

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 612 277**

51 Int. Cl.:

C12P 21/06 (2006.01)

C07K 1/00 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **19.09.2007** **E 12176853 (5)**

97 Fecha y número de publicación de la concesión europea: **28.12.2016** **EP 2548967**

54 Título: **Marcadores de aldehído, usos de los mismos en la modificación de proteínas en un sitio específico**

30 Prioridad:

21.09.2006 US 846644 P

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

16.05.2017

73 Titular/es:

**THE REGENTS OF THE UNIVERSITY OF
CALIFORNIA (100.0%)
1111 Franklin Street, 12th Floor
Oakland, CA 94607, US**

72 Inventor/es:

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

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 612 277 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

Marcadores de aldehído, usos de los mismos en la modificación de proteínas en un sitio específico

Antecedentes

- 5 La marcación en un sitio específico de las proteínas es importante como una herramienta para la disección de redes bioquímicas y celulares. Se ha desarrollado una variedad de tecnologías para abordar esta necesidad. Una de estas tecnologías utilizadas para la localización y seguimiento de proteínas es la marcación con proteínas fluorescentes, tales como la proteína fluorescente verde (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 colaboradores, Traffic 2003, 4, (7), 491-501).
- 10 Como resultado, muchos grupos han enfocado su atención en el uso de fusiones más pequeñas para dirigir reactivos específicos de marcación secundaria. F1AsH, desarrollado por Roger Tsien y colaboradores, utiliza la interacción entre motivos de tetracisteína específicamente dispuestos y biarsenil-fluoróforos (Chen y colaboradores, Science 1998, 281, (5374), 269-272). A pesar de la afinidad picomolar entre la secuencia mínima de 8 aminoácidos y las sondas biarsénicas (Adams y colaboradores, J. Amer. Chem. Soc. 2002, 124, (21), 6063-6076), el antecedente debido a motivos de cisteína nativa sigue siendo un problema (Stroffekova y colaboradores, Pflugers Archiv-Eur. J. Physiol., 2001, 442, (6), 859-866).
- 15 Para aumentar la especificidad, se han explorado los motivos de marcación de péptidos que dependen de la marcación secundaria por enzimas. Una de estas estrategias depende de la fusión con O⁶-alquilguanina-ADN transferasa (hAGT), que puede ligar una amplia variedad de moléculas pequeñas a una cisteína interna. Mientras que las fusiones de hAGT permiten una unión covalente muy específica de una amplia variedad de sondas de moléculas pequeñas, depende de una fusión de 207 aminoácidos (George y colaboradores, J. Amer. Chem. Soc. 2004, 126, (29), 8896-8897; Guignet y colaboradores, Nature Biotechnol 2004, 22, (4), 440-444). En un enfoque separado, las fusiones de proteínas con la proteína portadora de acilo de aproximadamente 80 aminoácidos pueden marcarse específicamente con sondas derivadas de CoA usando la enzima fosfopanteteína transferasa. Alternativamente, la biotina ligasa se ha utilizado 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 de oxima.
- 20 Prescher y colaboradores (2005) Nature Chemical Biology Vol. 1(1) páginas 13-21 se refiere a cómo se puede utilizar la enzima generadora de formilglicina (FGE) para convertir un residuo de cisteína en formilglicina que puede ser posteriormente marcada químicamente con reactivos aminooxi o hidrazida. La modificación de FGE del residuo de cisteína debe producirse en una secuencia de consenso conservada objetivo de 13 residuos.
- 25 Dierks y colaboradores (2003) Cell vol 113, páginas 435-444 describe un ensayo in vitro de FGE. La FGE se purifica a partir de testículos de bovino, se clona y se secuencian la FGE humana. La actividad de FGE en células transfectadas se detecta mediante inmunofluorescencia indirecta y transferencia tipo Western. Se usó un péptido natural de 23 aminoácidos de arilsulfatasa A (ASA) como sustrato para un ensayo de FGE. Se unió otro péptido de 16 aminoácidos a una matriz de cromatografía de afinidad y se usó en purificación por afinidad de FGE.
- 30 Dierks y colaboradores (1999) EMBO J., columna 18, páginas 2081-2091, describe experimentos y un sistema de ensayo para establecer determinantes con base en secuencias responsables de la conversión de cisteína en FGly en sulfatasas eucariotas. Se modificaron y analizaron las secuencias de arilsulfatasa (ASA) de origen natural y se encontró que el "motivo auxiliar" de los residuos de aminoácidos 74-80 de ASA tiene que estar presente.
- 35 Existe la necesidad de nuevos enfoques para la modificación de proteínas en un sitio específico.

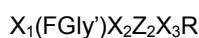
Literatura

- 45 Adams y colaboradores 2002 J. Amer. Chem. Soc. 2002, 124, (21), 6063-6076; Banghart y colaboradores 2004 Nat. Neurosci. 7 (12): 1381-6. Epub 2004 Nov 21; Berteau y colaboradores 2006 J Biol Chem. 281 (32): 22464-70 (Epub 2006 Junio 9); Chen y colaboradores 2005 Nature Methods 2005, 2, (2), 99-104; Cosma y colaboradores 2003 Cell 113, (4), 445-56; Dierks y colaboradores 1997 Proc Natl Acad Sci U S A 94, (22), 11963-8; Dierks y colaboradores 2003 Cell 113, (4), 435-44; Dierks y colaboradores 2005 Cell 121, (4), 541-52; George y colaboradores 2004 J. Amer. Chem. Soc. 126, (29), 8896-8897; Griffin y colaboradores 1998 Science 281, (5374), 269-272; Guignet y colaboradores 2004 Nature Biotechnol. 22, (4), 440-444; Landgrebe y colaboradores 2003 Gene 316: 47-56; Lemieux (1998) Trends Biotechnol 16, 506-13; Lisenbee y colaboradores 2003 Traffic 4, (7), 491-501; Mariappan y colaboradores 2005 J. Biol. Chem. 280 (15): 15173-9 (Epub 2005 Feb 11); Mougous y colaboradores 2004 Nat. Struc. Mol. Biol. 11, 721-729; Preusser y colaboradores 2005 J. Biol. Chem. 280 (15): 14900-10 (Epub 2005, enero 18); Roeser y colaboradores 2006 Proc Natl Acad Sci USA 103 (1): 81-6 (Epub 2005 diciembre 20); Rush y colaboradores (enero 5 de 2006) Org Lett. 8 (1): 131-4; Sardiello y colaboradores 2005 Human Mol. Genet. 14, 3203-3217; Schirmer y colaboradores 1998 Chemistry & Biology

5, R181-R186; Schmidt y colaboradores 1995 Cell 82, (2), 271-8; Stroffekova y colaboradores 2001 Archiv-Europ. J. Physiol. 442, (6), 859-866; Szameit y colaboradores 1999 J Biol Chem. 274, (22), 15375-81; Yin, J. y colaboradores 2005 Proc. Natl. Acad. Sci. USA 102, 15815-15820 (2005); uS20050026234; uS20030186229; y US 6.900.304.

Resumen

- 5 La invención presenta composiciones y métodos para la modificación en un sitio específico de proteínas mediante la incorporación de una etiqueta de aldehído. La modificación enzimática en un motivo de sulfatasa de la etiqueta de aldehído a través de la acción de una enzima generadora de formilglicina (FGE) genera una porción de formilglicina (FGly). La porción aldehído del residuo de FGly puede ser explotado como un mango químico para la fijación en un sitio específico de una porción de interés a un polipéptido.
- 10 Por consiguiente, la presente descripción también proporciona polipéptidos modificados, que comprenden un residuo de 2-formilglicina unido covalentemente a una porción de interés, en donde el polipéptido comprende una etiqueta de aldehído heterólogo modificado, en donde la etiqueta de aldehído consiste de hasta 8 residuos de aminoácidos y comprende:



- 15 donde

FGly' es un residuo de 2-formilglicina que tiene una porción unida covalentemente;

Z₂ es una porción de prolina o alanina;

X₁ está presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando la etiqueta de aldehído heteróloga está en un extremo terminal N del polipéptido, X₁ está presente; y

- 20 X₂ y X₃ son cada uno independientemente cualquier aminoácido.

- La porción de tales polipéptidos modificados puede ser un polímero soluble en agua, un marcador detectable, un fármaco, o una porción para inmovilización del polipéptido en una membrana o en una superficie. La porción modificada de sulfatasa de tales polipéptidos modificados puede posicionarse en el polipéptido modificado en un sitio de modificación postraducciona de un compuesto parental del polipéptido modificado. El sitio de modificación postraducciona puede ser, por ejemplo, un sitio de glicosilación. El sitio de modificación postraducciona puede ser nativo al polipéptido objetivo parental o el polipéptido objetivo puede ser modificado para incluir uno o más sitios no nativos de modificación postraducciona, y en donde el motivo heterólogo de sulfatasa se posiciona en dichos uno o más sitios no nativos de modificación postraducciona.
- 25

La invención incluye polipéptidos como se expone en cualquiera de las reivindicaciones 2 a 15.

- 30 La invención también proporciona polipéptidos modificados como se expone en la reivindicación 16 o la reivindicación 17.

Se describen en la presente invención métodos para modificar un polipéptido, comprendiendo el método poner en contacto un polipéptido que comprende un motivo de sulfatasa convertido con un compañero reactivo que comprende una porción de interés, en donde el motivo de sulfatasa convertido comprende:

- 35
$$X_1(\text{FGly})X_2Z_2X_3R$$

dónde

FGly es un residuo de formilglicina;

Z₂ es un residuo de prolina o alanina;

- 40 X₁ está presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando el motivo heterólogo de sulfatasa está en un extremo terminal N del polipéptido, X₁ está presente; y

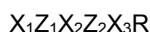
X₂ y X₃ son cada uno independientemente cualquier aminoácido;

En donde dicha puesta en contacto está bajo condiciones suficientes para la conjugación de la porción de interés del compañero reactivo con FGly del polipéptido, produciendo de este modo un polipéptido modificado.

El residuo de FGly puede estar situado en una secuencia interna del polipéptido, y/o posicionado en un bucle terminal, un extremo terminal C, o un extremo terminal N 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 al polipéptido parental, o el polipéptido puede ser modificado para incluir uno o más sitios no nativos de modificación postraducciona, y el motivo heterólogo de sulfatasa está posicionado en dichos uno o más sitios no nativos de modificación postraducciona.

De interés particular son los motivos de sulfatasa en los que X_1 , cuando está presente, es L, M, V, S o T. Otros motivos de sulfatasa de interés particular son aquellos en los que X_2 y X_3 son cada uno independientemente un aminoácido alifático, un aminoácido polar no cargado o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas realizaciones son cada uno independientemente S, T, A, V, G o C.

También se describen en la presente invención métodos para modificar un polipéptido, que comprende poner en contacto un polipéptido que comprende un motivo heterólogo de sulfatasa con una enzima generadora de formilglicina (FGE), en donde el motivo heterólogo de sulfatasa comprende:



dónde

Z_1 es cisteína o serina;

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 heterólogo de sulfatasa está en un extremo terminal N del polipéptido, X_1 está presente;

X_2 y X_3 son independientemente cualquier aminoácido,

en donde dicha puesta en contacto es bajo condiciones suficientes para la conversión de Z_1 en un residuo de formilglicina (FGly) en el polipéptido y produce un polipéptido convertido marcado con aldehído.

El polipéptido utilizado en este método puede tener al menos una de las siguientes propiedades: el motivo heterólogo de sulfatasa tiene menos de 16 residuos de aminoácidos de longitud, el motivo heterólogo de sulfatasa está situado en un extremo terminal N del polipéptido, el motivo heterólogo de sulfatasa está posicionado en un sitio interno de una secuencia de aminoácidos nativa del polipéptido, el motivo heterólogo de sulfatasa está situado en un bucle terminal del polipéptido, el motivo heterólogo de sulfatasa está posicionado en un sitio de modificación postraducciona 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 heterólogo de sulfatasa puede tener menos de 16 residuos de aminoácidos de longitud y puede estar situado en un extremo terminal C del polipéptido. El motivo heterólogo de sulfatasa puede estar presente en un sitio interno en un bucle terminal del polipéptido y/o estar presente en un sitio interno dentro de un bucle extracelular o un bucle intracelular. El motivo heterólogo de sulfatasa puede estar presente en un sitio interno o en el extremo terminal N, y/o puede ser accesible al disolvente cuando se pliega el polipéptido. El motivo heterólogo de sulfatasa puede estar presente en un sitio de modificación postraducciona, tal como un sitio de glicosilación. El sitio de modificación postraducciona puede ser nativo del polipéptido objetivo parental o el polipéptido objetivo puede modificarse por ingeniería genética para incluir uno o más sitios no nativos de modificación postraducciona y en donde el motivo heterólogo de sulfatasa está situado en dichos uno o más de los sitios no nativos de modificación postraducciona.

De interés particular son los motivos de sulfatasa en los que X_1 , cuando está presente, es un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar no cargado, (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y puede en ciertas realizaciones ser M, V, S o T. Otros motivos de sulfatasa de interés particular son aquellos en los que X_2 y X_3 son cada uno independientemente un aminoácido alifático, un aminoácido polar no cargado o un aminoácido que contiene azufre (es decir, distinto de un aminoácido aromático o un aminoácido cargado), y en ciertas realizaciones son cada uno independientemente S, T, A, VG o C. En una realización de interés, el polipéptido se expresa en una célula que contiene la FGE.

Los métodos descritos en la presente invención pueden comprender además poner en contacto el polipéptido convertido marcado con aldehído con un compañero reactivo que comprende una porción de interés; en donde dicho contacto es bajo condiciones que permitan la producción de un producto de reacción de un polipéptido modificado marcado con aldehído que tiene la porción de interés unida covalentemente al residuo de FGly del motivo heterólogo de sulfatasa. 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 en una superficie.

La descripción también contempla un polipéptido convertido marcado con aldehído producido por los métodos descritos en la presente invención, así como un polipéptido modificado marcado con aldehído producido por los métodos descrito en la presente memoria.

La descripción también contempla moléculas de ácido nucleico que comprenden una secuencia de nucleótidos que codifica polipéptidos de la invención, así como vectores y células huésped recombinantes que contienen tales moléculas de ácido nucleico.

La descripción también proporciona 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 situado 5' o 3' de la secuencia que codifica la etiqueta de aldehído, cuyo sitio de restricción permite la inserción de un segundo ácido nucleico que codifica un polipéptido de interés; y un promotor operativamente enlazado al casete de expresión para permitir la expresión de un polipéptido marcado con aldehído producido por la inserción del segundo ácido nucleico que codifica el polipéptido de interés en el sitio de restricción.

También se describe la enzima generadora de formilglucina (FGE) de *Mycobacterium tuberculosis* (FGE de Mtb), y por lo tanto abarca métodos para su uso. La invención también proporciona mezclas de reacción que comprenden una enzima aislada generadora de formilglucina (FGE) de *Mycobacterium tuberculosis*; y un polipéptido de la invención.

También se describen mezclas de reacción de un polipéptido de la invención que tiene un motivo heterólogo de sulfatasa y una FGE, que puede incluir además un polipéptido convertido marcado con aldehído en el que un motivo heterólogo de sulfatasa del polipéptido contiene un residuo de FGly. También se describen composiciones que comprenden una FGE y un polipéptido convertido marcado con aldehído en el que un motivo heterólogo de sulfatasa del polipéptido contiene un residuo de FGly. 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.

A continuación se proporcionan otras características de la invención y su divulgación relacionada, y serán fácilmente evidentes para el experto en la materia tras la lectura de la presente descripción.

Breve descripción de los dibujos

La invención se entiende mejor a partir de la siguiente descripción detallada cuando se lee conjuntamente con los dibujos adjuntos. Se hace hincapié en que, de acuerdo con la práctica común, las diversas características de los dibujos no son a escala. Por el contrario, las dimensiones de las diversas características se amplían o reducen arbitrariamente para mayor claridad. Se incluyen en los dibujos las siguientes figuras:

La Figura 1A es una figura esquemática que muestra un ejemplo de noción general de los métodos y composiciones de la invención. En este ejemplo, se coloca un ejemplo de motivo de sulfatasa ("LCTPSR") en un constructo que contiene un ácido nucleico que codifica una proteína de interés.

La Figura 1B es una figura esquemática que muestra una alineación de secuencia del motivo de sulfatasa a partir de una variedad de sulfatasas encontradas en diversos organismos. La secuencia de consenso contiene la secuencia de las dos etiquetas de aldehído usadas en este estudio. Se resaltan los residuos conservados.

La Figura 2 es un conjunto de gráficos que muestran el análisis del espectro de masas que confirma la presencia de FGly en un péptido tríptico de la sulfotransferasa micobacteriana con un motivo de consenso de 13 residuos de sulfatasa (ald₁₃-Stf0). Los espectros de masas que confirma la presencia de FGly en un péptido tríptico de ald₁₃-Stf0. (Panel a) Espectro de masas de fragmentos trípticos que incorporan FGly (M+2/2). Teórico: 378,2066 m/z Observado: 378,2065. (Panel b) Espectro M+2/2 de FT-ICR del fragmento tríptico que incorpora FGly después del tratamiento con metoxilamina. Teórico: 392,7198 m/z Observado: 392,7201. (Panel c) Secuenciación por MALDI-TOF/TOF del fragmento tríptico que incorpora FGly.

La Figura 3 ilustra los resultados de la cuantificación de la producción de FGly y marcación fluorescente selectiva de constructos marcadas con aldehído. (Panel a) Adición estándar de PL(FGly)TPSR sintético a la digestión tríptica ald₆-Stf0. (Panel b) Adición estándar de PLCTPSR sintético al digerido tríptico ald₆-Stf0. (Panel c) Marcación fluorescente selectiva de ald₁₃-Stf0, ald₆-Stf0 y ald₆-MBP con aminooxi-AlexaFluor 647 que forma imágenes directamente en un escáner de gel fluorescente. (Panel d) La carga de proteínas fue evaluada mediante tinción con Sypro Ruby.

La Figura 4 es un conjunto de imágenes que ilustran la modificación selectiva de proteínas marcadas con aldehído. (Panel a) Ensayo de conmutación de ald₆-MBP. Carril 1: proteína incubada con biotina-hidrazida. Carril 2: proteína incubada con biotina-hidrazida y posteriormente modificada con metoxilamina. Carril 3: proteína incubada con biotina-hidrazida y posteriormente modificada con aminooxiFLAG. La carga de proteínas (cajón inferior) se evaluó mediante tinción ponceau. (Panel b) PEGilación de ald₆-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 falta 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 del canal de K⁺ regulada por azobenceno fotoisomerizable sintética (SPARK) marcada con ald en células CHO (ovario de hámster chino) y HEK (riñón embrionario humano). (Panel a) Gel SDS-PAGE teñido para proteína usando 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 sólo de vector; P, C e I representan tres estrategias para insertar un ejemplo de etiqueta de ald de 6 residuos, a saber, la adición de la etiqueta de ald dentro de uno de los bucles extracelulares (I), la supresión de 6 residuos del bucle y el reemplazo con la etiqueta de ald (C), o supresión de 3 residuos del bucle y luego 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 marcada con myc de 17 kDa.

La Figura 6 ilustra la generación de enlaces hidrazona, oxima y semicarbazona. R y R' se refieren a sustituyentes adecuados que aparecen en el polipéptido marcado con aldehído como se describe en la presente memoria. R" se refiere a un sustituyente del reactivo que se transfiere al polipéptido marcado 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 sulfatasa propuesto. (Panel a) FGE activa sulfatasas oxidando una cisteína de sitio activo con un residuo de 2-formilglicilo (FGly). Las estructuras cristalinas de sulfatasa previamente determinadas indican que FGly de sitio activo está hidratado, lo que sugiere que la escisión de ésteres de sulfato está mediada por un mecanismo de eliminación de la transesterificación. (Panel b) El motivo de sulfatasa está situado hacia el extremo terminal N de las sulfatasas y dirige la cisteína apropiada (*) para modificación por FGE. Los residuos en cajones indican una coincidencia exacta de residuos; los residuos subrayados indican residuos conservados; los residuos con un punto (•) indican residuos similares.

La Figura 8 proporciona resultados que muestran la función de FGE de *Mtb* (Rv0712) in vitro e in vivo. (Panel a) Se trató un péptido sintético similar a un motivo de sulfatasa con FGE de *Mtb* recombinante y la oxidación resultante de cisteína con FGly se controló por espectrometría de masas. Se inactivó el mutante de FEG Cys263Ser sobre el sustrato peptídico. Los iones a 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 biotina, dando como resultado un desplazamiento de masa de +240 Da. (Panel c) Se ensayaron lisados de cepas de tipo silvestre (TS), Δfge y las cepas complementadas ($\Delta fge+fge$) de *Mtb* H37Rv para determinar la actividad de sulfatasa usando el sustrato fluorogénico sulfato de 4-metilumbeliferilo (4MUS) con y sin inhibidores de sulfatasa/fosfatasa. Se usó sulfatasa de lapa como control positivo. (Panel d) Se ensayaron los lisados de las cepas TS, Δfge y $\Delta fge + fge$ de *Mtb* H37Rv para determinar la actividad de la fosfatasa utilizando el sustrato fluorogénico fosfato de 4-metilumbeliferilo con y sin inhibidores de sulfatasa/fosfatasa. Se usó la fosfatasa recombinante de *Mtb* PtpA como control positivo.

La Figura 9 proporciona los resultados del análisis de transferencia tipo Southern del mutante Δfge de *Mtb*. El ADN genómico se digirió con FspI o NcoI, se separó mediante electroforesis en gel de agarosa y se transfirió a una mancha de nailon. La mancha se probó con un fragmento de ADN marcado con digoxigenina de 474 pb, identificando un fragmento de FspI de 4,8 kb y un fragmento de NcoI de 5,7 kb para el fragmento de FspI de tipo silvestre y de 5,5 kb y un fragmento de NcoI de 5,1 kb para el mutante.

La Figura 10 muestra datos que ilustran que Rv2407, Rv3406 y Rv3762c recombinantes no muestran actividad en el ensayo de 4MUS. Se utilizó sulfatasa de lapa como control positivo.

La Figura 11 es un esquema que muestra la estructura de la FGE de *Streptomyces*. (Panel a) Superposición estereoscópica de FGE de *Streptomyces* y FGE humana. Se indican elementos de la estructura secundaria de la FGE de *Streptomyces*. Los iones Ca²⁺ se representan como esferas. La desviación cuadrática media total es de 0,65 Å. (Panel b) Comparación de los residuos que rodean el sitio de unión de Ca²⁺ potencial de FGE de *Streptomyces* y el segundo sitio de unión a Ca²⁺ de FGE humano. La geometría de coordinación apropiada se pierde con una mutación Glu66Ala en FGE de *Streptomyces*. (Panel c) Representación superficial del sitio de unión secundario putativo de FGE de *Streptomyces*. La superficie, representada en escala de grises, se colorea de acuerdo con la conservación de residuos entre todas las FGE conocidas y putativas (con base en la alineación de secuencias de aminoácidos); el blanco representa residuos no conservados, el azul claro representa residuos débilmente conservados, el azul de color más intenso representa residuos conservados y el azul oscuro representa residuos idénticos. El sustrato peptídico de 6 residuos se modela a partir de la estructura compleja 24 de FGE-péptido humano (entrada de PDB 2AIK). Un sustrato hipotético de péptido extendido se representa como una cinta. (Paneles d, e) Las cisteínas del sitio activo 272 y 277

parecen existir en un enlace disulfuro parcial. Cys272 se muestra en dos conformaciones alternas. La densidad de electrones entre Cys272 y Ser269 puede modelarse como una molécula de agua (d) o un hidroperóxido con ocupación parcial (Panel e). Se muestra el monómero D. La densidad de electrones omitida es contorneada en 1σ .

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

5 La Figura 13 proporciona datos que ilustran que la actividad de FGE de *Mtb* y *estreptococo* depende del oxígeno molecular pero es independiente de los cofactores metálicos. (A) Se usó un péptido sintético similar al motivo de sulfatasa (masa = 1421,7 Da) como sustrato para FGE. La conversión de cisteína en FGly dio como resultado una pérdida de 18 Da, que se detectó por espectrometría de masas. (b, g) El quelante de metal EDTA no tuvo ningún efecto sobre la actividad. (c, h, i) La pérdida de cisteínas del sitio activo en FGE de *Mtb* y *estreptococo* eliminó la actividad. (d) La pérdida de Ser260 del sitio activo en FGE de *Mtb* redujo significativamente la actividad. (e) FGE de *Mtb* fue inactiva en ausencia de oxígeno molecular. (f) FGE de *Streptomyces silvestre* fue capaz de oxidar el péptido sintético. (j) Trp234 del sitio activo en FGE de *Streptomyces* no es esencial para la actividad catalítica. El ion con m/z 1427 es el aducto de sodio del péptido del producto que contiene FGly.

15 La Figura 14 proporciona datos que ilustran que las cisteínas 272 y 277 del sitio activo en la FGE de *Streptomyces* están acopladas en un enlace disulfuro parcial. (a) FGE de *Streptomyces* se marcó con NBD y los aductos se detectaron por espectrometría de masas. Se observaron dos poblaciones de FGE de *Streptomyces* modificadas con NBD. Una población tenía un solo aducto de NBD, que corresponde a disulfuro conectado a Cys272 y Cys277 (véase b). La otra población tenía tres aductos NBD correspondientes a tioles de Cys 272 y Cys277 (véase c). Se muestran estados de carga múltiple para ambas poblaciones. Los números por encima de cada pico de iones indican el número de aductos de NBD atribuidos a cada población. (b-d) Los aductos de NBD se mapearon para cada tiol expuesto en la superficie usando espectrometría de masas después de tripsinólisis. Se muestran los iones con estado de carga +2 correspondientes a Cys272 y Cys277 conectados por disulfuro (b, masa calculada = 1.471,58, masa observada = 1.471,58), aductos de NBD tanto en Cys272 como en 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 = 1.486,54). (e) Cys301 no se observó como un tiol libre (se muestra la posición esperada del ión con estado de carga +2). Cys301 fue un tiol adicional expuesto en la superficie que sirvió como control interno para evaluar la eficiencia de la marcación.

25 La Figura 15 proporciona datos que ilustran que una subpoblación de Cys277Ser de FGE de *Streptomyces* no contiene un residuo de Cys272 reducido. (a) Cys277Ser de FGE de *Streptomyces* se marcó con NBD y se detectaron los aductos por espectrometría de masas. Se observaron dos poblaciones de Cys277Ser de FGE de *Streptomyces* modificadas con NBD. Una población corresponde a Cys272 como un tiol libre con un solo aducto de NBD en Cys301 (véanse b y d). La otra población tenía aductos de NBD tanto en Cys272 como en Cys301 (véanse c y d). Se muestran estados de carga múltiples para ambas poblaciones. Los números por encima de cada pico de iones indican el número de aductos de NBD atribuidos a cada población. (b-d) Los aductos de NBD se mapearon a cada tiol expuesto en la superficie usando espectrometría de masas después de tripsinólisis. Se muestran los iones con estado de carga +2 correspondientes a Cys272 como 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) Cys301 no se observó como un tiol libre (se muestra la posición esperada del ión con estado de carga +2). Cys301 fue un tiol adicional expuesto en la superficie que sirvió como control interno para evaluar la eficiencia de la marcación.

40 La Figura 16 proporciona datos que ilustran que FGE de *Streptomyces* no exhibe una anomalía de masa indicativa de una modificación covalente estable de Cys272. La etiqueta de His6 se escindió de FGE de *Streptomyces* y se determinó la masa de la enzima por espectrometría de masas. Los números por encima de cada pico de iones representan estados de carga.

45 La Figura 17 proporciona un esquema de marcación en un sitio específico de Fc de IgG recombinante, que incluye un esquema de un anticuerpo y es el resultado de la modificación de un fragmento Fc que está o bien marcado en N (N-Ald₁₃-Fc) o marcado en C (C-Ald₁₃-Fc) con una etiqueta de aldehído de 13 mer (LCTPSRAALLTGR) o modificado en N o C con una etiqueta de control (LATPSRAALLTGR).

50 La Figura 18 proporciona un esquema de marcación en un sitio específico de un fragmento Fc de IgG usando una etiqueta de aldehído de 6 mer, e incluye los resultados de la modificación de un fragmento Fc que está o bien marcado en N (N-Ald₆-Fc) o marcado en C (C-Ald₆-Fc) con una etiqueta de aldehído de 6 mer (LCTPSR) o modificado en N o C con una etiqueta de control (LATPSR).

55 La Figura 19 proporciona los resultados de identificación de los péptidos que contienen formilglicina (FGly) a partir de Fc de IgG marcada en N, e incluye un conjunto de gráficos que muestran un análisis del espectro de masas que confirma la presencia de FGly en fragmentos tripticos del fragmento Fc marcado en N. (Panel a) Espectro de masas de fragmentos tripticos que incorporan FGly. Teórico: 429.7268 m/z; Observado: 429.7321 m/z. (Panel b) Espectro de

masas del fragmento tríptico del fragmento Fc marcado en N que incorpora FGly después del tratamiento con 2-yodoacetamida. Teórico: 467.2375 m/z Observado: 467.2410 m/z.

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

La Figura 21 se refiere a la marcación en un sitio específico de una proteína de la superficie celular y proporciona un esquema del vector pDisplay^{MR} usado para el constructo de una proteína de la superficie celular marcada con aldehído (usando una etiqueta de aldehído de 13 mer de LCTPSRAALTGR) y un gráfico que muestra un aumento de la fluorescencia media para la proteína de la superficie marcada con el aldehído de 13 mer (Ald₁₃-TM) en comparación con el control (LATPSRAALLTGR, denominado C-> A-TM).

La Figura 22 se refiere a la marcación en un sitio específico de una proteína citosólica, ejemplificada por His₆-Ald₁₃-AcGFP, y proporciona los resultados de la modificación de la proteína de fusión de proteína fluorescente verde (GFP) que contiene una etiqueta His y una etiqueta de aldehído de 13 mer (denominada como His₆-Ald₁₃-Ac-GFP o Ald-AcGFP) o una proteína de fusión GFP que contiene una etiqueta de control (LATPSRAALLTGR) (denominada C-> A-AcGFP).

La Figura 23 proporciona un esquema de la glicosilación que explica en términos generales la glicosilación en un sitio específico del interferón beta (IFN-Beta) usando la metodología de la etiqueta del aldehído.

Descripción detallada de las realizaciones de la invención

Antes de describir la presente invención, debe entenderse que esta invención no se limita a las realizaciones particulares descritas, ya que, por supuesto, pueden variar. También debe entenderse que la terminología utilizada en la presente memoria tiene el propósito de describir únicamente realizaciones particulares y no pretende ser limitativa, ya que el alcance de la presente invención estará limitado solamente por las reivindicaciones adjuntas.

Cuando se proporciona un intervalo de valores, se entiende que cada valor que interviene, hasta el décimo de la unidad del límite inferior, a menos que el contexto indique claramente lo contrario, entre los límites superior e inferior de ese intervalo también se describe específicamente. Cada intervalo más pequeño entre cualquier valor indicado o valor interviniente en un intervalo indicado y cualquier otro valor indicado o interviniente en ese intervalo indicado está comprendido dentro de la invención. Los límites superior e inferior de estos intervalos más pequeños pueden incluirse o excluirse independientemente en el intervalo, y cada intervalo en el que cualquiera, ninguno o ambos límites están incluidos en los rangos más pequeños también está incluido dentro de la invención, sujeto a cualquier límite específicamente excluido en el intervalo indicado. Cuando el intervalo indicado incluye uno o ambos de los límites, los intervalos que excluyen uno o ambos de los límites incluidos también se incluyen en la invención.

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que comúnmente entiende un experto en la técnica al que pertenece esta invención. Aunque se pueden utilizar en la práctica o análisis de la presente invención cualquiera de los métodos y materiales similares o equivalentes a los descritos en la presente invención, a continuación se describen algunos métodos y materiales potenciales y de ejemplo. Todas las publicaciones mencionadas aquí se incorporan aquí por referencia para divulgar y describir los métodos y/o materiales en relación con los cuales se citan las publicaciones. Se entiende que la presente descripción reemplaza cualquier divulgación de una publicación incorporada en la medida en que exista una contradicción.

Debe tenerse en cuenta que tal como se usa en el presente documento y en las reivindicaciones adjuntas, las formas singulares "un", "uno, una" y "el, la" incluyen referencia a los plurales a menos que el contexto indique claramente lo contrario. Por lo tanto, por ejemplo, la referencia a "una etiqueta de aldehído" incluye una pluralidad de dichas etiquetas y la referencia a "el polipéptido" incluye referencia a uno o más polipéptidos y equivalentes de los mismos conocidos por los expertos en la técnica, y así sucesivamente.

Se observa además que las reivindicaciones pueden redactarse para excluir cualquier elemento que pueda ser opcional. Por lo tanto, esta exposición pretende servir como base antecedente para el uso de la terminología exclusiva tal como "únicamente", "solamente" y similares en relación con la enumeración de los elementos de las reivindicaciones, o el uso de una limitación "negativa".

Las publicaciones aquí discutidas se proporcionan únicamente para su divulgación antes de la fecha de presentación de la presente solicitud. Ninguna parte de la presente invención debe interpretarse como una admisión de que la presente

invención no tiene derecho a anticipar tal publicación en virtud de la invención anterior. Además, las fechas de publicación proporcionadas pueden ser diferentes de las fechas reales de publicación que pueden necesitar ser confirmadas independientemente.

Definiciones

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

En la presente memoria se utiliza "polipéptido objetivo" para referirse a un polipéptido que va a ser modificado mediante el uso de una etiqueta de aldehído como se describe en el presente documento.

- 15 "Secuencia nativa de aminoácidos" o "secuencia parental de aminoácidos" se usan indistintamente aquí en el contexto de un polipéptido objetivo para referirse a la secuencia de aminoácidos del polipéptido objetivo antes de la modificación para incluir un marcador heterólogo de aldehído.

- 20 Por "etiqueta de aldehído" o "etiqueta de ald" se entiende una secuencia de aminoácidos que contiene una secuencia de aminoácidos derivada de un motivo de sulfatasa que puede ser convertido o que ha sido convertido, mediante la acción de una enzima generadora de formilglicina (FGE) para contener un residuo de 2-formilglicina (denominado en la presente memoria "FGly"). Aunque esto es técnicamente incorrecto, el residuo de FGly generado por una FGE se denomina a menudo en la literatura como una "formilglicina". Dicho de otra manera, el término "etiqueta de aldehído" se utiliza en la presente memoria para referirse a una secuencia de aminoácidos que comprende un motivo de sulfatasa "no convertido" (es decir, un motivo de sulfatasa en el que los residuos de cisteína o serina no han sido convertidos a FGly por una FGE, pero pueden ser convertidos), así como a una secuencia de aminoácidos que comprende un motivo de sulfatasa "convertido" (es decir, un motivo de sulfatasa en el que los residuos de cisteína o serina han sido convertidos en FGly por acción de una FGE).

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

- 35 "Modificación" abarca la adición, eliminación o alteración de una porción. Tal como se utiliza en el contexto de un polipéptido que tiene un motivo de sulfatasa convertido, se entiende por "modificación" 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 FGly con un compañero reactivo. Como se ha discutido 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.

- 40 Por "genéticamente codificable" tal como se utiliza en referencia a una secuencia de aminoácidos de polipéptido, péptido o proteína, significa que la secuencia de aminoácidos está compuesta de residuos de aminoácidos que son capaces de producir por transcripción y traducción de un ácido nucleico que codifica la secuencia de aminoácidos, donde la transcripción y/o traducción puede ocurrir en una célula o en un sistema de transcripción/traducción *in vitro* sin de células.

- 45 El término "secuencias de control" se refiere a secuencias de ADN para facilitar la expresión de una secuencia de codificación enlazada operativamente en un sistema de expresión particular, por ejemplo, una célula de mamífero, célula bacteriana, síntesis *in vitro*, etc. Las secuencias de control que son adecuadas para 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.

- 50 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 secuencia previa o líder secretora está enlazada operativamente al ADN por un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o potenciador está operativamente enlazado a una secuencia de codificación si afecta la transcripción de la secuencia; o un sitio de unión al ribosoma está operativamente enlazado a una secuencia codificante si está posicionado de manera que facilite el inicio de la traducción. Generalmente, "operativamente enlazado" significa que las secuencias de ADN que están enlazadas son contiguas y, en el caso de un líder secretor, contiguas y en el marco de lectura. La unión se logra mediante ligación o mediante reacciones de amplificación. Pueden utilizarse adaptadores o enlazadores sintéticos de oligonucleótidos para enlazar secuencias de acuerdo con la práctica convencional.

El término "casete de expresión" tal como se utiliza en la presente memoria se refiere a un segmento de ácido nucleico, normalmente ADN, que puede insertarse en un ácido nucleico (por ejemplo, mediante el uso de sitios de restricción compatibles con ligación en un constructo de interés o por recombinación homóloga en un constructo de interés o en un genoma de célula huésped). En general, el segmento de ácido nucleico comprende un polinucleótido que codifica un polipéptido de interés (por ejemplo, una etiqueta de aldehído, que puede estar operativamente enlazado 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 adecuado para la transcripción y la traducción. Los casetes de expresión también pueden comprender elementos que facilitan la expresión de un polinucleótido que codifica un polipéptido de interés en una célula huésped. Estos elementos pueden incluir, pero no se limitan a: un promotor, un promotor mínimo, un potenciador, un elemento de respuesta, una secuencia terminadora, una secuencia de poliadenilación y similares.

Tal como se usa en la presente memoria, el término "aislado" pretende describir un compuesto de interés que está en un entorno diferente de aquél en el que se produce naturalmente el compuesto. Se entiende que "aislado" incluye compuestos que están dentro de muestras que son sustancialmente enriquecidas por el compuesto de interés y/o en donde el compuesto de interés está parcial o sustancialmente purificado.

Tal como se utiliza en la presente memoria, el término "sustancialmente purificado" se refiere a un compuesto que se remueve de su entorno natural y está al menos 60% libre, usualmente 75% libre y más usualmente 90% libre de otros componentes con los que está naturalmente asociado.

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

Por "heterólogo" se entiende que se proporcionan una primera entidad y una segunda entidad en una asociación que normalmente no se encuentra en la naturaleza. Por ejemplo, una proteína que contiene un motivo "heterólogo" de sulfatasa o una etiqueta "heteróloga" de ald, es una proteína que normalmente no contiene un motivo de sulfatasa en esa posición dentro de su secuencia de aminoácidos (por ejemplo, proteínas que tienen un único motivo nativo de sulfatasa puede contener un segundo motivo de sulfatasa que es "heterólogo"; otras proteínas que contienen un motivo de sulfatasa pueden ser modificadas para reposicionar el motivo de sulfatasa, volviendo al motivo de sulfatasa reposicionado "heterólogo" a la proteína). En algunas realizaciones, un motivo heterólogo de sulfatasa está presente en un polipéptido que no contiene ningún motivo nativo de sulfatasa.

Por "compañero reactivo" se entiende una molécula o porción molecular que reacciona específicamente con otro compañero reactivo para producir un producto de reacción. Los ejemplos de compañeros reactivos incluyen un motivo de cisteína o de serina de sulfatasa y una enzima generadora de formilglicina (FGE), que reaccionan para formar un producto de reacción de una etiqueta de aldehído convertida que contiene un FGly en lugar de cisteína o serina en el motivo. Otros ejemplos de compañeros reactivos incluyen un aldehído de un residuo de formilglicina (FGly) de una etiqueta de aldehído convertida y un reactivo compañero reactivo que comprende una porción de interés, que reaccionan para formar un producto de reacción de un polipéptido marcado con aldehído modificado que tiene la porción de interés conjugada con el polipéptido etiquetado con aldehído en el residuo de FGly.

"Extremo terminal N" se refiere al residuo de aminoácido terminal de un polipéptido que tiene un grupo amina libre, cuyo grupo amina en los residuos de aminoácidos que no están en el extremo terminal N, normalmente forma parte de la cadena principal covalente del polipéptido.

"Extremo terminal C" 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 están en el extremo terminal C, normalmente forma parte de la cadena principal covalente del polipéptido.

Por "extremo terminal N" se entiende la región de un polipéptido que está más cerca del extremo terminal N que al extremo terminal C.

Por "extremo terminal C" se entiende la región de un polipéptido que está más cerca del extremo terminal C que al extremo terminal N.

Por "sitio interno" tal como se utiliza en referencia a un polipéptido o una secuencia de aminoácidos de un polipéptido significa una región del polipéptido que no está en el extremo terminal N o en el extremo terminal C, e incluye tanto regiones del extremo terminal N como del extremo terminal C del polipéptido.

50 Introducción

La presente invención explora un motivo de sulfatasa genéticamente codificable de origen natural para uso como una etiqueta de péptido, denominado en la presente memoria como una "etiqueta de aldehído" o "etiqueta de ald", para

dirigir la modificación a un sitio específico de un polipéptido. El motivo de sulfatasa de la etiqueta de aldehído, que se basa en un motivo encontrado en sitios activos de sulfatasas, contiene un residuo de serina o cisteína que puede ser convertido (oxidado) en formilglicina (FGly) por 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 una etiqueta de aldehído en una célula) o in vitro (por ejemplo, poniendo en contacto una proteína que contiene una etiqueta de aldehído con una FGE en un sistema libre de células). La porción aldehído del residuo de FGly resultante puede usarse como un "mango químico" para facilitar la modificación química de la proteína en un sitio específico.

La Figura 1A es un esquema que muestra ejemplos de métodos y composiciones de la invención. En este ejemplo, se coloca un ejemplo de un motivo de sulfatasa ("LCTPSR") en un constructo que contiene un ácido nucleico que codifica una proteína de interés. En este ejemplo, el motivo de sulfatasa se sitúa en el extremo terminal N de la proteína codificada después de la expresión; sin embargo, como se describe con más detalle a continuación, pueden insertarse motivos de sulfatasa en uno o más sitios deseados del polipéptido (por ejemplo, para proporcionar el motivo en el extremo terminal N, el extremo terminal C y/o el sitio interno del polipéptido codificado). El motivo de sulfatasa ejemplificado en la Figura 1A está dentro de un género de motivos de sulfatasa como se describe a continuación con más detalle. La Figura 1B es un esquema de una alineación de secuencias del motivo de sulfatasa a partir de una variedad de sulfatasas encontradas en diversos organismos. La secuencia de consenso contiene la secuencia de las dos etiquetas de aldehído usadas en este estudio. Se resaltan los residuos conservados.

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

Tanto la colocación de la etiqueta de aldehído dentro de la proteína objetivo que va a ser modificada como la modificación mediada por la etiqueta de aldehído como se describe en la presente memoria pueden generalizarse con respecto a una amplia variedad de proteínas. La capacidad de FGE para facilitar la conversión del motivo de sulfatasa para generar un residuo de FGly es independiente de la posición del motivo dentro de la proteína. Debido a que FGE puede convertir la cisteína/serina del motivo de sulfatasa de una manera que es o bien independiente del contexto de las secuencias o independiente del contexto estructural, las etiquetas de aldehído pueden posicionarse en cualquier sitio deseado dentro de un polipéptido objetivo a modificar, con la condición de que el motivo de sulfatasa sea accesible a la FGE en el momento de la conversión enzimática. Además, la reactividad única del aldehído permite la modificación bioortogonal y quimioselectiva de proteínas recombinantes, proporcionando así un medio para modificación química en un sitio específico de proteínas que se puede realizarse bajo condiciones fisiológicas y de una manera altamente selectiva.

Como se apreciará a partir de la presente descripción, las aplicaciones de etiquetas de aldehído son numerosas y pueden proporcionar una serie de ventajas. Por ejemplo, la etiqueta de aldehído es más pequeña que la mayoría si no todas las etiquetas de péptidos convencionales lo que permiten la modificación covalente de proteínas, requiriendo así cambios mínimos en la secuencia de aminoácidos de un polipéptido objetivo. En segundo lugar, la etiqueta de aldehído aprovecha las químicas de etiquetado secundario bien caracterizadas. En tercer lugar, la etiqueta de aldehído demuestra reversibilidad, y a través de la selección de compañeros reactivos que proporcionan la conjugación de porciones a través de enlaces covalentes de diferente estabilidad, permite la modificación secuencial y el reemplazo de una porción unida a una etiqueta de aldehído. Además, debido a que la etiqueta de aldehído se forma usando maquinaria biosintética ya presente en la mayoría de los sistemas celulares y es independiente de la naturaleza del objetivo o colocación dentro de la secuencia parental de aminoácidos, la etiqueta de aldehído puede usarse para facilitar la modificación de un gran número de polipéptidos utilizando un sistema de expresión fácilmente disponible.

La porción aldehído de una etiqueta convertida de aldehído puede usarse para una variedad de aplicaciones que incluyen, pero no se limitan a, visualización usando marcación de fluorescencia o de epítopos (por ejemplo, microscopía electrónica usando partículas de oro equipadas con grupos reactivos de aldehído), inmovilización de proteínas (por ejemplo, la producción de microarreglos de proteína), estudios y aplicaciones de dinámica y localización de proteínas, y conjugación de proteínas con una porción de interés (por ejemplo, porciones que mejoran el índice terapéutico de una proteína parental (por ejemplo, PEG), porciones objetivo (por ejemplo, para mejorar la biodisponibilidad en un sitio de acción), y porciones biológicamente activas (por ejemplo, una porción terapéutica).

De particular interés es el uso de etiquetas de aldehído para facilitar la fijación en un sitio específico de un polímero soluble en agua, tal como PEG. A pesar de los avances en la química de conjugación de proteínas, la modificación controlada en un sitio específico de las proteínas sigue siendo un desafío. Muchos métodos convencionales de PEGilación fijan porciones de PEG a través de la reacción, por ejemplo, con una lisina o cisteína como residuo objetivo. Debido a la presencia de múltiples residuos objetivo en una proteína, tales sistemas convencionales pueden resultar en PEGilación en múltiples sitios, creando una colección de conjugados discretos proteína-PEG con diferentes farmacocinéticas. Por el contrario, el uso de un residuo de FGly de una etiqueta de aldehído como residuo objetivo proporciona un sitio único para la unión del polímero covalente, y aumenta así tanto la especificidad como la

homogeneidad del producto modificado resultante. Estas y otras características y ventajas serán fácilmente evidentes para el especialista en la materia tras la lectura de la presente descripción.

A continuación se describirán con más detalle los métodos y composiciones para la práctica de la invención.

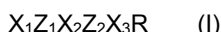
Etiquetas de aldehído

- 5 En general, una etiqueta de aldehído puede estar basada en cualquier secuencia de aminoácidos derivada de un motivo de sulfatasa (también denominado "dominio de sulfatasa") que puede ser convertido por acción de una enzima generadora de formilglicina (FGE) para contener una formilglicina (FGly). La acción de FGE se dirige en forma de una secuencia específica en la que la FGE actúa sobre un motivo de sulfatasa, pero este motivo de sulfatasa puede estar situado dentro de cualquier región de un polipéptido objetivo. Por lo tanto, la conversión mediada por FGE de un motivo de sulfatasa es en un sitio específico (es decir, ya que FGE actúa en la secuencia de aminoácidos de un motivo de sulfatasa) pero la capacidad de FGE para actuar tras el motivo de sulfatasa es independiente del contexto de la secuencia (es decir, la capacidad de la FGE para convertir una cisteína/serina de un motivo de sulfatasa es independiente del contexto de secuencia en el que se presenta el motivo de sulfatasa en el polipéptido objetivo).

Ejemplos de etiquetas de aldehído

- 15 Un motivo mínimo de sulfatasa de una etiqueta de aldehído es normalmente de aproximadamente 5 o 6 residuos de aminoácidos de longitud, usualmente no más de 6 residuos de aminoácidos de longitud. En general, es normalmente deseable minimizar la extensión de la modificación de la secuencia de aminoácidos nativa del polipéptido objetivo, para minimizar el número de residuos de aminoácidos que se insertan, eliminan, sustituyen (reemplazan) o se añaden (por ejemplo, en el extremo terminal N o C). Minimizar el grado de modificación de la secuencia de aminoácidos del polipéptido objetivo es usualmente deseable para minimizar el impacto que tales modificaciones pueden tener sobre la función y/o estructura del polipéptido objetivo. Por lo tanto, las etiquetas de aldehído de interés particular incluyen aquellas que requieren modificación (inserción, adición, eliminación, 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.
- 20 Debe observarse que aunque las etiquetas de aldehído de interés particular son aquellas basadas en un motivo mínimo de sulfatasa, se apreciará fácilmente que están contempladas y abarcadas las etiquetas de aldehído más largas por la presente descripción y pueden encontrar uso en las composiciones y métodos de la invención. Por lo tanto, las etiquetas de aldehído pueden comprender un motivo mínimo de sulfatasa de 5 o 6 residuos o pueden ser más largas y comprenden un motivo mínimo de sulfatasa que puede estar flanqueado en los lados del extremo extremo terminal N y/o C del motivo mediante residuos de aminoácidos adicionales. Se contemplan etiquetas de aldehído, por ejemplo, de 5 o 6 residuos de aminoácidos, así como secuencias más largas de aminoácidos de más de 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más residuos de aminoácidos.

En general, los motivos de sulfatasa útiles en las etiquetas de aldehído como se describe en la presente memoria son de la fórmula:



- 35 dónde

Z_1 es cisteína o serina (que también puede estar representada por (C/S));

Z_2 es o bien un residuo de prolina o de alanina (que también puede estar representado por (P/A));

- 40 X_1 está presente o ausente y, cuando está presente, puede ser cualquier aminoácido, aunque habitualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar no cargado, (es decir, que no sea un aminoácido aromático o un aminoácido cargado), usualmente L, M, V, S o T, más usualmente L, M, S o V, con la condición de que cuando el motivo de sulfatasa esté en el extremo terminal N del polipéptido objetivo, X_1 esté presente; y

- 45 X_2 y X_3 independientemente pueden ser cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido polar no cargado, 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 observarse que, después de la acción de una FGE sobre el motivo de sulfatasa, Z_1 se oxida para generar un residuo de formilglicina (FGly). Además, después tanto de la conversión mediada por FGE como la reacción con un compañero reactivo que comprende una porción de interés, la posición de FGly en Z_1 en la fórmula anterior está unida covalentemente a la porción de interés (por ejemplo, la etiqueta detectable, polímero soluble en agua, etc.).

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



donde X_1 y X_2 pueden ser independientemente cualquier aminoácido, aunque por lo general un aminoácido alifático, un aminoácido no cargado polar, o un aminoácido que contiene azufre (es decir, que no sea un aminoácido aromático o un aminoácido cargado), por lo general S, T, A, V o C, más usualmente S, T, A o V.

Como se señaló anteriormente, el motivo de sulfatasa puede contener residuos adicionales en uno o ambos de los extremos terminales N y C de la secuencia, por ejemplo, de manera que la etiqueta de aldehído incluye tanto un motivo de sulfatasa como un "motivo auxiliar". En una forma de realización, el motivo de sulfatasa incluye un motivo auxiliar en el extremo terminal C (es decir, tras el residuo de arginina en la fórmula anterior) 1, 2, 3, 4, 5, 6, o todos los 7 de los residuos contiguos de una secuencia de aminoácido AALLTGR, SLLTGR, AAFMTGR, AAFLTGR, SAFLTGR, ASILTGR, VSFLTGR, ASLLTGL, ASILITG, VSFLTGR, SAIMTGR, SAIVTGR, TNLWRG, TNLWRGQ, TNLCAAS, VSLWTGK, SMLLTG, SMLLTGN, SMLLTGT, ASFMAGQ, o ASLLTGL, (véase, por ejemplo, Dierks y colaboradores (1999) EMBO J 18 (8): 2084-2091), o de GSLFTGR. Sin embargo, como se indica en los Ejemplos a continuación, los presentes inventores han encontrado que no se requieren tales residuos adicionales de aminoácidos en el extremo terminal C para la conversión mediada por FGE del motivo de sulfatasa de la etiqueta de aldehído, y por lo tanto son sólo opcionales y pueden excluirse específicamente de las etiquetas de aldehído descritas en la presente memoria. En algunas realizaciones, 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 nativa de aminoácidos en fosfonato monoéster hidrolasas.

El motivo de sulfatasa de la etiqueta de aldehído se selecciona generalmente de modo que sea capaz de conversión mediante una FGE seleccionada, por ejemplo, una FGE presente en una célula huésped en la que se expresa el polipéptido marcado con aldehído o una FGE que se va a poner en contacto con el polipéptido marcado con aldehído en un método *in vitro* libre de células.

La selección de etiquetas de aldehído y de una FGE que proporcione compañeros reactivos adecuados para permitir la generación de un FGly en el polipéptido objetivo etiquetado con aldehído puede realizarse fácilmente a la luz de la información disponible en la técnica. En general, los motivos de sulfatasa susceptibles de conversión por una FGE eucariota contienen una cisteína y una prolina (es decir, una cisteína y una 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 son modificados por la FGE de tipo "SUMF1" (Cosma y colaboradores, Cell 2003, 113, (4), 445-56, Dierks y colaboradores, Cell 2003, 113, (4), 435-44). Los motivos de sulfatasa susceptibles a la conversión por una FGE procariota contienen una cisteína o una serina y una prolina en el motivo de 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 bien sea por la FGE de tipo "SUMF1" o por el FGE de tipo "AtsB", respectivamente (Szameit y colaboradores, J Biol Chem 1999, 274, (22), 15375-81). Otros motivos de sulfatasa susceptibles a la conversión por una FGE procariota contienen ya sea una cisteína o una serina y una prolina o una alanina en el motivo de 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, $X_1CX_2PX_3R$; $X_1SX_2PX_2R$; $X_1CX_2AX_3R$; $X_1SX_2AX_3R$); CX_1PX_2R ; SX_1PX_2R ; CX_1AX_2R ; SX_1AX_2R en la Fórmula II anterior), y son susceptibles de modificación, por ejemplo, pueden ser modificados por una FGE de un Firmicutes (por ejemplo, Clostridium perfringens) (véase Berteau y colaboradores, J. Biol. Chem., 2006; 281: 22464-2270).

Por lo tanto, por ejemplo, cuando la FGE es una FGE eucariota (por ejemplo, una FGE de mamífero, incluyendo una FGE humana), el motivo de sulfatasa es usualmente de la fórmula:



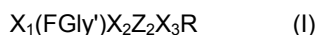
dónde

X_1 puede estar presente o ausente y, cuando está presente, puede ser cualquier aminoácido, aunque habitualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar no cargado (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 de la sulfatasa está en el extremo terminal N del polipéptido objetivo, X_1 está presente; y

X_2 y X_3 independientemente pueden ser cualquier aminoácido, aunque usualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar no cargado, (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.

Ejemplos específicos de motivos de sulfatasa incluyen LCTPSR, MCTPSR, VCTPSR, LCSPSR, LCAPSR, LCVPSR y LCGPSR. Otros motivos específicos de la sulfatasa son fácilmente evidentes a partir de la descripción proporcionada en este documento.

Como se describe con más detalle a continuación, se hace reaccionar un polipéptido convertido marcado con aldehído con un compañero reactivo que contiene una porción de interés para permitir la conjugación de la porción de interés con el residuo de FGly del polipéptido convertido marcado con aldehído, y la producción de un polipéptido modificado. Los polipéptidos modificados que tienen una etiqueta de aldehído modificada se describen generalmente porque comprenden un motivo modificado de sulfatasa de la fórmula:



dónde

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

Z₂ es o bien un residuo de prolina o de alanina (que también puede estar representado por (P/A));

X₁ 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 no cargado (es decir, distinto de un aminoácido aromático o un aminoácido cargado), usualmente L, M, V, S o T, más usualmente L, M o V, con la condición de que cuando el motivo de la sulfatasa esté en el extremo terminal N del polipéptido objetivo, X₁ esté presente; y

X₂ y X₃ pueden ser independientemente cualquier aminoácido, aunque habitualmente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar no cargado, (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.

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

Tal como se describe con más detalle a continuación, 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 en una superficie. Como es evidente a partir de la discusión anterior de polipéptidos marcados con aldehído, el motivo modificado de sulfatasa del polipéptido modificado puede posicionarse en cualquier sitio deseado del polipéptido. Por tanto, la presente descripción proporciona, por ejemplo, un polipéptido modificado que tiene un motivo de 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 posteriormente contendrá una porción en una posición correspondiente a este sitio de modificación postraducciona en el polipéptido parental). Por ejemplo, entonces, un polipéptido modificado puede ser producido de manera que tenga 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 objetivo parental. Así, por ejemplo, se puede producir un polipéptido PEGilado que tiene la porción PEG situada en la misma o casi la misma localización en que los residuos de azúcar se posicionarían en el polipéptido natural parental. De manera similar, cuando el polipéptido objetivo parental se modifica 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 corresponde a estos sitios no nativos de modificación postraducciona en el polipéptido parental.

Modificación de un polipéptido objetivo para incluir una etiqueta de aldehído

Las marcas de aldehído pueden posicionarse en cualquier lugar dentro de un polipéptido objetivo al que se desee permitir 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 modificación posterior en el FGly, o puede hacerse accesible (por ejemplo, desnaturalizando 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.

En algunas realizaciones es deseable colocar la etiqueta o etiquetas de aldehído en el polipéptido objetivo teniendo en cuenta su estructura cuando se pliega (por ejemplo, en un entorno libre de células, normalmente un entorno 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 transmembrana). Por ejemplo, una etiqueta de aldehído puede colocarse en un sitio accesible al disolvente en un polipéptido objetivo plegado. La etiqueta de aldehído accesible al disolvente en un polipéptido etiquetado de aldehído no convertido plegado es por lo tanto accesible a una FGE para la conversión de la serina o cisteína en un FGly. Del mismo modo, una etiqueta de aldehído accesible al disolvente de un polipéptido

convertido marcado con aldehído es accesible a un reactivo compañero reactivo para conjugación con una porción de interés para proporcionar un polipéptido marcado con aldehído modificado. Cuando se coloca una etiqueta de aldehído en un sitio accesible al disolvente, la conversión y modificación in vitro mediada por FGE con una porción por reacción con un compañero reactivo puede llevarse a cabo sin necesidad de desnaturalizar la proteína. Los sitios accesibles al disolventes también pueden incluir regiones de polipéptido objetivo que están expuestas a una superficie celular extracelular o intracelular cuando se expresan en una célula huésped (por ejemplo, distinta de una región transmembrana del polipéptido objetivo).

De acuerdo con ello, pueden proporcionarse una o más etiquetas de aldehído en sitios seleccionados independientemente, por ejemplo, en un extremo terminal N accesible al disolvente, una región del extremo terminal N accesible al disolvente, un extremo terminal C accesible al disolvente, una región del extremo terminal C 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 realizaciones, la etiqueta de aldehído está posicionada en un sitio distinto del extremo terminal C del polipéptido. En otras realizaciones, el polipéptido en el que está colocada la etiqueta de aldehído es un polipéptido de longitud completa.

En otras realizaciones, se coloca un sitio de etiqueta de aldehído en un sitio que se modifica después de la traducción en el polipéptido objetivo nativo. Por ejemplo, se puede introducir una etiqueta de aldehído 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 de consenso de una variedad de sitios modificados después de la traducción, y los métodos para la identificación de un sitio modificado después de la traducción en un polipéptido, son bien conocidos en la técnica. Se entiende que el sitio de modificación postraducciona puede ser de origen natural o tal sitio de un polipéptido que ha sido modificado genéticamente (por ejemplo, a través de técnicas recombinantes) para incluir un sitio de modificación postraducciona que es no nativo para el polipéptido (por ejemplo, como en un sitio de glicosilación de una variante hiperglicosilada de EPO). En esta última realización, los polipéptidos que tienen un sitio de modificación postraducciona no nativo y que han mostrado exhibir una actividad biológica de interés son de particular interés.

La divulgación también proporciona en la presente memoria métodos para identificar sitios adecuados para la modificación de un polipéptido objetivo para incluir una etiqueta de aldehído. Por ejemplo, pueden producirse uno o más constructos de polipéptidos objetivo marcados con aldehído, y los constructos expresados en una célula que expresa una FGE, o expuestas a FGE después del aislamiento de la célula (como se describe con más detalle a continuación). El polipéptido marcado con aldehído puede ponerse en contacto entonces con un compañero reactivo que, si la etiqueta de aldehído es accesible, permite la unión de una porción detectable al FGly de la etiqueta de aldehído. Entonces se determina la presencia o ausencia de la porción detectable. Si se detecta la porción detectable, entonces la colocación de la etiqueta de aldehído en el polipéptido fue exitosa. De esta manera, puede producirse y seleccionarse una biblioteca de constructos que tienen una etiqueta de aldehído posicionada en diferentes sitios en la secuencia codificante del polipéptido objetivo para facilitar la identificación de una posición óptima de una etiqueta de aldehído. Además o alternativamente, el polipéptido marcado con aldehído puede ser ensayado para determinar la actividad biológica normalmente asociada con el polipéptido objetivo y/o evaluada la estructura del polipéptido marcado con aldehído (por ejemplo, para evaluar si un epítipo normalmente presente en una superficie celular extracelular en el polipéptido objetivo nativo está también presente en el polipéptido marcado con aldehído).

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

La modificación de un polipéptido objetivo para incluir una o más etiquetas de aldehído puede llevarse a cabo usando técnicas genéticas moleculares recombinantes, de manera que produzcan ácido nucleico que codifica el polipéptido objetivo marcado con el aldehído deseado. Tales métodos son bien conocidos en la técnica, e incluyen métodos de clonación, métodos de mutación en un sitio específico y similares (véase, por ejemplo, Sambrook y colaboradores, En "Molecular Cloning: A Laboratory Manual", Cold Spring Harbor Laboratory Press, 1989), "Current Protocols in Molecular Biology" (eds., Ausubel y colaboradores, Greene Publishing Associates, Inc. y John Wiley & Sons, Inc., 1990 y suplementos.) Alternativamente, se puede añadir una etiqueta de aldehído utilizando técnicas no recombinantes, por ejemplo, usando ligación química nativa o ligación química pseudo-nativa, por ejemplo, para añadir una etiqueta de aldehído a un extremo terminal C del polipéptido objetivo (véase, por ejemplo, los documentos US 6.184.344, US 6.307.018, US 6.451.543, US 6.570.040, US 2006/0173159; US 2006/0149039). Véase también Rush y colaboradores (5 de enero de 2006) Org Lett. 8 (1): 131-4.

Polipéptidos objetivo

Cualquiera de una amplia variedad de polipéptidos puede modificarse para incluir una etiqueta de aldehído para facilitar la modificación del polipéptido. Los polipéptidos adecuados para la modificación basada en una etiqueta de aldehído incluyen tanto proteínas que tienen una secuencia de aminoácidos de origen natural, una secuencia de aminoácidos nativa que tiene una metionina en el extremo terminal N, fragmentos de polipéptidos de origen natural y polipéptidos de origen no natural y fragmentos de los mismos. En algunas realizaciones, el polipéptido objetivo es un polipéptido distinto de una sulfatasa o fragmento de la misma, distinto de una proteína informadora, o distinto de la preprolactina o la prolactina.

A continuación se describen ejemplos de clases y tipos de polipéptidos que son de interés para la modificación usando los métodos basados en etiquetas de aldehído descritos en la presente memoria.

Polipéptidos terapéuticos

En una realización, 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 aquellos polipéptidos para los cuales la unión a una porción puede proporcionar uno o más de, por ejemplo, un aumento en la semivida en suero, una disminución en una respuesta inmune adversa, actividad o funcionalidad biológica adicional o alternativa, y similares, u otro beneficio o reducción de un efecto secundario adverso. Cuando el polipéptido terapéutico es un antígeno para una vacuna, la modificación puede proporcionar una inmunogenicidad mejorada del polipéptido.

Ejemplos de clases de proteínas terapéuticas incluyen aquellas que son citoquinas, quimiocinas, factores de crecimiento, hormonas, anticuerpos y antígenos. Otros ejemplos incluyen eritropoyetina (EPO, por ejemplo, EPO nativa, EPO sintética (véase, por ejemplo, el documento US 2003/0191291), hormona de crecimiento humano (hGH), hormona de crecimiento bovino (bGH), hormona estimulante de folículos (FSH), interferón (por ejemplo, IFN-gamma, IFN-beta, IFN-alfa, IFN-omega, interferón de consenso, y similares), insulina, factor de crecimiento similar a insulina (por ejemplo, IGF-I, IGF-II), factores sanguíneos (por ejemplo, Factor VIII, Factor IX, Factor X, activador del plasminógeno tisular (TPA), y similares), factores estimuladores de colonias (por ejemplo, CSL de granulocitos (G-CSF), CSF de macrófagos (M-CSF), CSF de granulocitos-macrófagos (GM-CSF), y similares), factores de crecimiento transformadores (por ejemplo, TGF-beta, TGF-alfa), interleuquinas (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 una línea de células gliales (GDNF), factor de crecimiento, factor de crecimiento nervioso (NGF), RANTES, y similares.

Otros ejemplos incluyen anticuerpos, por ejemplo, anticuerpos policlonales, anticuerpos monoclonales, anticuerpos humanizados, fragmentos de unión al antígeno (por ejemplo, F(ab)', Fab, Fv), anticuerpos de cadena sencilla y similares. De particular interés son los anticuerpos que se unen específicamente a un antígeno tumoral, un antígeno de células inmunitarias (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.

Los métodos y composiciones descritos en la presente invención pueden aplicarse para proporcionar una porción (por ejemplo, un polímero soluble en agua) en un sitio nativo o modificado de glicosilación, tal como se encuentra en formas hiperglicosiladas de una proteína terapéutica, tal como, por ejemplo: un interferón (por ejemplo, IFN-γ, IFN-α, IFN-β, IFN-ω, IFN-τ); una insulina (por ejemplo, Novolina, Humulina, Humalog, Lantus, Ultralente, etc.); una eritropoyetina (por ejemplo, PROCIT®, EPREX® o EPOGEN® (epoetina-α), ARANESP® (darbepoietina-α), NEORECORMON®, EPOGIN® (epoetina-β) y similares); un anticuerpo (por ejemplo, un anticuerpo monoclonal) (por ejemplo, RITUXAN® (rituximab), REMICADE® (infliximab), HERCEPTIN® (trastuzumab), HUMIRA^{MR} (adalimumab), XOLAIR® (omalizumab), BEXXAR® (tositumomab), RAPTIVA^{MR} (efalizumab), ERBITUX^{MR} (cetuximab) y similares), incluyendo un fragmento de unión a antígeno de un anticuerpo monoclonal; un factor sanguíneo (por ejemplo, activador de plasminógeno de tejido ACTIVASE® (alteplasa), NOVOSEVEN® (factor recombinante humano VIIa), factor VIIa, factor VIII (por ejemplo, KOGENATE®), factor IX, β-globina, hemoglobina y similares); factor de estimulación de colonias (por ejemplo, NEUPOGEN® (filgrastim, G-CSF); Neulasta (pegfilgrastim); factor estimulador de colonias de granulocitos (G-CSF), factor de estimulación de colonias de granulocitos-monocitos, factor de estimulación de colonias de macrófagos, factor de estimulación de colonias de megacariocitos; y similares); una hormona de crecimiento (por ejemplo, una somatotropina, por ejemplo, GENOTROPIN®, NUTROPIN®, NORDITROPIN®, SAIZEN®, SEROSTIM®, HUMATROPE®, etc.), una hormona de crecimiento humana y similares; una interleuquina (por ejemplo, IL-1, IL-2, incluyendo, 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® (beclaperrina, PDGF), FIBLAST® (trafermina, bFGF), STEMGEN® (ancestima, factor de células madre), factor de crecimiento de queratinocitos, un factor de crecimiento de fibroblastos ácido, un factor de células madre, un factor de crecimiento básico de fibroblastos, un factor de crecimiento de hepatocitos, y similares); un receptor soluble (por ejemplo, un receptor soluble de unión a TNF-α tal como ENBREX® (etanercept), un receptor soluble de VEGF, un receptor de interleucina soluble, un receptor de células T γ/δ solubles, y similares); una enzima (por ejemplo, α-glucosidasa, CERAZYME® (imiglucarasa, β-glucocerebrosidasa, CEREDASE® (alglucerasa), un activador enzimático (por ejemplo, activador del plasminógeno tisular), una quimoquina (por ejemplo IP-10, Mig, Groα/IL-8, RANTES, MIP-1α,

MIP-1 β , MCP-1, PF-4, y similares); un agente angiogénico (por ejemplo, un factor de crecimiento endotelial vascular (VEGF), un agente antiangiogénico (por ejemplo, un receptor de VEGF soluble), una proteína vacuna, un péptido neuroactivo tal como bradiquinina, colecistoquinina, gastina, secretina, oxitocina, hormona liberadora de gonadotropina, beta-endorfina, encefalina, sustancia P, somatostatina, galanina, hormona liberadora de hormona de crecimiento, bombesina, warfarina, dinorfina, neurotensina, inotilina, tiotropina, neuropéptido Y, hormona luteinizante, calcitonina, insulina, glucagón, vasopresina, angiotensina II, hormona liberadora de tiotropina, péptido intestinal vasoactivo, un péptido de sueño, etc.; otras proteínas tales como un agente trombolítico, un péptido natriurético atrial, proteína morfogénica del hueso, trombopoyetina, relaxina, proteína ácida fibrilar glial, hormona estimulante de folículos, una antitripsina alfa-1 humana, un factor inhibidor de la leucemia, un factor de crecimiento transformante, un factor tisular, un factor de crecimiento similar a la insulina, una hormona luteinizante, una hormona estimulante de folículos, un factor activador de macrófagos, un factor de necrosis tumoral, un factor quimiotáctico de neutrófilos, un factor de crecimiento de nervios, un inhibidor tisular de metaloproteinasas; un péptido intestinal vasoactivo, angiogenina, angiotropina, fibrina; hirudina; un factor inhibidor de la leucemia; un antagonista del receptor de IL-1 (por ejemplo, Kineret® (Anakinra)); y similares. Se apreciará fácilmente que las formas nativas de las proteínas terapéuticas anteriores también son 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 deseada de la proteína parental correspondiente. Ejemplos de ensayos útiles para proteínas terapéuticas particulares incluyen, pero no se limitan a, GMCSF (Eaves, AC y Eaves CJ, Erythropoiesis in culture.) En: McCulloch EA (ed.) Cell culture Techniques--Clinics in hematology. WB Saunders, Eastbourne, páginas 371-91 (1984), Metcalf, D., International Journal of Cell Cloning 10: 116-25 (1992), Testa, NG, y colaboradores, Assays for hematopoietic growth factors. En: Balkwill FR (ed.) Cytokines A practical Approach, páginas 229-44; IRL Press Oxford 1991) EPO (bioensayo: Kitamura y colaboradores, J. Cell. Physiol., 140 página 323 (1989)); Hirudina (ensayo de agregación plaquetaria: Blood Coagul Fibrinolysis 7 (2): 259-61 (1996)); IFN α (ensayo antiviral: Rubinstein y colaboradores, J. Virol. 37 (2): 755-8 (1981); ensayo antiproliferativo: Gao Y, y colaboradores, Mol Cell Biol. 19 (11): 7305-13 (1999); y bioensayo: Czarniecki y colaboradores, J. Virol. 49 p 490 (1984)); GCSF (bioensayo: Shirafuji y colaboradores, Exp. Hematol, 17 p 116 (1989); proliferación de células NFS-60 murinas (Weinstein y colaboradores, Proc Natl Acad Sci 83: 5010-4 (1986); insulina (ensayo de captación de ^3H -glucosa: Steppan y colaboradores, Nature 409 (6818): 307-12 (2001)); hGH (ensayo de proliferación de Ba/F3-hGHR: J Clin Endocrinol Metab 85 (11): 4274-9 (2000); estándar Internacional para la hormona del crecimiento: Horm Res, 51 Suplemento 1: 7-12 (1999)); factor X (ensayo de actividad del factor X: Van Wijk y colaboradores, Thromb Res 22: 681-686 (1981); factor VII (ensayo de coagulación usando tiempo de coagulación de protrombina: Belaaouaj y colaboradores, J. Biol. Chem. 275: 27123-8 (2000); Diaz-Collier y colaboradores, Thromb Haemost 71: 339-46 (1994)).

Composiciones inmunogénicas

La tecnología basada en etiqueta de aldehído descrita en este documento también encuentra aplicación en la producción de componentes de composiciones inmunogénicas (por ejemplo, vacunas terapéuticas). Por ejemplo, se puede usar una etiqueta de aldehído para facilitar la unión de porciones que aumentan la semivida en suero de un antígeno polipeptídico, que aumentan la inmunogenicidad del polipéptido o que enlaza un antígeno no aminoácido con un portador polipeptídico. En este sentido, 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 provocar una respuesta inmune en el sujeto. También es de interés la modificación de antígenos que son útiles en la obtención de anticuerpos que pueden ser útiles como herramientas de investigación.

Otros ejemplos de polipéptidos de interés para la modificación usando una etiqueta o etiquetas de aldehído incluyen aquellos que son de interés para la detección o control funcional en un ensayo (por ejemplo, como una herramienta de investigación, en un ensayo de cribado de fármacos y similares). Los ejemplos de polipéptidos de este tipo incluyen receptores (por ejemplo, receptores acoplados a proteína G (GPCR, incluyendo GPCR huérfanos)), ligandos de receptor (incluyendo naturales y sintéticos), canales de proteína (por ejemplo, canales iónicos (por ejemplo, canales de potasio, canales de calcio, canales de sodio y similares) y otros polipéptidos. En una realización, la modificación de polipéptidos asociados a la superficie celular, tales como polipéptidos transmembrana) es de particular interés, particularmente cuando dicha modificación se lleva a cabo mientras el polipéptido está presente en una membrana. Los métodos para la modificación de un polipéptido marcado con aldehído en condiciones fisiológicas se describen más adelante.

Enzimas generadoras de formilglucina (FGE)

La enzima que oxida cisteína o serina en un motivo de sulfatasa a FGly se denomina aquí enzima generadora de formilglucina (FGE). Como se ha discutido anteriormente, en la presente memoria se utiliza "FGE" para referirse a enzimas generadoras de FGly que median la conversión de una cisteína (C) de un motivo de sulfatasa en FGly así como enzimas generadoras de FGly que median la conversión de serina (S) de un motivo de sulfatasa en FGly. Debe observarse que en general la literatura se refiere a enzimas generadoras de FGly que convierten una C en FGly en un

motivo de sulfatasa como FGE y se refiere a enzimas que convierten S en FGly en un motivo de sulfatasa similar a AtsB. Sin embargo, a los efectos de la presente descripción, se utiliza genéricamente "FGE" para referirse a ambos tipos de enzimas generadoras de FGly, entendiéndose que se seleccionará una FGE apropiada de acuerdo con el compañero reactivo objetivo que contiene el motivo de sulfatasa apropiado (es decir, que contiene C o que contiene S).

Como se pone de manifiesto por la presencia omnipresente de sulfatasas que tienen un FGly en el sitio activo, las FGE se encuentran en una amplia variedad de tipos celulares, incluyendo tanto eucariotas como procariotas. Hay al menos dos formas de FGE. Las sulfatasas eucariotas contienen una cisteína en su motivo de sulfatasa y son modificadas por la FGE de tipo "SUMF1" (Cosma y colaboradores, Cell 2003, 113, (4), 445-56; Dierks y colaboradores, Cell 2003, 113, (4), 435-44). La enzima generadora de FGly (FGE) está codificada por el gen *SUMF1*. Las sulfatasas procariotas pueden contener ya sea una cisteína o una serina en su motivo de sulfatasa y se modifican bien sea por la FGE de tipo "SUMF1" o bien por la FGE de tipo "AtsB", respectivamente (Szameit y colaboradores, J Biol Chem 1999, 274, (22), 15375-81). En los eucariotas, se cree que esta modificación ocurre en forma cotraduccional o poco después de la traducción en el retículo endoplásmico (ER) (Dierks y colaboradores, Proc Natl Acad Sci USA 1997, 94 (22): 11963-8). Sin estar sujeto a la teoría, en los procariotas se piensa que FGE de tipo SUMF1 actúa en el citosol y FGE de tipo AtsB actúa cerca o en la membrana celular. FGE de SUMF2 también se ha descrito en la deuterostomía, incluyendo vertebrados y equinodermos (véase, por ejemplo, Pepe y colaboradores (2003) Cell 113, 445-456; Dierks y colaboradores (2003) Cell 113, 435-444; Cosma y colaboradores (2004) Hum. Mutat, 23, 576-581).

En general, la FGE usada para facilitar la conversión de cisteína o serina en FGly en un motivo de sulfatasa de una etiqueta de aldehído de un polipéptido objetivo se selecciona de acuerdo con el motivo de sulfatasa presente en la etiqueta de aldehído. La FGE puede ser nativa de la célula huésped en la que se expresa el polipéptido marcado con aldehído, o la célula huésped puede modificarse genéticamente para expresar una FGE apropiada. En algunas realizaciones puede ser deseable utilizar un motivo de sulfatasa compatible con una FGE humana (por ejemplo, la FGE de tipo SUMF1, véase, por ejemplo, Cosma y colaboradores, Cell 113, 445-56 (2003); Dierks y colaboradores, Cell 113, 435-44 (2003)), y expresan la proteína marcada con aldehído en una célula humana que expresa la FGE o en una célula huésped, habitualmente una célula de mamífero, modificada genéticamente para expresar una FGE humana.

En general, una FGE para uso en los métodos descritos en la presente memoria se puede obtener a partir de fuentes naturales o ser producida sintéticamente. Por ejemplo, se puede derivar una FGE apropiada a partir de fuentes biológicas que producen naturalmente una FGE o que se modifican genéticamente para expresar un gen recombinante que codifica una FGE. Los ácidos nucleicos que codifican una serie de FGE son conocidos en la técnica y se encuentran fácilmente disponibles (véase, por ejemplo, Preusser y colaboradores, 2005, J. Biol. Chem. 280 (15): 14900-10 (Epub 2005, 18 de enero); Fang y colaboradores 2004 J Biol Chem. 79 (15): 14570-8 (Epub 2004, 28 de enero); Landgrebe y colaboradores, Gene, 2003 Oct 16, 316: 47-56; Dierks y colaboradores, 1998 FEBS Lett. 423 (1): 61-5; Dierks y colaboradores, Cell. 2003, 16 de mayo, 113 (4): 445-56; Baenziger (2003, 16 de mayo), Cell 113 (4): 421-2 (revisión); Dierks y colaboradores, Cell, 2005, 20 de mayo, 121 (4): 541-52; Roeser y colaboradores (2006, 3 de enero) Proc Natl Acad Sci USA 103 (1): 81-6; Sardiello y colaboradores (2005 Nov 1) Hum Mol Genet 14 (21): 3203-17; WO 2004/072275 y acceso al GenBank No. NM_182760. Por consiguiente, la divulgación proporciona aquí células huésped recombinantes genéticamente modificadas para expresar una FGE que es compatible para uso con una etiqueta de aldehído de un polipéptido objetivo etiquetado.

En una realización, se usa una FGE obtenida de *Mycobacterium tuberculosis* (Mtb) en los métodos descritos en la presente memoria. Un ejemplo de FGE de Mtb se describe en detalle en los Ejemplos más adelante. Un ejemplo de FGE de Mtb es aquella que tiene la secuencia de aminoácidos proporcionada en el GenBank con acceso No. NP_215226 (gi: 15607852):

```
mltelvdlpg gsfrmgstrf ypeeapihtv tvrafaverh pvtnaqfaef vsatgyvtva
eqpldpglyp gvdaadlcp amvfcptagp vldrdwrqww dwvpgacwrh pfggrdsdiad
raghpvvqva ypdavayarw agrrlpteae weyaarggtt atyawgdqek pggmlmantw
qgrfpyrndg algwvgtspv grfpangfgl ldmignvwew tttefyphhr idppstacca
pvklataadp tisqtlkggs hlcapeychr yrpaarspqs qdtatthigf rcvadpvsq
```

Por lo tanto, FGE de Mtb, y el ácido nucleico que codifica FGE de Mtb, se contemplan para uso en los presentes métodos. Además, los métodos utilizados para identificar y caracterizar la FGE de Mtb son aplicables a la identificación y caracterización de otras FGE útiles en los métodos descritos en el presente documento.

Con la amplia información sobre la secuencia de aminoácidos y la caracterización de las FGE proporcionados en la presente memoria, así como en el estado de la técnica, será fácilmente evidente para el especialista en la materia que las FGE incluyen FGE de origen natural así como FGE modificadas que comparten la identidad de secuencia con una FGE conocida (por ejemplo, una FGE de origen natural) y que conservan la función en la modificación específica de una serina o cisteína de un motivo de sulfatasa.

- En general, las FGE de interés incluyen aquellas que tienen al menos 60%, usualmente 75%, usualmente 80%, más usualmente 90%-95% de identidad de residuos de nucleótidos o de aminoácidos, cuando se comparan y se alinean para correspondencia máxima con una secuencia de nucleótidos o secuencia de aminoácidos de una FGE parental, medido utilizando un algoritmo de comparación de secuencias disponible en la técnica o mediante inspección visual.
- 5 Habitualmente, existe una identidad de secuencia mencionada sobre una región de las secuencias que tiene al menos aproximadamente 50 residuos de longitud, más habitualmente sobre una región de al menos aproximadamente 100 residuos, y más habitualmente sobre al menos aproximadamente 150 residuos hasta la longitud completa de la región de codificación o proteína, con la condición de que la región de comparación incluya un sitio activo de la FGE requerida para la actividad enzimática.
- 10 Para la comparación de secuencias, típicamente una secuencia actúa como una secuencia de referencia, con la que se comparan las secuencias de ensayo. Cuando se utiliza un algoritmo de comparación de secuencias, las secuencias de prueba y de referencia se introducen en un ordenador, se designan coordenadas de subsecuencia, si es necesario, y se designan parámetros de programa de algoritmo de secuencia. El algoritmo de comparación de secuencias calcula luego el porcentaje de identidad de secuencia para la secuencia o secuencias de prueba con respecto a la secuencia de
- 15 referencia, con base en los parámetros de programa designados.
- La alineación óptima de secuencias para comparación puede llevarse a cabo, por ejemplo, mediante el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2: 482 (1981), mediante el algoritmo de alineación de homología de Needleman y Wunsch, J. Mol. Biol. 48: 443 (1970), mediante la búsqueda del método de similitud de Pearson y Lipman, Proc. Nat'l. Acad. Sci. USA 85: 2444 (1988), mediante implementaciones informáticas de estos
- 20 algoritmos (GAP, BESTFIT, FASTA y TFASTA en el Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), o mediante inspección visual (véase, en general, Current Protocols in Molecular Biology, FM Ausubel y colaboradores, eds., Current Protocols, una empresa conjunta entre Greene Publishing Associates, Inc. y John Wiley & Sons, Inc. (suplemento de 1995) (Ausubel)).
- Ejemplos de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y similitud de
- 25 secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul y colaboradores (1990) J. Mol. Biol. 215: 403-410 y Altschul y colaboradores (1977) Nucleic Acids Res. 25: 3389-3402, respectivamente. El software para realizar análisis BLAST está disponible públicamente a través del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica primero identificar pares de secuencias de alta puntuación (HSP) mediante la identificación de palabras cortas de longitud W en la secuencia de consulta, que coinciden o satisfacen algún umbral T de valor positivo cuando están alineados con una palabra de la misma longitud en una secuencia de
- 30 base de datos. T se conoce como el umbral de puntuación de la palabra vecina (Altschul y colaboradores, citado más arriba).
- Estos aciertos iniciales de palabra vecina actúan como semillas para iniciar búsquedas para encontrar HSP más largas que las contienen. Los aciertos de palabras se extienden entonces en ambas direcciones a lo largo de cada secuencia, hasta que la puntuación de alineación acumulativa puede aumentarse. Las puntuaciones acumulativas se calculan usando, para secuencias de nucleótidos, los parámetros M (puntuación de recompensa para un par de residuos coincidentes, siempre > 0) y N (puntuación de penalización para los residuos que no coinciden, siempre < 0). Para las
- 35 secuencias de aminoácidos, se utiliza una matriz de puntuación para calcular la puntuación acumulada. La extensión de los aciertos de palabra en cada dirección se detienen cuando: la puntuación de alineación acumulativa disminuye en una cantidad X desde su valor máximo alcanzado; la puntuación acumulada baja hasta cero o por debajo, debido a la acumulación de una o más alineaciones de residuo de puntuación negativa; o se alcanza el final de cada secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y la velocidad de la alineación. El programa BLASTN (para secuencias de nucleótidos) utiliza de forma predeterminada una longitud de palabra (W) de 11, una expectativa (E) de 10, M = 5, N = -4, y una comparación de ambas cadenas. Para las secuencias de aminoácidos, el
- 40 programa BLASTP usa de forma predeterminada una longitud de palabra (W) de 3, una expectativa (E) de 10 y la matriz de puntuación BLOSUM62 (véase Henikoff y Henikoff, Proc. Nat'l. Acad. Sci. USA 89: 10915 (1989)).
- Además de calcular el porcentaje de identidad de secuencia, el algoritmo BLAST también realiza un análisis estadístico de la similitud entre dos secuencias (véase, por ejemplo, Karlin y Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5787 (1993)). Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P
- 50 (N)), que proporciona una indicación de la probabilidad por medio de la cual una coincidencia entre dos secuencias de nucleótidos o de aminoácidos se produciría por casualidad. Por ejemplo, un ácido nucleico se considera similar a una secuencia de referencia si la probabilidad de suma más pequeña en una comparación del ácido nucleico de prueba con el ácido nucleico de referencia es aproximadamente menor a 0,1, más usualmente aproximadamente menor a 0,01 y más usualmente aproximadamente menor a 0,001.
- 55 Las posiciones de residuos que no son idénticas pueden diferir por sustituciones de aminoácidos conservadoras, que serán fácilmente evidentes a partir del análisis de las alineaciones como se discutió anteriormente. Las sustituciones conservadoras de aminoácidos se refieren a la intercambiabilidad de residuos que tienen cadenas laterales similares. Por ejemplo, los grupos de aminoácidos que definen residuos que pueden ser intercambiados por otro residuo dentro del grupo y constituyen una sustitución conservativa de aminoácidos incluyen un grupo de aminoácidos que tienen

5 cadenas laterales alifáticas son glicina, alanina, valina, leucina, prolina e isoleucina ("aminoácido alifático"); un grupo de aminoácidos que tienen cadenas laterales de hidroxilo alifático es serina y treonina ("aminoácido de hidroxilo alifático", que también están incluidos dentro de "aminoácidos polares no cargados"); un grupo de aminoácidos que tienen cadenas laterales que contienen amida es asparagina y glutamina ("aminoácido que contiene amida", "que también
10 están comprendidos dentro de "aminoácidos polares no cargados"); un grupo de aminoácidos que tiene 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 no cargados (a pH fisiológico) incluye serina, treonina, asparagina y glutamina ("aminoácidos no cargados polares"); 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 de aminoácidos conservadores se ejemplifican por: valina-leucina-isoleucina, fenilalanina-tirosina, lisina-arginina, alanina-valina y asparagina-glutamina.

15 Cuando se usan métodos libres de células para convertir un polipéptido que contiene un motivo de sulfatasa, se puede usar una FGE aislada. Pueden usarse cualquiera de los procedimientos de purificación de proteínas convenientes para aislar una FGE; véase, por ejemplo, Guide to Protein Purification, (Deuthser ed.) (Academic Press, 1990). Por ejemplo, se puede preparar un lisado a partir de una célula que produce una FGE deseada, y se purifica usando HPLC, cromatografía de exclusión, electroforesis en gel, cromatografía de afinidad, y similares.

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

20 La descripción proporciona un ácido nucleico que codifica etiquetas de aldehído y polipéptidos marcados con aldehído, así como constructos y células huésped que contienen ácido nucleico. Dichos ácidos nucleicos comprenden una secuencia de ADN que tiene un marco de lectura abierto que codifica una etiqueta de aldehído o un polipéptido marcado con aldehído y, en la mayoría de las realizaciones, es capaz, en condiciones apropiadas, de ser expresado. "Ácido nucleico" abarca ADN, ADNc, ARNm, y vectores que comprenden dichos ácidos nucleicos.

25 En la presente invención se proporcionan ácidos nucleicos que codifican etiquetas de aldehído, así como polipéptidos marcados con aldehído. Dichos ácidos nucleicos incluyen ADN genómicos modificados por inserción de una secuencia que codifica una etiqueta de aldehído y ADNc de polipéptidos marcados con aldehído. El término "ADNc" tal como se usa en el presente documento, pretende incluir todos los ácidos nucleicos que comparten la disposición de elementos de secuencia encontrados en una especie de ARNm maduro nativo (incluyendo variantes de empalme), donde los
30 elementos de secuencia son exones y regiones no codificantes 3' y 5'. Normalmente, las especies de ARNm tienen exones contiguos, con los intrones que intervienen, cuando están presentes, siendo removidos mediante el empalme de ARN nuclear, para crear un marco de lectura abierto continuo que codifica una proteína de acuerdo con la presente invención.

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

45 Los ácidos nucleicos contemplados en la presente invención pueden proporcionarse como parte de un vector (también denominado constructo), una amplia variedad de los cuales son conocidos en la técnica y no necesitan ser elaborados aquí. Los ejemplos de vectores incluyen, pero no se limitan a, plásmidos; cósmidos; vectores virales (por ejemplo, vectores retrovirales); vectores no virales; cromosomas artificiales (de YAC, de BAC, etc.); minicromosomas; y similares.

50 La elección del vector dependerá de una diversidad de factores tales como el tipo de célula en la que se desea la propagación y el propósito de la propagación. Ciertos vectores son útiles para amplificar y elaborar grandes cantidades de la secuencia de ADN deseada. Otros vectores son adecuados para expresión en células en cultivo. Incluso otros vectores son adecuados para la transferencia y expresión en células en un animal completo. La elección del vector apropiado está dentro de los conocimientos del experto en la técnica. Muchos de estos vectores están disponibles comercialmente.

55 Para preparar los constructos, se inserta un polinucleótido en un vector, típicamente por medio de la unión de ADN ligasa a un sitio de enzima de restricción escindido en el vector. Alternativamente, la secuencia de nucleótidos deseada puede insertarse mediante recombinación homóloga o recombinación en el sitio específico. Típicamente, la recombinación homóloga se logra uniendo regiones de homología con el vector en los flancos de la secuencia de

nucleótidos deseada, mientras que la recombinación en un sitio específico puede lograrse mediante el uso de secuencias que facilitan la recombinación en el sitio específico (por ejemplo, sitios cre-lox, att, etc.). El ácido nucleico que contiene dichas secuencias puede añadirse, por ejemplo, mediante ligación de oligonucleótidos, o mediante reacción en cadena de la polimerasa usando cebadores que comprenden tanto la región de homología como una porción de la secuencia deseada de nucleótidos.

Los vectores pueden permitir el mantenimiento extracromosómico en una célula huésped o pueden permitir 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, incluyendo, por ejemplo, Short Protocols in Molecular Biology, (1999) F. Ausubel y colaboradores, eds., Wiley & Sons. Los vectores pueden permitir la expresión de los ácidos nucleicos que codifican un polipéptido de interés (por ejemplo, un polipéptido marcado con aldehído, una FGE, etc.), pueden permitir la propagación de los ácidos nucleicos objetivo, o ambos.

Los ejemplos de vectores que se pueden usar incluyen pero no se limitan a aquellos derivados de ADN recombinante de bacteriófago, ADN de plásmido o ADN de cósmido. Por ejemplo, pueden usarse vectores plasmídicos tales como pBR322, pUC 19/18, pUC 118, 119 y la serie mp M13 de vectores. Los vectores bacteriófagos pueden incluir λ gt10, λ gt11, λ gt18-23, λ ZAP/R y la serie EMBL de vectores bacteriófagos. Los vectores cósmidos que pueden utilizarse incluyen, pero no se limitan a, pJB8, pCV 103, pCV 107, pCV 108, pTM, pMCS, pNNL, pHSG274, COS202, COS203, pWE15, pWE16 y la serie de vectores charomid 9. Alternativamente, se pueden modificar vectores de virus recombinantes, incluyendo pero no limitándose a los derivados de virus tales como virus del herpes, retrovirus, virus vacuna, virus de la viruela, adenovirus, virus adeno-asociados o virus del papiloma bovino.

Para la expresión de un polipéptido de interés, se puede emplear un casete de expresión. Por lo tanto, la presente invención proporciona un vector de expresión recombinante que comprende un ácido nucleico sujeto. El vector de expresión proporciona secuencias reguladoras transcripcionales y traduccionales, y puede permitir 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 de traducción. Estas regiones de control pueden ser nativas del gen que codifica el polipéptido (por ejemplo, el polipéptido objetivo o la FGE), o pueden derivarse de fuentes exógenas. En general, las secuencias reguladoras de la transcripción y la traducción pueden incluir, pero no se limitan a, secuencias promotoras, sitios de unión ribosómicos, secuencias de inicio y terminación de la transcripción, secuencias de inicio y terminación de la traducción y secuencias potenciadoras o activadoras. Además de los promotores constitutivos e inducibles, los promotores fuertes (por ejemplo, T7, CMV y similares) encuentran uso en los constructos descritos en la presente memoria, particularmente cuando se desean altos niveles de expresión en un sistema de expresión *in vivo* (basado en células) o *in vitro*. Otros ejemplos de promotores incluyen promotores del virus del tumor mamario de ratón (MMTV), promotores del virus del sarcoma de Rous (RSV), promotores de adenovirus, el promotor del gen temprano inmediato del CMV humano (Boshart y colaboradores, Cell 41: 521-530, 1985), y el promotor de la repetición terminal larga (LTR) de RSV (Gorman y colaboradores, Proc. Natl. Acad. Sci. USA 79: 6777-6781, 1982). El promotor también puede proporcionarse, por ejemplo, con un 5'UTR de un retrovirus.

Los vectores de expresión generalmente tienen sitios de restricción convenientes situados cerca de la secuencia promotora para permitir la inserción de secuencias de ácido nucleico que codifican proteínas de interés. Puede estar presente un marcador seleccionable operativo en el huésped de expresión para facilitar la selección de las células que contienen el vector. Además, el constructo de expresión puede incluir elementos adicionales. Por ejemplo, el vector de expresión puede tener uno o dos sistemas de replicación, permitiendo así que se mantenga en organismos, por ejemplo en células de mamífero o de insecto para expresión y en un huésped procariota para clonación y amplificación. Además, el constructo de expresión puede contener un gen marcador seleccionable para permitir la selección de células huésped transformadas. Los genes de selección son bien conocidos en la técnica y variarán con la célula huésped utilizada.

También se proporciona aquí un casete de etiqueta de aldehído, que incluye un ácido nucleico que codifica una etiqueta de aldehído y sitios de restricción adecuados que flanquean la secuencia que codifica la etiqueta para la inserción en el marco de un ácido nucleico que codifica un polipéptido objetivo. Tal constructo de expresión puede permitir la adición de una etiqueta de aldehído en el extremo terminal N o en el extremo terminal C de un polipéptido objetivo. El casete de la etiqueta de aldehído puede estar operativamente enlazado a una secuencia promotora para permitir la expresión del polipéptido marcado con aldehído resultante y puede incluir además uno o más marcadores seleccionables.

La presente descripción también proporciona casetes de expresión para la producción de polipéptidos marcados con aldehído (por ejemplo, con una etiqueta de aldehído situada en un extremo terminal N, en un extremo terminal C). Tales casetes de expresión incluyen generalmente un primer ácido nucleico que comprende una secuencia que codifica una 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 situarse 5' y/o 3' de la secuencia que codifica la etiqueta de aldehído. La inserción de la secuencia que codifica el polipéptido en el marco con la secuencia que codifica la etiqueta del aldehído permite la producción de un ácido nucleico recombinante que codifica una proteína de fusión que es un polipéptido marcado con aldehído como se describe en la presente memoria. Los constructos que contienen dicho casete de expresión generalmente incluyen también un promotor operativamente enlazado al casete de expresión para

permitir la expresión del polipéptido producido marcado con aldehído. Otros componentes del constructo de expresión pueden incluir marcadores seleccionables y otros elementos adecuados.

Los constructos de expresión que codifican polipéptidos marcados con aldehído también pueden generarse usando métodos de amplificación (por ejemplo, reacción en cadena de la polimerasa (PCR)), donde al menos un cebador de amplificación (es decir, al menos uno de un cebador directo o inverso) incluye una secuencia de ácido nucleico que codifica una etiqueta de aldehído. Por ejemplo, un cebador de amplificación que tiene una secuencia codificadora de etiqueta de aldehído está diseñado para permitir la amplificación de un ácido nucleico que codifica un polipéptido objetivo de interés. El producto de extensión que resulta de la síntesis mediada por polimerasa a partir del cebador directo que contiene etiqueta de aldehído produce un producto de amplificación de ácido nucleico que codifica una proteína de fusión compuesta de un polipéptido objetivo marcado con aldehído. El producto de amplificación se inserta entonces en un constructo de expresión de elección para proporcionar un constructo de expresión de polipéptido etiquetado con aldehído.

Células huésped

Puede utilizarse cualquiera de un número de células huésped adecuadas en la producción de un polipéptido marcado con aldehído. La célula huésped utilizada para la producción de un polipéptido marcado con aldehído puede permitir opcionalmente la conversión mediada por FGE, de manera que el polipéptido producido contiene una etiqueta de aldehído que contiene FGly después de la expresión y modificación postraduccional por FGE. Alternativamente, la célula huésped puede permitir la producción de un polipéptido no convertido marcado con aldehído (por ejemplo, debido a la falta de expresión de una FGE que facilita la conversión de la etiqueta de aldehído).

En general, los polipéptidos descritos en la presente memoria pueden expresarse en procariotas o eucariotas de acuerdo con los procedimientos convencionales, dependiendo del propósito de la expresión. Por lo tanto, la presente invención proporciona además una célula huésped, por ejemplo, una célula huésped genéticamente modificada, que comprende un ácido nucleico que codifica un polipéptido marcado con aldehído. La célula huésped puede comprender además opcionalmente una FGE recombinante, que puede ser endógena o heteróloga a la célula huésped.

Las células huésped para la producción (incluyendo la producción a gran escala) de un polipéptido no convertido o (cuando la célula huésped expresa una FGE adecuada) convertido marcado con aldehído, o para la producción de una FGE (por ejemplo, para su uso en un método sin células) pueden seleccionarse de cualquiera de una variedad de células huésped disponibles. Ejemplos de células huésped incluyen aquellas de un organismo unicelular procariota o eucariota, tal como bacterias (por ejemplo, cepas de *Escherichia coli*, *Bacillus* spp. (por ejemplo, *B. subtilis*) y similares) levaduras u hongos (por ejemplo, *S. cerevisiae*, *Pichia* spp., y similares), y pueden usarse otras de tales células huésped. Como ejemplos de células huésped originalmente derivadas de un organismo superior tal como de insectos, vertebrados, particularmente mamíferos, (por ejemplo, CHO, HEK y similares), pueden usarse como las células huésped de expresión.

Los sistemas de expresión específicos de interés incluyen sistemas de expresión derivados de células bacterianas, levaduras, células de insectos y mamíferos. A continuación se presentan sistemas representativos de cada una de estas categorías.

Bacterias. Los sistemas de expresión en bacterias incluyen los descritos en Chang y colaboradores, *Nature* (1978) 275: 615; Goeddel y colaboradores, *Nature* (1979) 281: 544; Goeddel y colaboradores, *Nucleic Acids Res.* (1980) 8: 4057; el documento EP 0 036 776; la patente estadounidense No. 4.551.433; DeBoer y colaboradores, *Proc. Natl. Acad. Sci. (EE.UU.)* (1983) 80: 21-25; y Siebenlist y colaboradores, *Cell* (1980) 20: 269.

Levadura. Los sistemas de expresión en levadura incluyen los descritos en Hinnen y colaboradores, *Proc. Natl. Acad. Sci. (EE.UU.)* (1978) 75: 1929; Ito y colaboradores, *J. Bacteriol.* (1983) 153: 163; Kurtz y colaboradores, *Mol. Cell. Biol.* (1986) 6: 142; Kunze y colaboradores, *J. Basic Microbiol.* (1985) 25: 141; Gleeson y colaboradores, *J. Gen. Microbiol.* (1986) 132: 3459; Roggenkamp y colaboradores, *Mol. Gen. Genet.* (1986) 202: 302; Das y colaboradores, *J. Bacteriol.* (1984) 158: 1165; De Louvencourt y colaboradores, *J. Bacteriol.* (1983) 154: 737; Van den Berg y colaboradores, *Bio/Technology* (1990) 8: 135; Kunze y colaboradores, *J. Basic Microbiol.* (1985) 25: 141; Cregg y colaboradores, *Mol. Cell. Biol.* (1985) 5: 3376; las patentes estadounidenses Nos. 4.837.148 y 4.929.555; Beach y Nurse, *Nature* (1981) 300: 706; Davidow y colaboradores, *Curr. Genet.* (1985) 10: 380; Gaillardin y colaboradores, *Curr. Genet.* (1985) 10: 49; Ballance y colaboradores, *Biochem. Biophys. Res. Commun.* (1983) 112: 284-289; Tilburn y colaboradores, *Gene* (1983) 26: 205-221; Yelton y colaboradores, *Proc. Natl. Acad. Sci. (EE.UU.)* (1984) 81: 1470-1474; Kelly y Hynes, *EMBO J.* (1985) 4: 475479; el documento EP 0 244 234; y el documento WO 91/00357.

Células de insecto. La expresión de genes heterólogos en insectos se lleva a cabo como se describe en la patente estadounidense No. 4.745.051; Friesen y colaboradores, "The Regulation of Baculovirus Gene Expression", en: *The Molecular Biology of Baculovirus* (1986) (W. Doerfler, ed.); el documento EP 0 127 839; el documento EP 0 155.476; y Vlak y colaboradores, *J. Gen. Virol.* (1988) 69: 765-776; Miller y colaboradores, *Ann. Rev. Microbiol.* (1988) 42: 177;

Carbonell y colaboradores, *Gene* (1988) 73: 409; Maeda y colaboradores, *Nature* (1985) 315: 592-594; Lebacqz-Verheyden y colaboradores, *Mol. Cell. Biol.* (1988) 8: 3129; Smith y colaboradores, *Proc. Natl. Acad. Sci. (EE.UU.)* (1985) 82: 8844; Miyajima y colaboradores, *Gene* (1987) 58: 273; y Martin y colaboradores, *DNA* (1988) 7: 99. Numerosas cepas y variantes de baculovirus y las correspondientes células huésped permisivas de insecto de huéspedes se describen en Luckow y colaboradores, *Bio/Technology* (1988) 6: 47-55, Miller y colaboradores, *Generic Engineering* (1986) 8: 277-279, y Maeda y colaboradores, *Nature* (1985) 315: 592-594.

Células de mamífero. La expresión de mamífero se lleva a cabo como se describe en Dijkema y colaboradores, *EMBO J.* (1985) 4: 761, Gorman y colaboradores, *Proc. Natl. Acad. Sci. (EE.UU.)* (1982) 79: 6777, Boshart y colaboradores, *Cell* (1985) 41: 521 y la patente estadounidense No. 4.399.216. Otras características de la expresión en mamíferos se facilitan como se describe en Ham y Wallace, *Meth. Enz.* (1979) 58: 44, Barnes y Sato, *Anal. Biochem.* (1980) 102: 255, patentes estadounidenses Nos. 4.767.704, 4.657.866, 4.927.762, 4.560.655, documentos WO 90/103430, WO 87/00195 y US RE. 30.985.

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

El producto puede ser recuperado por cualquier medio apropiado conocido en la técnica. Además, se pueden emplear procedimientos convenientes de purificación de proteínas, donde se describen metodologías de purificación de proteínas adecuadas en *Guide to Protein Purification*, (Deutsher ed.) (Academic Press, 1990). Por ejemplo, puede prepararse un lisado a partir de una célula que comprende el vector de expresión que expresa un polipéptido de interés, y se purifica usando HPLC, cromatografía de exclusión, electroforesis en gel, cromatografía de afinidad y similares.

Porciones para la modificación de polipéptidos

Los polipéptidos marcados con aldehído, que contienen FGly pueden ser sometidos a modificación para permitir la fijación de una amplia variedad de porciones. Los ejemplos de moléculas de interés incluyen, pero no están necesariamente limitadas a, un marcador detectable, una molécula pequeña, un péptido y similares.

La porción de interés se proporciona como el componente de un compañero reactivo para la reacción con un aldehído del residuo de FGly de una etiqueta de aldehído convertida del polipéptido marcado. Dado que los métodos de modificación de polipéptidos marcados son compatibles con procedimientos químicos convencionales, los métodos de la invención pueden explotar una amplia gama de reactivos comercialmente disponibles para lograr la unión de una porción de interés a un residuo de FGly de un polipéptido marcado con aldehído. Por ejemplo, los derivados de aminooxi, hidrazida, hidracina o tiosemicarbazida de una cantidad de porciones de interés son compañeros reactivos adecuados y están fácilmente disponibles o pueden generarse usando métodos químicos estándar.

Por ejemplo, se puede generar un aminooxi-PEG a partir de los monoamino-PEG y aminooxiglicina usando protocolos estándar. El aminooxi-PEG puede entonces hacerse reaccionar con un polipéptido convertido marcado con aldehído para permitir la unión de la porción de PEG. La administración de una porción de biotina a un polipéptido convertido marcado con aldehído puede llevarse a cabo usando aminooxi biotina, biotina hidrazida o 2,4-dinitrofenilhidracina.

Dada la presente descripción, el experto en la materia pueda adaptar fácilmente cualquiera de una variedad de porciones para proporcionar un compañero reactivo para la conjugación con un polipéptido marcado con aldehído como se contempla en la presente memoria. Los expertos en la técnica apreciarán que son importantes factores tales como el pH y el impedimento estérico (es decir, la accesibilidad de la etiqueta de aldehído a la reacción con un compañero reactivo de interés). La modificación de las condiciones de reacción para proporcionar condiciones óptimas de conjugación están dentro de la habilidad del experto en la materia, y es rutinario en la técnica. En general, es normalmente deseable llevar a cabo las reacciones de conjugación a un pH inferior a 7, con un pH de aproximadamente 5,5, aproximadamente 6, aproximadamente 6,5, usualmente siendo óptimo aproximadamente 5,5. Cuando la conjugación se lleva a cabo con un polipéptido marcado con aldehído presente en o sobre una célula viva, las condiciones se seleccionan de manera que sean fisiológicamente compatibles. Por ejemplo, se puede dejar caer temporalmente el pH durante 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 30 minutos a 1 hora). Las condiciones fisiológicas para llevar a cabo la modificación de polipéptidos marcados con aldehído en una superficie celular pueden ser similares a las utilizadas en una reacción de cetona-azida en la modificación de células que portan azidas en la superficie de las células (véase, por ejemplo, la patente estadounidense No. 6.570.040).

En general, la porción o porciones pueden proporcionar una o más de una amplia variedad de funciones o características. Los ejemplos de fragmentos incluyen marcadores detectables (por ejemplo, marcadores colorantes (por ejemplo, cromóforos, fluoróforos), sondas biofísicas (etiquetas de spin, sondas de RMN), etiquetas de tipo FRET (por ejemplo, al menos un miembro de un par FRET, incluyendo al menos un miembro de un par fluoróforo/inactivador,

etiquetas 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 marcado al nivel del tejido o de las células moleculares (por ejemplo, asociación con un tipo de tejido, o una membrana celular particular) y similares); porciones dinámicas activadas por luz (por ejemplo, cierre de poros mediado por azobenceno, cambios estructurales mediados por azobenceno, motivos de reconocimiento de fotodecaimiento); polímeros solubles en agua (por ejemplo, PEGilación); etiquetas de purificación (por ejemplo, para facilitar el aislamiento por cromatografía de afinidad (por ejemplo, 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, incluyendo una unión selectiva); fármacos (por ejemplo, para facilitar el direccionamiento de fármacos, por ejemplo, a través de la unión del fármaco a un anticuerpo); porciones de suministro objetivo (por ejemplo, ligandos para la unión a un receptor objetivo) (por ejemplo, para facilitar la fijación viral, unión de una proteína objetivo presente en un liposoma, etc.), y similares.

A continuación se proporcionan ejemplos específicos no limitativos.

Etiquetas detectables

Las composiciones y métodos de la invención se pueden usar para suministrar un marcador detectable a un polipéptido marcado con aldehído. Ejemplos de marcadores detectables incluyen, pero no se limitan necesariamente, a moléculas fluorescentes (por ejemplo, moléculas autofluorescentes, moléculas que fluorescen al contacto con un reactivo, etc.), marcadores radiactivos (por ejemplo, ^{111}In , ^{125}I , ^{131}I , ^{212}B , ^{90}Y , ^{186}Rh , y similares); biotina (por ejemplo, que se va a detectar a través de la reacción de biotina y avidina); etiquetas fluorescentes; reactivos de formación de imágenes, y similares. Los marcadores detectables también incluyen péptidos o polipéptidos que pueden detectarse mediante unión a anticuerpos, por ejemplo, mediante la unión de un anticuerpo marcado en forma detectable o mediante la detección de anticuerpo unido a través de un ensayo tipo sándwich.

Unión de moléculas objetivo a un soporte.

Los métodos pueden permitir la conjugación de un polipéptido marcado 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 separación (por ejemplo, un hapteno reconocido por un anticuerpo unido a una perla magnética). En una realización, los métodos de la invención se usan para permitir la unión de una proteína a una matriz (por ejemplo, un chip) en una orientación definida. Por ejemplo, se puede generar un polipéptido que tiene una etiqueta de aldehído en un sitio seleccionado (por ejemplo, en o cerca del extremo terminal N) y los métodos y composiciones de la invención utilizados para suministrar una porción a la etiqueta convertida de aldehído. La porción puede usarse entonces como el sitio de unión para fijar el polipéptido a un soporte (por ejemplo, un soporte sólido o semisólido, particularmente un soporte adecuado para uso como microchip en ensayos de alto rendimiento).

Unión de moléculas para su suministro a un sitio objetivo

El compañero reactivo para el polipéptido marcado con aldehído puede comprender un fármaco de molécula pequeña, toxina u otra molécula para suministrar a la célula y que puede proporcionar una actividad farmacológica o puede servir como objetivo para el suministro de otras moléculas.

También se contempla el uso de un compañero reactivo que comprende uno de un par de compañeros de unión (por ejemplo, un ligando, una porción de unión a un ligando de un receptor, una porción de unión al receptor de un ligando, etc.). Por ejemplo, el compañero reactivo puede comprender un polipéptido que sirve como receptor viral y, al unirse con una proteína de envoltura viral o proteína de cápside viral, facilita la unión del virus a la superficie celular en la que se expresa la proteína modificada marcada con aldehído. Alternativamente, el compañero reactivo comprende un antígeno que está específicamente unido por un anticuerpo (por ejemplo, anticuerpo monoclonal), para facilitar la detección y/o separación de células huésped que expresan el polipéptido modificado marcado con aldehído.

Polímeros solubles en agua

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 es usualmente sustancialmente no inmunogénico, y usualmente tiene un peso molecular atómico mayor a aproximadamente 1.000 Daltons. Los métodos y composiciones descritos aquí se pueden usar para unir uno o más polímeros solubles en agua a un polipéptido marcado con aldehído. Puede ser deseable la unión de un polímero soluble en agua (por ejemplo, PEG) de un polipéptido, particularmente un polipéptido (terapéutico) farmacéuticamente activo, ya que dicha modificación puede aumentar el índice terapéutico aumentando la semivida en suero como resultado de una mayor estabilidad proteolítica y/o menor eliminación renal. Adicionalmente, la unión de uno o más polímeros (por ejemplo, PEGilación) puede reducir la inmunogenicidad de los productos farmacéuticos proteicos.

En algunas realizaciones, el polímero soluble en agua tiene un peso molecular hidrodinámico efectivo superior a aproximadamente 10.000 Da, superior a aproximadamente 20.000 a 500.000 Da, superior a aproximadamente 40.000 Da a 300.000 Da, superior a aproximadamente 50.000 Da a 70.000 Da, usualmente superior a aproximadamente 60.000 Da. Por "peso molecular hidrodinámico efectivo" se entiende el tamaño eficaz solvatado en agua de una cadena polimérica determinada por cromatografía de exclusión por tamaño de base acuosa (SEC). Cuando el polímero soluble en agua contiene cadenas poliméricas que tienen unidades de repetición de óxido de polialquileno, tales como unidades de repetición 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 de 2.000 a aproximadamente 20.000 Da. A menos que se haga referencia específicamente, se pretende que el peso molecular se refiera al peso molecular atómico. Los polímeros solubles en agua lineales, ramificados y terminalmente cargados (por ejemplo, PEG) son de particular interés.

Los polímeros útiles como porciones que se van a unir a un polipéptido marcado con aldehído pueden tener una amplia gama de pesos moleculares y subunidades poliméricas. Estas subunidades pueden incluir un polímero biológico, un polímero sintético, o una combinación de los mismos. Ejemplos de tales polímeros solubles en agua incluyen: dextrano y derivados de dextrano, incluyendo sulfato de dextrano, dextrina entrelazada con P-amino, carboximetil dextrina, celulosa y derivados de celulosa, incluyendo metilcelulosa y carboximetilcelulosa, almidón y dextrinas y derivados e hidrolizados de almidón, polialquilenglicol y sus derivados, incluyendo polietilenglicol, metoxi-polietilenglicol, homopolímeros de polietilenglicol, homopolímeros de polipropilenglicol, copolímeros de etilenglicol con propilenglicol, en donde dichos homopolímeros y copolímeros están sustituidos o no sustituidos en un extremo con un grupo alquilo, heparina y fragmentos de heparina, alcohol polivinílico y éteres etil polivinílicos, polivinilpirrolidona, aspartamida y polioles polioxietilados, con dextrano y derivados de dextrano, dextrina y derivados de dextrina. Se apreciará que también se contemplan diversos derivados de los polímeros hidrosolubles específicamente citados.

Los polímeros solubles en agua tales como los descritos anteriormente son bien conocidos, particularmente los polímeros a base de óxido de polialquileno tales como polietilenglicol "PEG" (véase, por ejemplo, "Poly(ethylene glycol) Chemistry: Biotechnical and Biomedical Applications", JM Harris, Ed., Plenum Press, New York, NY (1992) y "Poly(ethylene glycol) Chemistry and Biological Applications", JM Harris y S. Zalipsky, Eds., ACS (1997); y las solicitudes internacionales de patente WO90/13540, 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 WO 99/31371, 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 estadounidenses Nos. 4.179.337, 5.075.046, 5.089.261, 5.100.992, 5.134.192, 5.116.309, 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).

Ejemplos de polímeros de interés incluyen aquellos que contienen un óxido de polialquileno, óxido de alquilén poliamida, o derivados de los mismos, incluyendo óxido de polialquileno y óxido de alquilén poliamida que comprende una unidad de repetición de óxido de etileno de la fórmula $-(CH_2-CH_2-O)-$. Otros ejemplos de polímeros de interés incluyen una poliamida que tiene un peso molecular superior a aproximadamente 1.000 Daltons de la fórmula $-[C(O)-XC(O)-NH-Y-NH]_n-$ o $-[NH-Y-NH-C(O)-XC(O)]_n-$, donde X e Y son radicales divalentes que pueden ser iguales o diferentes y pueden ser ramificados o lineales y n es un número entero discreto de 2-100, usualmente de 2 a 50, y donde cualquiera o ambos de X e Y comprenden una unidad de repetición hidrosoluble biocompatible, sustancialmente no antigénica que puede ser lineal o ramificada. Otros ejemplos de unidades de repetición solubles en agua comprenden un óxido de etileno de la fórmula $-(CH_2-CH_2-O)-$ o $-(CH_2-CH_2-O)-$. El número de tales unidades repetitivas hidrosolubles puede variar significativamente, siendo el número usual de tales unidades de 2 a 500, 2 a 400, 2 a 300, 2 a 200, 2 a 100 y, más usualmente, 2 a 50. Un ejemplo de realización es aquel en el que uno o ambos de X e Y se seleccionan de: $-((CH_2)_{n1}-(CH_2-CH_2-O)_{n2}-(CH_2)-$ o $-(CH_2)_{n1}-(O-CH_2-CH_2)_{n2}-(CH_2)_{n-1}-$, donde n1 es 1 a 6, 1 a 5, 1 a 4 y más usualmente 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 usualmente 2 a 5. Un ejemplo de realización adicional es aquella en la que X es $-(CH_2-CH_2)-$, y donde Y es $-(CH_2-(CH_2-CH_2-O)_3-CH_2-CH_2)-$ o $-(CH_2-CH_2-CH_2-(O-CH_2-CH_2)-$.

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

La porción polimérica, o uno o más de los espaciadores o enlazadores de la porción polimérica cuando está presente, puede incluir cadenas o unidades poliméricas que son bioestables o biodegradables. Por ejemplo, los polímeros con enlaces repetidos tienen diferentes grados de estabilidad bajo condiciones fisiológicas dependiendo de la labilidad del enlace. Los polímeros con tales enlaces se pueden clasificar por sus velocidades relativas de hidrólisis bajo condiciones fisiológicas basadas en tasas de hidrólisis conocidas de análogos de bajo peso molecular, por ejemplo, de menos

- estables a más estables, por ejemplo poliuretanos (-NH-C(O)-O-) > poliortoésteres (-O-C((OR)(R'))-O-) > poliamidas (-C(O)-NH-). De forma similar, los sistemas de enlace que unen un polímero soluble en agua a una molécula objetivo pueden ser bioestables o biodegradables, por ejemplo, desde menos estables hasta más estables: carbonato (-O-C(O)-O-) > éster (-C(O)-O-) > uretano (-NH-C(O)-O-) > orto-éster (-O-C((OR)(R'))-O-) > amida (-C(O)-NH-). En general, puede ser deseable evitar el uso de polisacárido sulfatado, dependiendo de la labilidad del grupo sulfato. Además, puede ser menos deseable utilizar policarbonatos y poliésteres. Estos enlaces se proporcionan a título de ejemplo y no pretenden limitar los tipos de enlaces que pueden ser empleados en las cadenas de polímero o sistemas de enlace de los polímeros solubles en agua útiles en los polipéptidos modificados marcados con aldehído descritos en la presente memoria.
- 5
- 10 Métodos para la conversión y modificación de una etiqueta de aldehído
- Conversión de una etiqueta de aldehído presente en un polipéptido marcado con aldehído lograda mediante métodos basados en células (*in vivo*) o libres de células (*in vitro*). De forma similar, la modificación de una etiqueta convertida de aldehído de un polipéptido marcado con aldehído puede realizarse por métodos basados en células (*in vivo*) o libres de células (*in vitro*). Estos se describen con más detalle más abajo.
- 15 Conversión y modificación de células huésped "in vivo"
- La conversión de una etiqueta de aldehído de un polipéptido marcado con aldehído puede realizarse mediante la expresión del polipéptido marcado con aldehído en una célula que contiene una FGE adecuada. En esta realización, 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 realización, la FGE de la célula huésped puede ser endógena a la célula huésped, o la célula huésped puede ser recombinante para una FGE adecuada que sea heteróloga para la célula huésped. La expresión de FGE puede proporcionarse mediante un sistema de expresión endógeno del gen FGE (por ejemplo, la expresión es proporcionada por un promotor y otros elementos de control presentes en el gen nativo de FGE de la célula huésped), o puede proporcionarse mediante un sistema de expresión recombinante en el que la secuencia de codificación de FGE está operativamente enlazada a un promotor heterólogo para proporcionar una expresión constitutiva o inducible. El uso de un promotor fuerte para proporcionar altos niveles de expresión de FGE puede ser de particular interés en alguna realización.
- 20
- 25
- Dependiendo de la naturaleza del polipéptido objetivo que contiene la etiqueta de aldehído, después de la conversión, el polipéptido convertido marcado con aldehído es o bien retenido en la célula huésped intracelularmente, es secretado, o está asociado con la membrana extracelular de la célula huésped. Cuando la etiqueta de aldehído del polipéptido marcado con aldehído está presente en la superficie celular, la modificación de la etiqueta convertida de aldehído puede llevarse a cabo utilizando un compañero reactivo para unir una porción del compañero reactivo a un residuo de FGly de una etiqueta de aldehído accesible en la superficie bajo condiciones fisiológicas. Las condiciones adecuadas de uso para lograr la conjugación de una porción de compañero reactivo con un polipéptido marcado con aldehído son similares a las descritas en Mahal y colaboradores (1997, 16 de mayo) Science 276 (5315): 1125-8.
- 30
- 35 Conversión y modificación "in vitro" (libre de células)
- La conversión *in vitro* (libre de células) de una etiqueta de aldehído de un polipéptido marcado con aldehído puede realizarse poniendo en contacto un polipéptido marcado con aldehído con una FGE en condiciones adecuadas para la conversión de una cisteína o serina de un motivo de sulfatasa de la etiqueta de aldehído a una FGly. Por ejemplo, el ácido nucleico que codifica un polipéptido marcado con aldehído puede ser expresión en un sistema de transcripción/traducción *in vitro* en presencia de una FGE adecuada para permitir la producción de polipéptidos convertidos marcados con aldehído.
- 40
- Alternativamente, el polipéptido aislado, no convertido marcado con aldehído puede ser aislado después de la producción recombinante en una célula huésped que carece de una FGE adecuada o mediante producción sintética. El polipéptido aislado marcado con aldehído se pone entonces en contacto con una FGE adecuada bajo condiciones que permitan la conversión de la etiqueta de aldehído. En esta realización, si la etiqueta de aldehído puede no ser fácilmente accesible por el disolvente en el polipéptido aislado, el polipéptido marcado con aldehído puede ser desplegado por métodos conocidos en la técnica (por ejemplo, usando calor, ajuste de pH, agentes caotrópicos, (urea y similares), disolventes orgánicos (por ejemplo, hidrocarburos: octano, benceno, cloroformo), etc.) y la proteína desnaturalizada en contacto con una FGE adecuada. El polipéptido convertido marcado con aldehído puede entonces replegarse bajo condiciones adecuadas.
- 45
- 50
- Con respecto a la modificación del polipéptido convertido marcado con aldehído, la modificación se lleva a cabo normalmente *in vitro*. El polipéptido convertido marcado con aldehído se aísla de una fuente de producción (por ejemplo, producción de células huésped recombinantes, producción sintética) y se pone en contacto con un compañero reactivo en condiciones adecuadas para permitir la conjugación de una porción del compañero reactivo con el FGly de la

etiqueta de aldehído. Si la etiqueta de aldehído no es accesible al disolvente, el polipéptido marcado con aldehído puede desplegarse por métodos conocidos en la técnica antes de la reacción con el compañero reactivo.

Porciones conmutables unidas a etiqueta de aldehído

5 En algunas realizaciones, los polipéptidos marcados con aldehído pueden ser modificados de una forma que facilite la remoción 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 explora la estabilidad termodinámica relativa de los conjugados formados con diferentes compañeros reactivos.

10 Por ejemplo, como se ilustra en la Figura 6, los aldehídos reaccionan fácilmente con porciones de hidrazida y aminooxi para producir hidrazonas y oximas, respectivamente. Aunque ambos conjugados son robustos en condiciones fisiológicas, las oximas son más estables termodinámicamente. Además, las tiosemicarbazidas también reaccionan fácilmente con aldehídos para formar conjugados de tiosemicarbazona, que son menos estables termodinámicamente que las oximas. Estas diferencias en la estabilidad termodinámica pueden explotarse para conmutar el conjugado de hidrazona de estabilidad más baja a un conjugado de oxima más estable y para conmutar el conjugado de oxima de menor 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 siguiente.

Polipéptidos modificados marcados con aldehído

20 Los productos de reacción producidos mediante reacción de un polipéptido marcado con aldehído con un compañero reactivo que comprende una porción de interés se modifican generalmente en un sitio específico (es decir, en el residuo de FGly), proporcionando una población sustancialmente homogénea de polipéptidos modificados marcados con aldehído. Las poblaciones heterogéneas de tales productos de reacción pueden generarse mediante el uso de dos o más compañeros reactivos que comprenden diferentes porciones, según se desee.

25 Por ejemplo, cuando el polipéptido objetivo se modifica por PEGilación, se pueden adaptar los métodos adaptarse para permitir 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

30 Se proporcionan kits y sistemas para facilitar y, cuando se desea, estandarizar las composiciones de la invención y sus usos. Los kits contemplados en la presente memoria pueden incluir uno o más de un constructo que codifica una etiqueta de aldehído para su inserción en un polipéptido objetivo; un constructo que codifica un polipéptido marcado con aldehído para su expresión en una célula huésped (por ejemplo, como un casete de expresión para permitir la inserción de una secuencia codificante de un polipéptido objetivo como una fusión en el extremo terminal N o en el extremo terminal C 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 genéticamente modificada para expresar un polipéptido marcado con aldehído de interés, cuya célula huésped puede expresar además una FGE endógena, recombinante o heteróloga compatible para la conversión de la etiqueta de aldehído del polipéptido marcado; y un compañero reactivo para la modificación química de la etiqueta convertida de aldehído del polipéptido marcado.

40 Además, el kit puede contener instrucciones para usar los componentes del kit, particularmente las composiciones de la invención que están contenidas en el kit.

Ejemplos

45 Los siguientes ejemplos se presentan para proporcionar a los expertos en la técnica una divulgación y descripción completa de cómo elaborar y usar la presente invención, y no pretenden limitar el alcance de lo que los inventores consideran como su invención ni pretenden representar que los experimentos siguientes son todos o los únicos experimentos realizados. Se han hecho esfuerzos para asegurar la exactitud con respecto a los números utilizados (por ejemplo, cantidades, temperatura, etc.), pero algunos errores experimentales y desviaciones deben tenerse en cuenta. A menos que se indique lo contrario, las partes son partes en peso, el peso molecular es el peso molecular medio en peso, la temperatura es en grados centígrados y la presión está en o cerca de la atmosférica.

Métodos y materiales:

50 Se usaron los siguientes materiales y métodos en los Ejemplos 1-6 expuestos a continuación.

Construcción de plásmidos. Se usaron los siguientes oligonucleótidos en los ejemplos a continuación:

Cebador	Secuencia (5' → 3') ^a
ald ₆ - <i>stf0</i> inicio ^b	CCACTGTGCACACCATCGCGGATGTCCGACCACCCACCGCC
ald ₁₃ - <i>stf0</i> sentido	CATGGCACCACTGTGCACACCATCGCGGGGCTCGCTGTTACCGGCCG CGACGTCCA
ald ₁₃ - <i>stf0</i> antisentido	TATGGACGTCGCGGCCGGTGAACAGCGAGCCCCGCGATGGTGTGCACA GTGGTGC
ald- <i>stf0</i> final	GCACCACCACCACCACCACTGAGATCCGGCTGC
ald- <i>stf0</i> Cys5Ala ^b	CCATGGCACCACTGGCCACACCATCGCGG
<i>stf0</i> final ^b	CGGCCGCGATGTGCGCCTTGAAGATCTGC
<i>mbp</i> sentido	GATCCCTGTGCACACCATCGCGGTGAGC
<i>mbp</i> sin sentido	GGCCGCTCACCGCGATGGTGTGCACAGG
ald- <i>mbp</i> Cys5Ala ^b	CCGCGTGGATCCCTGGCCACACCATCGCGG
ald ₆ - <i>hGH</i> inicio	CTATGCTACCATGGCGCTGTGCACACCATCGCGGACCATTCCCTTATCC AGGC
<i>hGH</i> final	CTATGCTAGCGGCCGCGAAGCCACAGCTGCCCTCCAC

(continuación)

Cebador	Secuencia (5' → 3') ^a
ald ₆ - <i>hGH</i> Cys5Ala ^b	TATACCATGGCGCTGGCCACACCATCGCGGACC
<i>fge</i> inicio	CTATGCTACCATGGCTGACCGAGTTGGTTGACCTGC
<i>fge</i> final	TAGCATAGCTCGAGCTACCCGGACACCGGGTCG
<i>fge</i> en marco ^b	GAGGAATTAACCATGCTGACCGAGTTGGTT,G
^a Las bases que codifican aldehído están subrayadas. ^b Cebador de mutagénesis dirigida al sitio. Cuando sea apropiado, se enumera desde el comienzo del respectivo codón de partida. Se usó un par de cebadores complementarios para cada mutante. Los complementos inversos no se muestran.	

Los motivos de sulfatasa de los constructos se proporcionan a continuación:

ald₁₃-Stf0: LCTPSRGSFLTGR- (sulfotransferasa micobacteriana)

ald₁₃-Stf0 (C5A): LATPSRGSFLTGR- (sulfotransferasa micobacteriana)

ald₆-Stf0: LCTPSR- (sulfotransferasa micobacteriana)

ald₆-MBP: LCTPSR-(proteína de unión a maltosa)

5 ald₆-hGH: LCTPSR- (hormona de crecimiento humano)

El ácido nucleico que codifica ald₁₃-Stf0 se construyó ligando oligonucleótidos hibridados en un vector¹⁹ pET28-Stf0 construido previamente entre los sitios de restricción NcoI y NdeI. El codón de detención *stf0* se eliminó mediante mutagénesis QuikChange^{MR} (Stratagene) para permitir construir una etiqueta His₆ del extremo terminal C. Se construyó ald₆-Stf0 utilizando QUICKCHANGE^{MR} (Stratagene) para eliminar los nucleótidos que codifican los últimos 7 residuos de la etiqueta de aldehído de 13 aminoácidos. Se construyó el gen que codifica ald₆-MBP ligando los oligonucleótidos hibridados en el vector¹⁹ pMALC-H entre los sitios de restricción XhoI y HindIII. Se amplificó el gen que codifica hGH (variante 1 del transcripto de hormona 1 de crecimiento humano, que codifica los residuos 29-217) a partir de pCMV-SPORT6.1.ccdB (Open Biosystems) usando un cebador 5' que codificó la etiqueta de aldehído de 6 aminoácidos y se ligó en pET28b entre los sitios de restricción NcoI y NotI. El gen que codifica FGE de *Mycobacterium tuberculosis* (Rv0712, que codifica los residuos 2-299) se amplificó a partir de un plásmido pET14b previamente preparado (que contenía FGE¹⁴ y se ligó en pBAD/myc-his A (Invitrogen) entre sitios de restricción NcoI y XhoI. Se colocó el gen de FGE en marco con el codón de inicio utilizando kit de mutagénesis de PCR QuikChange (Stratagene). Los mutantes Cys → Ala de ald-Stf0, ald-MBP, y ald-hGH se produjeron utilizando mutagénesis con QuikChange^{MR}. Se realizó la secuenciación del ADN para confirmar la fidelidad de cada producto génico. Los plásmidos que codifican la proteína se transformaron en células BL21 (DE3) *E. coli* (Invitrogen).

Expresión y purificación de proteínas. Las poblaciones clonales de células BL21 (DE3) de *E. coli* que albergan solamente un plásmido que codifica una proteína marcada con aldehído se incubaron en medio LB con kanamicina con agitación a 37°C hasta OD₆₀₀ = 0,5, momento en el que la temperatura se redujo a 18°C y se añadió IPTG 100 µM. Se incubaron las células BL21 (DE3) de *E. coli* que albergan un plásmido que codifica la proteína marcada con aldehído y un plásmido que codifica FGE en medio LB con kanamicina y ampicilina con agitación a 37°C hasta OD₆₀₀ = 0,5, momento en el que se indujo la expresión de FGE con arabinosa al 0,02%. Después de 30 min, la temperatura se redujo a 18°C y se añadió IPTG 100 µM para inducir la expresión de la proteína marcada con aldehído. Después de 12-16 h, se recogieron las células y se resuspendieron en 20 ml de regulador de lisis (Tris 50 mM, de NaCl 500 mM, glicerol al 10%, imidazol al 20 mM, DTT 1 mM, TCEP 1 mM, metionina 1 mM, pH 7,5; Ald₆-hGH o NaH₂PO₄ 50 mM, NaCl 300 mM, imidazol 10 mM, pH 7,4, para ald-Stf0 y ald₆-MBP) por litro de cultivo y se lisó mediante sonicación.

Los lisados celulares se trataron con DNasa (10 µg/ml), se aclararon por centrifugación y se aplicaron a una columna HisTrap de 1 ml (GE Healthcare). La columna se lavó con regulador de lisis con imidazol 35 mM y se eluyó la proteína marcada con His₆ usando regulador de lisis con imidazol 250 mM. Se purificó adicionalmente ald₆-hGH en una columna Sefadex 16/60 S300 (GE Healthcare).

35 Ensayo de digestión triptica y adición de estándar. Se digirieron 10 µg de proteína con 0,4 µg de tripsina (Promega) a 37°C durante 16 horas en NH₄HCO₃ 50 mM pH 8. Este protocolo se consideró suficiente para la digestión completa ya que no se detectaron péptidos que contuvieran sitios de escisión perdidos mediante espectrometría de masas MALDI-TOF después de 3 horas de digestión en condiciones idénticas. Los ensayos de adición de estándar se realizaron en agua con aproximadamente 0,6 µg de digestión de proteína por corrida. Se añadieron péptidos sintéticos que contenían ya sea la cisteína o el aldehído (FGly) en cantidades equimolares seguido por la adición de DTT 100 mM. Esta solución se dejó incubar a TA durante 1 hora antes del análisis por espectrometría de masas (Agilent MSD). Se añadieron corridas de blanco entre corridas seleccionadas aleatoriamente y no se detectó señal residual. No se observó oxidación de cisteína.

45 Etiquetado de moléculas pequeñas. Las reacciones de marcado fluorescente se realizaron con 10 µg de proteína objetivo con colorante aminooxi 300 µM (Alexa Fluor 647 C5-aminooxiacetamida, Invitrogen) en regulador de marcación (MES 100 mM pH 5,5, SDS al 1%) a 37°C durante 2 horas. Se añadió metoxilamina 166 mM a las reacciones de control. Las mezclas de reacción se separaron mediante SDS-PAGE y se detectó la fluorescencia usando un escáner Typhoon 9410 (GE Healthcare). La carga de proteínas se determinó mediante tinción con Sypro Ruby (Sigma). Se proporcionó la biotilación incubando 10 µg de proteína objetivo con biotina-hidrazida 30 µM (Sigma) en regulador de marcación durante 2 horas a 37°C. El desplazamiento subsiguiente de biotina-hidrazida se hizo mediante la adición de metoxilamina 166 mM o aminooxiFLAG 1 mM a 37°C durante 2 horas. La transferencia tipo Western de α-biotina se realizó usando un protocolo estándar. La transferencia de µ-FLAG se obtuvo pelando la membrana y probando nuevamente con µ-FLAG M2 (Sigma). Se sintetizó aminooxi-FLAG utilizando protocolos estándar de síntesis de péptidos en fase sólida basados en FMOC. El residuo final añadido, extremo terminal C, fue ácido (t-Boc-aminooxi) acético seguido por escisión en condiciones estándar.

5 PEGilación. Se crearon aminooxi-PEG a partir de monoamino-PEG y aminooxiglicina utilizando protocolos estándar. Más específicamente, se produjeron los aminooxiPEG añadiendo amino-PEG (Shearwater Polymers) al ácido (t-Boc-aminooxi)acético activado usando condiciones estándar de acoplamiento peptídico. En resumen, se logró la formación del enlace amida mediante la adición de aminoPEG al éster 8-hidroxibenzotriazol preformado del ácido (t-Boc-aminooxi)acético (5 equivalentes) en acetonitrilo. La purificación del producto se produjo por precipitación a partir del éter, seguido de trituración. La desprotección se realizó por tratamiento con una solución acuosa de ácido trifluoroacético (95% de TFA, 5% de H₂O) durante 3 horas a TA. La precipitación en éter y la trituración proporcionaron un producto puro según se juzgó por RMN ¹H. La conjugación con proteínas marcadas con aldehído se produjo mediante incubación de 10 µg de proteína objetivo y aminooxiPEG 10 mM en solución de acoplamiento (49,95% de CH₃CN, 49,95% de H₂O, 0,1% de TFA) durante 1 hora seguido de liofilización.

10 La conjugación con proteínas marcadas con aldehído se produjo mediante incubación de 10 µg de proteína objetivo y aminooxiPEG 10 mM en solución de acoplamiento (49,95% de CH₃CN, 49,95% de H₂O, 0,1% de TFA) durante 1 hora seguido de liofilización. Las mezclas de reacción se resuspendieron en agua, se separaron mediante SDS-PAGE, se tiñeron con Sypro Orange (Invitrogen) y se detectaron usando un escáner Typhoon 9410 (GE Healthcare).

15 Ejemplo 1: modificación en un sitio específico de un motivo de sulfatasa en proteína expresada en *E. Coli*

20 Los constructos de proteínas con etiquetas de aldehído en el extremo terminal C o N se expresaron en *E. coli*. Se exploraron tres objetivos de proteína, una proteína de unión a maltosa marcada en el extremo terminal C (MBP), una hormona de crecimiento humana (hGH) marcada en el extremo terminal N y una sulfotransferasa micobacteriana marcada en el extremo terminal N (Stf0). Además, se probaron dos variantes de la etiqueta de aldehído - una etiqueta de 13 residuos (ald₁₃-Stf0) que incluía todo el motivo de consenso de sulfatasa y una etiqueta de 6 residuos (ald₆-Stf0) que incluía una secuencia más corta que contenía un motivo de consenso de sulfatasa:

LCTPSRGSFLTGR-Stf0 (ald₁₃-Stf0)

LCTPSR-Stf0 (ald₆-Stf0)

25 Con el fin de asegurar la formación eficiente de FGly, se coexpresaron las proteínas marcadas con una FGE procariota de *Mycobacterium tuberculosis* (*Mtb*) (descrito en los Ejemplos a continuación).

La digestión triptica del péptido que contiene el motivo de consenso de sulfatasa de 13 residuos (ald₁₃-Stf0) permitió la identificación espectral de masa directa de FGly (Figura 2). Aunque el péptido que contenía FGly podría ser fácilmente identificado, no se observó el péptido que contenía cisteína, indicando una oxidación eficaz de la etiqueta de aldehído.

30 Ejemplo 2: Los péptidos que contienen el motivo de consenso de 6 residuos de sulfatasa demuestran altas tasas de conversión

Para cuantificar el grado de conversión de Cys a FGly, se realizó un ensayo de adición estándar. Los niveles relativos de cisteína y FGly dentro de los péptidos típicamente derivados de las proteínas objetivo se compararon con una curva de adición estándar, que se produjo dopando péptidos sintéticos en digestos tripticos a diversas concentraciones (Figura 3, panel a).

35 Inesperadamente, el péptido Stf0 que contiene el motivo de consenso de 6 residuos de sulfatasa (ald₆-Stf0), demostró una conversión ligeramente superior a la del péptido Stf0 que contiene la secuencia conservada de 13 aminoácidos (ald₁₃-Stf0), con niveles de conversión de 92 ± 3% y 86 ± 5%, respectivamente. Este resultado contrasta con los estudios anteriores de sulfatasa que indican que la secuencia distal de treonina-glicina-arginina (TGR) es importante para la oxidación eficaz de cisteína (Dierks y colaboradores, EMBO J. 18 (8): 2084-91). El péptido de hGH que contiene el motivo de consenso de 6 residuos de sulfatasa (ald₆-hGH) demostró una conversión significativamente mayor a 99 ± 5%.

Ejemplo 3: La conversión de los motivos de sulfatasa por FGE es independiente del contexto de la secuencia primaria, permitiendo por lo tanto el posicionamiento de las etiquetas de aldehído ya sea en las posiciones del extremo terminal N o el extremo terminal C dentro de un polipéptido

45 Estudios anteriores de sulfatasa han indicado que la secuencia distal de treonina-glicina-arginina (TGR) en el motivo consenso de 13 residuos de sulfatasa (subrayado a continuación) es importante para la oxidación de cisteína de alto nivel:

LCTPSRGSFLTTGR

Puesto que se piensa que la formación de FGly ocurre cotraduccionalmente (Dierks y colaboradores, Proc Natl Acad Sci EE.UU. 1997, 28 de octubre, 94 (22): 11963-8), se razonó que los constructos del extremo terminal C podrían experimentar menor formación de FGly debido a la inaccesibilidad de la etiqueta de aldehído. Esto se probó generando un polipéptido marcado en el extremo terminal C que contenía proteína de unión a maltosa y el motivo de consenso de 6 residuos de sulfatasa (ald₆-MBP):

LCTPSR- (proteína de unión a maltosa)

Sorprendentemente, el ald₆-MBP marcado en extremo terminal C también demostró una conversión casi cuantitativa del 99 ± 2%. Teniendo en cuenta que el motivo de sulfatasa se encuentra en forma nativa en el interior de las sulfatasas, estos resultados indican que la formación de aldehído no está limitada con respecto a la posición de la secuencia primaria de la etiqueta.

Ejemplo 4: marcación fluorescente selectiva de proteínas etiquetadas con aldehído

Para demostrar la especificidad proporcionada por la introducción de FGly, se marcó un panel de proteínas marcadas con aldehído con colorante aminooxiacetamida ALEXA FLUOR® 647 (Invitrogen):

ald₁₃-Stf0 (C5A): LATPSRGSFLTGR- (sulfotransferasa micobacteriana)

ald₁₃-Stf0: LCTPSRGSFLTGR- (sulfotransferasa micobacteriana)

ald₆-Stf0: LCTPSR- (sulfotransferasa micobacteriana)

ald₆-MBP: LCTPSR-(proteína de unión a la maltosa)

Las proteínas etiquetadas con aldehído demostraron una marcación robusta de fluoróforo (Figura 3, panel b): ald₁₃-Stf0 (-), ald₆-Stf0 (-) y ald₆-MBP (-). Por el contrario, las proteínas de 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 marcación de fondo (Figura 3, paneles c y d: ald₁₃-Stf0 (C5A) (-)). Las proteínas etiquetadas con aldehído incubadas con un exceso de metoxilamina, un nucleófilo competidor, no demostraron marcación (Figura 3, paneles c y d, ald₁₃-Stf0 (C5A) (+), ald₁₃-Stf0 (+), ald₆-Stf0 (+), y ald₆-MBP (+)). Curiosamente, aunque el genoma de *E. coli* no contiene una FGE anotada, la proteína etiquetada con aldehído expresada sin FGE exógena todavía muestra marcación fluorescente, aunque con menor intensidad. Esto indica que *E. coli* debe expresar en forma nativa una enzima o enzimas que son capaces de oxidación del motivo de sulfatasa.

Ejemplo 5: La modificación de una proteína etiquetada con aldehído puede proporcionar porciones "conmutables"

Los aldehídos reaccionan fácilmente con porciones hidrazida y aminooxi para producir hidrazonas y oximas, respectivamente. Aunque ambos conjugados son robustos en condiciones fisiológicas, las oximas son más termodinámicamente estables. Esta diferencia puede ser explotada para conmutar el conjugado de hidrazona de estabilidad más baja por el conjugado de 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 ejemplifica mediante la conjugación de una etiqueta de purificación seguida por el reemplazo de la etiqueta de purificación conjugada para proporcionar un fluoróforo conjugado.

Para evaluar la viabilidad de esta técnica, primero se marcó un polipéptido que contenía proteína de unión a maltosa y el motivo de consenso de 6 residuos de sulfatasa (ald₆-MBP) con biotina-hidrazida y posteriormente se incubó con metoxilamina o una etiqueta de epítipo de aminooxi (aminooxi-FLAG). El etiquetado con biotina-hidrazida condujo a una señal robusta por α -biotina en una transferencia tipo western (Figura 4, panel a, carril 1). La incubación subsiguiente con metoxilamina o aminooxi-FLAG condujo a una pérdida completa de la señal de α -biotina (Figura 4, panel a, carril 2) o a una señal robusta de α -FLAG (Figura 4, panel a, carril 3), respectivamente.

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

Ejemplo 6: Creación de una pegilación en un sitio específico para producir conjugados de proteína PEG en una proteína terapéutica objetivo

Para ilustrar el uso de etiquetas de aldehído en la mediación de la PEGilación en un sitio específico, se usaron etiquetas de aldehído para unir en un sitio específico cadenas de polietilenglicol (PEG) a ald₆-Stf0 expresado de forma

recombinante. Se expresó de forma recombinante ald₆-Stf0 y se modificó con una serie de aminooxi-PEG con diferentes longitudes de cadena. El análisis SDS-PAGE de los conjugados Stf0-PEG demostró cambios de masa inequívocos consistentes con el peso molecular y la carga de las moléculas de PEG añadidas (Figura 4, Panel b). Estos resultados demuestran la facilidad de obtener conjugados PEG-proteína en un sitio específico independientemente del número de

Lo anterior proporciona una prueba del principio para la aplicación de etiquetas de aldehído para mediar la PEGilación en un sitio específico, por ejemplo, de proteínas terapéuticas. La PEGilación de proteínas farmacéuticas es deseable ya que puede incrementar el índice terapéutico incrementando la estabilidad proteolítica y disminuyendo la eliminación renal. Además, la PEGilación puede ser explotada para reducir la inmunogenicidad de los productos farmacéuticos proteicos. A pesar de los avances en la química de conjugación de proteínas, la modificación en un sitio específico de las proteínas sigue siendo problemática. La obtención de derivados de residuos de cisteína o lisina es actualmente el método más utilizado para PEGilar proteínas, pero este método de etiquetado no específico da lugar a PEGilación de múltiples sitios, creando una colección no deseada de conjugados discretos proteína-PEG con diferentes farmacocinéticas. La tecnología de etiqueta de aldehído descrita en la presente invención puede utilizarse para satisfacer necesidades tales como estas.

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

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

Se introdujo la etiqueta de aldehído de 6 residuos descrita anteriormente (ald₆ (LCTPSR)) en un constructo que codifica una proteína del canal SPARK. Se usaron tres estrategias: 1) adición del motivo de consenso de 6 residuos de sulfatasa de la etiqueta de ald dentro de uno de los bucles extracelulares de la proteína (denominada como "I" en la Figura 5); 2) suprimiendo 6 residuos del bucle y luego reemplazando estos residuos con la etiqueta ald de 6 residuos (denominada como "C" en la Figura 5), y 3) suprimiendo 3 residuos del bucle y luego añadiendo la etiqueta ald de 6 residuos (denominada como "P" en la Figura 5). También se corrió un control negativo sólo de vector ("V" en la Figura 5).

Los plásmidos que codifican cada una de las tres variantes del canal SPARK marcado con aldehído recombinante se transfectaron en células de ovario de hámster chino (CHO) y en células de riñón embrionario humano (HEK). Tanto las células CHO como HEK expresan una FGE endógena. Sin embargo, con el fin de aumentar la conversión de la Cys de la etiqueta de aldehído, se cotransfectó un plásmido que codifica SPARK marcado con aldehído con un constructo pcDNA3.1 que codifica FGE humana. La FGE humana utilizada tiene la secuencia de aminoácidos:

```
maapalglvc grcpelglvl lllllslcgc aagsqeagtg agagslagsc
gcgtpqrpga hgssaaahry sreanapgpv pgerqlahsk mvpipagvft
mgtdpdpqikg dgeaparrvt idafymdaye vsntefekfv nstgylteae
kfgdsfvfeg mlseqvktnt qqavaapww lpvkganwrh pegpdstilh
rpdhplvhlvs wndavayctw agkrlpteae weyscrgglh nrlfpwgnkl
qpkgqhyani wqgefptvnt gedgfggtap vdaftpngyg lynivgnawe
wtsdwwtvhh sveetlnpkg ppsgkdrvkk ggsymchrsy cyrrycaars
qntpdssasn lgfrcaadr1 ptmd
```

Después de un día, se lisaron las células y se probó el lisado mediante transferencia tipo Western para detectar la presencia de un epítipo myc (que está presente en la proteína del canal SPARK, y por lo tanto demuestra una transfección y traducción exitosas) y la presencia del aldehído por reacción con el uso de aminooxi-FLAG, seguido de sondeo con un anticuerpo anti-FLAG. Una transferencia de ponceau demostró una carga igual de muestras del mismo tipo de célula en la transferencia.

Como se muestra en la Figura 5, la estrategia que implica la supresión de 6 residuos del bucle extracelular de SPARK y el reemplazo con la etiqueta ald de 6 residuos fue exitosa (véase la flecha en la transferencia anti-FLAG en la Figura 5, panel c). Este resultado demuestra la capacidad de modificar los residuos de la superficie celular de una proteína marcada con aldehído en células de mamífero.

La presencia de FLAG de SPARK marcada con aldehído sobre la superficie de la célula puede ser confirmada usando citometría de flujo.

Ejemplo 8: Uso de etiquetas de aldehído para modificar el fragmento del anticuerpo Fc

- 5 Con el fin de demostrar adicionalmente las aplicaciones de etiquetas de aldehído, se modificó un fragmento Fc de IgG soluble para contener una etiqueta de aldehído en el extremo terminal N o el C. En resumen, se introdujo una etiqueta de aldehído de 13 residuos (ald₁₃) (LCTPSRAALLTGR) para situar la etiqueta de aldehído en el extremo terminal N o el extremo terminal C del fragmento Fc de IgG soluble codificado en el vector pFuse-Fc disponible comercialmente (Invitrogen). Con el fin de aumentar la conversión de la Cys de la etiqueta de aldehído, se cotransfectaron las células CHO con el constructo de codificación de Fc y un constructo pcDNA3.1 que codifica la FGE humana.
- 10 Se aislaron fragmentos Fc del sobrenadante celular y se realizó la detección del fragmento Fc de IgG marcado con aldehído en el que se convirtió la Cys en FGly haciendo reaccionar la proteína aislada con una sonda aminooxi-FLAG (DYKDDDDK) (FLAG-ONH₂), seguido de SDS-PAGE y análisis de transferencias tipo Western. Las proteínas que no reaccionaron con la sonda FLAG-ONH₂ sirvieron como un control adicional.
- 15 Mientras que las fusiones de Fc que contenían la etiqueta de aldehído de 12 mer produjeron una marcación robusta cuando estaban presentes ya sea en el extremo terminal N (N-Fc-Ald13) o en el extremo terminal C (C-Fc-Ald13), la proteína de control, en la que la cisteína crítica había sido mutada por alanina (mutación de C por A), no produjo una señal detectable (Figura 17).
- 20 Con el fin de evaluar si una etiqueta de aldehído de 6 mer es suficiente para mediar la modificación de una proteína, se generaron los fragmentos Fc de IgG que tienen una etiqueta de aldehído de 6 mer (Fc-Ald) o una etiqueta de control (Fc-C-> A) en el extremo terminal C usando el vector pFuse-Fc. Los fragmentos Fc de IgG marcados con aldehído se detectaron haciendo reaccionar la proteína aislada con una sonda de aminooxi-FLAG (FLAG-ONH₂), seguido por SDS-PAGE y transferencia tipo Western. Las proteínas que no reaccionaron con la sonda FLAG-ONH₂ sirvieron como un control adicional. Como se muestra en la Fig. 18, la etiqueta de aldehído de 6 mer facilitó la marcación robusta del Fc-Ald, mientras que no se observó marcación detectable con fragmentos Fc modificados para incluir la etiqueta de control.
- 25 Los constructos que codifican fragmentos Fc de IgG que tienen la posición de la etiqueta de aldehído de 6 mer en el extremo terminal N produjeron resultados similares (datos no mostrados).
- 30 Con el fin de confirmar la modificación de formilglicina (FGly) de los fragmentos Fc, se sometieron los fragmentos ald13-Fc marcados en el extremo terminal N o C a digestión trípica para permitir la identificación espectral directa de masa de FGly. Como se muestra en las Figs. 19 y 20, el péptido que contiene FGly y el péptido que contiene cisteína podrían identificarse fácilmente a partir tanto de fragmentos Fc modificados en el extremo terminal N como en el extremo terminal C.
- También se realizó un marcación específica de Fc marcado con aldehído sometiendo el medio exento de suero directamente a la sonda de aminooxi-FLAG (datos no mostrados).

Ejemplo 9: Eficiencia de la conversión de Cys en FGly en proteínas marcadas con aldehído

- 35 Con el fin de cuantificar el grado de conversión de Cys en FGly, se desarrolló un ensayo para analizar la eficacia de conversión de proteínas objetivo digeridas con tripsina. Se determinó la cantidad del péptido no modificado que contiene cisteína a partir de una curva estándar, que fue producida dopando péptidos sintéticos en digeridos trípticos a diversas concentraciones. La cantidad del péptido que contenía FGly se calculó restando la cantidad del péptido que contiene cisteína de la cantidad total de proteína, determinada usando el ensayo de proteína BCA.
- 40 Cuando se aplicó este ensayo al fragmento Fc marcado en el extremo terminal N y C descrito en el Ejemplo anterior, se encontró que en presencia de FGE humana exógena (hFGE) la eficiencia de conversión de Cys en FGly fue del $86 \pm 1\%$ para el ald13-Fc marcado con extremo terminal N y del $58 \pm 2\%$ para Fc-ald13 marcado en el extremo terminal C. Por el contrario, en ausencia de hFGE exógena, la eficiencia de conversión fue sólo de aproximadamente 25% y aproximadamente 23% para el fragmento Fc marcado con el extremo terminal N y C, respectivamente. El fragmento Fc
- 45 modificado en el extremo terminal C que contenía una etiqueta de aldehído de 6 mer mostró una eficiencia de conversión de aproximadamente 92% en presencia de hFGE exógena.

Ejemplo 10: Modificación medida por etiqueta de aldehído de las proteínas de la superficie celular

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

La etiqueta de aldehído de 13 mer (LCTPSRAALLTGR) o una etiqueta de control (LATPSRAALLTGR) se introdujo en un constructo de expresión pDisplay^{MR} (Invitrogen, Figura 21) entre los sitios Bgl II y Sal I. Las proteínas de fusión resultantes se denominan aquí Ald13-TM (que contienen la etiqueta aldehído de 13 mer) y C-> A-TM (que contienen la etiqueta de control). Este constructo de expresión y uno que expresada FGE humana (hFGE) se transfectaron transitoriamente en células HEK293-T para permitir la expresión.

La marcación de las células se realizó por reacción con una oxiamina biotina y se sondeó con conjugados de estreptavidina AlexaFluor 488. Las células se sometieron a continuación 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 Ald13-TM fue significativamente mayor (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 etiquetas de aldehído para marcación de proteína citosólica

Para ilustrar el uso de la etiqueta de aldehído en la marcación específica de proteínas citosólicas, se generaron constructos que codifican proteína fluorescente verde etiquetada con aldehído o etiquetada con control derivada de *Aequorea coerulescens* (AcGFP). Usando el vector pAcGFP1-N1 comercialmente disponible (Clontech), se generó un constructo de expresión que codifica una proteína de fusión AcGFP compuesta de una etiqueta His (seis residuos de histidina, representada por His₆) seguido de una etiqueta de aldehído de 13 mer (LCTPSRAALLTGR) o una etiqueta de control (LATPSRAALLTGR) posicionada en el extremo terminal N de AcGFP mediante la inserción de las secuencias de codificación de la etiqueta His y de la etiqueta de aldehído de 13 mer entre los sitios de restricción *Kpn I* y *Xma I*.

Se clonó un homólogo bacteriano de FGE derivado de *Streptomyces coelicolor* (StrepFGE) en un vector de expresión de mamífero (pcDNA 3.1, Invitrogen) por 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 utilizaron como un control adicional.

La detección de la AcGFP etiquetada con aldehído que contenía FGly se llevó a cabo haciendo reaccionar la proteína aislada con una sonda de aminooxi-FLAG (FLYON-ONH₂), seguido de SDS-PAGE y análisis de transferencias tipo Western. Las proteínas que no reaccionaron con la sonda FLAG-ONH₂ sirvieron como un control adicional.

En presencia del homólogo de FGE citosólico, se convirtió eficientemente el residuo de cisteína dentro de la secuencia de consenso con una formilglicina (FGly) (Figura 22), mientras que la AcGFP marcada de control no mostró un marcación detectable que indicaba que no había FGly detectable. Además, Ald-GFP producido en células HEK que no expresó StrepFGE también produjo una señal fuerte (Fig. 22). Esto puede deberse al método de aislamiento de proteínas utilizado en el que las células HEK se lisan y, de este modo, pueden liberar la hFGE de la ER de estas células, permitiendo así el contacto entre la hFGE y la etiqueta de aldehído que conduce a la conversión de cisteína con el aldehído.

Ejemplo 12: Modificación de la etiqueta de aldehído de IFN-beta

Las etiquetas de aldehído pueden usarse para facilitar la modificación de una variedad de proteínas. Los ejemplos de proteínas de interés para modificación incluyen interferón beta (IFN-beta). El IFN-beta está compuesto por cinco hélices alfa (A-E) con un único sitio de glicosilación existente en el residuo Asn-80. El IFN-beta puede modificarse para permitir la modificación en el sitio de glicosilación y/o en otros sitios accesibles al disolvente de la proteína. Por ejemplo, la secuencia de aminoácidos de IFN-beta que facilita la glicosilación puede modificarse para proporcionar una etiqueta de aldehído. Por ejemplo, usando técnicas recombinantes, la secuencia de IFN-beta DSSSTGWNE presente en un bucle de IFN-beta puede ser reemplazada por la secuencia que contiene la etiqueta de aldehído GSLCTPSRG. La etiqueta de aldehído puede explotarse para unir una porción de interés, como se ejemplifica en la Fig. 23.

Ejemplos 13-14: Identificación y caracterización de un FGE de *M. tuberculosis*

En los ejemplos siguientes, se identifica funcionalmente una FGE procariota en *Mycobacterium tuberculosis* (*Mtb*). Como se discutió anteriormente, las sulfatasas son miembros de una familia de enzimas en expansión que emplean nuevos cofactores derivados cotraduccionalmente o postraduccionalmente para facilitar la catálisis, y contienen un residuo de FGly en el sitio activo. Se cree que el residuo de FGly experimenta hidratación con el diol geminal, después de lo cual uno de los grupos hidroxilo actúa como un nucleófilo catalítico para iniciar la escisión del éster sulfato (Figura 7, Panel a). El residuo de FGly se encuentra dentro de una secuencia de consenso de sulfatasa, que define la familia de enzimas sulfatasa y está altamente conservada en todos los dominios de vida (Figura 7, Panel b). Mientras que FGly se forma a partir de residuos de cisteína en las sulfatasas eucariotas, tanto la cisteína (dentro del motivo central CXPXR) como la serina (SXPXR) se pueden oxidar a FGly en las sulfatasas procariotas. Algunos procariotas, tales como *Mtb*, codifican sólo las sulfatasas de tipo cisteína, mientras que otras especies sólo tienen 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 con 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 de FGE que aún no se han identificado. Definir el repertorio completo de sulfatasas de *Mtb* (y otros procariontes) es un importante objetivo futuro y proporcionará la plataforma para definir el papel de estas enzimas en el ciclo de vida y la patogénesis del organismo.

Métodos y Materiales

Se usaron los siguientes métodos y materiales en los ejemplos relacionados con la identificación de una FGE en *Mycobacterium tuberculosis* (*Mtb*), y la producción de una cepa de *Mtb* deficiente en FGE.

Preparación de vectores de expresión de proteínas. La siguiente tabla enumera los oligonucleótidos usados en los ejemplos siguientes. El gen que codifica FGE de *Mtb* (Rv0712, que codifica los residuos 2-299) se amplificó a partir del ADN genómico H37Rv de *Mtb* y se clonó en pET14b (Novagen) usando sitios de restricción NdeI y XhoI. El gen que codifica FGE de *Streptomyces* (SCO7548, que codifica los residuos 2-314) se amplificó a partir del ADN genómico de A3(2) de *Strep* y se clonó en pET151/D-TOPO (Invitrogen). Los marcos de lectura abierta 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*. Rv2407 se ligó en pMAL-C2X (New England Biolabs) usando sitios de restricción BamHI y PstI, y tanto Rv3406 como Rv3762c se ligaron en pET28b (Novagen) usando sitios de restricción NdeI y XhoI. Se realizó la secuenciación del ADN para confirmar la fidelidad de cada producto génico. Los plásmidos codificantes de proteínas se transformaron en células BL21 (DE3) (Invitrogen).

20 Cebadores oligonucleótidos

Cebador	Secuencia de cebador (5' → 3')
<i>Mtb fge</i> Inicio	CTATGCTACATATGCTGACCGAGTTGGTTGACCTGC
<i>Mtb fge</i> Final	TAGCATAGCTCGAGCTACCCGAGAGCGGGTTCG
<i>Strep fge</i> Inicio	CACCGCCGTGGCCGCCCGTCCCC
<i>Strep fge</i> Final	TCACTCAGCGGCTGATCCGG
<i>Mtb</i> Rv2407 Inicio	CTATGCTAGGATCCCTTGAGATCACGTTGCTCGG
<i>Mtb</i> Rv2407 Final	CTATGCTACTGCAGCTAGCGCCGCGGGTGACCTC
<i>Mtb</i> Rv3406 Inicio	CTATGCTACATATGACAGATCTGATTACCGTGAAG
<i>Mtb</i> Rv3406 Final	CTATGCTACTCGAGTCAGCCAGCGATCTCCATCG
<i>Mtb</i> Rv3762c Inicio	CTATGCTACATATGCCGATGGAACACAAACCTCC
<i>Mtb</i> Rv3762c Final	CTATGCTACTCGAGCTACGGCGTCACGATGTTGAAG
<i>Mtb fge</i> Ser260Ala	GACCCTCAAGGGCGGCGCACACCTGTGCGCGCCG
<i>Mtb fge</i> Cys263Sera	TCGCACCTGAGCGCGCCGGAGTACTGC
<i>Mtb fge</i> Cys268Sera	GCGCCGGAGTACAGCCACCGCTACCGC
<i>Strep fge</i> Trp234Ala	CACCGCGGGCAACGTGGCGGAATGGTGCTCCGAC
<i>Strep fge</i> Trp234Phea	CACCGCGGGCAACGTGTTTGAATGGTGCTCCGAC
<i>Strep fge</i> Cys272Sera	GGCGGCTCCTACCTGTCCCACGACTCCTACTGC
<i>Strep fge</i> Cys277Sera	GTGCCACGACTCCTACTCCAACCGCTACCGGGTTCG
<i>Mtb</i> Δfge en dirección 5'	CTATGCTAAAGCTTGAATCGAGTGAGATATrGCC

Mtb *Δfge* en dirección 3' TAGCATAGTCTAGAATGACGCTCGATCGAGAACG

Mtb *Δfge* en dirección 5' CTATGCTATCTAGATCCTCACAGTCGAGGACAGC

Mtb *Δfge* en dirección 3' TAGCATAGTTAATTAATGCACCATCTCGTTGCTCTCG

^a Numerado desde el principio del codón de inicio de FGE respectivo. Se utilizó un par de cebadores complementarios para cada mutante. No se muestran los complementos inversos; los cambios en la secuencia están subrayados.

- 5 Mutagénesis dirigida al sitio. Las mutaciones en un sitio específico en FGE de *Mtb* y FGE de *Strep* se produjeron utilizando un kit de mutagénesis de PCR QuikChange (Stratagene). Se usaron los plásmidos pET14b *Mtb* FGE y pET151 *Strep* FGE y los oligonucleótidos apropiados de la tabla anterior en las reacciones de mutagénesis. Las mutaciones se confirmaron mediante secuenciación de ADN y se transformaron los plásmidos en células BL21 (DE3) para la expresión de proteínas como se describe a continuación.
- 10 Expresión y purificación de proteínas. Las poblaciones clonales de células BL21 (DE3) que albergan un plásmido que codifica una proteína marcada con His₆ se incubaron en medio LB con ampicilina o kanamicina con agitación a 37°C hasta OD₆₀₀ = 0,5, momento en el que la temperatura se redujo a 18°C y se añadió IPTG 250 μM. Después de 12-16 h, se recogieron las células y se resuspendieron en 20 ml de regulador de lisis (Tris 50 mM, NaCl 500 mM, glicerol al 10%, imidazol 20 mM, DTT 1 mM, TCEP 1 mM, metionina 1 mM, pH 7,5) por litro de cultivo y se lisó por sonicación. Se trató el lisado celular con DNasa (10 μg/ml), se eliminó por centrifugación y se aplicó a una columna HisTrap de 1 ml (GE Healthcare). Se lavó la columna con regulador de lisis con imidazol 35 mM y se eluyó la proteína etiquetada con His₆ usando regulador de lisis con imidazol 250 mM. El volumen de elución se concentró hasta menos de 2 ml si fuera necesario y se purificó adicionalmente en una columna Sefadex 16/60 S300 (GE Healthcare). Se concentró posteriormente la proteína recombinante purificada hasta aproximadamente 20 mg/ml.
- 15
- 20 Se evaluó la identidad y pureza de FGE de *Mtb* y *Strep* por espectrometría de masas con ionización por electroaspersión (Bruker/Agilent Esquire). Rv2407 no era soluble en la forma etiquetada con His₆ y se fusionó alternativamente a la proteína de unión a la maltosa (MBP). Las condiciones de cultivo y lisis de las células productoras de MBP-Rv2407 fueron las mismas que las anteriores excepto con la ausencia de imidazol en el regulador de lisis. Se aplicó el lisado removido a la resina de amilosa (New England Biolabs) en regulador de lisis, se lavó en regulador de lisis adicional y se eluyó MBP-Rv2407 en regulador de lisis con maltosa 10 mM y posteriormente se concentró. La MBP se escindió y se removió de Rv2407 usando Factor Xa (New England Biolabs) y resina de amilosa, respectivamente.
- 25
- 30 Cristalización de FGE de *Strep*. Los intentos de cristalizar los homólogos de FGE de *Mtb*, *Mycobacterium smegmatis* y *Mycobacterium avium* no tuvieron éxito debido a la inestabilidad de la proteína. Se dializó FGE de *Strep* en Tris 10 mM pH 7,5, NaCl 150 mM y TCEP 1 mM. Los cristales de FGE de *Streptomyces* marcados con His₆ se obtuvieron usando difusión de vapor mezclando 1 μl de proteína dializada con 1 μl de solución de cristalización (Tris 100 mM pH 8,0, formato de amonio 2,4 M, 0,3% de β-octilglucósido, 3,2% de 2-butanol) a temperatura ambiente (TA). Los cristales crecieron durante un período de dos semanas y se transfirieron posteriormente a un crioprotector que consistía en una solución de cristalización con glicerol al 20%.
- 35
- 40 Determinación de la estructura de FGE de *Strep*. Los datos se recogieron en la línea de luz 8.2.2 en el Advanced Light Source utilizando un detector ADSC Quantum-Q315 CCD. Los datos de difracción se procesaron usando HKL2000 (Otwinowski y colaboradores (1997) Métodos Enzymol: Macromol Crystallogr Parte A 276, 307-326) Las fases iniciales se determinaron mediante reemplazo molecular usando la FGE humana (PDB entrada 1Y1E) como modelo de búsqueda en PHASER (Storoni y colaboradores (2004) Acta Crystallogr D Biol Crystallogr 60, 432-8). La unidad asimétrica contenía cinco monómeros de FGE de *Strep* en el grupo espacial P3₁2₁. Las etapas iniciales de refinamiento del modelo incluyeron ciclos de hibridación simulados con dinámica de ángulo de torsión y refinamiento del factor B restringido usando CNS (Brunger y colaboradores (1998) Acta Crystallogr D Biol Crystallogr 54, 905-21), seguido por la reconstrucción manual del modelo usando O (Jones y colaboradores (1991) Acta Crystallogr A 47 (Pt 2), 110-9). Los ciclos finales de refinamiento se llevaron a cabo con restricciones de TLS (Winn y colaboradores (2001) Acta Crystallogr D Biol Crystallogr 57, 122-33) como se implementa en REFMAC5 (Murshudov (1997) Acta Crystallogr D Biol Crystallogr 53, 240-55) utilizando 5 grupos TLS (correspondientes a cada monómero de FGE en la unidad asimétrica). Se añadieron moléculas de agua con ARP/WARP (Lamzin y colaboradores (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 de R_{trabajo} y R_{libre} fueron 19,5% y 23,3%, respectivamente. Las estadísticas de recopilación y procesamiento de datos se resumen en la siguiente tabla. Todas las cifras fueron generadas con PyMOL (www.pymol.org).
- 45
- 50

Recopilación de datos

	Resolución (Å) ^a	20-2.1 (2.1-2.17)
	Longitud de onda (eV)	12,398.4
	Grupo espacial	P3121
	Dimensiones de la celda unitaria (a = b, c) (Å)	142,444, 217,067
5	Reflexiones medidas	123276
	Integridad (%)	83,4 (88,2)
	Redundancia	2.6 (2.6)
	Mosaisicidad (°)	0,32
	I/σ	15,8 (1,9)
10	R _{sim} (%) ^b	5,7 (23,3)
	Refinamiento	
	R _{trabajo} (%) ^c	19,5
	R _{libre} (%) ^c	23,3
	Número de residuos/aguas	1438/1017
15	Rms enlaces (Å) / ángulos (°)	0,008/1,062
	gráfico de amachandran (%) ^d	87,9/11,2/0,5/0,6e
	Valores promedio de B	41,5

^aLos valores entre paréntesis corresponden al de mayor resolución.

20 ^bR_{sim} = 100*Σ_hΣ_i|I_i(h) - <1(h)>|/Σ_hΣ_iI_i(h), donde I_i(h) es la I_{ava} medida de reflexión h y <1(h)> es el valor promedio de la intensidad de la reflexión.

^cR_{trabajo} = 100*Σ||F_{obs}|-|F_{calc}||/|F_{obs}|, donde F_{obs} y F_{calc} son las amplitudes del factor de estructura de los datos y del modelo, respectivamente. R_{libre} es R_{trabajo} con 5% de las reflexiones dejadas de lado durante el refinamiento.

^dLos números corresponden al porcentaje de residuos de aminoácidos en las regiones favorecidas, permitidas, generosamente permitidas y rechazadas, respectivamente. Calculado usando PROCHECK₃₉.

25 ^eSe observaron siete residuos en conformaciones estereoquímicamente forzadas ya sea debido a los contactos de empaquetamiento del cristal (Tyr219 en los monómeros A y C) o interacciones de enlace de hidrógeno (Asn232 en los monómeros A-E).

30 Ensayo de actividad de FGE. Las FGE de tipo silvestre y mutante de *Mtb* y *Strep* se purificaron como se describió anteriormente. El sustrato peptídico se sintetizó mediante métodos de síntesis en fase sólida Fmoc estándar y consistió en la secuencia de 13 residuos LCSPSRGSLFTGR, un motivo de consenso de sulfatasa. El extremo terminal N fue acetilado, el extremo terminal C fue amidado y la secuencia fue confirmada por espectrometría de masas. Las condiciones de ensayo fueron similares a las informadas previamente por Dierks y colaboradores en estudios de FGE humana (Dierks y colaboradores (2003) Cell 113, 435-4). Los experimentos anaeróbicos se realizaron de la misma manera excepto que las soluciones se hicieron anaeróbicas utilizando un colector de gas depurado de oxígeno y las reacciones se iniciaron mezclando la enzima con el sustrato en una caja de guantes anaeróbica. Se añadió EDTA a las reacciones apropiadas a una concentración de 100 mM. La confirmación de la formación de FGly se realizó incubando 1 µl de producto desalado con 1 µl de biotina-hidrazida 5 mM (Sigma) durante 30 minutos a TA. Las muestras se mezclaron 1 : 1 (v/v) con solución matricial (10 mg/ml de ácido α-ciano-4-hidroxi-cinámico con citrato de amonio 2 mM) y se analizaron mediante desorción/ionización láser asistida por matriz - espectrometría de masas de tiempo de vuelo

40 (Applied Biosystems Voyager DE Pro).

Detección de metales. Se preparó una solución estándar de múltiples elementos mediante la dilución apropiada de los estándares de ICP de Ca, Cu, Fe, Mn, Mg y Zn (Sigma). El contenido de metales de FGE de *Mtb* y *Strep* se analizó mediante ICP-AES utilizando un Perkin Elmer Optima 3000 DV. La ausencia de Fe, Cu y Zn en FGE de *Strep* se confirmó en la línea de luz 8.3.1 en la Fuente de Luz Avanzada. Los bordes de absorción de estos metales se examinaron utilizando un monocrómetro de doble cristal y el detector de fluorescencia de rayos X de la línea de luz.

Producción de la cepa deficiente en FGE de *Mtb*. Se creó una supresión genética no marcada en el marco del marco de lectura abierto Rv0712 que codifica FGE en H37Rv de *Mtb* usando sustitución alélica (Parish y colaboradores (2000) Microbiology 146 (Pt 8), 1969-75). Se amplificó una región de 2 kb secuencia arriba de Rv0712 y se insertó en el vector de suministro micobacteriano p2NILX entre los sitios de restricción HindIII y XbaI. p2NILX se deriva de pNIL (Parish y otros (2000) Microbiology 146 (Pt 8), 1969-75) y se modifica con la adición de un sitio de restricción XbaI entre sitios de restricción KpnI y NotI. Se amplificó e insertó una región de 2 kb secuencia abajo de Rv0712 en p2NILX entre los sitios de restricción XbaI y PaeI. Los marcadores de selección lacZ y sacB se digirieron a partir de pGOAL17 y se ligaron en p2NILX usando el sitio de restricción PaeI. El vector de suministro completado se trató con luz UV (120 mJ cm⁻²) y se sometió a electroporación en H37Rv de *Mtb* electrocompetente como se describió anteriormente (Hatfull, GF & Jacobs, WRJ (eds.) Molecular Genetics of Mycobacteria (ASM Press, Washington DC, 2000)). La selección del mutante se realizó como se describió anteriormente (Parish y colaboradores (2000) Microbiology 146 (Pt 8), 1969-75), y el genotipo se confirmó mediante análisis tipo Southern (Figura 9). Se produjo la cepa complementada transformando la cepa Δ fge con el vector integrador pMV306.kan que contenía todo el marco de lectura abierto Rv0712 bajo el control del promotor de la glutamina sintasa.

Ensayo de sulfatasa/fosfatasa. Las cepas de *Mtb* H37Rv se cultivaron en medio 7H9 suplementado con ADC (Becton Dickinson) a 37°C hasta una OD₆₀₀ = 1,0. Las células se lisaron mediante ruptura mecánica utilizando perlas de circonio de 0,1 mm (FastPrep, MP Biomedicals) y se removió el lisado crudo por centrifugación y se filtró a través de una membrana de 0,22 µm. Las muestras de lisado removidas se normalizaron para la concentración de proteína total (kit de ensayo de proteína Biorad AC/DC) y se añadieron 50 µg de proteína lisada al regulador (Tris 50 mM pH 7,5, NaCl 500 mM, MgCl₂ 100 µM, MnCl₂ 100 µM, CaCl₂ 100 µM), inhibidores de proteasa (grupo cóctel del inhibidor de proteasa III, EMD Bioscience), y sulfato de 4-metilumbeliferilo 8 mM (4MUS). Se usó sulfatasa de lapa (Sigma) a una concentración final de 1 µg/ml como control positivo. Las reacciones se incubaron a 37°C durante 3 h y se detuvieron añadiendo 4 volúmenes de Na₂CO₃/NaHCO₃ 0,5 M pH 10,5. Se midió la actividad de la sulfatasa usando un fluorímetro (Gemini XL, Molecular Devices) usando longitudes de onda de excitación y emisión de 360 nm y 460 nm, respectivamente. Se usaron los inhibidores de sulfatasa/fosfatasa de acuerdo con las instrucciones del fabricante y se incluyeron microcistina, cantaridina, p-bromotetramisol, vanadato de sodio, molibdato de sodio, tartrato de sodio e imidazol (cóctel inhibidor de fosfatasa 1&2, Sigma). La actividad de sulfatasa de Rv2407, Rv3406 y Rv3762c recombinante se determinó usando las mismas condiciones mencionadas anteriormente, con la adición de α -cetoglutarato 1 mM, ascorbato 200 µM y FeCl₂ 100 µM al regulador. La actividad de la fosfatasa se controló como se ha descrito anteriormente, excepto con la sustitución de 4-metilumbeliferil fosfato por 4MUS.

Etiquetado de NBD. Se trató FGE de *Strep* etiquetado con His₆ con una proteasa TEV en proporción 1:50 (p/p) para remover la etiqueta His₆ del extremo terminal N antes de la marcación de NBD y los análisis de espectrometría de masas. Se incubó FGE de *Strep* (45 µM) en regulador (fosfato de potasio 25 mM pH 7,0, NaCl 150 mM) con 4-cloro-7-nitrobenz-2-oxa-1,3-diazol 1 mM (NBD-Cl, Invitrogen) durante 30 min a TA (Ellis y colaboradores, (1997) Biochemistry 36, 15013-8). La muestra se desaló mediante cromatografía de fase inversa C₁₈ y se detectaron los aductos de proteína-NBD por espectrometría de masas (Bruker/Agilent Esquire). El mapeo de los aductos de NBD se realizó digiriendo FGE de *Strep* que reaccionó con NBD con tripsina 1:50 (p/p), desalando mediante cromatografía en fase inversa C₁₈ y analizando los fragmentos peptídicos resultantes usando espectrometría de masas de resonancia de ciclotrón de iones de transformadas de Fourier con ionización por electroaspersión (Bruker 9,4T Apex III).

Ejemplo 13: Identificación y clonación de una FGE de *M. tuberculosis*

Se identificó el marco de lectura abierto Rv0712 de *Mycobacterium tuberculosis* (*Mtb*) H37Rv por análisis BLAST (Altschul y colaboradores (1997) Nucleic Acids Res 25, 3389-402) que era más del 30% idéntico a la FGE SUMF1 humana (Cosma y colaboradores (2003) Cell 113, 445-56 (2003); Dierks y colaboradores (2003) Cell 113, 435-44). Rv0712 recombinante fue capaz de modificar un péptido sintético que contenía el motivo de sulfatasa determinado por espectrometría de masas (Figura 8, Panel a). La presencia de FGly dentro del sustrato se confirmó tratando el péptido modificado con biotina-hidrazida, que formó un aducto covalente con el péptido a través de la formación de hidrazona (Figura 8, Panel b). Juntos, estos datos implican Rv0712 como FGE de *Mtb*.

Similar al genoma humano, el genoma de *Mtb* parece codificar sólo una copia funcional de FGE. Por lo tanto, se esperaba que la interrupción de Rv0712 en *Mtb* para producir una cepa deficiente en sulfatasa Rv0712 se interrumpiera en *Mtb* H37Rv utilizando recombinación homóloga y confirmado por el análisis tipo Southern (Figura 9). *Mtb* Δ fge fue viable y no demostró defectos de crecimiento evidentes *in vitro*.

La actividad de sulfatasa de la cepa Δ fge se comparó con la de H37Rv de tipo silvestre (TS) y con el mutante Δ fge en el que se restableció la expresión de FGE mediante complementación. Se generaron lisados sin purificar a partir de estas

tres cepas de *Mtb* y se determinó la actividad global de sulfatasa utilizando el sustrato general sulfato de 4-metilumbeliferilo (4MUS). La cepa Δfge mostró una pérdida substancial, aunque sorprendentemente incompleta, de actividad de sulfatasa (Figura 8, Panel c). Aunque pudo haber sido posible que la actividad de sulfatasa residual fuera el resultado de las fosfatasa que actúan sobre el 4MUS, cuando se controló la actividad de sulfatasa en presencia de un cóctel de inhibidores de sulfatasa/fosfatasa de amplio espectro, Δfge no fue afectada. De hecho, la actividad en lisados de TS y Δfge complementada se redujo en aproximadamente un 40% en presencia del cóctel inhibitor, ajustando la actividad de sulfatasa de Δfge en ausencia de inhibidores (Figura 8, Panel c). Debido a que se conoce que los inhibidores aplicados inhiben las sulfatasas activadas con FGE (Stankiewicz y colaboradores (1988) *Biochemistry* 27, 206-12), estos datos sugieren que *Mtb* posee sulfatasas independientes de FGE.

Para verificar adicionalmente que las fosfatasa promiscuas no eran responsables de la actividad residual de sulfatasa residual, se controló la actividad de fosfatasa de lisados sin purificar de cada cepa utilizando fosfato de 4-metilumbeliferilo. Las tres cepas exhibieron el mismo nivel de actividad de fosfatasa en ausencia de inhibidores, pero la actividad se abolió en todas las cepas en presencia de los inhibidores (Figura 8, Panel d). Estos datos indican además que las fosfatasa no son responsables de la actividad de hidrólisis de 4MUS residual observada en la cepa Δfge y que las sulfatasas activadas con FGE son responsables de aproximadamente el 40% de la actividad total de sulfatasa en el lisado de *Mtb*.

Se examinó el genoma de *Mtb* en busca de fuentes potenciales de actividad de sulfatasa independiente de FGE. La mayoría de las sulfatasas procariotas conocidas o putativas son homólogas a las sulfatasas eucariotas y contienen el motivo de sulfatasa. Sin embargo, algunos procariotas también tienen sulfatasas independientes de FGE que no requieren FGly y presumiblemente operan a través de diferentes mecanismos. Estas enzimas pueden no ser sensibles a los inhibidores de sulfatasa/fosfatasa de amplio espectro. Las sulfatasas independientes de FGE no son homólogas a las sulfatasas activadas con FGE y se han clasificado en una de las dos familias de enzimas, las metalo- β -lactamasas y las dioxigenasas¹⁸⁻²⁰ dependientes de Fe(II) α -cetoglutarato. Con base en la similitud de secuencias con las sulfatasas independientes de FGE conocidas de otros procariotas, *Mtb* tiene al menos tres sulfatasas putativas independientes de FGE codificadas por los marcos de lectura abiertos Rv2407, Rv3406 y Rv3762c. Las formas recombinantes de Rv2407, Rv3406 y Rv3762c se expresaron en *E. coli*, pero las proteínas purificadas no mostraron actividad en el ensayo de 4MUS, lo que indica que estas supuestas sulfatasas probablemente no son responsables de la actividad de sulfatasa residual en *Mtb* Δfge (Figura 10). Teniendo en cuenta la falta de similitud de secuencia entre sulfatasas independientes de FGE, *Mtb* puede tener otras sulfatasas no detectables por análisis BLAST.

Ejemplo 14: Estructura de FGE de *Mtb*

Con el fin de comprender mejor el mecanismo enzimático único y las características de unión al sustrato de las FGE procariotas, se determinó la estructura del ortólogo de FGE de *Mtb* de *Streptomyces coelicolor* (*Strep*) con una resolución de 2,1 Å. La topología global de la FGE bacteriana es notablemente similar a la estructura de FGE humana recientemente determinada (Dierks, y colaboradores (2005) *Cell* 121, 541-52). (Figura 11A). Al igual que la FGE humana, FGE de *Strep* tiene un bajo contenido en estructura secundaria, que contiene 16% de hélices α y 12% de láminas β . Ambas comparten el nuevo "plegado de FGE", pero la variante de FGE de *Strep* contiene solo un ion Ca^{2+} determinado por la geometría de coordinación y la espectroscopia de emisión atómica de plasma inductivamente acoplado (ICP-AES) (Figura 12). La variante humana se estabiliza mediante dos iones Ca^{2+} ; esta diferencia se debe aparentemente a una sustitución de Glu66Ala en FGE de *Strep* que interrumpe un entorno de coordinación apropiado (Figura 3, panel b). Los datos de ICP-AES indican que FGE de *Mtb* carece de ambos iones Ca^{2+} (Figura 12), lo que sugiere que el plegado de FGE no requiere estabilización por cationes divalentes.

Los sitios activos de la FGE procariota y humana son notablemente similares. Ambos tienen aproximadamente 20 Å de longitud, 12 Å de ancho y 10 Å de profundidad y pueden acomodar sólo 6 de los 13 aminoácidos que definen el motivo de sulfatasa. Teniendo en cuenta que el motivo de sulfatasa se extiende hacia el extremo terminal C del sustrato peptídico por otros ocho residuos más allá de la secuencia central de consenso (CXPXR) (Figura 7, panel b), es posible que FGE haya desarrollado una región de unión secundaria para ayudar en el reconocimiento del sustrato, de forma similar a otras proteínas tales como la trombina y la neurotoxina botulínica (Hageman y colaboradores, (1974) *Arch Biochem Biophys* 164: 707-15, Breidenbach y colaboradores (2004) *Nature* 432: 925-9). De hecho, cuando se mapean los residuos conservados entre las FGE de *Strep*, *Mtb*, humanas y otras FGE putativas en la superficie de la molécula de FGE de *Strep*, se observa una región de alta conservación donde la sección del extremo terminal C del motivo de sulfatasa podría enlazarse (Figura 11C).

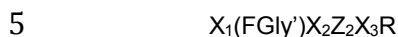
Se cree que FGE cataliza la oxidación de un tiol a un aldehído usando dos residuos de cisteína conservados dentro de su sitio activo (Dierks y colaboradores (2005) *Cell* 121, 541-52; Roeser y colaboradores (2006) *Proc Natl Acad. Sci EE.UU.* 103, 81-6). Se requieren estas cisteínas en FGE de *Mtb* (Cys263 y Cys268) y FGE de *Strep* (Cys272 y Cys277) para la renovación del sustrato ya que los mutantes de serina eran 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. La omisión de los mapas mostró que Cys272 y Cys277 estaban acoplados 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 fue proporcionada tratando las

tres cisteínas expuestas al disolvente de FGE de *Strep* nativa con el reactivo 4-cloro-7-nitrobenz-2-oxa-1,3-diazol (NBD) marcado con tiol. Dos poblaciones distintas que corresponden a FGE de *Strep* con 1 o 3 aductos de NBD fueron detectados por espectrometría de masa de proteína intacta (Figura 14), lo que confirma que aproximadamente un tercio de estas dos cisteínas proximales están unidas en un enlace disulfuro.

- 5 Además de las dos cisteínas de sitio activo, FGE de *Strep* y *Mtb* también requieren de oxígeno molecular para la catálisis ya que no se observó formación FGly en las reacciones realizadas en un entorno anaeróbico (Figura 13, panel e, y datos no mostrados). Como miembro de la familia de la oxigenasa, se podría esperar que 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 la exploración del borde de absorción de rayos X indicó que FGE activa de *Mtb* y de *Strep* carecen de todos los metales redox activos (Figura 12 y datos no mostrados). Además, estas FGE no requieren la adición de metales para funcionar en ensayos *in vitro* y pueden funcionar en presencia de EDTA (Figura 13, paneles b, g). De forma similar, la espectroscopia de absorción de UV-visible no reveló la presencia de cofactores orgánicos cromóforos (datos no mostrados). Junto con la información de densidad electrónica para FGE de *Strep*, estos datos indican que las FGE procariotas, en forma similar a la FGE humana, no utilizan cofactores exógenos para la catálisis.
- 10
- 15 Como medio alternativo de activación del oxígeno molecular, la FGE puede funcionar de forma similar a otras oxigenasas sin cofactor y hacer uso único de los residuos convencionales²⁵. Los residuos absolutamente conservados dentro del intervalo reactivo del par de cisteínas catalíticas de FGE de *Strep* incluyen Trp234 y Ser269. Roeser y colaboradores han teorizado que Trp234 puede funcionar para activar el oxígeno molecular (Roeser y colaboradores (2006) Proc Natl Acad Sci EE.UU. 103: 81-6), de forma similar al mecanismo propuesto de reducción de O₂ por anticuerpos catalíticos²⁶. Sin embargo, la mutación de Trp234 a Phe no abolió la actividad (Figura 13, panel j), lo que indica que la activación del oxígeno molecular debe ser lograda por otra vía. La actividad fue severamente reducida por la mutación equivalente de Ser269Ala en FGE de *Mtb* (Ser260Ala) (Figura 13, panel d), pero actualmente se desconoce cómo este residuo desempeña un papel en el ciclo catalítico de FGE.
- 20
- 25 Curiosamente, la propia Cys272 puede estar implicada en la activación del oxígeno molecular. Los cinco residuos de Cys272 modelados dentro de la unidad asimétrica de FGE de *Streptomyces* tienen una densidad de electrones extra que se extiende más allá de sus conformaciones alternativas, no unidas por disulfuro. La omisión de los mapas indica que esta densidad adicional podría modelarse como una molécula de agua o una porción de hidroperóxido con ocupación parcial (Figura 13, paneles d, e). Los experimentos de marcación con NBD utilizando el mutante de FGE de *Strep* Cys277Ser indican que Cys272 no es un tiol reactivo en una subpoblación de moléculas enzimáticas (Figura 15), lo que sugiere que la densidad adicional corresponde a una porción covalentemente unida a Cys272, tal como hidroperóxido. Las estructuras publicadas anteriormente de la FGE humana también han sugerido que esta densidad adicional podría ser hidroperóxido o ácido sulfénico de cisteína combinado con una molécula de agua unida²¹. Sin embargo, este último se ajusta mal a la densidad de electrones observada en FGE de *Streptomyces*. Además, no se detectó ácido sulfénico cuando se trató FGE de *Strep* con cloruro de NBD (Figura 15 y Figura 14). La presencia de hidroperóxido modelado con ocupación parcial no puede descartarse con base en la densidad electrónica observada de FGE de *Strep*. Sin embargo, el análisis espectrométrico de masas de FGE intacta de *Strep* y *Mtb* no reveló ninguna anomalía de masa (Figura 16 y datos no mostrados), lo que sugiere que si Cys272 se modifica, la modificación es transitoria o sensible al ácido.
- 30
- 35

Reivindicaciones

1. Un polipéptido modificado que comprende un residuo de 2-formilglicina unido covalentemente a una porción de interés, en donde el polipéptido comprende una etiqueta de aldehído heterólogo modificado, la etiqueta de aldehído consiste de hasta 12 residuos de aminoácidos y comprende:



dónde

FGly' es un residuo de 2-formilglicina que tiene una porción unida covalentemente;

Z₂ es una porción de prolina o alanina;

10 X₁ está presente o ausente y, cuando está presente, es cualquier aminoácido, con la condición de que cuando la etiqueta de aldehído heteróloga está en un extremo terminal N del polipéptido, X₁ está presente; y

X₂ y X₃ son cada uno independientemente cualquier aminoácido.

2. Un polipéptido como el reivindicado en la reivindicación 1, en el que la etiqueta de aldehído consta de hasta 11 residuos de aminoácido.

15 3. Un polipéptido como el reivindicado en la reivindicación 1, en el que la etiqueta de aldehído consta de hasta 10 residuos de aminoácidos.

4. Un polipéptido como el reivindicado en la reivindicación 1, en el que la etiqueta de aldehído consta de hasta 9 residuos de aminoácidos.

5. Un polipéptido como el reivindicado en la reivindicación 1, en el que la etiqueta de aldehído consta de hasta 8 residuos de aminoácidos.

20 6. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones 1 a 5, en el que la etiqueta de aldehído está situado en un extremo terminal C del polipéptido.

7. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones 1 a 5, en el que el polipéptido tiene al menos una de las siguientes propiedades:

(a) la etiqueta de aldehído está situada en el extremo terminal N del polipéptido;

25 (b) la etiqueta de aldehído está situada en un sitio interno de una secuencia de aminoácidos nativa del polipéptido;

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

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

30 8. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones 1 a 5, en el que el polipéptido es una proteína transmembrana y la etiqueta de aldehído está presente en un sitio interno dentro de un bucle extracelular o un bucle intracelular.

9. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones 1 a 5, en el que la etiqueta de aldehído está presente en un sitio interno o en el extremo terminal N del polipéptido, y es accesible al disolvente cuando se pliega el polipéptido.

35 10. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones 1 a 5, en el que se modifica el polipéptido para incluir dos o más etiquetas de aldehído.

11. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones anteriores, en el que X₁, cuando está presente, X₂ y X₃ son cada uno independientemente un aminoácido alifático, un aminoácido que contiene azufre, o un aminoácido polar no cargado.

40 12. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones anteriores, en el que X₁, cuando está presente, es L, M, V, S o T.

13. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones anteriores, en el que X_2 y X_3 son cada uno independientemente S, T, A, V, G o C.

14. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones 1 a 5, en el que la etiqueta de aldehído comprende L(FGly')TPSR, M(FGly')TPSR, V(FGly')TPSR, L(FGly')SPSR, L(FGly')APSR, L(FGly')VPSR, o L(FGly')GPSR.

15. Un polipéptido como el reivindicado en cualquiera de las reivindicaciones anteriores, en el que el polipéptido está presente en un anticuerpo.

16. Un polipéptido modificado como el reivindicado en cualquiera de las reivindicaciones anteriores, en el que la porción es un polímero soluble en agua, una etiqueta detectable, un fármaco o una porción para la inmovilización del polipéptido en una membrana o sobre una superficie.

17. Un polipéptido modificado como el reivindicado en la reivindicación 16, en el que la porción es un fármaco.

FIG. 1A

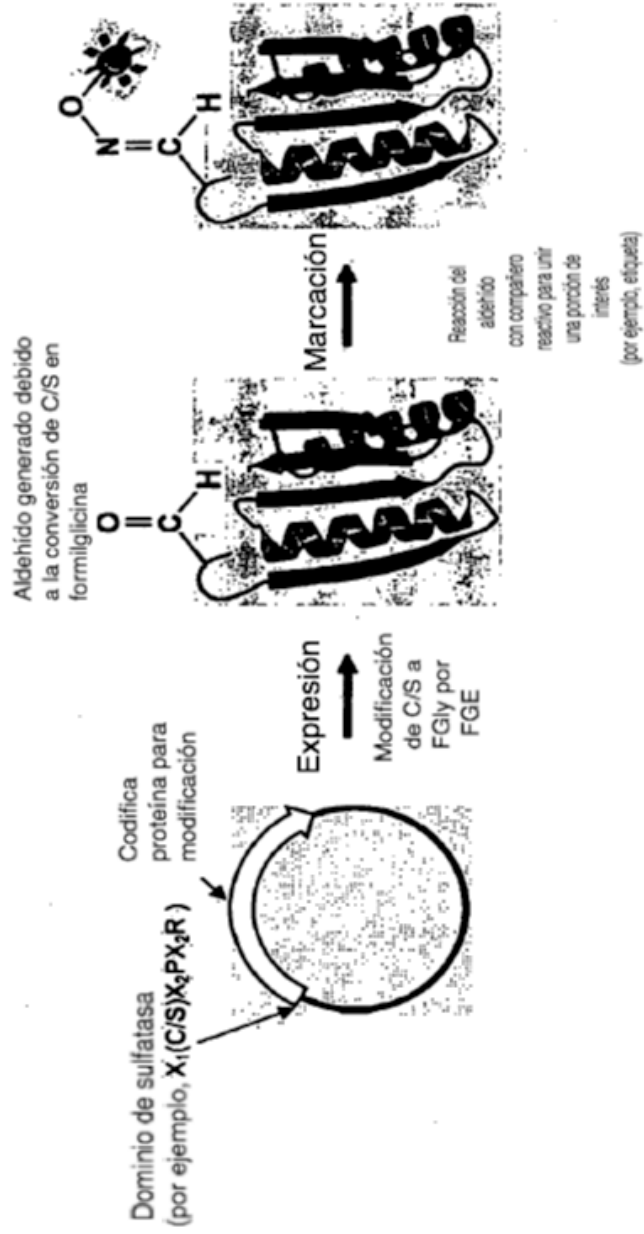


FIG. 1B

<i>H. sapiens</i> Arylsulfatase E	(82)	AASLC	CTPSRAAFL	TGRYPV
<i>M. musculus</i> Sulf2	(84)	TPMC	OPSRSSIL	TGKYVH
<i>C. elegans</i> Sul3	(75)	VNQL	CTPIRSAF	MTGYYPF
<i>M. tuberculosis</i> AtsG	(54)	TAPLC	TPSRGSLF	TGRYPQ
<i>E. coli</i> YidJ	(48)	CSPVC	TPPARAGL	ETGIYAN
<i>S. coelicolor</i> SC07547	(84)	PTAIC	TPPARASLL	TGQAPF
<i>M. barkeri</i> A3073	(80)	TTALC	SPSRSCIL	TGRNHH

etiqueta de 6 mer

Consenso

LCTPSRGSFLTG

R

etiqueta de 13 mer

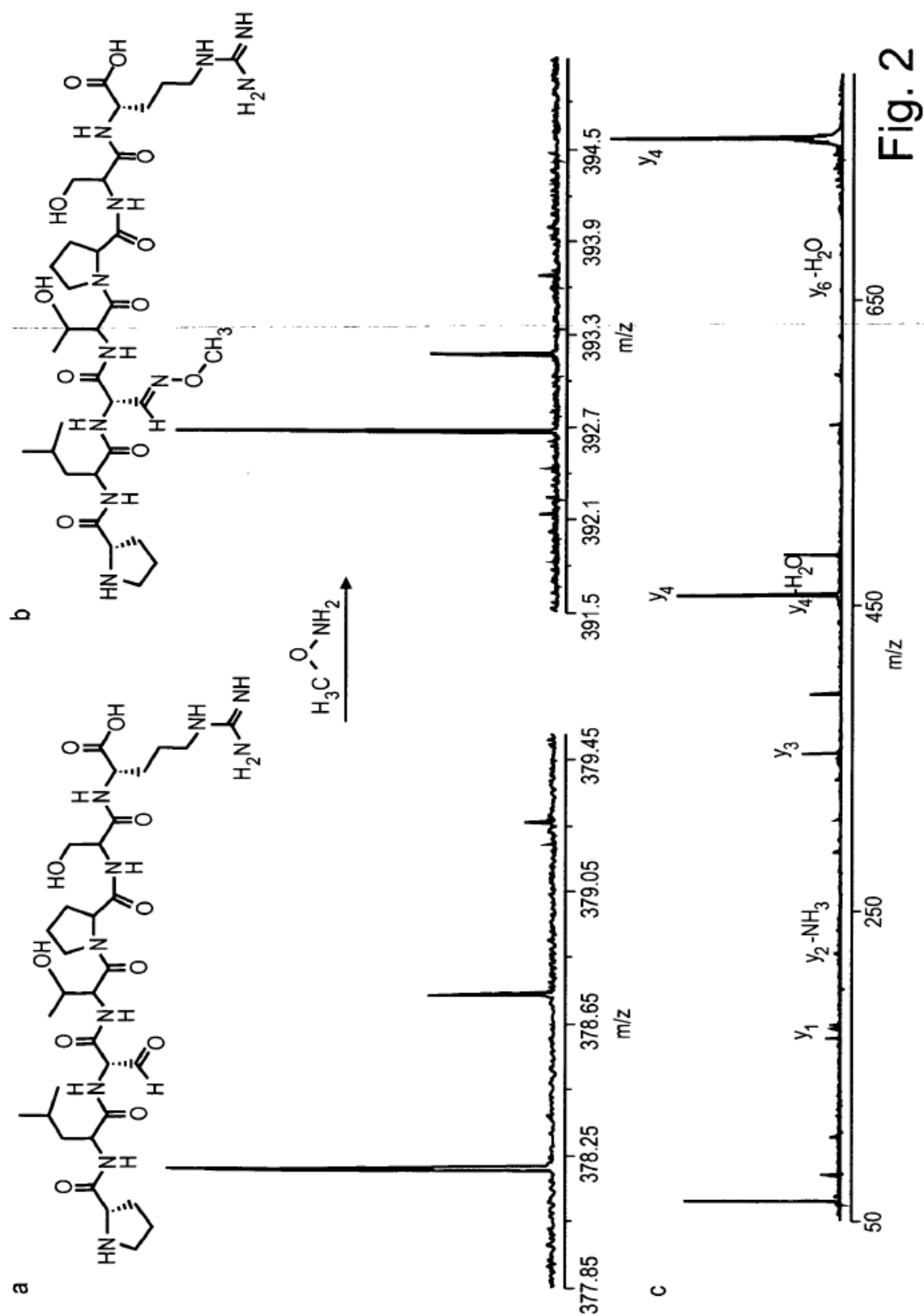


FIG. 3

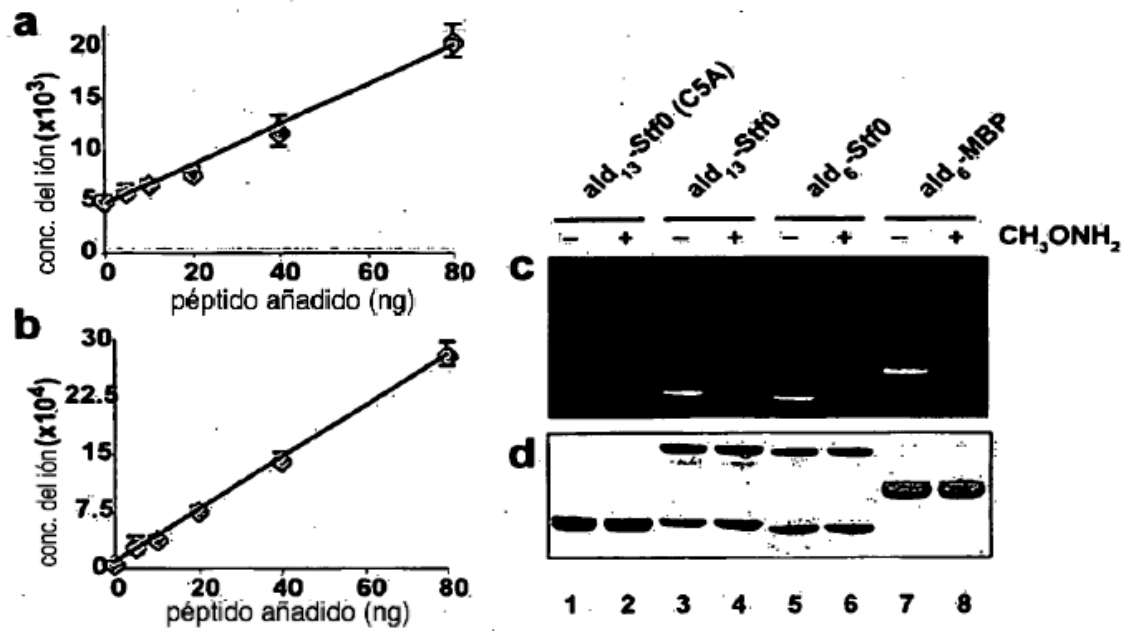


FIG. 4

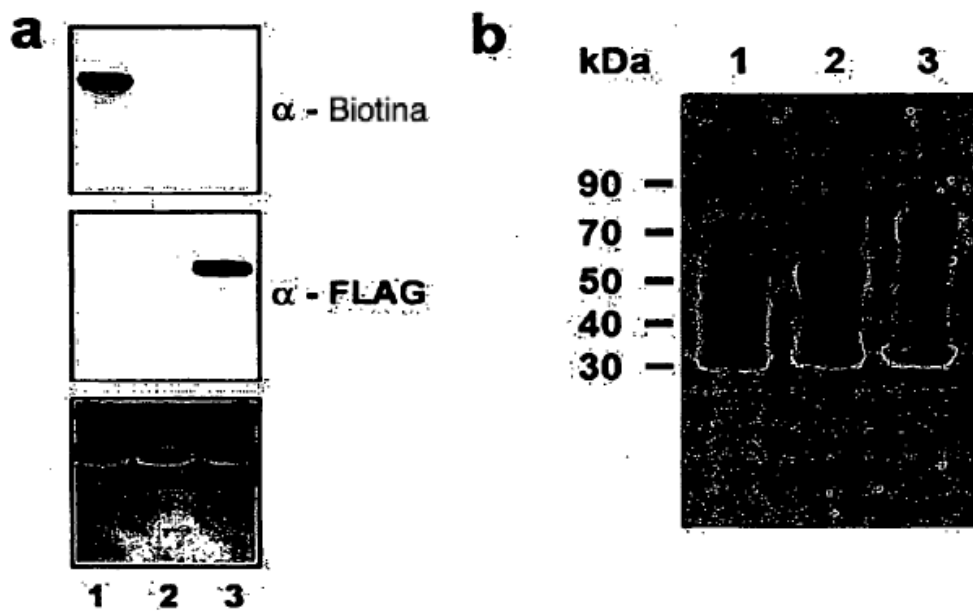


FIG. 5

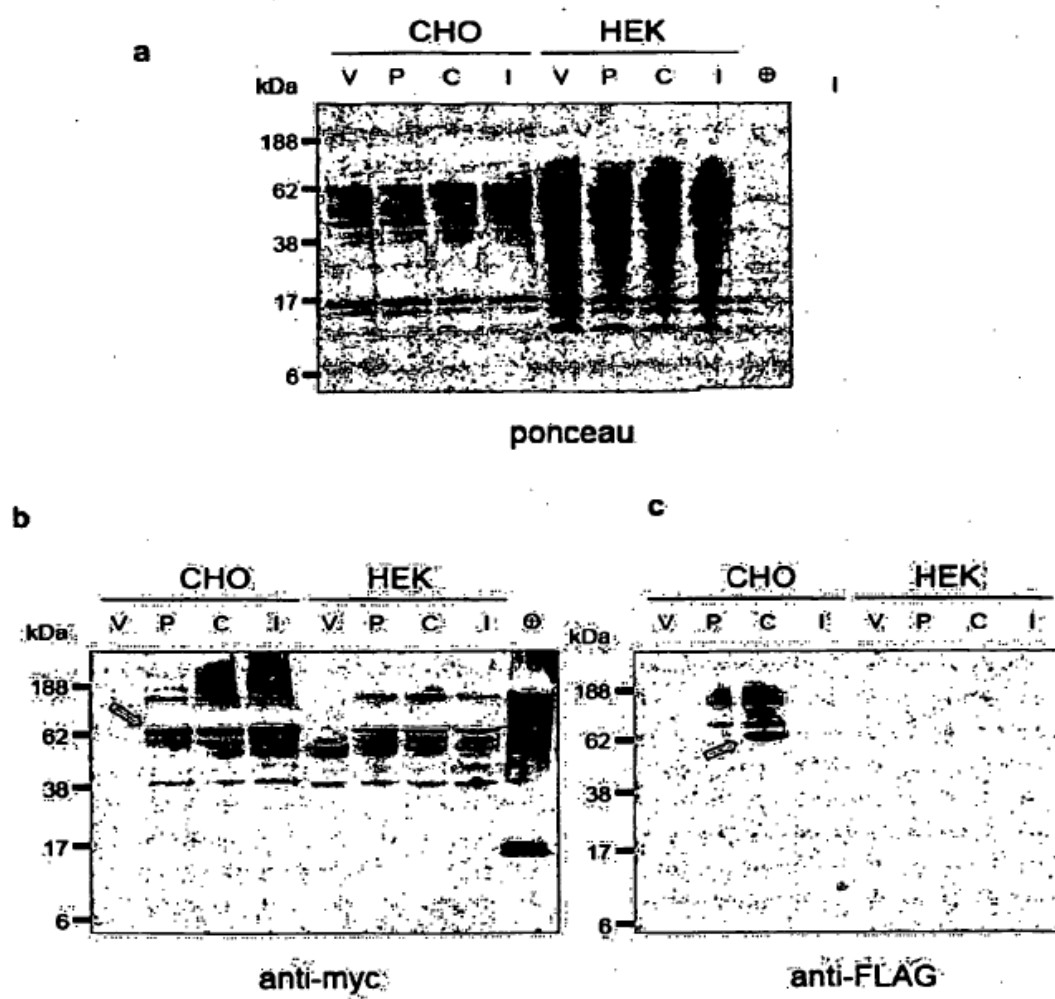


FIG. 6

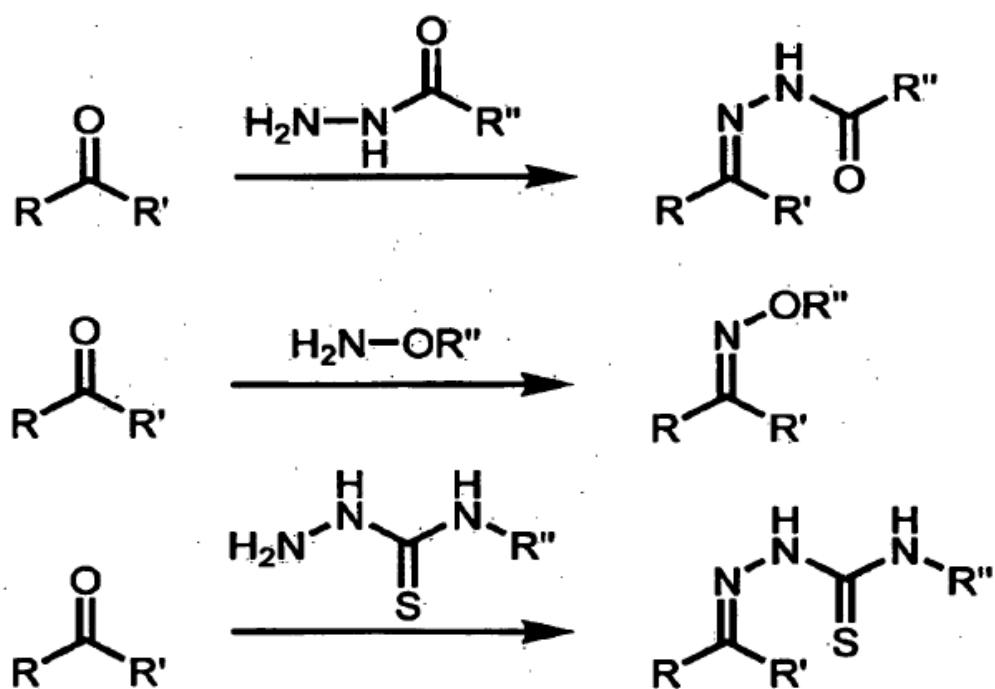
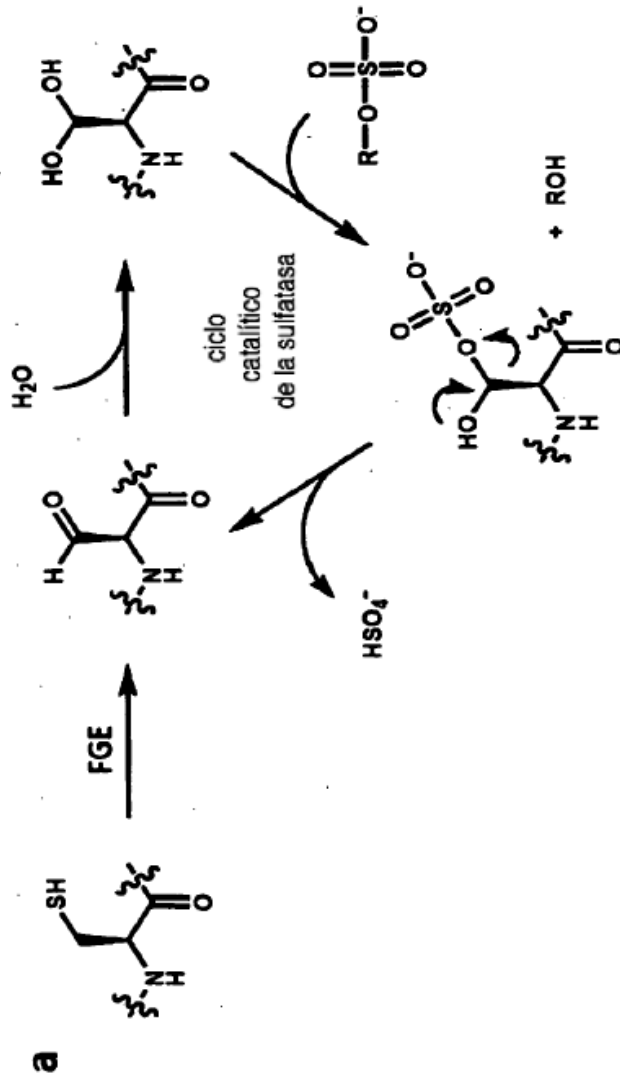


FIG. 7



H. sapiens Arilsulfatasa E (82) AASLCTPSRAAFITGRYPV
M. musculus Sulf2 (84) TTPMCCPSRSISITGKYVH
C. elegans Sulf3 (75) VNQLCTPISAFMTGTYPF
M. tuberculosis Atsg (54) TAPLCTPSRGSIFTGRYPQ
E. coli YidJ (48) CSPVCTPANAAGIFTGIYAN
S. coelicolor SC07547 (84) PTAICTPAPASLITGQAPF
M. barkeri A3073 (80) TTAICTSPSPSCITGRNHH

FIG. 9

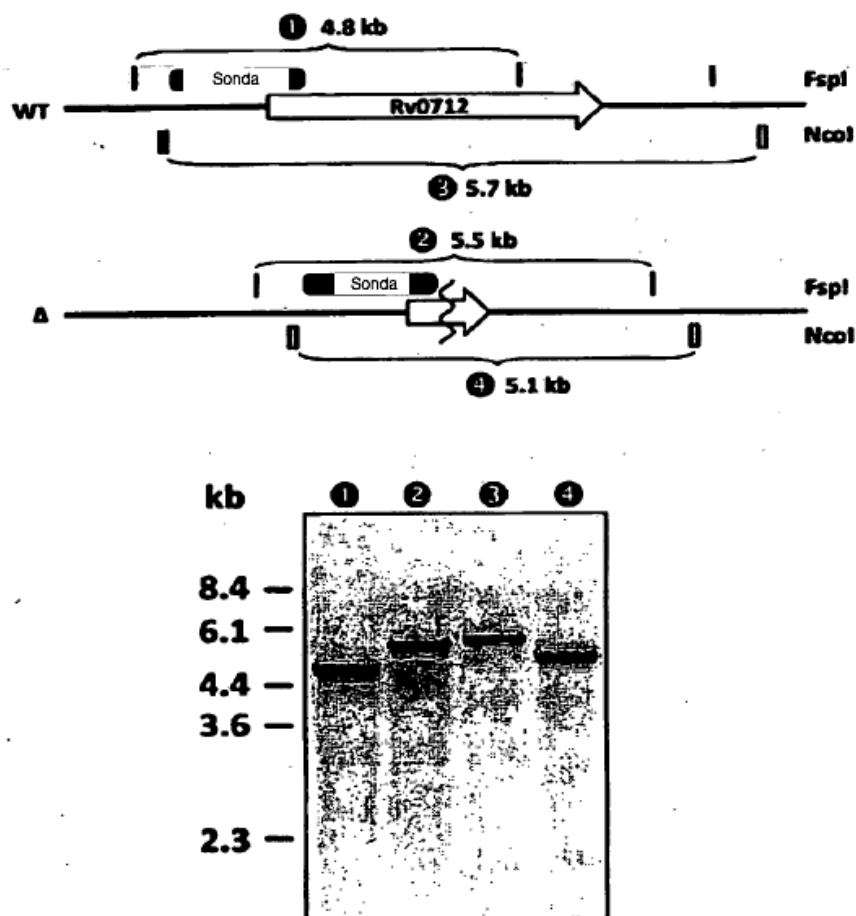


FIG. 10

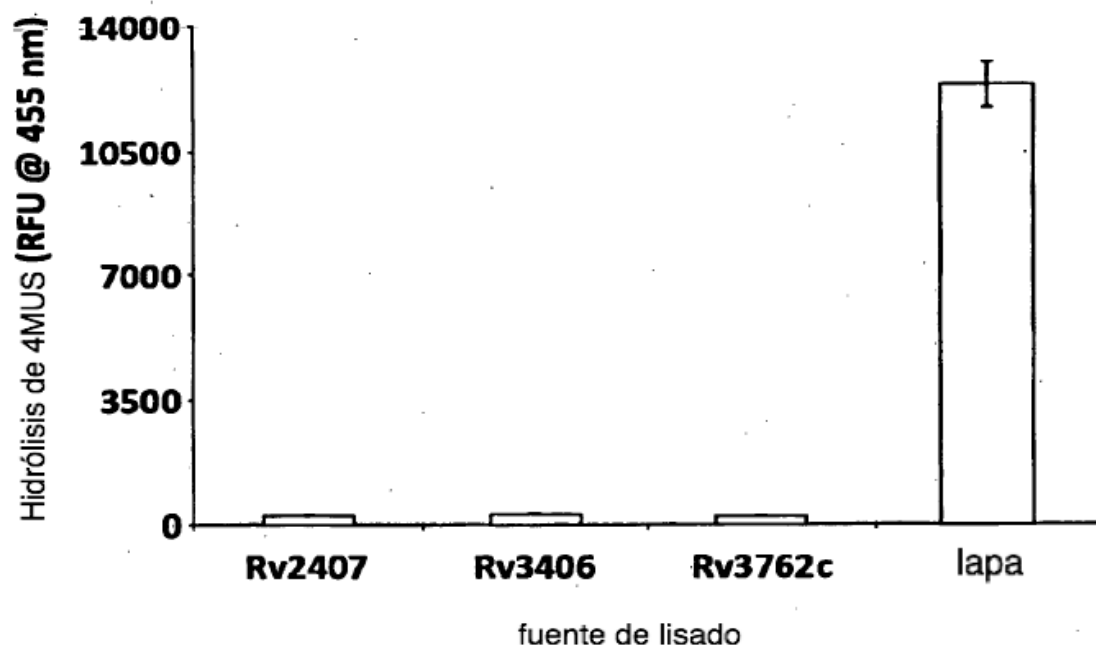


FIG. 11A

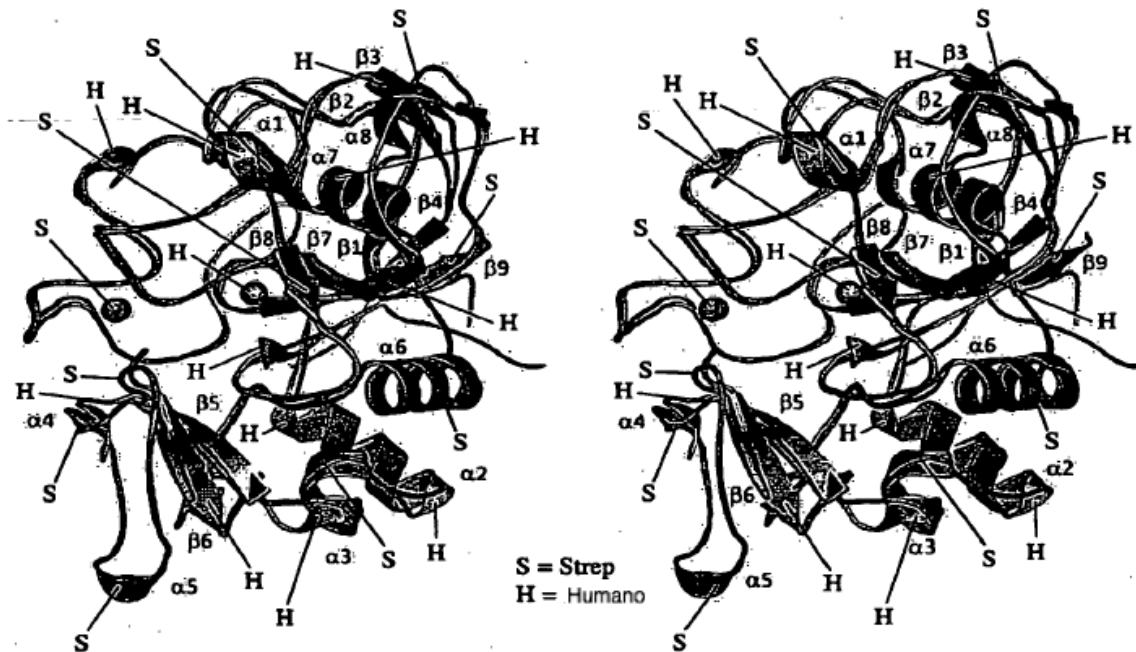


FIG. 11B

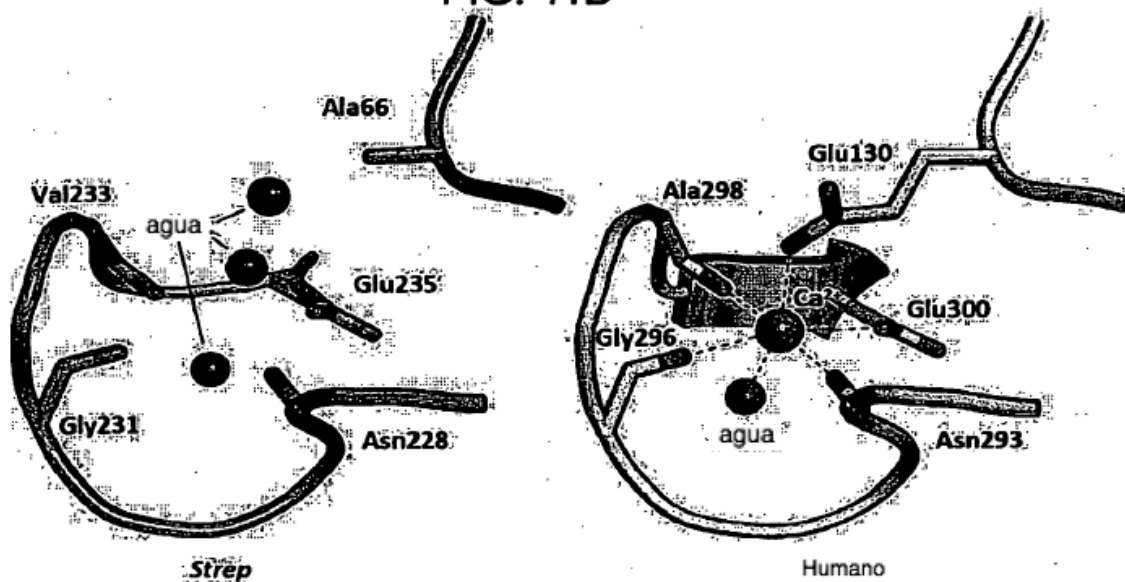


FIG. 11C

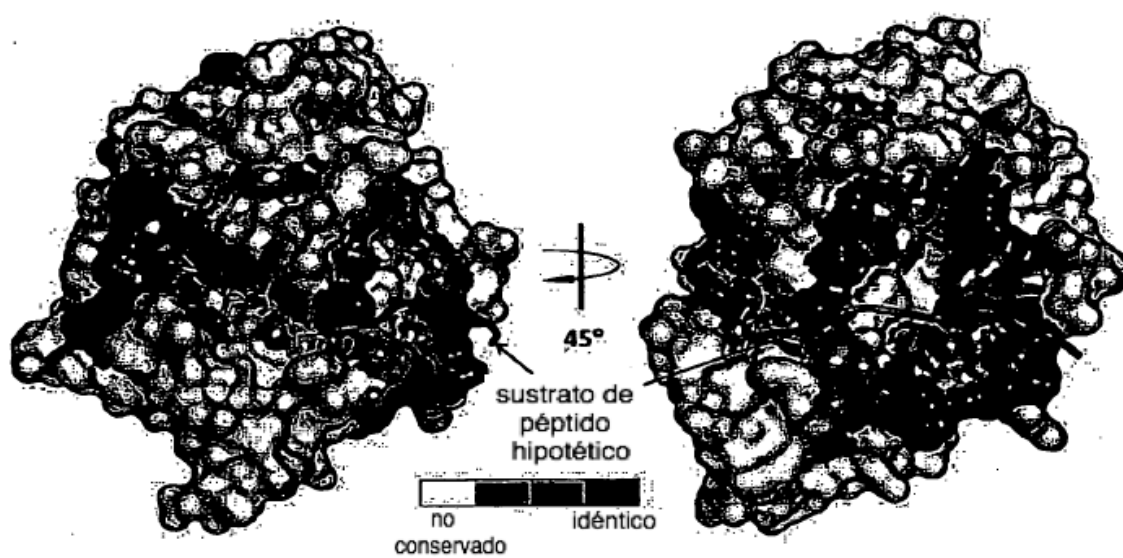


FIG. 11D

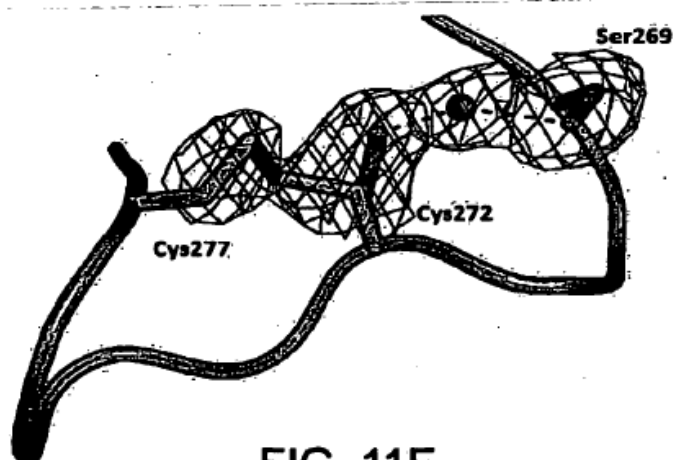


FIG. 11E

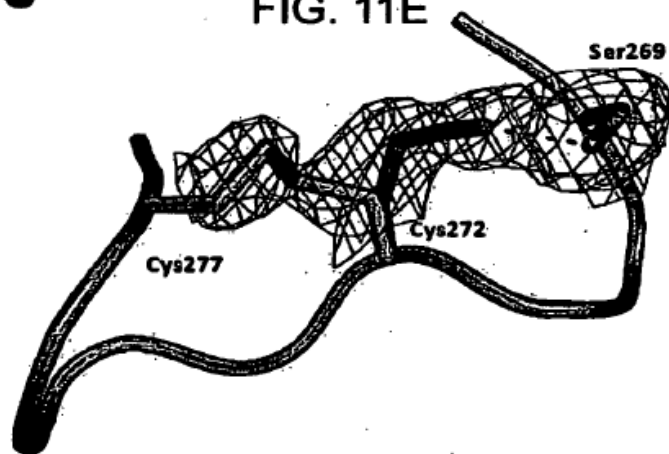


FIG. 12

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

^aPeso molecular de FGE de *Strep* etiquetado con His6 = 37.540,35

^bPeso molecular de FGE de *Mtb* etiquetado con His6 = 35.015,44

FIG. 13

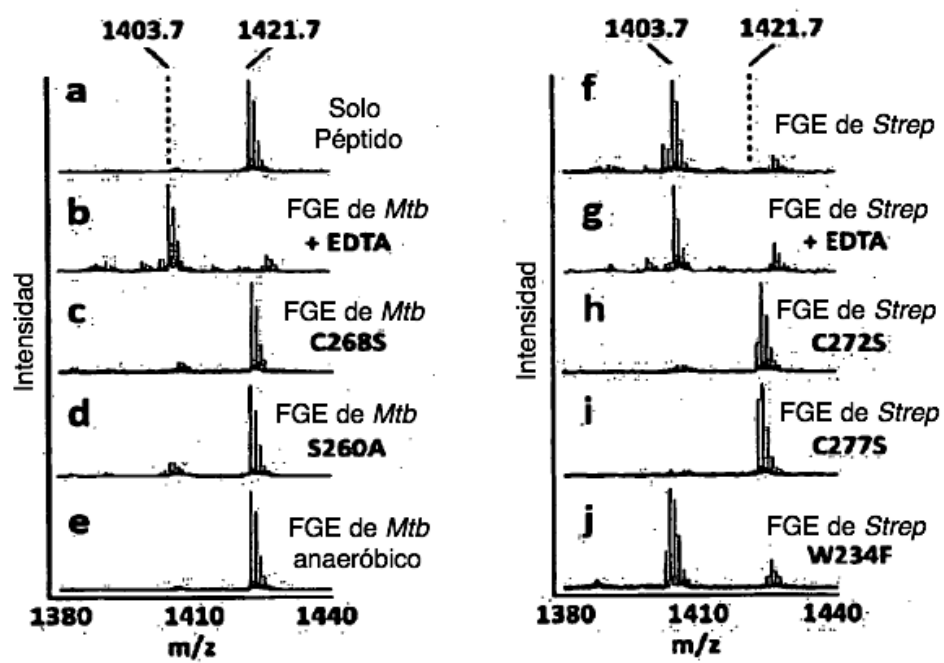


FIG. 14

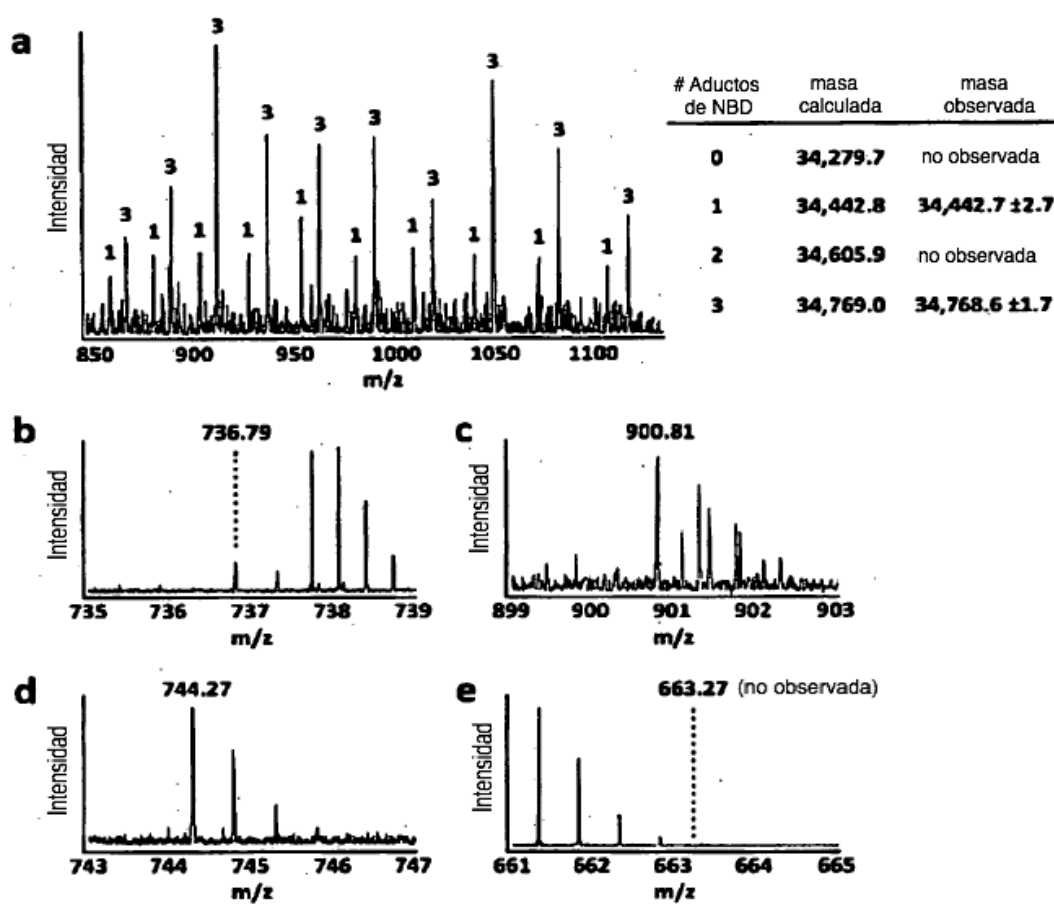


FIG. 15.

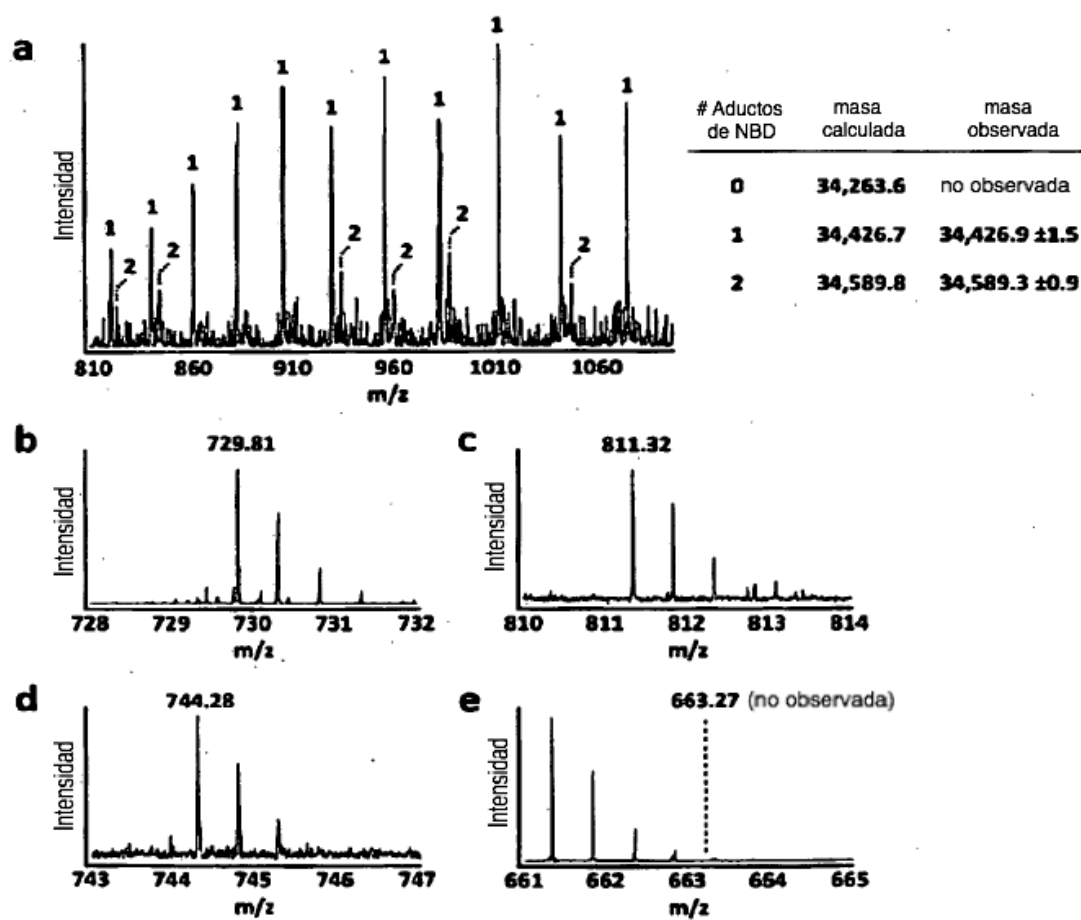


FIG. 16

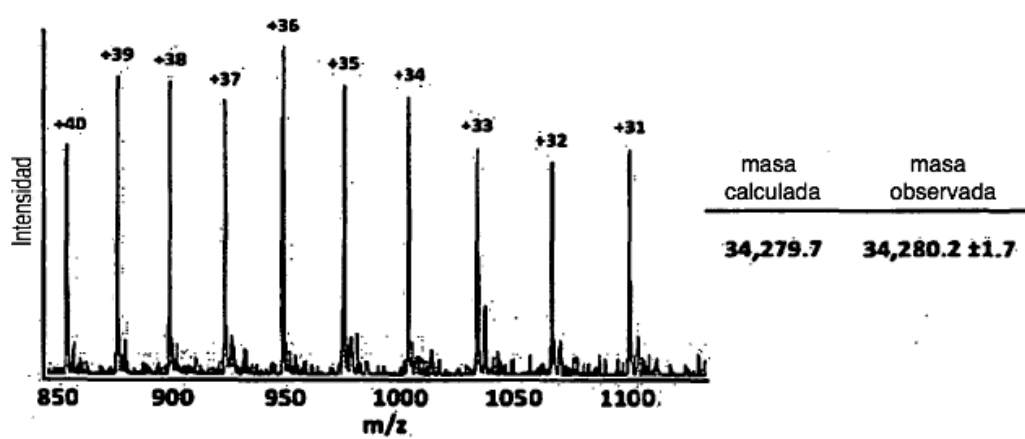


FIG. 17

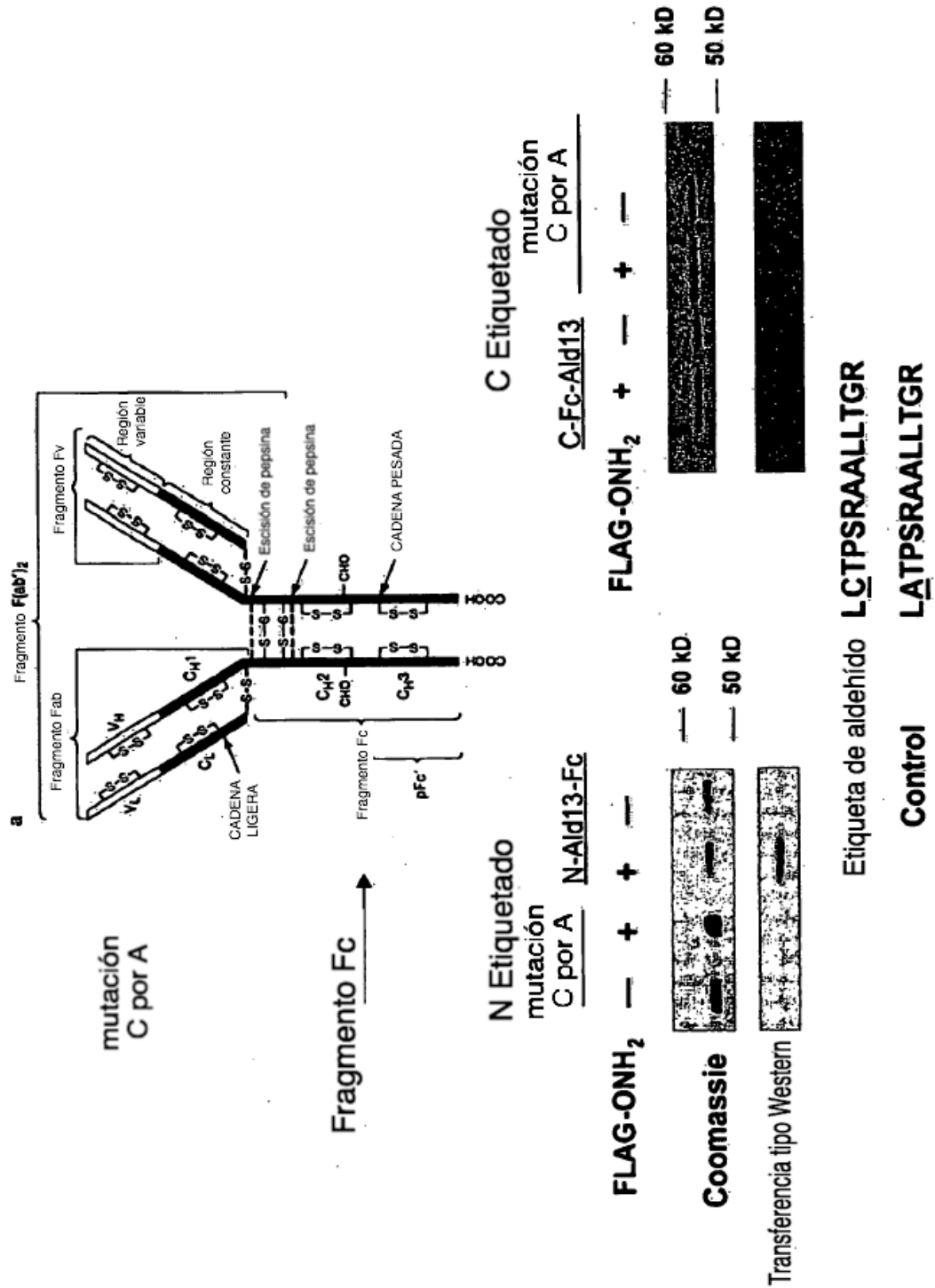
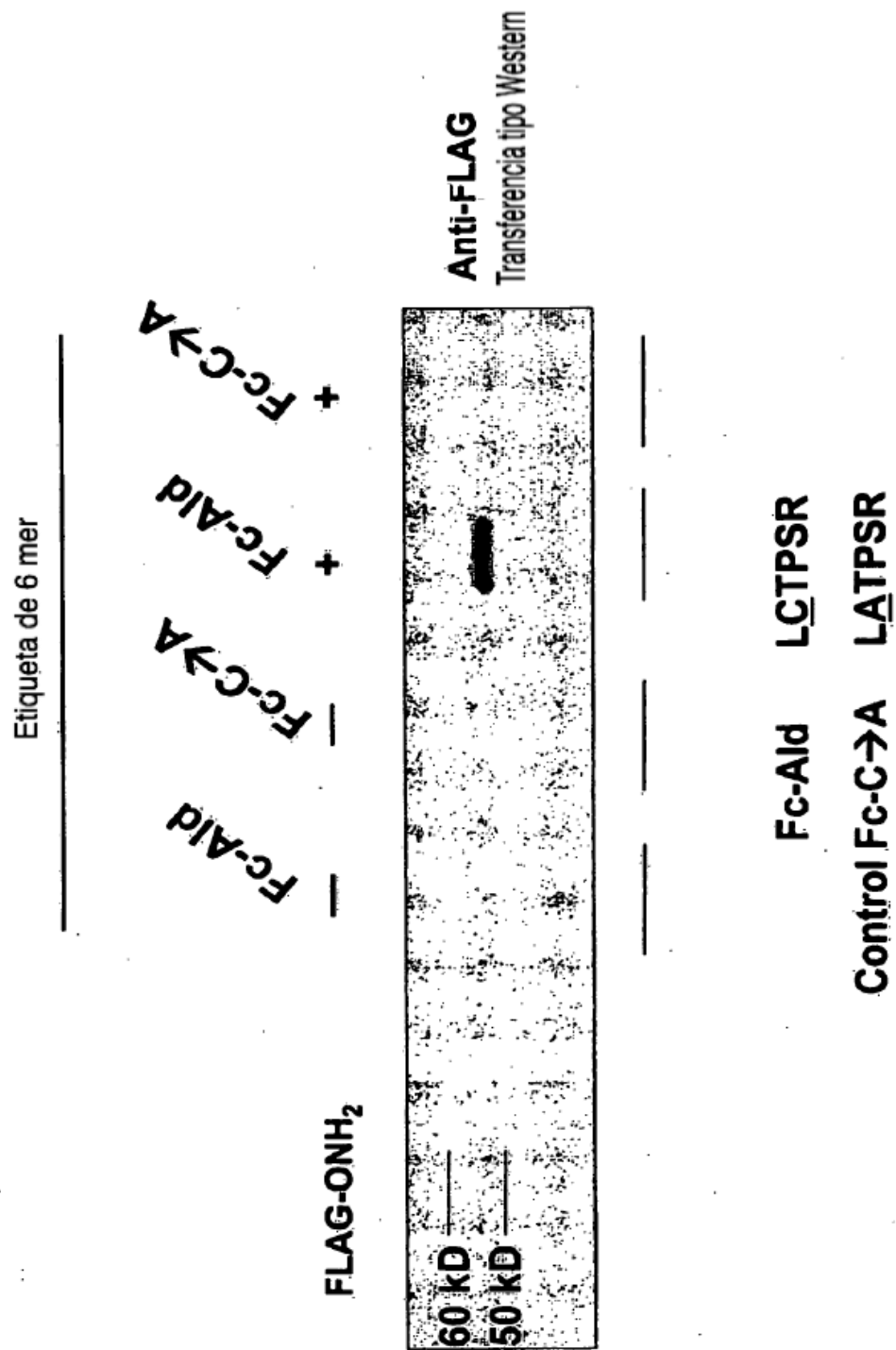
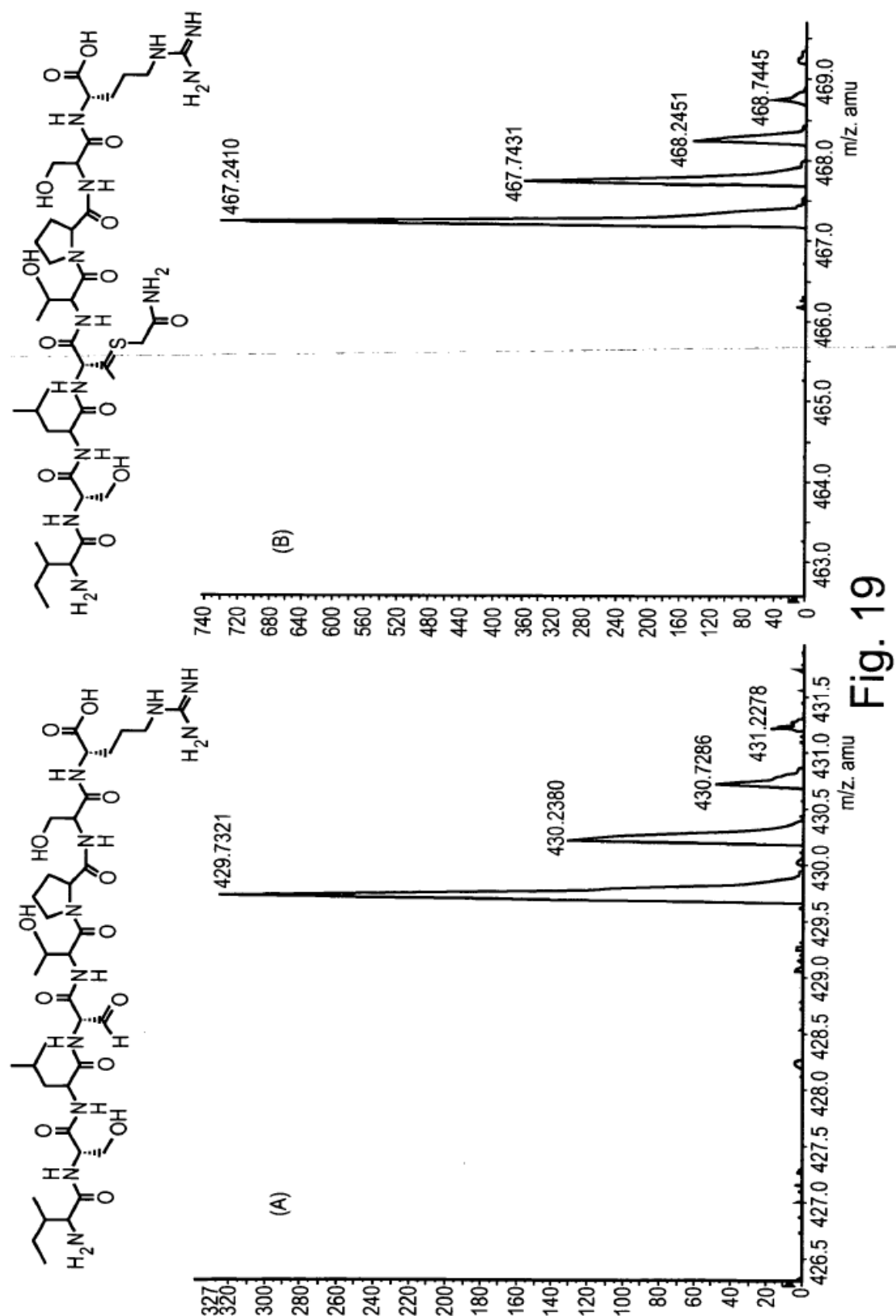


FIG. 18





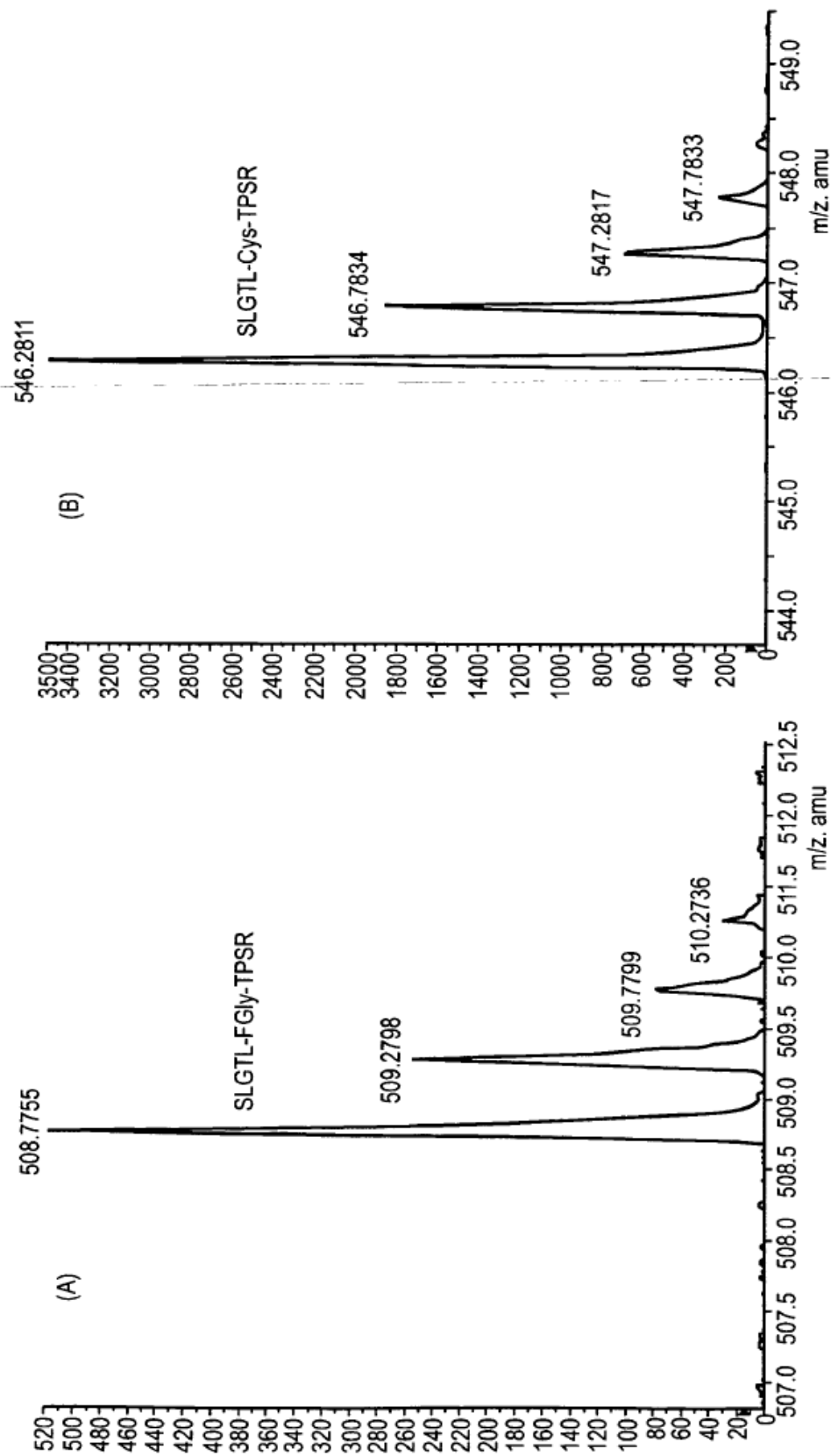


Fig. 20

FIG. 21

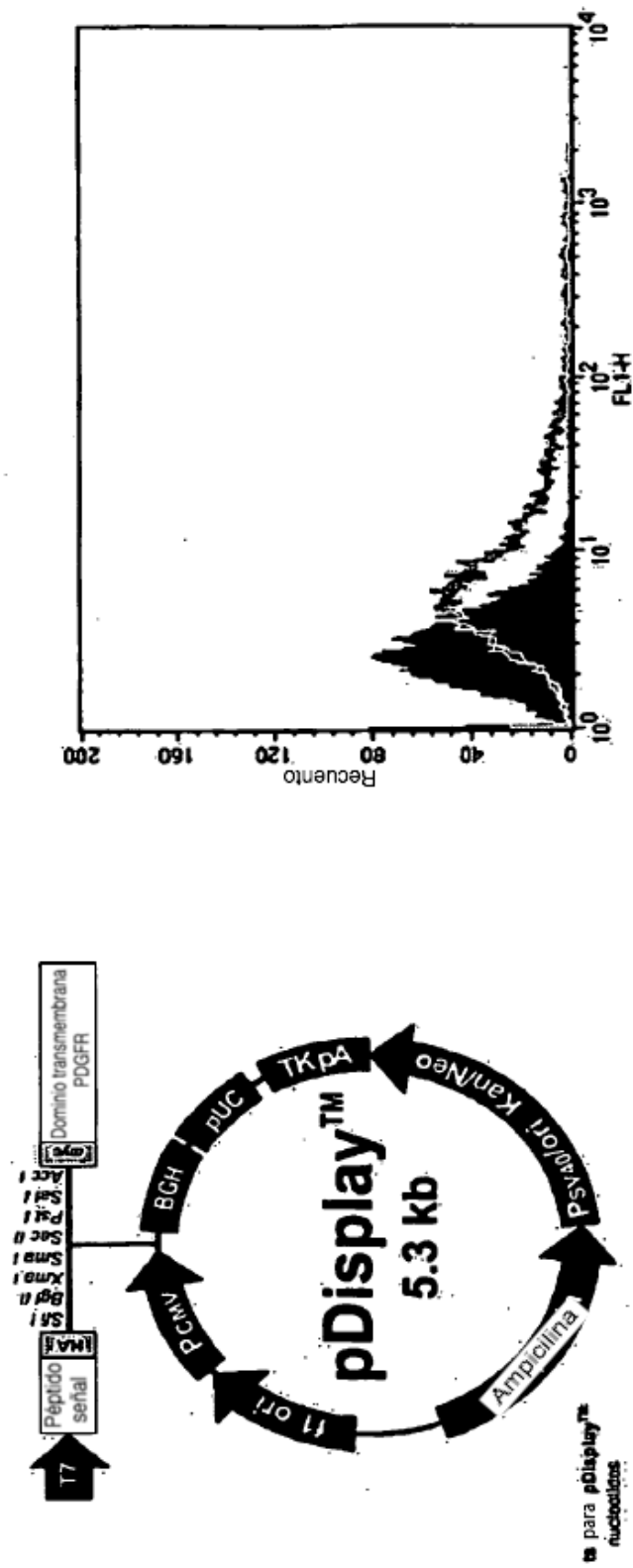


FIG. 22

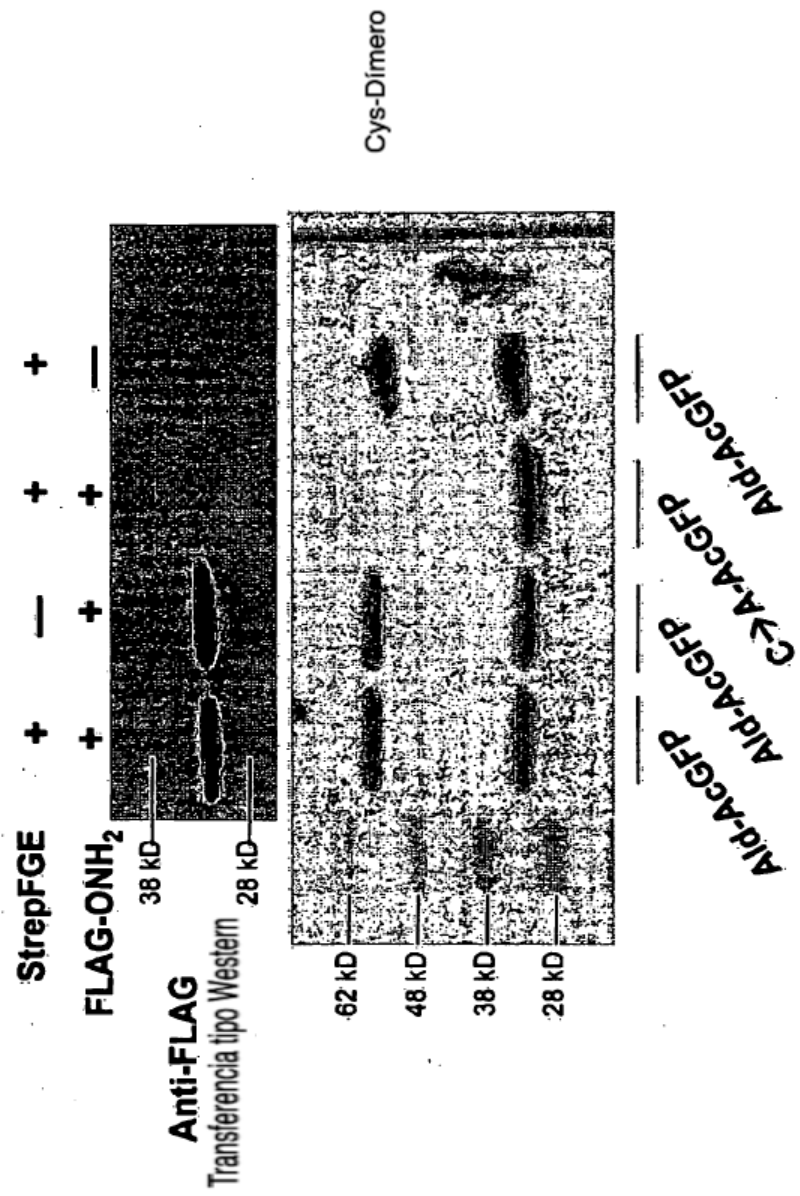
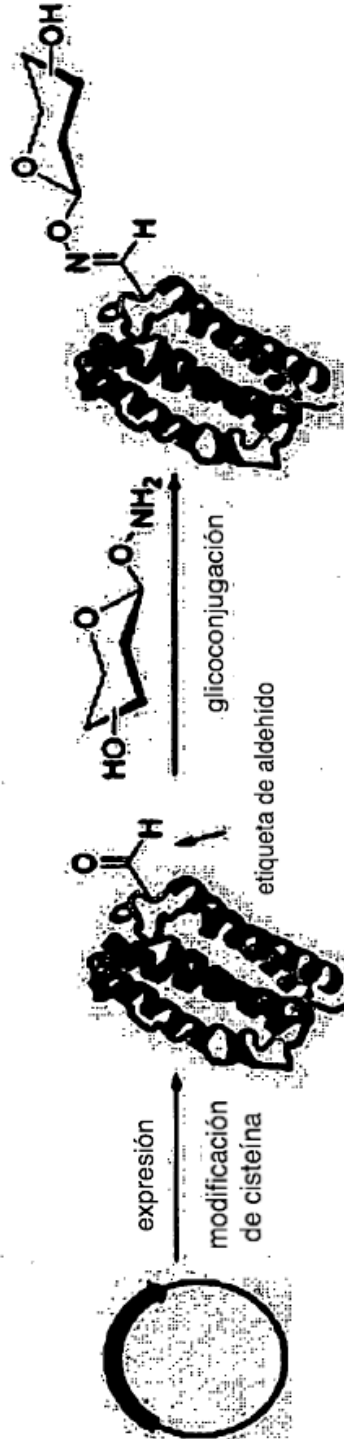


FIG. 23



- IFN-Beta: cinco helices α (A-E) con un solo sitio de glicosilación existente en el residuo Asn-80