinámica relativa de los conjugados formados con diferentes parejas reactivas.

60 Por ejemplo, como se ilustra en la Figura 6, los aldehídos reaccionan fácilmente con porciones hidrazida y aminooxi para rendir hidrazonas y oximas, respectivamente. Aunque ambos de estos conjugados son robustos bajo condiciones fisiológicas, las oximas son termodinámicamente más estables. Por otra parte, las tiosemicarbazidas reaccionan además fácilmente con aldehídos para formar conjugados de tiosemicarbazona, que son termodinámicamente menos estables que las oximas. Estas diferencias en la estabilidad termodinámica pueden explotarse para el intercambio del conjugado de hidrazona de inferior estabilidad a un conjugado de oxima más estable, y para el intercambio del

conjugado de oxima de inferior estabilidad a un conjugado de semicarbazona más estable. Esta característica de la etiqueta de aldehído permite la modificación de la proteína objetivo con dos reactivos en secuencia (es decir, secuencialmente), como se ilustra en el ejemplo más abajo.

5 Polipéptidos modificados etiquetados con aldehído

10 Los productos de reacción producidos mediante la reacción de un polipéptido etiquetado con aldehído con una pareja reactiva que comprende una porción de interés generalmente se modifican de una manera sitio específica (es decir, en el residuo de FGly), que proporciona una población prácticamente homogénea de polipéptidos modificados etiquetados con aldehído. Las poblaciones heterogéneas de tales productos de reacción pueden generarse mediante el uso de dos o más parejas reactivas que comprenden porciones diferentes, cuando se prefiera.

15 Por ejemplo, cuando se modifica el polipéptido objetivo mediante la pegilación, los métodos pueden adaptarse para proporcionar la producción de una población homogénea de polipéptidos pegilados (en la que los polipéptidos se modifican con las mismas porciones de PEG) o una población heterogénea de polipéptidos pegilados (en la que los polipéptidos en la composición se modifican con diferentes tipos de moléculas de PEG).

Kits y sistemas

20 Los kits y sistemas se proporcionan para facilitar y, cuando se prefiera, estandarizar las composiciones de la invención y los usos de estas. Los kits contemplados en la presente pueden incluir uno o más de una construcción que codifica una etiqueta de aldehído para la inserción en un polipéptido objetivo; una construcción que codifica un polipéptido etiquetado con aldehído para la expresión en una célula huésped (por ejemplo, como un casete de expresión para proporcionar la inserción de una secuencia codificante de un polipéptido objetivo como una fusión N-terminal o C-terminal con la etiqueta de aldehído); una célula huésped que produce una FGE compatible con una etiqueta de aldehído del kit, donde la FGE puede ser endógena, recombinante, o heteróloga; una célula huésped modificada genéticamente para expresar un polipéptido de interés etiquetado con aldehído, cuya célula huésped puede expresar además una FGE compatible endógena, recombinante, o heteróloga para la conversión de la etiqueta de aldehído del polipéptido etiquetado; y una pareja reactiva para la modificación química de la etiqueta de aldehído convertida del polipéptido etiquetado.

30 Adicionalmente, el kit puede contener instrucciones para el uso de los componentes del kit, particularmente las composiciones de la invención que están contenidas en el kit.

Ejemplos

35 Los siguientes ejemplos se exponen para proporcionar a los expertos en la técnica una divulgación y descripción completas de cómo hacer y usar la presente invención, y no pretende limitar el alcance de lo que los inventores consideran como su invención ni pretende representar que los experimentos más abajo son todos o los únicos experimentos realizados. Se han hecho esfuerzos para asegurar la exactitud con respecto a los números usados (por ejemplo, cantidades, temperatura, etc), pero deben tomarse en cuenta algunos errores y desviaciones experimentales. A menos que se indique lo contrario, las partes son las partes en peso, el peso molecular es el peso molecular promedio, la temperatura es en grados centígrados, y la presión es la atmosférica o cercana.

Métodos y Materiales:

45 Se usaron los siguientes materiales y métodos en los Ejemplos 1-6 expuestos más abajo.

Construcción del plásmido. Se usaron los siguientes oligonucleótidos en los

50 Ejemplos más abajo:

Iniciador	Secuencia (5' → 3') <sup>a</sup>
ald <sub>6</sub> - <i>stf0</i> inicio <sup>b</sup>	CCACTGTGCACACCATCGCGGATGTCCGACCACCCACCGCC
ald <sub>13</sub> - <i>stf0</i> sentido	<b>CATGGCACCACTGTGCACACCATCGCGGGGCTCGCTGTTACCGGCCG CGACGTCCA</b>
ald <sub>13</sub> - <i>stf0</i> antisentido	<b>TATGGACGTCGCGGCCGGTGAACAGCGAGCCCCGCGATGGTGTGCACA GTGGTGC</b>
ald- <i>stf0</i> final	GCACCACCACCACCACACTGAGATCCGGCTGC

Iniciador	Secuencia (5' → 3') <sup>a</sup>
ald- <i>stf0</i> Cys5Ala <sup>b</sup>	CCATGGCACCACCTGGCCACACCATCGCGG
<i>stf0</i> final <sup>b</sup>	CGGCCGCGATGTGCGCCTTGAAGATCTGC
<i>mbp</i> sentido	GATCCCTGTGCACACCATCGCGGTGAGC
<i>mbp</i> sin sentido	GGCCGCTCACCGCGATGGTGTGCACAGG
ald- <i>mbp</i> Cys5Ala <sup>b</sup>	CCGCGTGGATCCCTGGCCACACCATCGCGG
ald <sub>6</sub> - <i>hGH</i> inicio	<b>CTATGCTACCATGGCGCTGTGCACACCATCGCGGACCATTCCCTTATCC AGGC</b>
<i>hGH</i> final	CTATGCTAGCGGCCGCGAAGCCACAGCTGCCCTCCAC
ald <sub>6</sub> - <i>hGH</i> Cys5Ala <sup>b</sup>	TATACCATGGCGCTGGCCACACCATCGCGGACC
<i>fge</i> inicio	CTATGCTACCATGGCTGACCGAGTTGGTTGACCTGC
<i>fge</i> final	TAGCATAGCTCGAGCTACCCGGACACCGGGTTCG
<i>fge</i> en el marco <sup>b</sup>	GAGGAATTAACCATGCTGACCGAGTTGGTTG

<sup>a</sup> Se subrayan las bases que codifican aldehído. <sup>b</sup> Iniciador de mutagénesis dirigida. Cuando sea adecuado, numerados desde el comienzo del codón de inicio de proteína respectivo. Un par de iniciadores complementarios se usó para cada mutante. No se muestran los complementos reversos.

Los motivos sulfatasa de las construcciones se proporcionan más abajo:

ald <sub>13</sub> -Stf0:	LCTPSRGSFLTGR-(sulfotransferasa micobacteriana)
ald <sub>13</sub> -Stf0 (C5A):	LATPSRGSFLTGR-(sulfotransferasa micobacteriana)
ald <sub>6</sub> -Stf0:	LCTPSR-(sulfotransferasa micobacteriana)
ald <sub>6</sub> -MBP:	LCTPSR-(proteína de unión a maltosa)
ald <sub>6</sub> -hGH:	LCTPSR-(hormona del crecimiento humana)

5 El ácido nucleico que codifica el ald<sub>13</sub>-Stf0 se construyó mediante oligonucleótidos hibridados que se ligan en un vector pET28-Stf0 construido previamente<sup>19</sup> entre los sitios de restricción NcoI y NdeI. El codón de terminación *stf0* se eliminó mediante mutagénesis con QuickChange™ (Stratagene) para permitir una etiqueta de His<sub>6</sub> en C-terminal. El ald<sub>6</sub>Stf0 se construyó con el uso de QuickChange (Stratagene) para eliminar los nucleótidos que codifican los últimos 7 residuos de la etiqueta de aldehído de 13 aminoácidos. El gen que codifica el ald<sub>6</sub>-MBP se construyó mediante oligonucleótidos hibridados que se ligan en el vector pMALc-H<sup>19</sup> entre los sitios de restricción XhoI y HindIII. El gen que codifica la hGH (variante de transcripción 1 de la hormona del crecimiento humana 1, que codifica los residuos 29-217) se amplificó a partir del pCMV-SPORT6.1.ccdB (Open Biosystems) con el uso de un iniciador 5' que codificó la etiqueta de aldehído de 6 aminoácidos y se ligó en el pET28b entre los sitios de restricción NcoI y NotI. El gen que codifica la FGE (Rv0712, que codifica los residuos 2-299) de *Mycobacterium tuberculosis* se amplificó a partir de un plásmido pET14b preparado previamente que contiene la FGE<sup>14</sup> y se ligó en el pBAD/myc-his A (Invitrogen) entre los sitios de restricción NcoI y XhoI. El gen de la FGE se ubicó en el marco con el codón de iniciación con el uso del kit de mutagénesis por PCR QuickChange (Stratagene). Los mutantes Cys→Ala de ald-Stf0, ald-MBP y ald-hGH se produjeron con el uso de mutagénesis con QuickChange™. La secuenciación del ADN se realizó para confirmar la fidelidad de cada producto génico. Los plásmidos que codifican proteínas se transformaron en células BL21(DE3) de *E. coli* (Invitrogen).

20 **Expresión y purificación de proteínas.** Las poblaciones clonales de células BL21(DE3) de *E. coli* que albergan sólo un plásmido que codifica una proteína etiquetada con aldehído se incubaron en medio LB con kanamicina con agitación a 37 °C hasta que la OD<sub>600</sub>= 0.5, momento en el que la temperatura se bajó a 18 °C y se añadió IPTG 100 µM. Las células BL21(DE3) de *E. coli* que albergan un plásmido que codifica una proteína etiquetada con aldehído y un plásmido que codifica la FGE se incubaron en medio LB con kanamicina y ampicilina con agitación a 37 °C hasta que la OD<sub>600</sub>= 0.5, momento en el que la expresión de la FGE se indujo con arabinosa 0.02%. Luego de 30 min, la temperatura se bajó a 18 °C y se añadió IPTG 100 µM para inducir la expresión de la proteína etiquetada con aldehído. Luego de 12-16 h, las células se cosecharon y resuspendieron en 20 ml de amortiguador de lisis (Tris 50 mM, NaCl 500 mM, glicerol 10%, imidazol 20 mM, DTT 1 mM, TCEP 1 mM, metionina 1 mM, pH 7.5, para ald<sub>6</sub>-hGH o NaH<sub>2</sub>PO<sub>4</sub> 50 mM, NaCl 300 mM, imidazol 10 mM, pH 7.4, para ald-Stf0 y ald<sub>6</sub>-MBP) por litro de cultivo y se lisaron por sonicación.

Los lisados celulares se trataron con ADNasa (10 µg/ml), se aclararon mediante centrifugación y se aplicaron a un 1 ml

de columna HisTrap (GE Healthcare). La columna se lavó con amortiguador de lisis con imidazol 35 mM y la proteína etiquetada con His<sub>6</sub> se eluyó con el uso de amortiguador de lisis con imidazol 250 mM. El ald<sub>6</sub>-hGH se purificó adicionalmente en una columna Sephadex 16/60 S300 (GE Healthcare).

5 **Digestión trípica y ensayo de adición de estándar.** Se digirieron 10 µg de proteína con 0.4 µg de tripsina (Promega) a 37°C por 16 horas en NH<sub>4</sub>HCO<sub>3</sub> 50 mM pH 8. Este protocolo se consideró suficiente para la digestión completa ya que no se detectaron péptidos que contenían sitios de clivaje errados mediante espectrometría de masa MALDI-TOF luego de 3 horas de digestión bajo idénticas condiciones. Los ensayos de adición de estándar se ejecutaron en agua con aproximadamente 0.6 µg de la digestión proteica por corrida. Los péptidos sintéticos que contienen ya sea la cisteína o el aldehído (FGly) se añadieron en cantidades equimolares seguido de la adición de DTT 100 mM. Esta solución se dejó incubarse a temperatura ambiente por 1 hora previo al análisis por espectrometría de masa (Agilent MSD). Se añadieron corridas blanco entre las corridas seleccionadas aleatoriamente y no se detectó señal residual. No se observó oxidación de la cisteína.

15 **Marcaje de molécula pequeña.** Las reacciones de marcaje fluorescente se ejecutaron con 10 µg de la proteína objetivo con colorante aminooxi 300 µM (C5-aminooxiacetamida Alexa Fluor 647, Invitrogen) en amortiguador de marcaje (MES 100 mM pH 5.5, SDS 1%) a 37 °C por 2 horas. Se añadió metoxilamina 166 mM para controlar las reacciones. Las mezclas de reacción se separaron mediante SDS-PAGE y la fluorescencia se detectó con el uso de un escáner Typhoon 9410 (GE Healthcare). La carga proteica se determinó mediante tinción con Sypro Ruby (Sigma). La biotilación se facilitó mediante la incubación de 10 µg de proteína objetivo con biotina hidrazida (Sigma) 30 µM en amortiguador de marcaje por 2 horas a 37 °C. El desplazamiento subsecuente de la biotina hidrazida se facilitó mediante la adición ya sea de metoxilamina 166 mM o aminooxiFLAG 1 mM a 37 °C por 2 horas. La transferencia Western de la α-biotina se realizó con el uso del protocolo estándar. La membrana α-FLAG se obtuvo mediante la estabilización de la membrana y el sondeo con α-FLAG M2 (Sigma). El aminooxiFLAG se sintetizó con el uso de los protocolos estándar de síntesis de péptidos en fase sólida basados en Fmoc. El residuo final añadido, en C-terminal, fue el ácido acético(t-Boc-aminooxi) seguido por el clivaje bajo condiciones estándar.

20 **Pegilación.** Los aminooxiPEG se crearon a partir de los monoamino-PEG y aminooxiglicina con el uso de protocolos estándar. Más específicamente, los aminooxiPEG se produjeron mediante la adición de aminoPEG (Shearwater Polymers) a ácido acético(t-Boc-aminooxi) activado con el uso de condiciones de acoplamiento estándar de péptidos. Brevemente, la formación del enlace amida se logra mediante la adición de aminoPEG al éster 8-hidroxibenzotriazol de ácido acético(t-Boc-aminooxi) (5 equivalentes) preformado en acetonitrilo. La purificación del producto se facilitó mediante la precipitación a partir de éter, seguido por la trituración. La desprotección se logró mediante el tratamiento con una solución acuosa de ácido trifluoroacético (TFA 95%, H<sub>2</sub>O 5%) por 3 horas a temperatura ambiente. La precipitación en éter y la trituración facilitó un producto puro a juzgar mediante NMR <sup>1</sup>H. La conjugación a proteínas etiquetadas con aldehídos se facilitó mediante la incubación de 10 µg de la proteína objetivo y aminooxiPEG 10 mM en solución de acoplamiento (CH<sub>3</sub>CN 49.95%, H<sub>2</sub>O 49.95%, TFA 0.1%) por 1 hora seguida por la liofilización.

35 La conjugación a proteínas etiquetadas con aldehídos se facilitó mediante la incubación de 10 µg de la proteína objetivo y aminooxiPEG 10 mM en solución de acoplamiento (CH<sub>3</sub>CN 49.95%, H<sub>2</sub>O 49.95%, TFA 0.1%) por 1 hora seguida por la liofilización. Las mezclas de reacción se resuspendieron en agua, se separaron mediante SDS-PAGE, se tiñeron con naranja Sypro (Invitrogen) y se detectaron con el uso de un escáner Typhoon 9410 (GE Healthcare).

40 **Ejemplo 1: modificación sitio específica de un motivo sulfatasa en la proteína expresada en *E. Coli***

45 Las construcciones proteicas ya sean con etiquetas de aldehído en N-o C-terminal se expresaron en *E. coli*. Tres proteínas objetivo se exploraron, una proteína de unión a maltosa (MBP) etiquetada en el C-terminal, una hormona de crecimiento humana (hGH) etiquetada en el N-terminal, y una sulfotransferasa micobacteriana (Stf0) etiquetada en el N-terminal. Además, se probaron dos variantes de la etiqueta de aldehído, una etiqueta de 13 residuos (ald<sub>13</sub>-Stf0) que incluyó la totalidad del motivo consenso sulfatasa, y una etiqueta de 6 residuos (ald<sub>6</sub>-Stf0) que incluyó una secuencia más corta que contiene un motivo consenso sulfatasa:

LCTPSRGSFLTGR-Stf0	(ald <sub>13</sub> -Stf0)
LCTPSR-Stf0	(ald <sub>6</sub> -Stf0).

50 Con el objetivo de asegurar la formación eficiente de FGly, las proteínas etiquetadas se coexpresan con una FGE procarionota de *Mycobacterium tuberculosis (Mtb)* (descrito en los Ejemplos más abajo).

55 La digestión trípica del péptido que contiene el motivo consenso sulfatasa de 13 residuos (ald<sub>13</sub>-Stf0) permitió la identificación directa en el espectro de masa de la FGly (Figura 2). Si bien el péptido que contiene FGly podría identificarse fácilmente, el péptido que contiene cisteína no se observó lo que indica la oxidación eficiente de la etiqueta de aldehído.

Ejemplo 2: los péptidos que contienen el motivo consenso sulfatasa de 6 residuos demuestran superiores velocidades de conversión

5 Para cuantificar el grado de conversión de Cys a FGly, se realizó un ensayo de adición de estándar. Los niveles relativos de cisteína y FGly dentro de péptidos derivados trópicamente de proteínas objetivos se compararon con una curva de adición de estándar, que se produjo mediante la implantación de péptidos sintéticos en las digestiones trópticas a varias concentraciones (Figura 3, panel a).

10 Inesperadamente, el péptido Stf0 que contiene el motivo consenso sulfatasa de 6 residuos (ald<sub>6</sub>-Stf0), demostró una conversión ligeramente superior a la del péptido Stf0 que contiene la secuencia conservada de 13 aminoácidos (ald<sub>13</sub>-Stf0), con niveles de conversión de 92 ± 3% y 86 ± 5%, respectivamente. Este resultado está en contraste a los estudios previos de la sulfatasa que indicaron que la secuencia distal treonina-glicina-arginina (TGR) es importante para la oxidación eficiente de cisteína (Dierks y otros (1999) EMBO J. 18 (8):2084-91). El péptido de hGH que contiene el motivo consenso sulfatasa de 6 residuos (ald<sub>6</sub>-hGH) demostró una conversión significativamente superior a 99 ± 5%.

15 Ejemplo 3: la conversión de los motivos sulfatasa por fge es independiente del contexto de la secuencia primaria, lo que permite así el posicionamiento de las etiquetas de aldehído en las posiciones ya sean N-terminal o C-terminal dentro de un polipéptido

20 Los estudios previos de la sulfatasa han indicado que la secuencia distal treonina-glicina-arginina (TGR) en el motivo consenso sulfatasa de 13 residuos (subrayado más abajo) es importante para la oxidación de cisteína a nivel superior:

LCTPSRGS**LFTGR**

25 Debido a que se cree que la formación de FGly se produce cotraduccionalmente (Dierks y otros Proc Natl Acad Sci Estados Unidos A. 1997 28 de octubre; 94(22):11963-8), se razonó que las construcciones C-terminal podrían experimentar inferior formación de FGly debido a la inaccesibilidad de la etiqueta de aldehído. Esto se probó mediante la generación de un polipéptido etiquetado en el C-terminal que contiene la proteína de unión a maltosa y el motivo consenso sulfatasa de 6 residuos (ald<sub>6</sub>-MBP):

30 LCTPSR-(proteína de unión a maltosa)

35 Sorprendentemente, el ald<sub>6</sub>-MBP etiquetado en el C-terminal demostró además una conversión casi cuantitativa en 99 ± 2%. Teniendo en cuenta que el motivo sulfatasa se encuentra de forma nativa dentro del interior de las sulfatasas, estos resultados indican que la formación del aldehído no se limita con respecto a la posición de la secuencia primaria de la etiqueta.

Ejemplo 4: marcaje fluorescente selectivo de las proteínas etiquetadas con aldehído

40 Para demostrar la especificidad facilitada por la introducción de la FGly, se marcó un panel de proteínas etiquetadas con aldehído con el colorante aminoacetamida 647 Alexa Fluor® (Invitrogen):

ald <sub>13</sub> -Stf0 (C5A):	LATPSRGS <b>LFTGR</b> -(sulfotransferasa micobacteriana)
ald <sub>13</sub> -Stf0:	LCTPSRGS <b>LFTGR</b> -(sulfotransferasa micobacteriana)
ald <sub>6</sub> -Stf0:	LCTPSR-(sulfotransferasa micobacteriana)
ald <sub>6</sub> -MBP:	LCTPSR-(proteína de unión a maltosa)

45 Las proteínas etiquetadas con aldehído demostraron un marcaje robusto con fluoróforo (Figura 3, panel b): ald<sub>13</sub>-Stf0 (-), ald<sub>6</sub>-Stf0 (-), y ald<sub>6</sub>-MBP (-). En contraste, las proteínas control en las que la cisteína crítica en el motivo de la etiqueta de aldehído se mutó a alanina demostraron sólo una pequeña cantidad de marcaje de fondo (Figura 3, paneles c y d: ald<sub>13</sub>-Stf0 (C5A) (-)). Las proteínas etiquetadas con aldehído incubadas con un exceso de metoxilamina, un nucleófilo que compite, demostraron no marcaje (Figura 3, paneles c y d, ald<sub>13</sub>-Stf0 (C5A) (+), ald<sub>13</sub>-Stf0 (+), ald<sub>6</sub>-Stf0 (+), y ald<sub>6</sub>-MBP (+)). Curiosamente, aunque el genoma de *E. coli* no contiene una FGE anotada, la proteína etiquetada con aldehído expresada sin la FGE exógena aún demostró el marcaje fluorescente, aunque con inferior intensidad. Esto indica que *E. coli* tiene que expresar de forma nativa una enzima o enzimas que son capaces de la oxidación del motivo sulfatasa.

55 Ejemplo 5: la modificación de una proteína etiquetada con aldehído puede proporcionar porciones "intercambiables"

60 Los aldehídos reaccionan fácilmente con porciones hidrazida y aminooxi para rendir hidrazonas y oximas, respectivamente. Aunque ambos de estos conjugados son robustos bajo condiciones fisiológicas, las oximas son termodinámicamente más estables. Esta diferencia puede explotarse para el intercambio del conjugado de hidrazona de inferior estabilidad al conjugado oxima más estable. Esta característica de la etiqueta de aldehído permite la modificación de la proteína objetivo con dos reactivos en secuencia (es decir, secuencialmente), como se ilustró

mediante la conjugación de una etiqueta de purificación seguida por el reemplazo de la etiqueta de purificación conjugada para proporcionar un fluoróforo conjugado.

5 Para evaluar la viabilidad de esta técnica, un polipéptido que contiene la proteína de unión a maltosa y el motivo consenso sulfatasa de 6 residuos (ald<sub>6</sub>-MBP) primero se marcó con biotina hidrazida y subsecuentemente se incubó con metoxilamina o una etiqueta de epítipo aminooxi (aminooxi-FLAG). El marcaje con biotina hidrazida condujo a una señal robusta de α-biotina en una transferencia Western (Figura 4, panel a, carril 1). La incubación subsecuente con metoxilamina o aminooxi-FLAG condujo a una pérdida completa de la señal de la α-biotina (Figura 4, panel a, carril 2) o una señal robusta de α-FLAG (Figura 4, panel a, carril 3), respectivamente.

10 Cuando la proteína marcada con aminooxi-FLAG se expuso subsecuentemente a metoxilamina, sólo se observó una pérdida parcial de la señal (datos no mostrados), presumiblemente debido a las estabilidades similares de los conjugados. Estos resultados indican que la conjugación secuencial a una proteína etiquetada con aldehído puede programarse basado en la estabilidad del enlace químico.

15 Ejemplo 6: creación de la pegilación sitio específica para producir conjugados de proteína con peg en una proteína terapéutica objetivo

20 Para ilustrar el uso de etiquetas de aldehído en la mediación de la pegilación sitio específica, se usaron etiquetas de aldehído para unir sitio específicamente las cadenas de polietilenglicol (PEG) al ald<sub>6</sub>-Stf0 expresado recombinantemente. El ald<sub>6</sub>-Stf0 se expresó recombinantemente y se modificó con una serie de aminooxiPEG con longitudes variables de cadena. El análisis por SDS-PAGE de los conjugados de Stf0-PEG demostró cambios inequívocos en la masa consistentes con el peso molecular y la carga de las moléculas de PEG anexas (Fig. 4, panel b). Estos resultados demuestran la facilidad de obtención de conjugados sitio específicos de PEG-proteína independientemente del número de cisteínas o lisinas nativas.

30 Lo anterior proporciona la prueba de principio para la aplicación de etiquetas de aldehído para mediar la pegilación sitio específica de, por ejemplo, proteínas terapéuticas. La pegilación de proteínas farmacéuticas es preferible ya que puede aumentar el índice terapéutico mediante el aumento de la estabilidad proteolítica y la disminución del aclaramiento renal. Además, la pegilación puede explotarse para reducir la inmunogenicidad de los productos farmacéuticos proteicos. A pesar de los avances en la química de conjugación proteica, la modificación sitio específica de proteínas sigue siendo problemática. La derivatización de residuos de cisteína o lisina actualmente es el método más utilizado para pegar proteínas, pero este método de marcaje no específico resulta en la pegilación de múltiples sitios, creando una colección no preferible de distintos conjugados de proteína-PEG con diferentes farmacocinéticas. La tecnología de etiqueta de aldehído descrita en la presente invención puede usarse para abordar necesidades tales como estas.

35 Ejemplo 7: uso de etiquetas de aldehído para modificar los residuos accesibles en la superficie celular de un polipéptido expresado en células de mamífero

40 Para demostrar la introducción de la FGly en una proteína integral de la membrana celular, se produjo una proteína canal de K<sup>+</sup> sintética regulada por azobenceno fotoisomerizable (SPARK) etiquetada con aldehído. Las proteínas canales SPARK, que son canales de iones de K<sup>+</sup> activados por la luz, se desarrollaron para el control no invasivo de la actividad neuronal (Banghart y otros Nat. Neurosci. 2004).

45 La etiqueta de aldehído de 6 residuos descritos anteriormente (ald<sub>6</sub>(LCTPSR)) se introdujo en una construcción que codifica una proteína canal SPARK. Se usaron tres estrategias: 1) la adición del motivo consenso sulfatasa de 6 residuos de la etiqueta ald dentro de uno de los bucles extracelulares de la proteína (referidos como "I" en la Figura 5); 2) la delección de 6 residuos del bucle y después el reemplazo de estos residuos con la etiqueta ald de 6 residuos (referido como "C" en la Figura 5), y 3) la delección de 3 residuos del bucle y después la adición de la etiqueta ald de 6 residuos (referido como "P" en la Figura 5). Se ejecutó además un control negativo de vector solo ("V" en la Figura 5).

50 Los plásmidos que codifican cada una de las tres variantes del canal SPARK recombinante, etiquetado con aldehído se transfectaron en células de ovario de hámster chino (CHO) y en células humanas embrionarias de riñón (HEK). Ambas células CHO y HEK expresan una FGE endógena. Sin embargo, con el objetivo de aumentar la conversión de la Cys de la etiqueta de aldehído, se cotransfectó un plásmido que codifica el SPARK etiquetado con aldehído con una construcción pcDNA3.1 que codifica la FGE humana. La FGE humana usada tiene la secuencia de aminoácidos:

maapalglvc grcpelglvl lllllsllcg aagsqeagtg agagslagsc  
 gcgtpqrpga hgssaaahry sreanapgpv pgerqlahsk mvpipagvft  
 mgtddpqikq dgeaparrvt idafymdaye vsntefekfv nstgylteae  
 kfgdsfvfeg mlseqvktni qqavaapww lpvkganwrh pegpdstilh  
 rpdhplhvs wndavayctw agkrlpteae weyscrgglh nrlfpwgnkl  
 qpkqghyani wqgefvtnt gedgfggtap vdafppngyg lynivgnawe  
 wtsdwwtvhh sveetlnpkg ppsgkdrvkk ggsymchrsy cyryrcaars  
 qntpdssasn lgfrcaadr1 ptmd

- 5 Luego de un día, las células se lisaron y el lisado se probó mediante una transferencia Western para la presencia de un epítipo myc (que está presente en la proteína canal SPARK, y así demuestra la transfección y traducción exitosas) y para la presencia del aldehído mediante la reacción con el uso de aminooxiFLAG, seguido por el sondeo con un anticuerpo anti-FLAG. Una membrana Ponceau demostró cargas iguales de muestras del mismo tipo celular en la membrana.
- 10 Como se muestra en la Figura 5, fue exitosa la estrategia que incluye la delección de 6 residuos del bucle extracelular del SPARK y el reemplazo con la etiqueta ald de 6 residuos (ver la flecha en la membrana anti-FLAG en la Figura 5, panel c). Este resultado demuestra la capacidad para modificar los residuos de la superficie celular de una proteína etiquetada con aldehído en células de mamífero.
- 15 La presencia del FLAG del SPARK etiquetado con aldehído en la superficie celular puede confirmarse con el uso de la citometría de flujo.

#### Ejemplo 8: uso de etiquetas de aldehído para modificar el fragmento Fc de un anticuerpo

- 20 Con el objetivo de demostrar además las aplicaciones de las etiquetas de aldehído, un fragmento soluble Fc de IgG se modificó para contener una etiqueta de aldehído ya sea en el N-o C-terminal. Brevemente, una etiqueta de aldehído de 13 residuos (ald<sub>13</sub>) (LCTPSRAALLTGR) se introdujo para posicionar la etiqueta de aldehído ya sea en el N-terminal o C-terminal del fragmento soluble Fc de IgG codificado en el vector pFuse-Fc (Invitrogen) disponible comercialmente. Con el objetivo de aumentar la conversión de la Cys de la etiqueta de aldehído, las células CHO se cotransfectaron con la construcción que codifica la Fc y una construcción pcDNA3.1 que codifica la FGE humana.
- 25

30 Los fragmentos Fc se aislaron a partir del sobrenadante celular, y la detección del fragmento Fc de IgG etiquetado con aldehído en el que la Cys se convirtió a FGly se logró mediante la reacción de la proteína aislada con una sonda (FLAG-OH<sub>2</sub>) de aminooxiFLAG (DYKDDDDK), seguido por análisis en SDS-PAGE y transferencia Western. Las proteínas que no reaccionaron con la sonda FLAG-HNO<sub>2</sub> sirvieron como un control adicional.

35 Mientras que las fusiones de Fc que contienen la etiqueta de aldehído de 12mer dieron un marcaje robusto cuando están presentes ya sea en el N-terminal (N-Fc-Ald13) o C-terminal (C-Fc-Ald13), la proteína control, en la que la cisteína crítica se había mutado a alanina (mutación C a A), dio una señal no detectable (Figura 17).

40 Con el objetivo de evaluar si una etiqueta de aldehído de 6mer es suficiente para mediar la modificación de una proteína, se generaron los fragmentos Fc de IgG que tienen una etiqueta de aldehído de 6mer (Fc-Ald) o una etiqueta control (Fc-C->A) en el C-terminal con el uso del vector pFuse-Fc. Los fragmentos de Fc de IgG etiquetados con aldehído se detectaron mediante la reacción de la proteína aislada con una sonda (FLAG-HNO<sub>2</sub>) de aminooxiFLAG, seguido por SDS-PAGE y transferencia Western. Las proteínas que no reaccionaron con la sonda FLAG-HNO<sub>2</sub> sirvieron como un control adicional. Como se muestra en la Fig. 18, la etiqueta de aldehído de 6mer facilitó el marcaje robusto del Fc-Ald, mientras que se observó marcaje no detectable con fragmentos Fc modificados para incluir la etiqueta control. Las construcciones que codifican fragmentos Fc de IgG que tienen la posición de la etiqueta de aldehído de 6mer en el N-terminal rindieron resultados similares (datos no mostrados).

45

50 Con el objetivo de confirmar la modificación de formilglicina (FGly) de los fragmentos Fc, los fragmentos etiquetados ald13-Fc en el N-o el C-terminal se sometieron a digestión trípica para permitir la identificación directa en el espectro de masa de la FGly. Como se muestra en las Figs. 19 y 20, el péptido que contenía FGly y el péptido que contenía cisteína pudieron identificarse fácilmente a partir de los dos fragmentos Fc modificados en el N-terminal y en el C-terminal.

El marcaje específico del Fc etiquetado con aldehído se realizó además sometiendo el medio libre de suero directamente a la sonda aminooxiFLAG (datos no mostrados).

#### Ejemplo 9: eficiencia de conversión de cys a FGly en proteínas etiquetadas con aldehído



Con el objetivo de cuantificar el grado de conversión de Cys a FGly, se desarrolló un ensayo para analizar la eficiencia de conversión de proteínas objetivo digeridas con tripsina. La cantidad del péptido no modificado que contenía cisteína se determinó a partir de una curva estándar, que se produjo mediante la implantación de péptidos sintéticos en las digestiones tríplicas a varias concentraciones. La cantidad del péptido que contiene FGly se calculó restando la cantidad del péptido que contiene cisteína de la cantidad total de proteína, determinada con el uso del ensayo de proteína BCA.

Cuando este ensayo se aplicó al fragmento Fc etiquetado en N-y C-terminal descrito en el ejemplo anterior, se encontró que en presencia de la FGE humana exógena (hFGE) la eficiencia de conversión de Cys a FGly fue  $86 \pm 1\%$  para el ald13-Fc etiquetado en el N-terminal y  $58 \pm 2\%$  para el Fc-ald13 etiquetado en el C-terminal. En contraste, en la ausencia de la hFGE exógena, la eficiencia de conversión fue sólo aproximadamente 25% y aproximadamente 23% para el fragmento Fc etiquetado en el N-y el C-terminal, respectivamente. El fragmento Fc modificado en el C-terminal que contenía una etiqueta de aldehído de 6mer exhibió una eficiencia de conversión de aproximadamente 92% en presencia de hFGE exógena.

#### Ejemplo 10: modificación mediada por etiqueta de aldehído de proteínas de superficie celular

Este ejemplo demuestra que las etiquetas de aldehído pueden usarse para facilitar la modificación sitio específica de las proteínas de superficie celular, el dominio transmembrana (PDGFR) del receptor del factor de crecimiento derivado de plaquetas (codificado por el vector pDisplay de Invitrogen), en células HEK vivas con el uso del mismo enfoque.

La etiqueta de aldehído de 13mer (LCTPSRAALLTGR) o una etiqueta control (LATPSRAALLTGR) se introdujo en una construcción de expresión pDisplay™ (Invitrogen; Fig. 21) entre los sitios Bgl II y Sal I. Las proteínas de fusión resultantes se refieren aquí como Ald13-TM (que contiene la etiqueta de aldehído de 13mer) y C->A-TM (que contiene la etiqueta control). Esta construcción de expresión y una construcción que expresa la FGE humana (hFGE) se transfectaron transitoriamente en células HEK293-T para proporcionar la expresión.

El marcaje de las células se logró mediante la reacción con una oxiaminobiotina y se sondearon con conjugados de estreptavidina Alex fluoro 488. Las células se sometieron después a análisis por citometría de flujo.

Como se ilustra en la Fig. 21, la fluorescencia media de las células que expresan la proteína de superficie Ald 13-TM fue significativamente superior (fluorescencia media de aproximadamente 24.42) que las células que expresan el control C->A-TM (fluorescencia media de aproximadamente 3.31).

#### Ejemplo 11: modificación de la etiqueta de aldehído para el marcaje de proteína citosólica

Para ilustrar el uso de la etiqueta de aldehído en el marcaje específico de proteínas citosólicas, se generaron las construcciones que codifican la proteína verde fluorescente derivada de *Aequorea coerulea* (AcGFP) etiquetada con aldehído o etiquetada con control. Con el uso del vector pAcGFP1-N1 disponible comercialmente (Clontech), una construcción de expresión que codifica una proteína de fusión AcGFP compuesta por una etiqueta de His (seis residuos de histidina, representado por His<sub>6</sub>) seguida por una etiqueta de aldehído de 13mer (LCTPSRAALLTGR) o una etiqueta control (LATPSRAALLTGR) posicionada en el N terminal de la AcGFP se generó mediante la inserción y las secuencias codificantes de la etiqueta de His y de la etiqueta de aldehído de 13mer entre los sitios de restricción *Kpn I* y *Xma I*.

Un homólogo de la FGE bacteriana derivado de *Streptomyces coelicolor* (StrepFGE) se clonó en un vector de expresión de mamífero (pcDNA 3.1, Invitrogen) para la cotransfección de células HEK con un plásmido que codifica GFP etiquetada con aldehído (Ald-AcGFP) o GFP etiquetada con control (C->A-AcGFP). Las células que carecen de expresión de StrepFGE se usaron como un control adicional.

La detección de la AcGFP etiquetada con aldehído que contenía FGly se logró mediante la reacción de la proteína aislada con una sonda (FLAG-ONH<sub>2</sub>) aminooxiFLAG (DYKDDDDK), seguido por análisis de SDS-PAGE y transferencia Western. Las proteínas que no reaccionaron con la sonda FLAG-HNO<sub>2</sub> sirvieron como un control adicional.

En presencia del homólogo de la FGE citosólica, el residuo de cisteína dentro de la secuencia consenso se convirtió eficientemente a una formilglicina (FGly) (Fig. 22), mientras que la AcGFP etiquetada con control no exhibió marcaje detectable lo que indica una FGly no detectable. Adicionalmente, el Ald-GFP producido en células HEK que no expresan StrepFGE produjo además una fuerte señal (Fig. 22). Esto puede deberse al método de aislamiento de proteínas usado en el que las células HEK se lisan, y así pueden liberar la hFGE desde el ER de estas células, lo que permite así el contacto entre la hFGE y la etiqueta de aldehído resultante en la conversión de la cisteína al aldehído.

#### Ejemplo 12: modificación de IFN-beta con la etiqueta de aldehído

Las etiquetas de aldehído pueden usarse para facilitar la modificación de una variedad de proteínas. Las proteínas ilustrativas de interés para la modificación incluyen el interferón beta (IFN-beta). El IFN-beta está compuesto por cinco hélices alfa (A-E) con un único sitio de glicosilación que existe en el residuo Asn-80. El IFN-beta puede modificarse para

proporcionar la modificación en el sitio de glicosilación y/o en otros sitios de la proteína accesibles al disolvente. Por ejemplo, la secuencia de aminoácidos del IFN-beta que facilita la glicosilación puede modificarse para proporcionar una etiqueta de aldehído. Por ejemplo, con el uso de técnicas recombinantes, puede reemplazarse la secuencia DSSSTGWNE del IFN-beta presente en un bucle del IFN-beta por la secuencia GSLCTPSRG que contiene la etiqueta de aldehído. La etiqueta de aldehído puede explotarse después para unir a una porción de interés, como se ilustró en la Fig. 23.

#### Ejemplos 13-14: identificación y caracterización de una FGE de *M. Tuberculosis*

En los siguientes Ejemplos, una FGE procariota se identifica funcionalmente en *Mycobacterium tuberculosis* (*Mtb*). Como se discutió anteriormente, las sulfatasas son miembros de una familia en expansión de enzimas que emplean cofactores novedosos derivados co o postraduccionalmente para facilitar la catálisis, y contienen un residuo de FGly en el sitio activo. Se cree que el residuo de FGly experimenta la hidratación del diol geminal, luego de lo que uno de los grupos hidroxilo actúa como un nucleófilo catalítico para iniciar el clivaje éster sulfato (Figura 7, panel a). El residuo de FGly se localiza dentro de una secuencia consenso sulfatasa, lo que define a la familia de enzimas sulfatasa y es altamente conservada a lo largo de todos los dominios de la vida (Figura 7, panel b). Mientras que la FGly se forma a partir de residuos de cisteína en sulfatasas eucariotas, en sulfatasas procariotas pueden oxidarse ya sea la cisteína (dentro del motivo núcleo CXPXR) o la serina (SXPXR). Algunos procariotas, tales como *Mtb*, codifican sólo sulfatasas de tipo cisteína, mientras que otras especies tienen sólo sulfatasas de tipo serina o una combinación de ambas.

Los ejemplos 8-9 describen la caracterización de una FGE procariota de *Mtb* y resuelven la estructura del ortólogo de *Strep*. Nuestros estudios indican que las sulfatasas activadas por FGE representan aproximadamente la mitad de la actividad total de sulfatasa en el lisado de *Mtb*, lo que sugiere que este organismo posee sulfatasas independientes a FGE que aún no se han identificado. La definición del repertorio completo de sulfatasas de *Mtb* (y otros procariotas) es una meta importante para el futuro y proporcionará la plataforma para la definición del rol de estas enzimas en el ciclo de vida y la patogénesis del organismo.

#### Material y métodos

Los siguientes métodos y materiales se usaron en los ejemplos que se refieren a la identificación de una FGE en *Mycobacterium tuberculosis* (*Mtb*), y la producción de una cepa de *Mtb* deficiente de FGE.

**Preparación de los vectores de expresión de proteína.** La tabla más abajo enumera los oligonucleótidos usados en los ejemplos más abajo. El gen que codifica la FGE de *Mtb* (Rv0712, que codifica los residuos 2-299) se amplificó a partir del ADN genómico H37Rv de *Mtb* y se clonó en el pET14b (Novagen) con el uso de los sitios de restricción NdeI y XhoI. El gen que codifica la FGE de *Strep* (SCO7548, que codifica los residuos 2-314) se amplificó a partir del ADN genómico A3(2) de *Strep* y se clonó en pET151/D-TOPO (Invitrogen). Los marcos abiertos de lectura de Rv2407 (que codifica los residuos 2-273), Rv3406 (que codifica los residuos 2-295), y Rv3762c (que codifica los residuos 2-626) se amplificaron a partir del ADN genómico H37Rv de *Mtb*. El Rv2407 se ligó en el pMAL-C2X (New England Biolabs) con el uso de sitios de restricción BamHI y PstI, y ambos Rv3406 y Rv3762c se ligaron en el pET28b (Novagen) con el uso de los sitios de restricción NdeI y XhoI. La secuenciación del ADN se realizó para confirmar la fidelidad de cada producto génico. Los plásmidos que codifican proteína se transformaron en células BL21(DE3) (Invitrogen).

#### Iniciadores de oligonucleótidos

Iniciador	Secuencia (5' → 3')
<i>Mtb fge</i> Inicio	CTATGCTACATATGCTGACCGAGTTGGTTGACCTGC
<i>Mtb fge</i> Fin	TAGCATAGCTCGAGCTACCCGGACACCGGGTCCG
<i>fge Strep</i> Inicio	CACCGCCGTGGCCGCCCGTCCCC
<i>fge Strep</i> Fin	TCACTCAGCGGCTGATCCGG
<i>Mtb Rv2407</i> Inicio	CTATGCTAGGATCCCTTGAGATCACGTTGCTCGG
<i>Mtb Rv2407</i> Fin	CTATGCTACTGCAGCTAGCGCCGCGGGTGCACCTC
<i>Mtb Rv3406</i> Inicio	CTATGCTACATATGACAGATCTGATTACCGTGAAG
<i>Mtb Rv3406</i> Fin	CTATGCTACTCGAGTCAGCCAGCGATCTCCATCG
<i>Mtb Rv3762c</i> Inicio	CTATGCTACATATGCCGATGGAACACAAACCTCC
<i>Mtb Rv3762c</i> Fin	CTATGCTACTCGAGCTACGGCGTCACGATGTTGAAG
<i>Mtb fge Ser260Alaa</i>	GACCCTCAAGGGCGGCGCACACCTGTGCGCGCCC
<i>Mtb fge Cys263Sera</i>	TGCGACCTGAGCGCGCCGGAGTACTGC
<i>Mtb fge Cys268Sera</i>	GCGCCGGAGTACAGCCACCGCTACCGC

Iniciador	Secuencia (5' → 3')
<b>Strep fge Trp234Alaa</b>	CACCGCGGGCAACGTGGCGGAATGGTGCTCCGAC
<b>Strep fge Trp234Phea</b>	CACCGCGGGCAACGTGTTTGAATGGTGCTCCGAC
<b>Strep fge Cys272Sera</b>	GGCGGCTCCTACCTGTCCCACGACTCCTACTGC
<b>Strep fge Cys277Sera</b>	GTGCCACGACTCCTACTCCAACCGCTACCGGGTCCG
<b>Mtb Δfge corriente arriba 5'</b>	CTATGCTAAAGCTTGAATCGAGTGAGATATTGCC
<b>Mtb Δfge corriente arriba 3'</b>	TAGCATAGTCTAGAATGACGCTCGATCGAGAACG
<b>Mtb Δfge corriente abajo 5'</b>	CTATGCTATCTAGATCCTCACAGTCGCAGGACAGC
<b>Mtb Δfge corriente abajo 3'</b>	TAGCATAGTTAATTAATGCACCATCTCGTTGCTCTCG

"Numerado desde el comienzo del codón de iniciación respectivo de la FGE. Un par de iniciadores complementarios se usó para cada mutante. Los complementos reversos no se muestran; se subrayan los cambios en la secuencia.

5 **Mutagénesis dirigida.** Las mutaciones sitio específicas en la FGE de *Mtb* y la FGE de *Strep* se produjeron con el uso del kit de mutagénesis por PCR QuikChange (Stratagene). Los plásmidos pET14b FGE de *Mtb* y pET151 FGE de *Strep* y los oligonucleótidos adecuados de la tabla anterior se usaron en las reacciones de mutagénesis. Las mutaciones se confirmaron mediante secuenciación de ADN y los plásmidos se transformaron en células BL21(DE3) para la expresión de proteína como se describe más abajo.

10 **Expresión y purificación de proteínas.** Las poblaciones clonales de células BL21(DE3) que albergan un plásmido que codifica proteína etiquetada con His<sub>6</sub> se incubaron en medio LB con ampicilina o kanamicina con agitación a 37 °C hasta que la OD<sub>600</sub> = 0.5, momento en el cual la temperatura se bajó a 18 °C y se añadió IPTG 250 μM. Luego de 12-16 h, las células se cosecharon y resuspendieron en 20 ml de amortiguador de lisis (Tris 50 mM, NaCl 500 mM, glicerol 10%, imidazol 20 mM, DTT 1 mM, TCEP 1 mM, metionina 1 mM, pH 7.5) por litro de cultivo y se lisaron por sonicación. El lisado celular se trató con ADNasa (10 μg/ml), se aclaró por centrifugación y se aplicó a 1 ml de columna HisTrap (GE Healthcare). La columna se lavó con amortiguador de lisis con imidazol 35 mM y la proteína etiquetada con His<sub>6</sub> se eluyó con el uso de amortiguador de lisis con imidazol 250 mM. El volumen de elución se concentró a menos de 2 ml si fue necesario y se purificó además en una columna Sephadex 16/60 S300 (GE Healthcare). La proteína recombinante purificada se concentró subsecuentemente a aproximadamente 20 mg/ml.

20 La identidad y pureza de la FGE de *Mtb* y *Strep* se evaluó mediante espectrometría de masa de ionización por electronebulización (Bruker/Agilent Esquire). EL Rv2407 no fue soluble en la forma etiquetada con His<sub>6</sub> y se fusionó alternativamente a la proteína de unión a maltosa (MBP). Las condiciones de crecimiento y de lisis para las células productoras de MBP-Rv2407 fueron las mismas que anteriormente excepto con la ausencia del imidazol en el amortiguador de lisis. El lisado aclarado se aplicó a resina de amilosa (New England Biolabs) en amortiguador de lisis, se lavó en amortiguador de lisis adicional, y el MBP-Rv2407 se eluyó en amortiguador de lisis con 10 mM de maltosa y se concentró subsecuentemente. La MBP se clivó y se eliminó del Rv2407 con el uso del Factor Xa (New England Biolabs) y resina de amilosa, respectivamente.

30 **Cristalización de la FGE de Strep.** Los intentos para cristalizar homólogos de la FGE de *Mtb*, *Mycobacterium smegmatis* y *Mycobacterium avium* no fueron exitosos debido a la inestabilidad de la proteína. La FGE de *Strep* se dializó en Tris 10 mM pH 7.5, NaCl 150 mM, y TCEP 1 mM. Los cristales de la FGE de *Strep* etiquetada con His<sub>6</sub> se obtuvieron con el uso de la difusión en vapor mediante la mezcla de 1 μl de proteína dializada con 10 μl de solución de cristalización (Tris 100 mM pH 8.0, formiato amónico 2.4 M, β-octilglucósido 0.3%, 2-butanol 3.2%) a temperatura ambiente (RT). Los cristales crecieron por un período de dos semanas y se transfirieron subsecuentemente a un crioprotector que consiste de solución de cristalización con glicerol 20%.

35 **Determinación de la estructura de la FGE de Strep.** Los datos se recolectaron en el beamline 8.2.2 en Advanced Light Source con el uso de un detector CCD ADSC Quantum-Q315. Los datos de difracción se procesaron con el uso de HKL2000 (Otwinowski y otros (1997) Methods Enzymol: Macromol Crystallogr Parte A 276, 307-326). Las fases iniciales se determinaron mediante el reemplazo molecular con el uso de la FGE humana (PDB entrada 1Y1E) como un modelo de búsqueda en PHASER (Storoni y otros (2004) Acta Crystallogr D Biol. Crystallogr 60, 432-8). La unidad asimétrica contuvo cinco monómeros de la FGE de *Strep* en el grupo espacial P3<sub>1</sub>2<sub>1</sub>. Las etapas iniciales del refinamiento del modelo incluyeron el refinamiento de los ciclos de hibridación simulada con las dinámicas de ángulo de torsión y el factor B restringido con el uso de CNS (Brunger y otros (1998) Acta Crystallogr D Biol Crystallogr 54, 905-21), seguido por la reconstrucción manual del modelo con el uso de O (Jones y otros (1991) Acta Crystallogr A 47 (Pt 2), 110-9). Los ciclos finales del refinamiento se llevaron a cabo con restricciones de TLS (Winn y otros (2001) Acta Crystallogr D Biol Crystallogr 57, 122-33) como se implementa en REFMAC5 (Murshudov (1997) Acta Crystallogr D Biol. Crystallogr 53, 240-55) con el uso de 5 grupos TLS (que corresponde a cada monómero de FGE en la unidad asimétrica). Las moléculas de agua se añadieron con ARP/WARP (Lamzin y otros (1993) Acta Crystallogr D Biol. Crystallogr 49, 129-47). El modelo final contenía los residuos 18-305 en el monómero A, los residuos 19-306 en el monómero B, los residuos 20-306 en el monómero C, los residuos 19-305 en el monómero D, y los residuos 19-307 en el monómero E. Los valores

finales  $R_{\text{trabajo}}$  y  $R_{\text{libre}}$  fueron 19.5% y 23.3%, respectivamente. La recolección y el procesamiento estadístico de los datos se resumen en la tabla más abajo. Todas las figuras se generaron con PyMOL (www.pymol.org).

<b>Recolección de datos</b>	
<b>Resolución (Å)<sup>a</sup></b>	20-2.1 (2.1-2.17)
<b>Longitud de onda (eV)</b>	12,398.4
<b>Grupo espacial</b>	P3121
<b>Dimensiones de la célula unitaria (a=b, c) (Å)</b>	142.444, 217.067
<b>Reflexiones medidas</b>	123276
<b>Completamiento (%)</b>	83.4 (88.2)
<b>Redundancia</b>	2.6 (2.6)
<b>Mosaicidad (°)</b>	0.32
<b><math>I/\sigma</math></b>	15.8 (1.9)
<b>Rsym (%)<sup>b</sup></b>	5.7 (23.3)
<b>Refinamiento</b>	
<b>Rtrabajo (%)<sup>c</sup></b>	19.5
<b>Rlibre (%)<sup>c</sup></b>	23.3
<b>Cantidad de residuos/aguas</b>	1438/1017
<b>Enlaces Rms (Å) / ángulos (°)</b>	0.008/1.062
<b>Diagrama de Ramachandran (%)<sup>d</sup></b>	87.9/11.2/0.5/0.6e
<b>Valores promedio de B</b>	41.5

<sup>a</sup>Los valores en paréntesis corresponden a la bin de superior resolución.

<sup>b</sup> $R_{\text{sym}} = 100 \cdot \frac{\sum_h \sum_l |I_i(h) - \langle I(h) \rangle|}{\sum_h \sum_l I_i(h)}$ , donde  $I_i(h)$  es la  $i_{\text{th}}$  medición de la reflexión  $h$  e  $\langle I(h) \rangle$  es el valor promedio de la intensidad de la reflexión.

<sup>c</sup> $R_{\text{trabajo}} = 100 \cdot \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$ , donde  $F_{\text{obs}}$  y  $F_{\text{calc}}$  son las amplitudes del factor de la estructura de los datos y el modelo, respectivamente  $R_{\text{libre}}$  es la  $R_{\text{trabajo}}$  con el 5% de las reflexiones ajustadas en todo el refinamiento.

<sup>d</sup>Los números corresponden al porcentaje de residuos de aminoácidos en las regiones favorecidas, permitidas, generosamente permitidas y no permitidas, respectivamente. Calculado con el uso de PROCHECK<sub>39</sub>.

<sup>e</sup>Se observaron siete residuos en conformaciones estereoquímicamente tensas ya sea debido a los contactos de cristales empaquetados (Tyr219 en monómeros A y C) o interacciones por uniones de hidrógeno (Asn232 en monómeros A-E).

5

**Ensayo de actividad de la FGE.** La FGE silvestre y mutante de *Mtb* y *Strep* se purificaron como se describe anteriormente. El sustrato peptídico se sintetizó mediante los métodos de síntesis estándar en fase sólida Fmoc y consistió de la secuencia de 13 residuos LCSPSRGSLFTGR, un motivo consenso sulfatasa. El N-terminal se acetiló, el C-terminal se amidó y la secuencia se confirmó mediante espectrometría de masa. Las condiciones del ensayo fueron similares a aquellas reportadas previamente por Dierks y otros en los estudios de la FGE humana (Dierks y otros (2003) Cell 113, 435-4). Los experimentos anaeróbicos se realizaron de la misma manera excepto las soluciones que se hicieron anaeróbicas con el uso de un distribuidor de gas de oxígeno depurado y las reacciones se iniciaron mediante la mezcla de la enzima con el sustrato en una caja de guantes anaeróbica. El EDTA se añadió a las reacciones adecuadas a una concentración de 100 mM. La confirmación de la formación de FGly se realizó mediante la incubación de 1  $\mu$ l de producto desalado con 1  $\mu$ l de biotina hidrazida 5 mM (Sigma) por 30 min a temperatura ambiente. Las muestras se mezclaron 1:1 (v/v) con una solución matriz (10 mg/ml de ácido  $\alpha$ -ciano-4-hidroxi-cinámico con citrato de amonio 2 mM) y se analizaron mediante espectrometría de masa de tiempo de vuelo de desorción/ionización láser asistida por matriz (Applied Biosystems Voyager DE Pro).

10

15

20

**Detección de metal.** Una solución estándar multielemento se preparó mediante la dilución adecuada de estándares ICP de Ca, Cu, Fe, Mn, Mg, y Zn (Sigma). El contenido metálico de la FGE de *Mtb* y *Strep* se analizó mediante ICP-AES con el uso de un Optima 3000 DV de Perkin Elmer. La ausencia de Fe, Cu y Zn en la FGE de *Strep* se confirmó en el beamline 8.3.1 en Advanced Light Source. Los bordes de absorción de estos metales se examinaron con el uso de un monocromador de cristal doble y un detector de fluorescencia de rayos X del beamline.

25

**Producción de la cepa de *Mtb* deficiente de la FGE.** Una deleción en el marco genético, sin marcar del marco de lectura abierto de Rv0712 que codifica la FGE se creó en el H37Rv de *Mtb* con el uso de un reemplazo alélico (Parish y otros (2000) Microbiology 146 (Pt 8), 1969-75.). Una región corriente arriba de 2 kb de Rv0712 se amplificó e insertó en el vector de suministro micobacteriano p2NILX entre los sitios de restricción HindIII y XbaI. El p2NILX se deriva de pNIL

(Parish y otros (2000) *Microbiology* 146 (Pt 8), 1969-1975) y se modifica con la adición de un sitio de restricción XbaI entre los sitios de restricción KpnI y NotI. Una región corriente abajo de 2 kb de Rv0712 se amplificó e insertó en p2NILX entre los sitios de restricción XbaI y PaeI. Los marcadores de selección lacZ y sacB se digirieron a partir del pGOAL17 y se ligaron en p2NILX con el uso del sitio de restricción PaeI. El vector de suministro completado se trató con luz UV (120 mJ cm<sup>-2</sup>) y se electroporó en el H37Rv electrocompetente de *Mtb* como se describe previamente (Hatfull, G. F. Jacobs, W. R. J. (eds). *Molecular Genetics of Mycobacteria* (ASM Press, Washington, D. C, 2000)). La selección del mutante se realizó como se describe previamente (Parish y otros (2000) *Microbiology* 146 (Pt 8), 1969-1975), y el genotipo se confirmó mediante análisis Southern (Figura 9). La cepa complementada se produjo mediante la transformación de la cepa  $\Delta fge$  con el vector integrador pMV306.kan que contenía el marco de lectura abierto de Rv0712 bajo el control del promotor de la glutamina sintasa.

**Ensayo sulfatasa/fosfatasa.** Las cepas H37Rv de *Mtb* se crecieron en medio 7H9 suplementado con ADC (Becton Dickinson) a 37 °C hasta que la OD<sub>600</sub> = 1.0. Las células se lisaron mediante rotura mecánica con el uso de perlas de zirconia de 0.1 mm (FastPrep, MP Biomedicals) y el lisado crudo se aclaró mediante centrifugación y se filtró a través de una membrana de 0.22  $\mu$ m. Las muestras de lisado aclarado se normalizaron para la concentración de proteína total (kit de ensayo de proteína Biorad AC/DC) y se añadieron 50  $\mu$ g de lisado de proteína al amortiguador (Tris 50 mM pH 7.5, NaCl 500 mM, MgCl<sub>2</sub> 100  $\mu$ M, MnCl<sub>2</sub> 100  $\mu$ M, CaCl<sub>2</sub> 100  $\mu$ M), inhibidores de proteasas (coctel inhibidor de proteasa Set III, EMD Bioscience), y 4-metilumbeliferil sulfato (4MUS) 8 mM. La sulfatasa de lapa (Sigma) se usó como control positivo a una concentración final de 1  $\mu$ g/ml. Las reacciones se incubaron a 37 °C por 3 h y se pararon mediante la adición de 4 volúmenes de Na<sub>2</sub>CO<sub>3</sub>/NaHCO<sub>3</sub> 0.5 M pH 10.5. La actividad sulfatasa se midió con el uso de un fluorímetro (Gemini XL, Molecular Devices) con el uso de longitudes de onda de excitación y emisión de 360 nm y 460 nm, respectivamente. Los inhibidores de sulfatasa/fosfatasa se usaron según las instrucciones del fabricante y se incluyeron microcistina, cantaridina, *p*-bromotetramisol, vanadato sódico, molibdato sódico, tartrato sódico, e imidazol (coctel inhibidor de la fosfatasa 1 y 2, Sigma). La actividad sulfatasa de Rv2407, Rv3406 y Rv3762c recombinantes se determinó con el uso de las mismas condiciones mencionadas anteriormente, con la adición de  $\alpha$ -cetoglutarato 1 mM, ascorbato 200  $\mu$ M y FeCl<sub>2</sub> 100  $\mu$ M al amortiguador. La actividad fosfatasa se monitoreó como se describe anteriormente, excepto con la sustitución de 4-metilumbeliferil fosfato por 4MUS.

**Marcaje con NBD.** La FGE de *Strep* etiquetada con His<sub>6</sub> se trató con 1:50 (p/p) de la proteasa de TEV para eliminar la etiqueta de His<sub>6</sub> del N-terminal antes del marcaje con NBD y el análisis por espectrometría de masa. La FGE de *Strep* (45  $\mu$ M) se incubó en amortiguador (fosfato potásico 25 mM pH 7.0, NaCl 150 mM) con 4-cloro-7-nitrobenz-2-oxa-1,3-diazol (NBD-Cl, Invitrogen) 1 mM por 30 min a temperatura ambiente (Ellis y otros (1997) *Biochemistry* 36, 15013-8). La muestra se desaló mediante cromatografía de fase reversa C<sub>18</sub> y los aductos de proteína con NBD se detectaron mediante espectrometría de masa (Bruker/Agilent Esquire). La correlación de aductos con NBD se realizó mediante la digestión de la FGE de *Strep* que reaccionó con NBD con tripsina 1:50 (p/p), la desalinización mediante cromatografía de fase reversa C<sub>18</sub> y el análisis de los fragmentos peptídicos resultantes mediante espectrometría de masa de resonancia ciclónica de iones con transformada de Fourier con ionización por electronebulización (Bruker 9.4T Apex III).

#### 40 Ejemplo 13. Identificación y clonación de un fge de m. Tuberculosis

El marco de lectura abierto de Rv0712 de *Mycobacterium tuberculosis* (*Mtb*) H37Rv se identificó mediante el análisis BLAST (Altschul, y otros (1997) *Nucleic Acids Res.* 25, 3389-402) para ser idéntico más del 30% a la FGE SUMF1 humana (Cosma y otros (2003) *Cell* 113, 445-56 (2003); Dierks y otros (2003) *Cell* 113, 435-44). El Rv0712 recombinante fue capaz de modificar un péptido sintético que contenía el motivo sulfatasa como se determinó mediante espectrometría de masa (Figura 8, panel a). La presencia de la FGly dentro del sustrato se confirmó mediante el tratamiento del péptido modificado con biotina hidrazida, que formó un aducto covalente con el péptido a través de la formación de hidrazona (figura 8, panel b). Estos datos juntos implican al Rv0712 como la FGE de *Mtb*.

50 Similar al genoma humano, el genoma de *Mtb* parece codificar sólo una copia funcional de la FGE. Por lo tanto, se esperó que la interrupción de Rv0712 en *Mtb* produjera una cepa deficiente de sulfatasa. El Rv0712 se interrumpió en el H37Rv de *Mtb* con el uso de la recombinación homóloga y se confirmó mediante análisis Southern (Figura 9). El  $\Delta fge$  de *Mtb* fue viable y demostró defectos de crecimiento no evidentes *in vitro*.

55 La actividad sulfatasa de la cepa  $\Delta fge$  se comparó con la silvestre (WT) H37Rv y la  $\Delta fge$  mutante en la que la expresión de la FGE se restauró por complementación. Los lisados crudos se generaron a partir de estas tres cepas de *Mtb* y la actividad sulfatasa total se determinó con el uso del sustrato general 4-metilumbeliferil sulfato (4MUS). La cepa de  $\Delta fge$  exhibió una pérdida sustancial, pero sorprendentemente incompleta de la actividad sulfatasa (Figura 8, panel c). Si bien pudo ser posible que la actividad sulfatasa residual resultara de las fosfatasas que actúan sobre 4MUS, la  $\Delta fge$  no se afectó cuando se monitoreó la actividad sulfatasa en presencia de un coctel de inhibidores de sulfatasa/fosfatasa de amplio espectro. Claramente, la actividad en lisados de WT y  $\Delta fge$  complementada se redujo en aproximadamente un 40% en presencia del coctel de inhibidores, igualando la actividad sulfatasa de  $\Delta fge$  en ausencia de inhibidores (Figura 8, panel c). Debido a que se conoce que los inhibidores aplicados inhiben las sulfatasas activadas por FGE (Stankiewicz y otros (1988) *Biochemistry* 27, 206-12), estos datos sugieren que *Mtb* posee sulfatasas independientes a FGE.

65

Para verificar además que las fosfatasas promiscuas no fueron responsables de la actividad sulfatasa residual, se monitoreó la actividad fosfatasa de los lisados crudos de cada cepa con el uso de 4-metilumbeliferil fosfato. Las tres cepas exhibieron el mismo nivel de actividad fosfatasa en ausencia de inhibidores, pero la actividad se suprimió en todas las cepas en presencia de los inhibidores (Figura 8, panel d). Estos datos indican además que las fosfatasas no son responsables de la actividad de hidrólisis residual de 4MUS observada en la cepa  $\Delta fge$  y que las sulfatasas activadas por FGE son responsables de aproximadamente el 40% de la actividad sulfatasa total del lisado de *Mtb*.

El genoma de *Mtb* se investigó para las fuentes potenciales de la actividad sulfatasa independiente de FGE. La mayoría de las sulfatasas procariotas conocidas o supuestas son homólogas a las sulfatasas eucariotas y contienen el motivo sulfatasa. Sin embargo, algunos procariotas tienen además sulfatasas independientes de FGE que no requieren de FGly y presumiblemente operan a través de diferentes mecanismos. Estas enzimas pueden no ser sensibles a inhibidores de sulfatasa/fosfatasa de amplio espectro. Las sulfatasas independientes de FGE no son homólogas a sulfatasas activadas por la FGE y se han clasificado en una de dos familias de enzimas, las metalo- $\beta$ -lactamasas y las dioxigenasas dependientes de Fe(II)  $\alpha$ -cetoglutarato<sub>18-20</sub>. Basado en la similitud de secuencia con sulfatasas independientes de FGE conocidas de otros procariotas, el *Mtb* tiene al menos tres supuestas sulfatasas independientes de FGE codificada por los marcos abiertos de lectura Rv2407, Rv3406 y Rv3762c. Las formas recombinantes de Rv2407, Rv3406 y Rv3762c se expresaron en *E. coli*, pero las proteínas purificadas no exhibieron actividad en el ensayo 4MUS, lo que indica que estas supuestas sulfatasas probablemente no son responsables de la actividad residual sulfatasa en la  $\Delta fge$  de *Mtb* (Figura 10). Teniendo en cuenta la carencia de similitud de secuencia entre sulfatasas independientes de FGE, *Mtb* puede tener otras sulfatasas no detectables por análisis BLAST.

#### Ejemplo 14: estructura de la FGE de *Mtb*

Con el objetivo de comprender mejor el mecanismo enzimático único y las características de unión al sustrato de las FGE procariotas, se determinó la estructura de la FGE de *Mtb* ortóloga de *Streptomyces coelicolor* (*Strep*) con una resolución de 2.1 Å. La topología general de la FGE bacteriana es sorprendentemente similar a la estructura de la FGE humana recientemente determinada (Dierks, y otros (2005) Cell 121, 541-52). (Figura 11A). Similar a la FGE humana, la FGE de *Strep* tiene inferior contenido de estructura secundaria, que contiene  $\alpha$ -hélice 16% y hoja  $\beta$  12%. Ambos comparten la novedosa "FGE plegada," pero la variante FGE de *Strep* contiene sólo un ión  $Ca^{2+}$  como se determina por espectroscopía de emisión atómica de plasma acoplado inductivamente (ICP-AES) y geometría de coordinación (Figura 12). La variante humana se estabiliza por dos iones  $Ca^{2+}$ ; esta diferencia se debe aparentemente a una sustitución de Glu66Ala en la FGE de *Strep* que interrumpe un medio ambiente adecuado de coordinación (Figura 3, panel b). Los datos de la ICP-AES indican que la FGE de *Mtb* carece de ambos iones  $Ca^{2+}$  (Figura 12), lo que sugiere que la FGE plegada no requiere la estabilización por cationes divalentes.

Los sitios activos de la FGE procariota y humana son sorprendentemente similares. Ambos son de aproximadamente 20 Å de longitud, 12 Å de ancho y 10 Å de profundidad y tienen capacidad para sólo 6 de los 13 aminoácidos que definen el motivo sulfatasa. Teniendo en cuenta que el motivo sulfatasa se extiende hacia el C-terminal del sustrato peptídico por otros ocho residuos más allá de la secuencia consenso núcleo (CXPXR) (Figura 7, panel b), es posible que la FGE haya evolucionado a una región de unión secundaria para ayudar en el reconocimiento del sustrato, similar a otras proteínas tales como trombina y neurotoxina botulínica (Hageman y otros (1974) Arch. Biochem Biophys 164:707-15; Breidenbach y otros (2004) Nature 432:925-9). Claramente, cuando se correlacionan los residuos conservados entre FGE de *Strep*, *Mtb*, humana y otras supuestas en la superficie de la molécula de la FGE de *Strep*, se observa una región de superior conservación donde la sección del C-terminal del motivo puede unirse posiblemente (Figura 11C).

Se cree que la FGE cataliza la oxidación de un tiol a un aldehído con el uso de dos residuos conservados de cisteína dentro de su sitio activo (Dierks y otros (2005) Cell 121, 541-52; Roeser y otros (2006) Proc Natl Acad Sci Estados Unidos 103, 81-6). Estas cisteínas en FGE de *Mtb* (Cys263 y Cys268) y *Strep* (Cys272 y Cys277) se requieren para la renovación del sustrato ya que los mutantes de serina fueron incapaces de generar FGly *in vitro* (Figura 8, panel a, Figura 13, paneles c, h, i). Curiosamente, el estado de oxidación de estos residuos es diferente entre los cinco monómeros dentro de la unidad asimétrica. Los mapas Omit mostraron que Cys272 y Cys277 se comprometieron en disulfuros parciales en tres de los cinco monómeros de FGE de *Strep* dentro de la unidad asimétrica (Figuras 11D y 11E, y datos no mostrados). La confirmación bioquímica de estos disulfuros parciales se proporcionó mediante el tratamiento de las tres cisteínas expuestas al disolvente de la FGE nativa del *Strep* con el reactivo 4-cloro-7-nitrobenz-2-oxa-1,3-diazol (NBD) con marcaje tiol. Se detectaron dos poblaciones distintas correspondientes a la FGE de *Strep* ya sea con 1 o 3 aductos de NBD mediante espectrometría de masa de proteína intacta (Figura 14), que confirma que aproximadamente un tercio de estas dos cisteínas proximales están enlazadas en un enlace disulfuro.

Adicionalmente a las dos cisteínas del sitio activo, la FGE de *Strep* y *Mtb* requiere además de oxígeno molecular para la catálisis ya que se observó la no formación de la FGly en reacciones realizadas en un medio ambiente anaeróbico (Figura 13, panel e, y datos no mostrados). Como un miembro de la familia oxigenasa, podría esperarse que la FGE contenga un metal de transición, tal como Fe o Cu, o un cofactor orgánico, tal como FADH, para la activación de oxígeno molecular. Sin embargo, el análisis por ICP-AES y de escaneo de borde de absorción de rayos X indicó que la FGE activa de *Mtb* y *Strep* carece de todos los metales redox activos (Figura 12 y datos no mostrados). Además, estas FGE no requieren la adición de metales para funcionar en los ensayos *in vitro* y pueden funcionar en presencia de

EDTA (Figura 13, paneles b, g). Similarmente, la espectroscopia de absorción en el UV visible no reveló la presencia de cofactores orgánicos cromóforos (datos no mostrados). Junto con la información de la densidad electrónica para la FGE de *Strep*, estos datos indican que las FGE procariotas, similares a la FGE humana, no usan cofactores exógenos para la catálisis.

5

Como un medio alternativo de activación de oxígeno molecular, la FGE puede funcionar de manera similar a otras oxigenasas sin cofactor y hacer uso único de los residuos<sup>25</sup> convencionales. Los residuos absolutamente conservados dentro de la distancia reactiva del par cisteína catalítica de la FGE de *Strep* incluyen Trp234 y Ser269. Roeser y otros han teorizado que el Trp234 puede funcionar para activar el oxígeno molecular (Roeser y otros (2006) Proc Natl Acad Sci Estados Unidos 103:81-6), similar al mecanismo propuesto de reducción de O<sub>2</sub> de los anticuerpos catalíticos<sup>26</sup>. Sin embargo, la mutación de Trp234 a Phe no suprimió la actividad (Figura 13, panel j), lo que indica que la activación del oxígeno molecular debe alcanzarse por otra vía. La actividad se redujo drásticamente por la mutación equivalente Ser269Ala de la FGE (Ser260Ala) en *Mtb* (Figura 13, panel d), pero actualmente se desconoce cómo este residuo juega un papel en el ciclo catalítico de la FGE.

10

Curiosamente, la Cys272 en sí puede estar incluida en la activación del oxígeno molecular. Todos los cinco residuos Cys272 modelados dentro de la unidad asimétrica de la FGE de *Strep* tienen densidad electrónica de más que se extiende fuera de sus conformaciones de enlace alternativas, no disulfuro. Los mapas Omit indican que esta densidad de más podría modelarse como una molécula de agua o una porción de hidropéroxido con una ocupación parcial (Figura 13, paneles d, e). Los experimentos de marcaje con NBD con el uso de la FGE mutante Cys277Ser de *Strep* indican que Cys272 no es un tiol reactivo en una subpoblación de moléculas de enzima (Figura 15), lo que sugiere que la densidad de más corresponde a una porción unida covalentemente a Cys272, tal como el hidropéroxido. Las estructuras previamente publicadas de la FGE humana han sugerido además que esta densidad de más podría ser hidropéroxido o ácido sulfénico de cisteína combinado con una molécula de agua unida<sup>21</sup>. Sin embargo, este último se ajusta pobremente a la densidad electrónica observada en la FGE de *Strep*. Además, no se detectó ácido sulfénico cuando la FGE de *Strep* se trató con cloruro de NBD (Figura 15 y Figura 14). La presencia de hidropéroxido modelado con ocupación parcial no puede descartarse basándose en la densidad electrónica observada en la FGE de *Strep*. Sin embargo, el análisis por espectrometría de masa de la FGE intacta de *Strep* y *Mtb* reveló ninguna anomalía en la masa (Figura 16 y datos no mostrados), lo que sugiere que si la Cys272 se modifica, la modificación es transitoria o lábil al ácido.

15

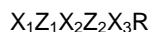
20

25

30

## REIVINDICACIONES

1. Un método de modificación de un polipéptido objetivo, el polipéptido objetivo que comprende una etiqueta heteróloga de aldehído, la etiqueta aldehído que consiste de hasta 12 residuos de aminoácidos y que comprende:



en donde

$Z_1$  es cisteína o serina;

$Z_2$  es un residuo de prolina o alanina;

$X_1$  está presente o ausente y, cuando está presente, es cualquier aminoácido con la condición de que cuando la etiqueta de aldehído está en un N-terminal del polipéptido,  $X_1$  está presente;

$X_2$  y  $X_3$  son cada uno independientemente cualquier aminoácido;

el método que comprende poner en contacto el polipéptido objetivo con una enzima generadora de formilglicina (FGE) de manera que  $Z_1$  se convierte a un residuo de formilglicina (FGly) en el polipéptido y se produce un polipéptido convertido etiquetado con aldehído.

2. Un método como el reivindicado en la reivindicación 1, en donde el polipéptido tiene al menos una de las siguientes propiedades:

(a) la etiqueta de aldehído está posicionada en el N-terminal del polipéptido;

(b) la etiqueta de aldehído está posicionada en un sitio interno de una secuencia de aminoácidos nativa para el polipéptido;

(c) la etiqueta de aldehído está posicionada en un bucle terminal del polipéptido; o

(d) la etiqueta de aldehído está posicionada en un sitio de modificación postraduccional del polipéptido.

3. Un método como el reivindicado en la reivindicación 1, en donde la etiqueta de aldehído está posicionada en un C-terminal del polipéptido.

4. Un método como el reivindicado en la reivindicación 1, en donde

(a) la etiqueta de aldehído está presente en un sitio interno en un bucle terminal del polipéptido; o

(b) la etiqueta de aldehído está presente en un sitio interno o en el N-terminal y es accesible al disolvente cuando el polipéptido se pliega; o

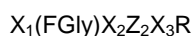
(c) la etiqueta de aldehído está posicionada en un sitio de modificación postraduccional del polipéptido que es un sitio de glicosilación; preferentemente en donde el polipéptido objetivo se diseña para incluir uno o más sitios no nativos de modificación postraduccional, y en donde la etiqueta de aldehído está posicionada en dicho uno o más sitios no nativos de modificación postraduccional.

5. Un método como el reivindicado en la reivindicación 1, en donde el polipéptido es una proteína de membrana y la etiqueta de aldehído está presente en un sitio interno dentro de un bucle extracelular o un bucle intracelular.

6. Un método como el reivindicado en cualquier reivindicación precedente, en donde el polipéptido se expresa en una célula que contiene la FGE.

7. Un método como el reivindicado en cualquier reivindicación precedente, en donde el método comprende además poner en contacto el polipéptido convertido etiquetado con aldehído con una molécula reactiva que comprende una porción de interés; en donde dicho contacto es bajo las condiciones que proporcionan la producción de un producto de reacción de un polipéptido modificado etiquetado con aldehído que tiene la porción de interés unida covalentemente al residuo de FGly del motivo sulfatasa heterólogo.

8. Un método de modificación de un polipéptido objetivo, el polipéptido objetivo que comprende una etiqueta heteróloga convertida de aldehído, la etiqueta de aldehído que consiste de hasta 12 residuos de aminoácidos y que comprende:



en donde

FGly es un residuo de formilglicina;

$Z_2$  es un residuo de prolina o alanina;

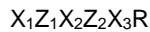


$X_1$  está presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando la etiqueta aldehído es al menos un N- terminal del polipéptido,  $X_1$  está presente; y  $X_2$  y  $X_3$  son cada uno independientemente cualquier aminoácido; el método que comprende poner en contacto el polipéptido objetivo con una molécula reactiva que comprende una porción de interés de manera que la porción de interés se conjuga a la FGly del polipéptido, produciendo de esta manera un polipéptido modificado.

9. Un método como el reivindicado en la reivindicación 7 o la reivindicación 8, en donde la porción de interés es un polímero soluble en agua, un marcador detectable, un fármaco, un péptido, una toxina, una porción para la inmovilización del polipéptido en una membrana o sobre una superficie.

10. Un método como el reivindicado en la reivindicación 8, en donde el motivo sulfatasa comprende L(FGly)TPSR, M(FGly)TPSR, V(FGly)TPSR, L(FGly)SPSR, L(FGly)APSR L(FGly)VPSR, o L(FGly)GPSR.

11. Un método para la producción de un polipéptido modificado, que comprende el cultivo de una célula que comprende un polinucleótido que codifica un polipéptido objetivo, el polipéptido objetivo que comprende una etiqueta heteróloga de aldehído, y un polinucleótido que codifica una enzima generadora de formilglicina (FGE), en donde la etiqueta heteróloga de aldehído consiste de hasta 12 residuos de aminoácidos y comprende:



en donde

$Z_1$  es cisteína o serina;

$Z_2$  es un residuo de prolina o alanina;

$X_1$  está presente o ausente y, cuando está presente, es cualquier aminoácido con la condición de que cuando la etiqueta aldehído está en un N-terminal del polipéptido,  $X_1$  está presente; y,  $X_2$  y  $X_3$  independientemente puede ser cualquier aminoácido;

y en donde dicho cultivo resulta en la expresión del polipéptido y la FGE y la conversión de un residuo de cisteína o serina de la etiqueta heteróloga de aldehído por la FGE a un residuo de 2-formilglicina (FGly) para producir un polipéptido convertido etiquetado con aldehído.

12. Un método como el reivindicado en la reivindicación 11, en donde la célula es una célula de un mamífero.

13. Un método como el reivindicado en la reivindicación 11, en donde la célula es una célula procariota.

14. Un método como el reivindicado en la reivindicación 11 o la reivindicación 12, en donde el método comprende además:

poner en contacto el polipéptido convertido etiquetado con aldehído con una molécula reactiva que comprende una porción de interés; en donde dicho contacto es bajo las condiciones para proporcionar la producción de un producto de reacción de un polipéptido modificado etiquetado con aldehído que tiene la porción de interés unida covalentemente a un residuo de FGly del polipéptido convertido etiquetado con aldehído.

15. Un método como el reivindicado en la reivindicación 11 o la reivindicación 12, en donde el residuo de FGly está posicionado en una secuencia interna del polipéptido.

16. Un método como el reivindicado en la reivindicación 11 o la reivindicación 12, en donde el residuo de FGly está posicionado en un bucle terminal, un C-terminal, o un N-terminal del polipéptido; o en una región del polipéptido accesible al disolvente cuando se pliega.

17. Un método como el reivindicado en la reivindicación 11 o la reivindicación 12, en donde el residuo de FGly está presente en un sitio de modificación postraducciona del polipéptido; preferentemente en donde el sitio de modificación postraducciona es un sitio de glicosilación; o en donde el polipéptido objetivo está diseñado para incluir uno o más sitios no nativos de modificación postraducciona, y en donde la etiqueta de aldehído está posicionada en dichos uno o más sitios no nativos de modificación postraducciona.

18. Un método como el reivindicado en cualquiera de las reivindicaciones 1 a 7 o 11 a 17, en donde la etiqueta de aldehído comprende LCTPSR, MCTPSR, VCTPSR, LCSPSR, LCAPSR LCVPSR, o LCGPSR.

19. Un método como el reivindicado en cualquier reivindicación precedente, en donde  $X_1$ , cuando está presente,  $X_2$ , y  $X_3$  son cada uno independientemente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar, no cargado.
- 5 20. Un método como el reivindicado en cualquiera de las reivindicaciones 1 a 18, en donde  $X_1$ , (cuando está presente) es L, M, V, S o T y/o  $X_2$  y  $X_3$  son cada uno independientemente S, T, A, V, G, o C.
- 10 21. Un método como el reivindicado en cualquier reivindicación precedente, en donde el polipéptido objetivo comprende dos o más etiquetas de aldehído.
- 15 22. Un método como el reivindicado en cualquiera de las reivindicaciones precedentes, en donde el polipéptido es un anticuerpo.
- 20 23. Un método como el reivindicado en cualquiera de las reivindicaciones 1 a 22, en donde la etiqueta de aldehído consiste de hasta 11 residuos de aminoácidos.
- 25 24. Un método como el reivindicado en cualquiera de las reivindicaciones 1 a 22, en donde la etiqueta de aldehído consiste de hasta 10 residuos de aminoácidos.
- 25 25. Un método como el reivindicado en cualquiera de las reivindicaciones 1 a 22, en donde la etiqueta de aldehído consiste de hasta 9 residuos de aminoácidos.
- 25 26. Un método como el reivindicado en cualquiera de las reivindicaciones 1 a 22, en donde la etiqueta de aldehído consiste de hasta 8 residuos de aminoácidos.

FIG. 1A

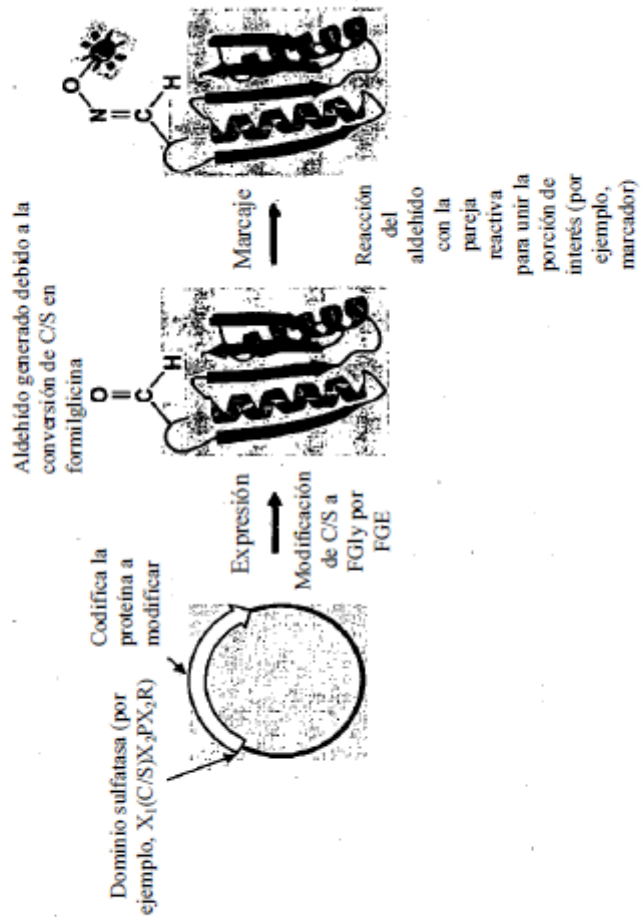




FIG. 3

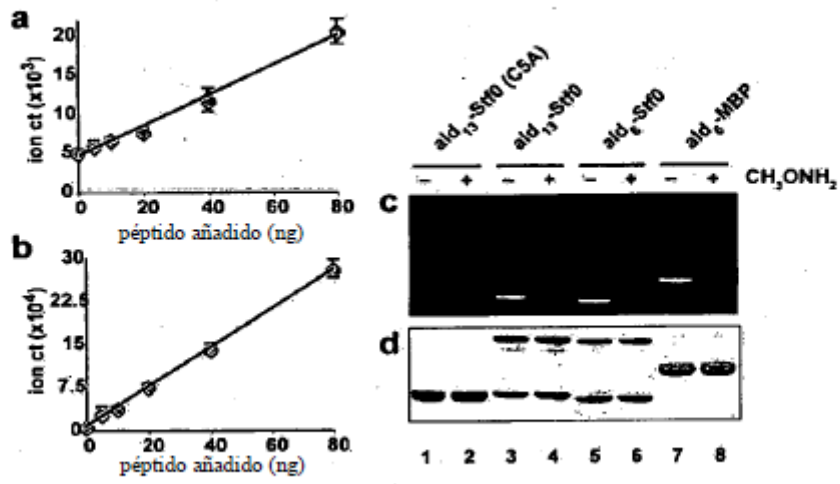


FIG. 4

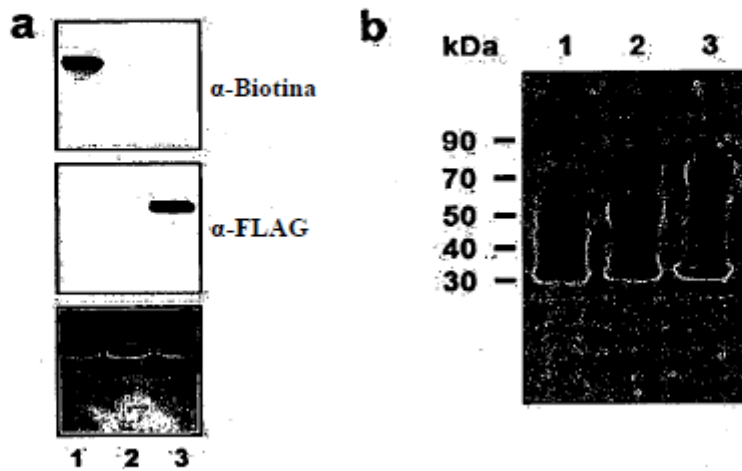


FIG. 5

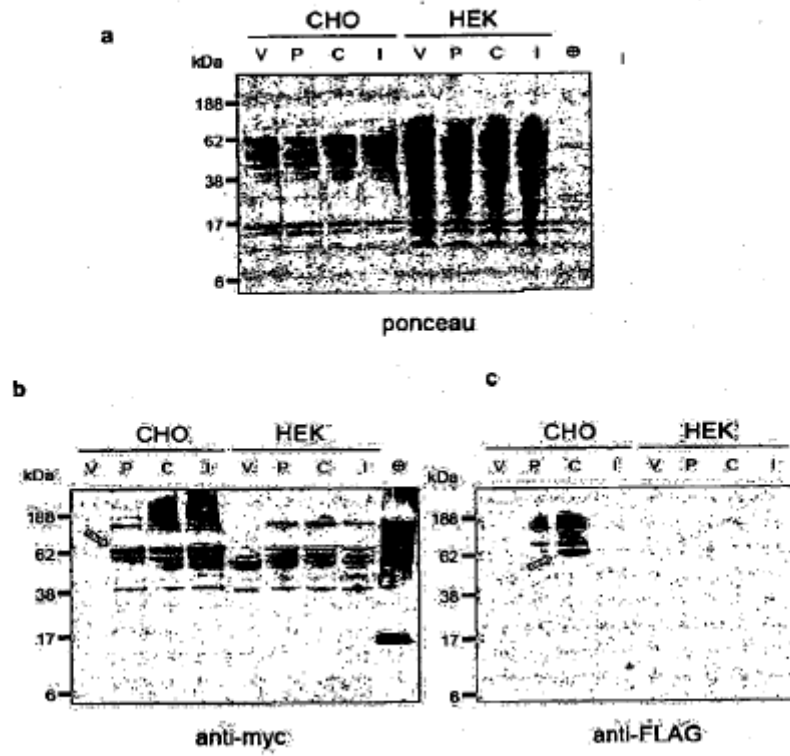


FIG. 6

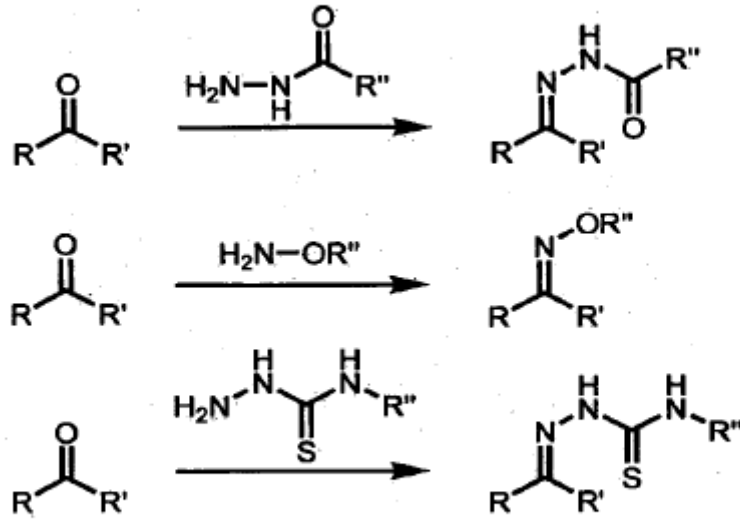
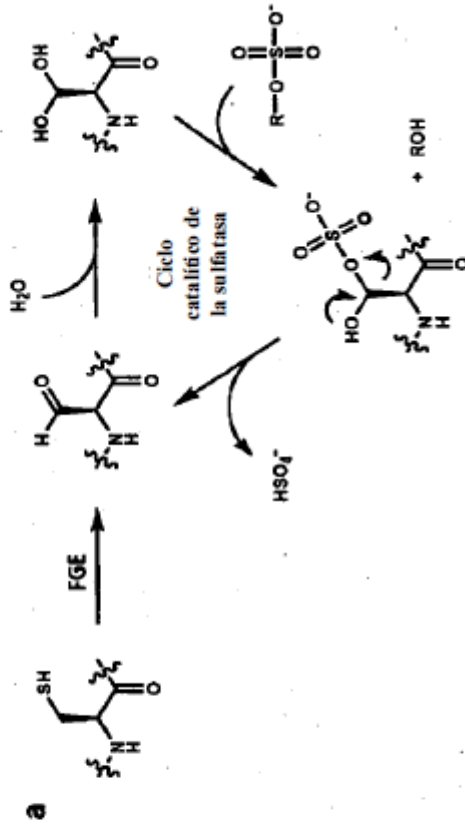


FIG. 7



- \*
- H. sapiens* Arylsulfatase E (82) AA<sup>1</sup>SI<sup>1</sup>GF<sup>1</sup>GG<sup>1</sup>RAA<sup>1</sup>FL<sup>1</sup>FG<sup>1</sup>IPV
  - M. musculus* Sul12 (84) TTPMCCPSMSIIIFGKIYH
  - C. elegans* Sul3 (75) YNQLCIPINSARFMTGTFP
  - M. tuberculosis* Ats6 (54) TAPLIP<sup>1</sup>IP<sup>1</sup>NG<sup>1</sup>IF<sup>1</sup>FG<sup>1</sup>IFPQ
  - E. coli* YidJ (48) CSPV<sup>1</sup>IP<sup>1</sup>FR<sup>1</sup>GI<sup>1</sup>FT<sup>1</sup>GI<sup>1</sup>YAN
  - S. coelicolor* SC07547 (84) FTAL<sup>1</sup>CP<sup>1</sup>PA<sup>1</sup>NS<sup>1</sup>LI<sup>1</sup>FG<sup>1</sup>QAPF
  - M. barbari* A3073 (80) TPTAL<sup>1</sup>SP<sup>1</sup>ERS<sup>1</sup>CI<sup>1</sup>AC<sup>1</sup>GRNHE



FIG. 8

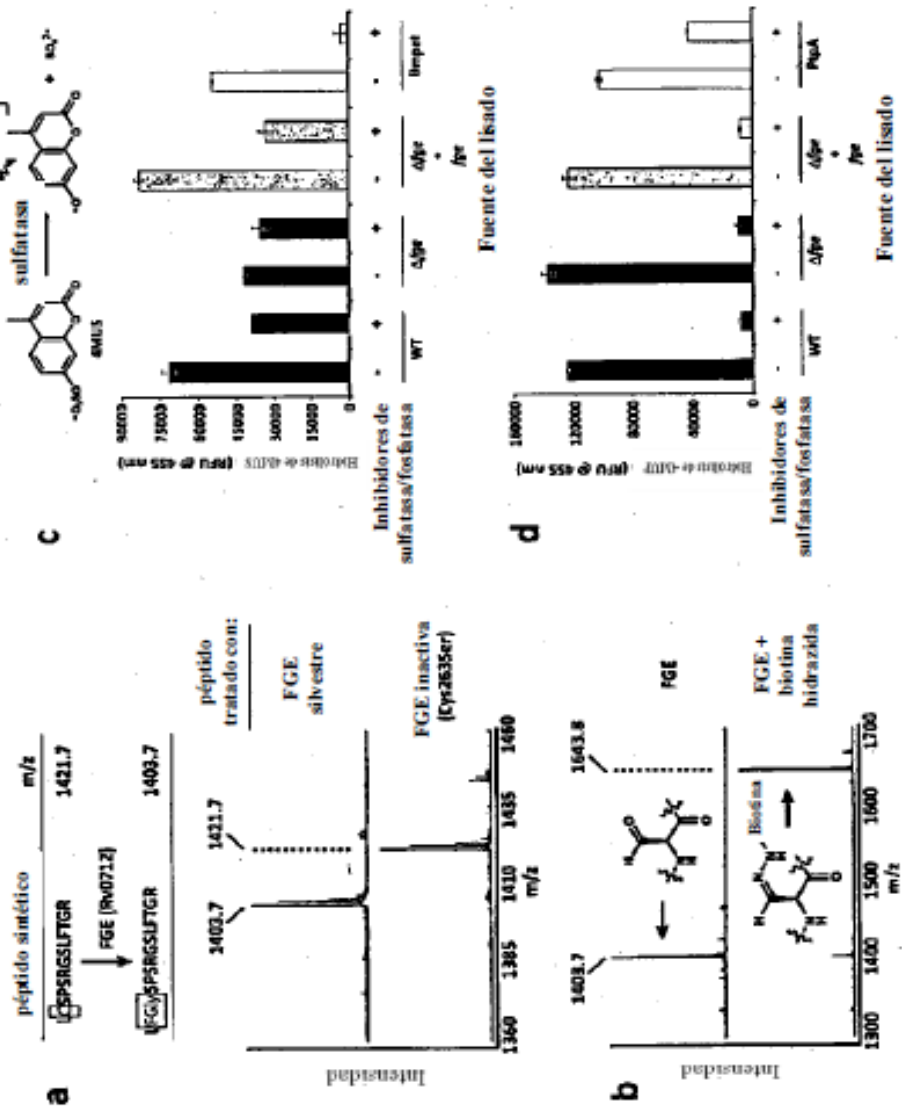


FIG. 9

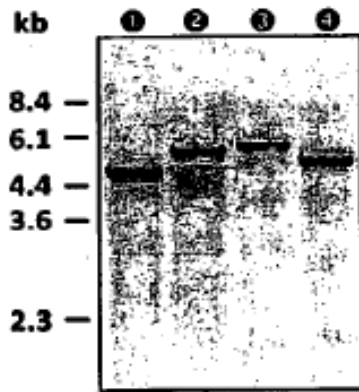
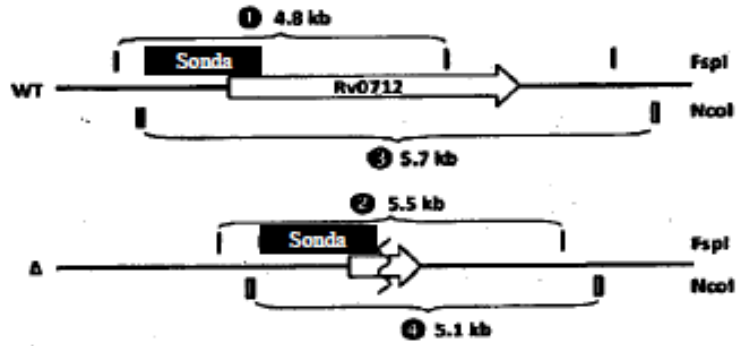


FIG. 10

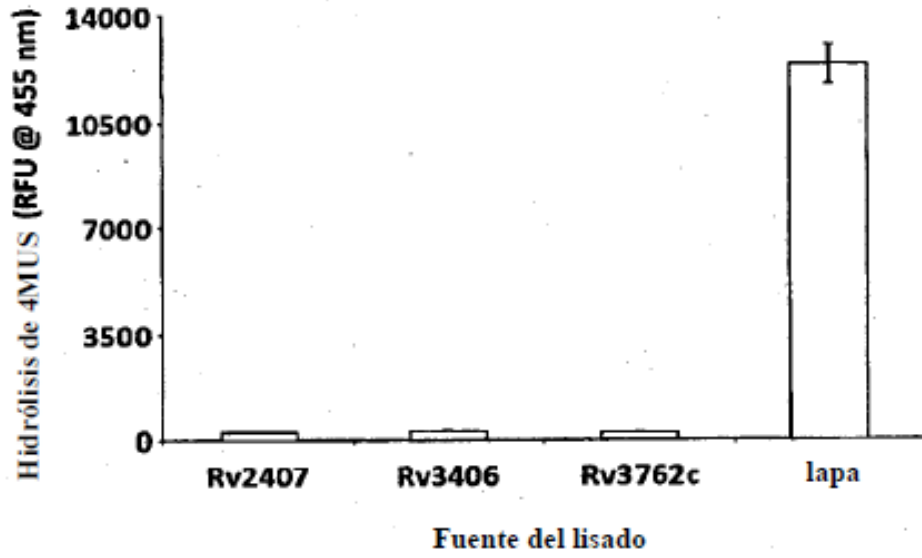


FIG. 11A

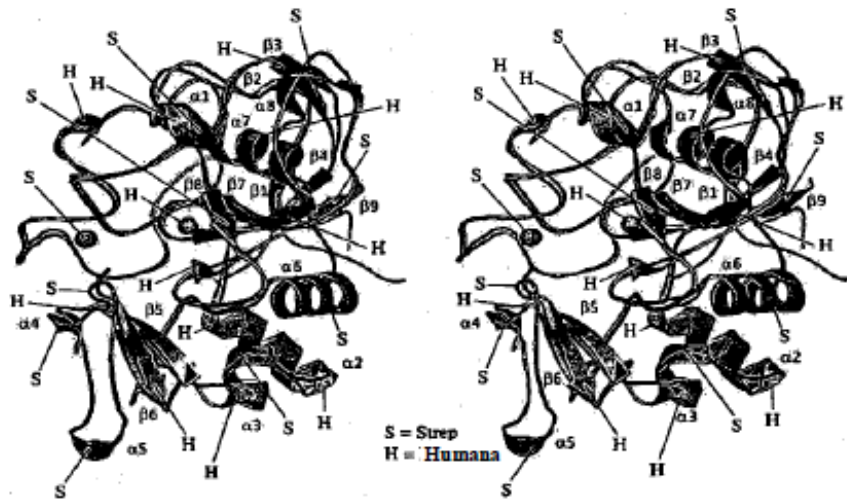


FIG. 11B

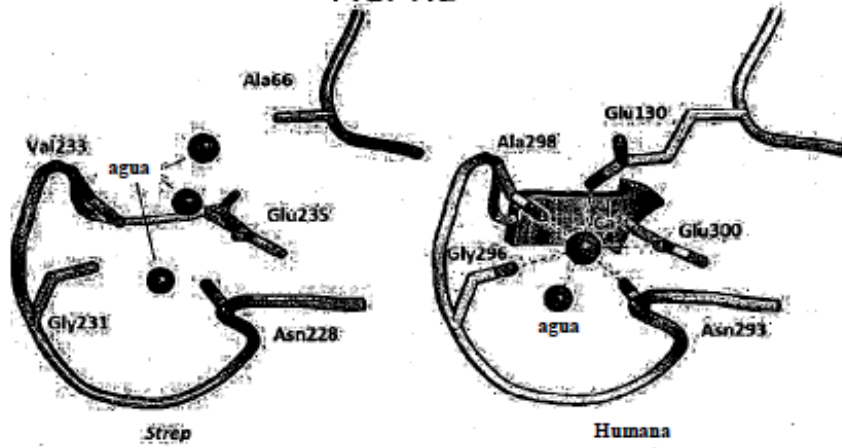


FIG. 11C

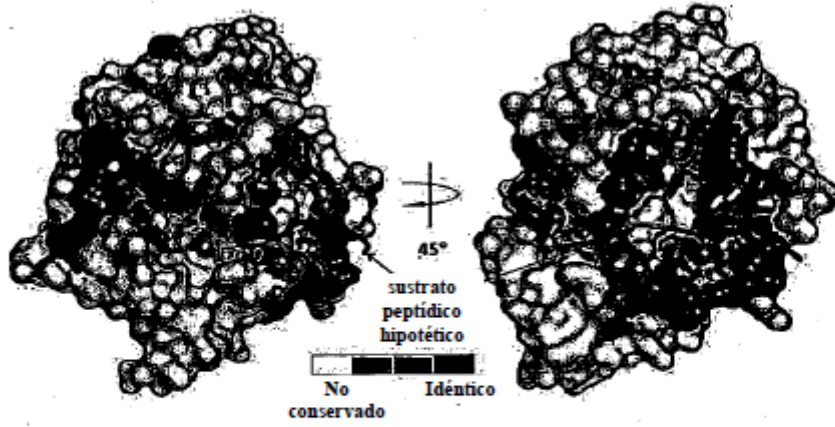


FIG. 11D

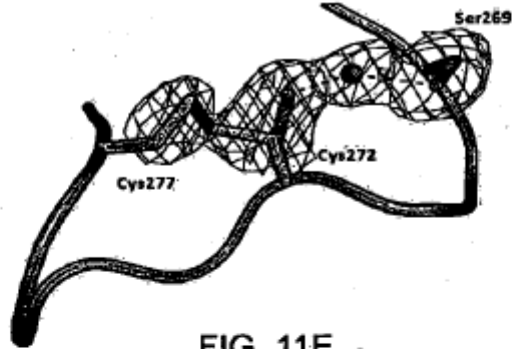


FIG. 11E

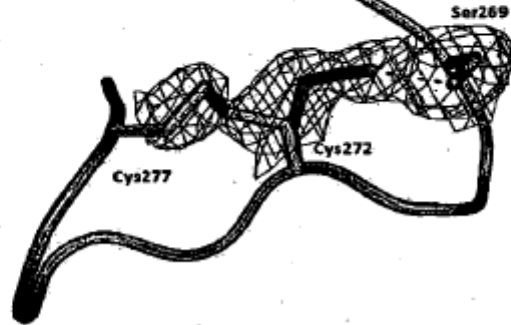


FIG. 12

	Concentración (ppb)	Desviación estándar (ppb)	Molaridad ( $\mu\text{M}$ )	Relación Metal/FGE
FGE <sup>a</sup> de Strep	382,522	6,317	10.19	--
<b>Ca</b>	496.56	16.69	12.39	1.22 $\pm$ 0.04
<b>Cu</b>	3.81	1.87	0.06	0
<b>Fe</b>	0.52	1.60	0.01	0
<b>Mg</b>	-1.34	0.34	-0.06	0
<b>Mn</b>	-1.39	1.04	-0.03	0
<b>Zn</b>	0.97	0.43	0.01	0
FGE <sup>b</sup> de Mtb	214,071	6,843	6.54	--
<b>Ca</b>	37.58	39.91	0.93	0.14 $\pm$ 0.15
<b>Cu</b>	7.67	2.93	0.12	0
<b>Fe</b>	2.64	1.70	0.04	0
<b>Mg</b>	-1.30	0.41	-0.05	0
<b>Mn</b>	-0.95	1.16	-0.02	0
<b>Zn</b>	5.93	0.55	0.09	0

<sup>a</sup>Peso molecular de la FGE de Strep marcada con His<sub>6</sub> = 37,540.35

<sup>b</sup>Peso molecular de la FGE de Mtb marcada con His<sub>6</sub> = 35,015.44

FIG. 13

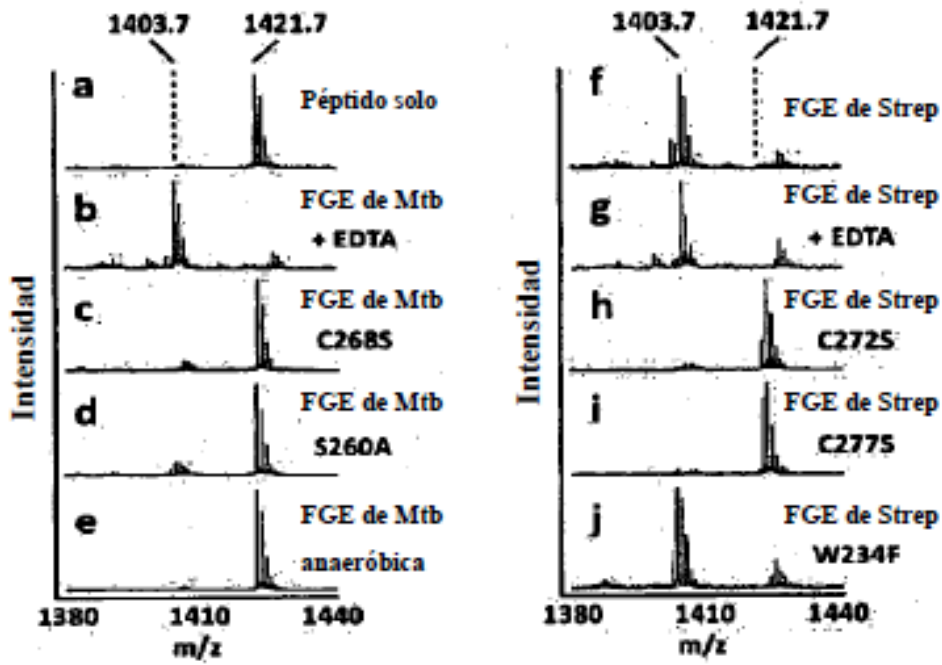


FIG. 14

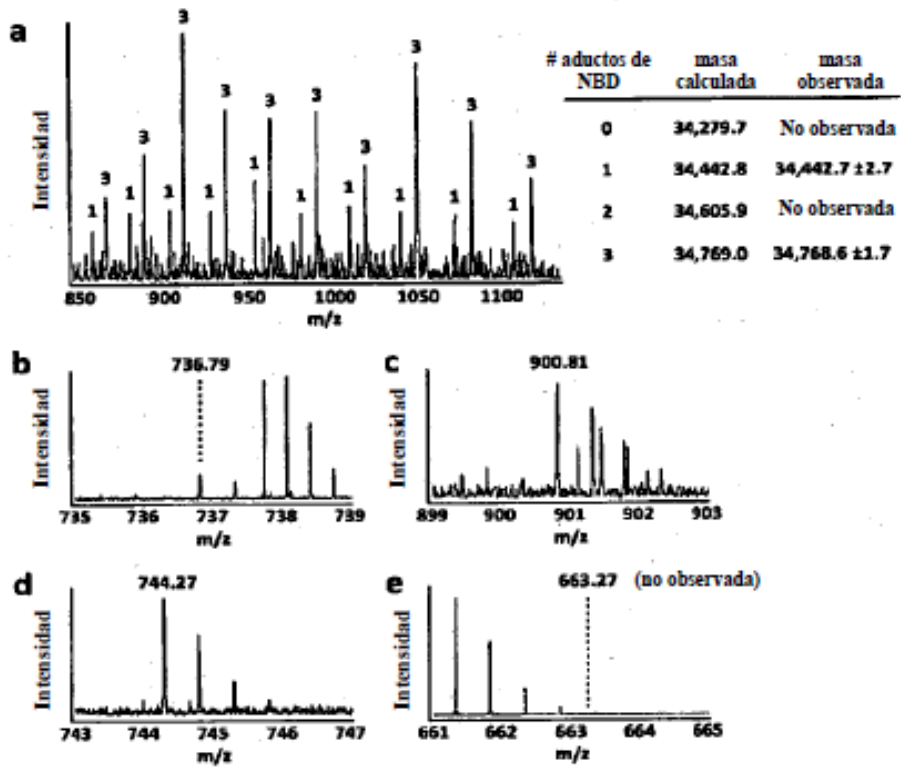




FIG. 15

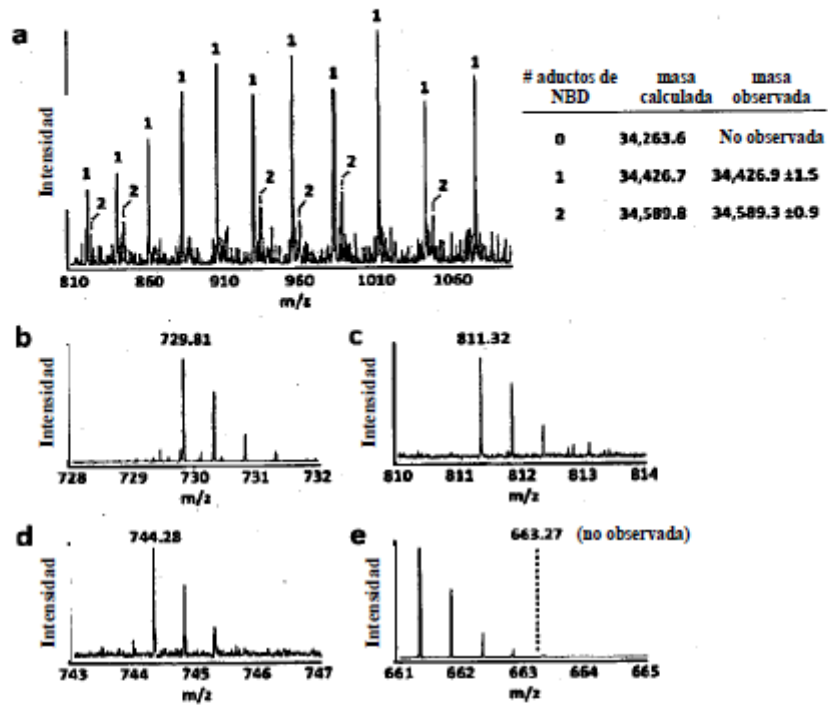


FIG. 16

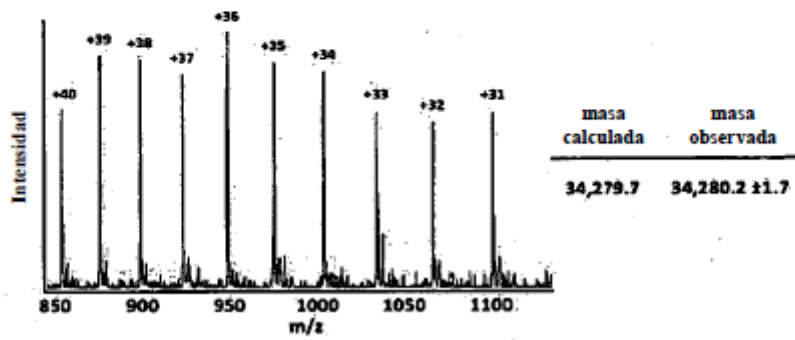


FIG. 17

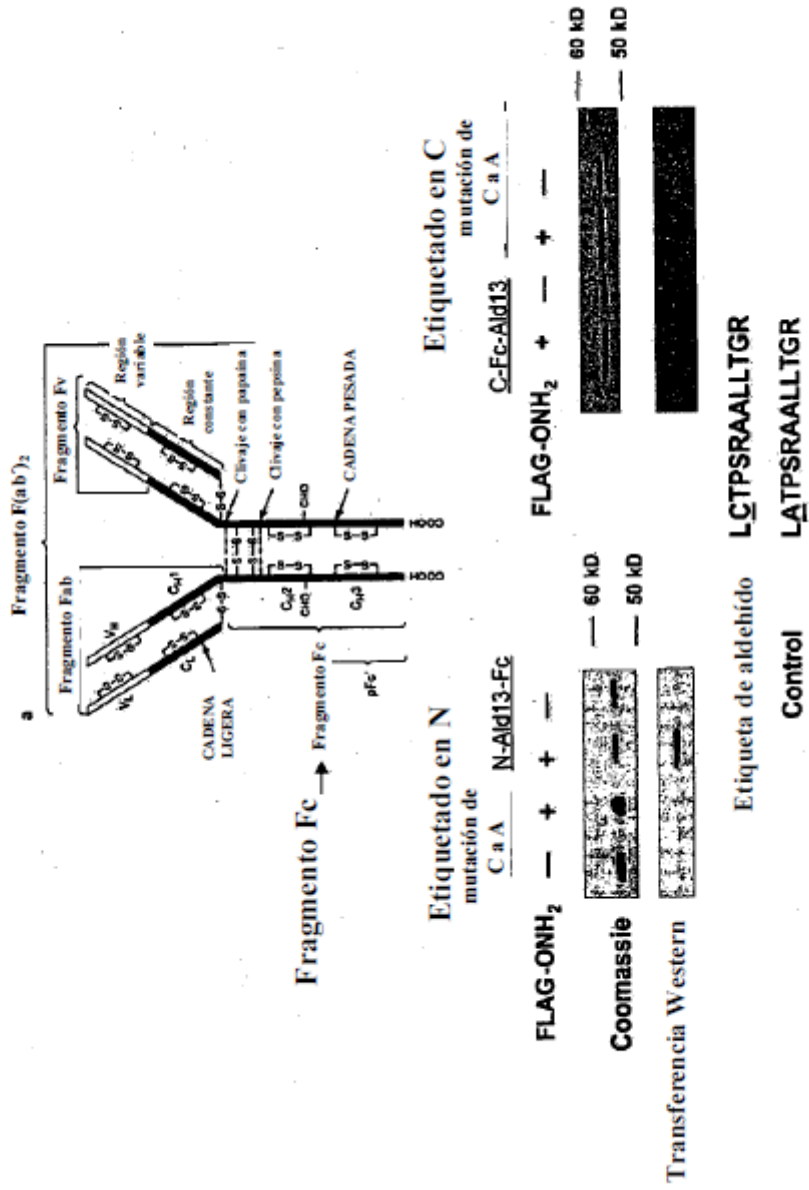
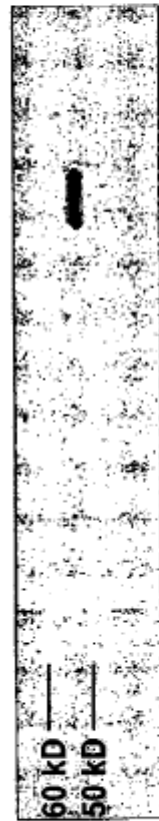


FIG. 18

Etiqueta de 6 mer

FLAG-OH<sub>2</sub>  
- Fc-Ald  
- Fc-C→A  
+ Fc-Ald  
+ Fc-C→A



Fc-Ald    LCTPSR  
Control Fc-C→A    LATPSR

FIG. 19

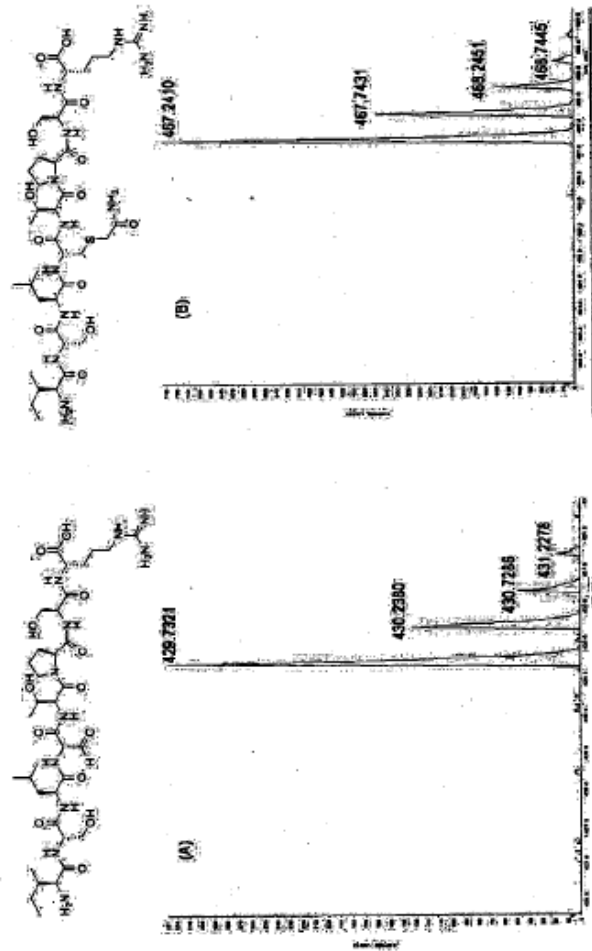


FIG. 20

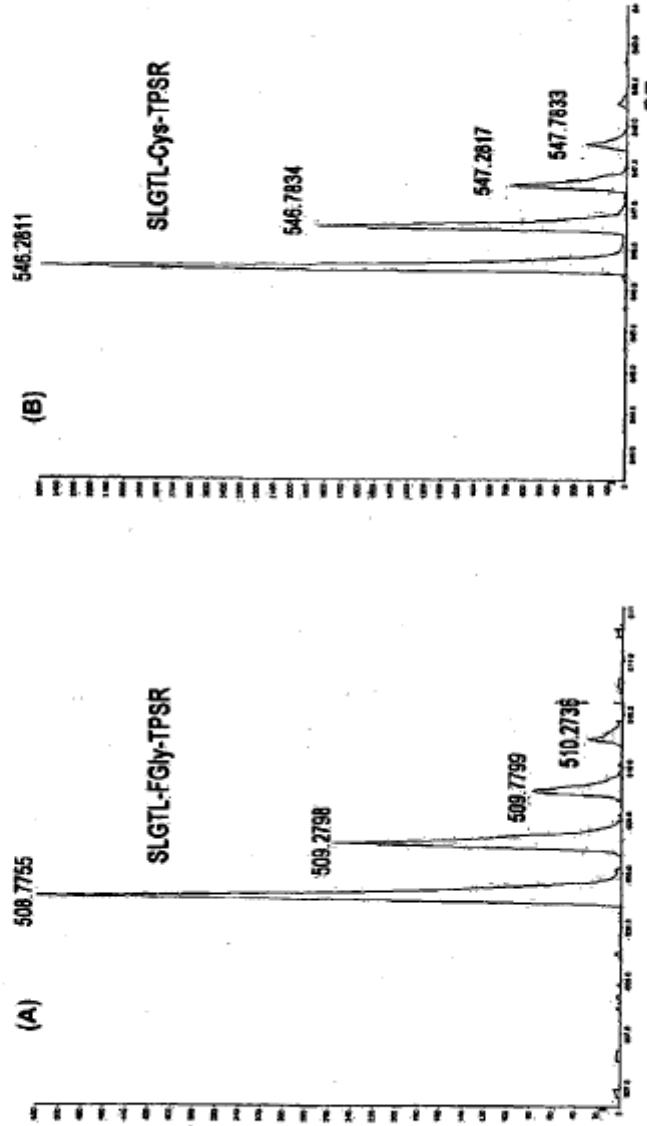
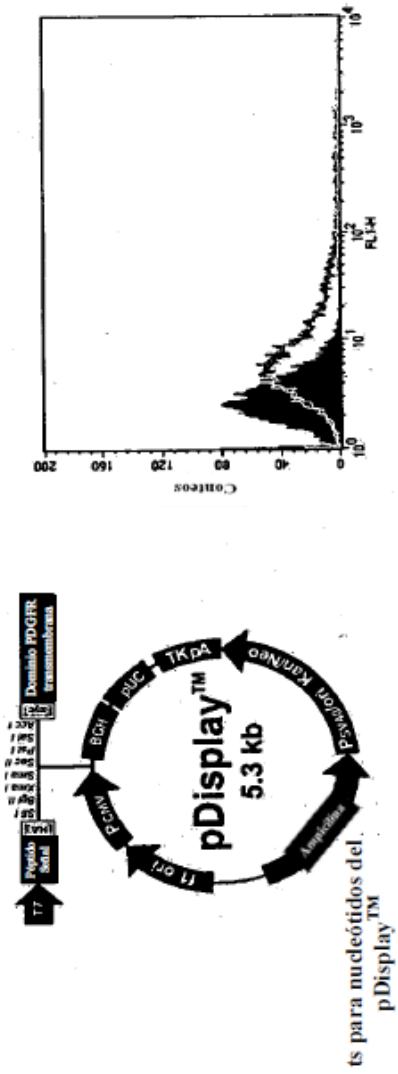


FIG. 21



ts para nucleótidos del pDisplay™

FIG. 22

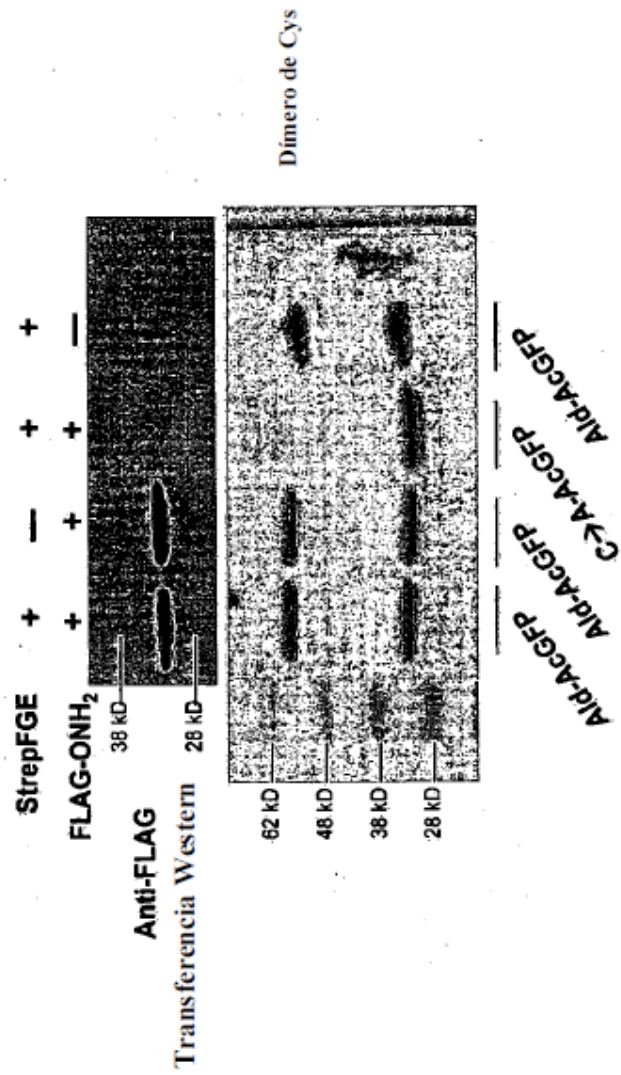
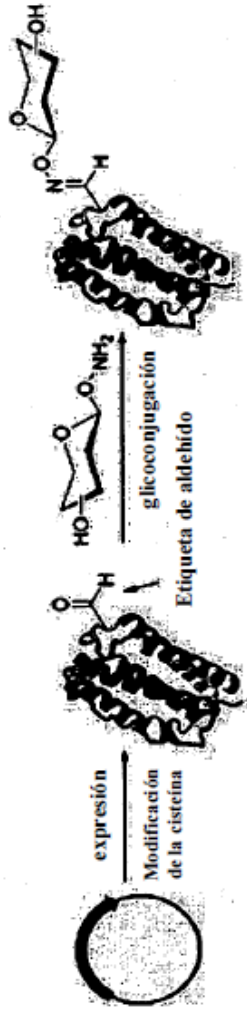


FIG. 23



- IFN-Beta: cinco  $\alpha$ -hélices (A-E) con un único sitio de glicosilación que existe en el residuo Asn-80