

(19)



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS
ESPAÑA



(11) Número de publicación: **2 986 910**

(51) Int. Cl.:

C12N 9/24 (2006.01)
C12Q 1/37 (2006.01)
G01N 33/68 (2006.01)
C12N 9/52 (2006.01)

(12)

TRADUCCIÓN DE PATENTE EUROPEA

T3

(86) Fecha de presentación y número de la solicitud internacional: **25.05.2018 PCT/EP2018/063832**

(87) Fecha y número de publicación internacional: **29.11.2018 WO18215656**

(96) Fecha de presentación y número de la solicitud europea: **25.05.2018 E 18728576 (2)**

(97) Fecha y número de publicación de la concesión europea: **12.06.2024 EP 3630967**

(54) Título: **Proteasa y polipéptido de unión para o-glicoproteínas**

(30) Prioridad:

26.05.2017 GB 201708471
26.05.2017 GB 201708476
24.04.2018 GB 201806655

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
13.11.2024

(73) Titular/es:

GENOVIS AB (100.0%)
Box 4
244 21 Kaevlinge, SE

(72) Inventor/es:

LEO, FREDRIK;
LOOD-ALAYÓN, ROLF;
BJORK, STEPHAN;
MEJARE, MALIN y
OLSSON, FREDRIK

(74) Agente/Representante:

CARVAJAL Y URQUIJO, Isabel

Observaciones:

Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 986 910 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

Proteasa y polipéptido de unión para o-glicoproteínas

5 **Campo de la invención**

La presente invención se refiere a una nueva endoproteasa, mutantes de la misma que tienen actividad de unión pero carecen de actividad hidrolizante o la tienen reducida, y su uso en métodos de estudio y aislamiento de glicoproteínas O-ligadas.

10 **Antecedentes de la invención**

Recientemente, se ha prestado mayor atención al impacto de la glicosilación en las funciones biológicas, en particular en lo que se refiere a los glicanos O-ligados. Sin embargo, aunque se ha renovado el interés por estas 15 importantes modificaciones de las proteínas, han faltado herramientas para estudiar eficazmente los glicanos y las glicoproteínas.

Se han desarrollado varias exo- y endoglicosidasas muy útiles tanto para la eliminación de glicanos O-ligados de 20 proteínas nativas como para la secuenciación de glicanos. Ambos enfoques pueden utilizarse individualmente para reducir la heterogeneidad de las glicoproteínas, facilitando así el análisis de la proteína y sus péptidos fragmentados en espectrometría de masas. También se puede llevar a cabo un análisis más eficaz del efecto biológico de los glicanos mediante un análisis posterior de las funciones afectadas por la hidrólisis. Sin embargo, 25 estas herramientas no son eficaces, por ejemplo, para facilitar la identificación de glicoproteínas O-ligadas, la determinación del lugar de glicosilación y la purificación de glicopéptidos O-ligados.

La primera endoproteasa específica de la O-glicoproteína, que se une a los O-glicanos e hidroliza principalmente 30 los enlaces R-N cercanos al glicano, se notificó en 1991/1992 (Abdullah et al., J Bacteriol 173, 5597-5603 (1991); Abdullah et al., Infect Immun 60, 56-62 (1992). Sin embargo, esta enzima tiene una utilidad limitada para la medicina y la biotecnología porque sólo es específica para los O-glicanos que comprenden ácidos siálicos (la mayoría pero no todos los glicanos O-ligados) y tiene demandas específicas de aminoácidos, lo que resulta en bajos niveles de hidrólisis en general. Se necesitan mejores herramientas para estudiar las glicoproteínas O-ligadas.

35 **Breve descripción de la invención**

Los presentes inventores han identificado, purificado y caracterizado un nuevo polipéptido de *Akkermansia muciniphila*, denominado en el presente documento como el LS. Este polipéptido actúa como una endoproteasa, escindiendo/hidrolizando específicamente enlaces de aminoácidos N terminal a y en proximidad de un glicano O-ligado, sin mostrar ninguna especificidad o limitación a una secuencia particular de aminoácidos.

40 Los inventores también han modificado la secuencia del LS y han identificado mutantes capaces de unirse a glicanos O-ligados pero que carecen o tienen una capacidad reducida para hidrolizar las glicoproteínas. Estos mutantes pueden utilizarse para la eliminación selectiva, el enriquecimiento o la purificación de O-glicanos libres, O-glicopéptidos y/o O-glicoproteínas.

45 Por consiguiente, en un primer aspecto de la invención, se proporciona una composición que comprende un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas que comprende:

50 (a) una secuencia de aminoácidos de SEQ ID NO: 1;

55 (b) una secuencia de aminoácidos que sea idéntica en al menos un 85% a la secuencia de aminoácidos de SEQ ID NO: 1 o

60 (c) una secuencia de aminoácidos que es un fragmento de la secuencia de SEQ ID NO: 1 o un fragmento de una secuencia de aminoácidos que sea idéntica en un 85% a la secuencia de aminoácidos de SEQ ID NO: 1;

en el que la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en el que Am1757 es un polipéptido de cualquiera de los SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de los SEQ ID NOs: 12 a 14.

65 La invención también proporciona un método de hidrólisis de una O-glicoproteína, en el que el método comprende poner en contacto una muestra que comprende la O-glicoproteína con una composición de la invención y opcionalmente comprende además la detección o análisis de los productos de hidrólisis.

65 Además, se proporciona un método para evaluar el estado de glicosilación de una proteína, que comprende poner en contacto una muestra, que comprende la proteína con una composición de la invención y detectar y/o analizar

los productos producidos, opcionalmente en el que la presencia o ausencia de productos de escisión se utilizan para determinar la presencia o ausencia de una O-glicoproteína en la muestra, y/o en el que dicho análisis se lleva a cabo para identificar el tipo de una cadena de O-glicano y/o su posición de unión a una O-glicoproteína.

- 5 En un segundo aspecto de la invención, se proporciona una composición que comprende un polipéptido que es capaz de unirse a un O-glicano, O-glicopéptido y/o O-glicoproteína y que carece o tiene una actividad endoproteasa reducida específica para proteínas O-glicosiladas, que comprenden una secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20; donde la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en la que Am1757 es un polipéptido de cualquiera de los SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de los SEQ ID NOs: 12 a 14.

La invención también proporciona un método de unión a un O-glicano, O-glicopéptido y/o O-glicoproteína, en el que el método comprende poner en contacto una muestra que comprende el O-glicano, O-glicopéptido y/o O-glicoproteína con una composición de la invención, y opcionalmente determinar si se ha unido o no un O-glicano, O-glicopéptido u O-glicoproteína y/o separar el O-glicano y cualquier glicoproteína enlazada, el O-glicopéptido o la O-glicoproteína de la mezcla resultante.

Además, se proporciona un método para evaluar el estado de glicosilación de una proteína, que comprende poner en contacto una muestra, que comprende la proteína con una composición de la invención y determinar si la proteína está unida o no por el polipéptido dentro de dicha composición.

También se proporciona un método para detectar O-glicopéptidos y/o O-glicoproteínas en una muestra, en el que el método comprende:

- 25 (a) poner en contacto dicha muestra con una composición de la invención para permitir así la formación de un complejo entre el polipéptido dentro de la composición de la invención y el glicopéptido O-ligado y/o la O-glicoproteína (un complejo glicopéptido/proteína-polipéptido O-ligado);
- 30 (b) opcionalmente, separar dicho polipéptido de la muestra contactada; y
- (c) determinar si el polipéptido separado está unido a un glicopéptido o glicoproteína O-ligado, determinando así la presencia o ausencia de glicopéptidos o glicoproteínas O-ligados en la muestra.

35 Breve descripción de las figuras

40 **Figura 1:** Expresión y purificación del LS. El LS se expresó como proteína de fusión con una etiqueta His C-terminal en el vector pET21(a)+. Tras la transformación en BL21(DE3) Star, se expresaron cuatro clones individuales y se purificaron en columnas His GravityFlow hasta alcanzar la homogeneidad. Según la cantidad total de proteína en las muestras purificadas, así como la pureza basada en SDS-PAGE, los cuatro clones investigados se expresaron igual de bien.

45 **Figura 2:** El LS actúa específicamente sobre las proteínas que contienen O-glicanos - la figura muestra los productos analizados por SDS-PAGE. La incubación del LS con IgG o IgA produjo una degradación específica de IgA, pero ninguna actividad visible contra IgG (Herceptina/trastuzumab). Todas las incubaciones se realizaron a 37 °C en PBS. La adición de sialidasa (Am0707) no fue necesaria para la actividad del LS en estas condiciones.

50 **Figura 3:** Condiciones enzimáticas óptimas. El LS es activo en un amplio intervalo de pH (A), tolera bien el NaCl (B), pero es muy sensible al EDTA (C, D) y está parcialmente inhibida por el Zn²⁺ (D). Todos los experimentos (excepto el ensayo de pH) se realizaron en PBS o/n, a 37°C. Para la determinación del pH óptimo, la enzima se incubó en Tris-HCl 20 mM (pH 6.8-8.8) o ácido acético 50 mM (pH 5.6).

55 **Figura 4:** La actividad del LS está regulada por la composición de los glicanos. (A) La eliminación secuencial de glicanos específicos antes de la hidrólisis con el LS durante 30 minutos dio como resultado una actividad muy baja en una proteína sialilada, una actividad alta en una proteína asialilada y ninguna actividad en una muestra con galactosas eliminadas. (S) sialidasa, (SG) sialidasa y galactosidasa, (LS) LS. (B) La incubación prolongada (o/h) de glicoproteínas totalmente glicosiladas (Enbrel) o tratadas con sialidasa (Enbrel(S)) dio lugar a una hidrólisis completa, en ambas muestras Enbrel, también puede denominarse en el presente documento etanercept. (C) La parte de unión al TNFα del etanercept (TNFαR) se pretrató con sialidasa ("sialidasa"), O-glicosidasa/sialidasa ("O-glic"), o con PNGaseF ("N-glic"), para eliminar los ácidos sílicos, los O-glicanos y los N-glicanos, respectivamente. Se añadió el LS a las muestras y se dejó que la incubación continuara o/h antes del análisis. El LS tenía actividad en todas las muestras excepto en las tratadas con O-glicosidasa.

Figura 5: Resultados de la búsqueda que muestran que el LS hidroliza el N-terminal de O-glicanos de la glicoproteína. El etanercept hidrolizado a fragmentos con el LS, y posteriormente deglicosilado con tratamiento con O-glicosidasa, se sometió a análisis de espectrometría de masas (LC/MS y MS/MS). Los péptidos identificados (recuadros blancos y sombreados) se ajustaron a la secuencia de Etanercept basándose en los valores m/z y los datos MS/MS, con los iones y' y b' marcados como pequeños recuadros grises. Todos los recuadros blancos y sombreados (p. ej. péptidos) comienzan directamente en una T o una S, donde se han unido O-glicanos. El aminoácido precedente varía (P, S, H, T, G), y parece probable que no influya en la hidrólisis. (A) Análisis utilizando un enfoque sesgado, buscando específicamente péptidos generados con una S/T-peptidasa. (B) Análisis utilizando un enfoque no sesgado.

Figura 6: El LS inactivado se une específicamente a las glicoproteínas ligadas a O. Se mutó el sitio activo de la metaloproteasa para eliminar la capacidad catalítica sin afectar a la afinidad o interacción con el sustrato. En concreto, esto se hizo cambiando una E por una A, creando así el clón "LS_{mut}" (también denominado LS_{E206A}). (A) Mientras que el LS fue capaz de hidrolizar Enbrel en presencia de sialidasa, el LSmut inactivado no pudo hidrolizar Enbrel en las condiciones probadas. La pérdida de actividad se verificó en SDS-PAGE. (B) A pesar de haber perdido la actividad hidrolítica, el LS_{mut} seguía siendo capaz de unirse a la O-glicoproteína. La unión específica se verificó en columnas de centrifugación con el LS_{E206A} inmovilizado, demostrando una afinidad específica por las glicoproteínas O-ligadas. Inmovilizando el LSmut en sepharose pudimos purificar IgA por afinidad. El Herceptin (Trastuzumab), que carece de O-glicanos, así como la IgA tratada con O-glicosidasa, no se unió a la columna, pero pudo detectarse en el flujo (FT). Neur = Neuraminidasa/Sialidasa 0707.

Figura 7: Los ácidos siálicos enlazados α 2-3 limitan la eficacia del LS. La incubación simultánea del LS con un conjunto de sialidasas diversas durante 30 min - 20 h, utilizando Enbrel como sustrato de glicoproteína, reveló la mayor eficacia en presencia de la sialidasa específica α 2-3 1757, o con la mezcla (0707 + 1757), mientras que la sialidasa de amplio espectro 0707 no era necesaria para la actividad aparentemente completa del LS, sugiriendo así que los enlaces α 2-6 (y α 2-8) no son un problema para la actividad del LS.

Figura 8: Presentación esquemática de la actividad del LS. El LS se une preferentemente a galactosas terminales unidas a GalNAc O-ligado, dando lugar a una hidrólisis N-terminal de la serina o treonina a la que está unido el glicano. La presencia de ácidos siálicos reducirá la eficacia del LS, pero no la inhibirá. No se observa actividad sobre los glicanos ligados a N.

Figura 9: Resultados de un experimento en el que la eritropoyetina se escinde con diferentes combinaciones del LS, PNGasaF para eliminar los N-glicanos, Sialidasas para eliminar el ácido siálico, y con una O-glicosidasa para eliminar los O-glicanos. Los productos de la reacción se analizaron mediante SDS-PAGE, RPLC y espectrometría de masas ESI (A) Resultados del análisis SDS-PAGE: Carril 1 = EPO tratada con PNGasaF y sialidasa; Carril 2 = EPO tratada con PNGasaF y sialidasa + LS; Carril 3 = EPO tratada con PNGasaF + LS; Carril 4 = EPO tratada con PNGasaF, sialidasa, O-glicosidasa antes del LS; Carril 5 = control enzimático. Banda X = EPO no escindida; Banda Y = fragmento terminal N de EPO digerido por el LS; Banda Z = fragmento terminal C de EPO digerido por el LS. Los carriles 2 y 3 muestran que el LS escinde la EPO cuando se han eliminado los ácidos siálicos, así como cuando están intactos. Los carriles 2-3 muestran que el LS también escinde la EPO cuando se han eliminado los N-glicanos con PNGasaF. El carril 4 muestra que el LS no escinde la EPO cuando se han eliminado los O-glicanos. (B) El cromatograma UV muestra los resultados de la separación por RPLC de la EPO tratada con PNGasaF y sialidasa + LS. Se identificaron dos picos principales como se muestra. El pico 1 es el fragmento terminal C de la EPO digerido por el LS; el pico 2 es el fragmento terminal N de la EPO digerido por el LS; (C,D) muestran los resultados del análisis por espectrometría de masas. La Fig. C muestra las masas del fragmento C terminal de la EPO con el O-glicano todavía unido a la serina (ahora N terminal) (Cuadrado = GlcNAc, Círculo = Galactosa). Las diferencias de masa se deben a diferencias en la degradación del O-glicano (pérdida de una Galactosa terminal) en algunas partes de la muestra, probablemente causadas por la energía de ionización en el instrumento MS; la Fig D muestra el fragmento N terminal de EPO carente de glicano, más EPO no digerida con glicano todavía unido.

Figura 10: Resultados de experimentos que muestran que el LS_{E206A} conserva cierta actividad mientras que el LS_{H205A/E206A} (también puede denominarse LS_{HE206AA}) es completamente inactivo. La actividad de los mutantes LS_{E206A} (A) y LS_{H205A/E206A} (B) se evaluó frente a sustratos O-glucosilados asialilados, incluyendo la parte de unión al TNFa del etanercept (TNFaR2; también puede denominarse en el presente documento TNFaR) y el propio etanercept (Etanercept), en comparación con la enzima LS de tipo salvaje. Se añadieron diferentes concentraciones de los mutantes LS a 1 μ g de sustrato (1:1 - 15:1, enzima: sustrato), incubando en PBS a 37°C durante la noche antes del análisis en SDS-PAGE.

5
1015
2025
30

35

40
4550
55

60

65

A) Carril 1: Sólo sustrato asialilado; Carril 2: Sólo LS, carril 3: 0.5 µg LS_{E206A}, carril 4: 5 µg LS_{E206A}, carril 5: TNFaR2 + LS (proporción 1:1), carril 6: TNFaR2 + LS_{E206A} (proporción 1:1).

5 B) Carril 1: Sólo sustrato asialilado; Carril 2: LS_{H205A/E206A} + Etanercept (proporción 15:1), carril 3: LS_{H205A/E206A} + Etanercept (proporción 5:1), carril 4: LS_{H205A/E206A} + Etanercept (proporción 1:1), carril 5: LS + Etanercept (proporción 1:1), carril 6: LS_{H205A/E206A}

10 **Figura 11:** Resultados de experimentos que demuestran que el LS_{H205A/E206A} inmovilizado en resina se une específicamente a proteínas que contienen O-glicanos. Las figuras muestran los análisis SDS-PAGE para el material de partida/carga, el flujo (FT) y el eluido (E) en cada caso. (A) Las muestras incluían BSA (albúmina de suero bovino), Etanercept, IgA e IgG, nativas o pretratadas con mezclas de Sialidasa +/- O-glicosidasa como se muestra. (B) La muestra incluía una mezcla de proteínas O-glicosiladas (TNFaR y ApoE), proteínas N-glicosiladas (afibbercept, AGP (glicoproteína alfa-1-ácido), IgG Fc (dominio Fc de IgG) y proteínas no glicosiladas (BSA), pretratadas con mezcla de Sialidasa. (C) La muestra incluía una mezcla de proteínas N-glicosiladas (cetuximab, afibbercept, AGP) y proteínas no glicosiladas (BSA, anhidrasa carbónica), pretratadas con la mezcla Sialidasa.

15 **Figura 12:** Resultados del experimento que muestra que el LS_{H205A/E206A} inmovilizado tiene una capacidad dependiente de la concentración para unirse a O-glicoproteínas. Se añadió Etanercept asialilado (50-250 µg; en 100 µl de PBS) a 50 µl de resina LS_{H205A/E206A} equilibrada con PBS con diferentes condiciones de inmovilización de LS_{H205A/E206A} (5-15 mg/ml). Se dejó que las proteínas se unieran a la resina durante 2 horas a temperatura ambiente con rotación de extremo a extremo. La resina se lavó tres veces con PBS (350 µl) y después se eluyó con la adición de 8 M de urea (50 µl, 5 min de incubación; 2 repeticiones). Para estudiar el efecto de la urea y el clorhidrato de guanidina (GHCl) en la unión, se incluyó en el tampón de unión junto con 50 µg de Etanercept asialilado, pero por lo demás se manipuló de forma idéntica. (A) Todas las muestras se separaron en SDS-PAGE y la intensidad de las bandas se determinó mediante densitometría utilizando GelDoc EZ y el programa ImageLab. (B) La capacidad de unión de proteínas, determinada por la intensidad de la banda, se representó gráficamente frente a la cantidad del LS_{H205A/E206A} inmovilizado.

20 **Figura 13:** Resultados del experimento que muestra que el LS_{H205A/E206A} puede purificar por afinidad alrededor de 3 mg de etanercept /ml de resina. Se añadió etanercept asialilado (10-200 µg; 100 µl en PBS) a 50 µl de resina LS_{H205A/E206A} equilibrada con PBS. Se dejó que las proteínas se unieran a la resina durante 2 horas a temperatura ambiente con rotación de extremo a extremo. La resina se lavó tres veces con PBS (350 µl) y después se eluyó con la adición de 8 M de urea (50 µl, 5 min de incubación; 2 repeticiones). Todas las muestras se separaron en SDS-PAGE, y las intensidades de banda se determinaron mediante densitometría utilizando GelDoc EZ y el software ImageLab. Los porcentajes indican la intensidad de la banda en comparación con el control.

25 **Figura 14:** Resultados del experimento que demuestra que la interacción del LS_{H205A/E206A}-sustrato es insensible a elevadas fuerzas iónicas y diferencias en el volumen/tipo de tampón, y funciona en un amplio rango de pH. (A) Etanercept asialilado, 50 µg; 100 µl en PBS con NaCl añadido a las concentraciones mostradas (B) Etanercept asialilado, 50 µg; 100-300 µl en PBS como se muestra; (C) Flujo de muestras de etanercept asialilado (50 µg) y BSA (50 µg) en diferentes tampones a diferentes pHs como se muestra; (D) Eluatos de las muestras de C.

30 **Figura 15:** Resultados del experimento que demuestra que la desnaturización o la adición de detergentes eluyen las O-glicoproteínas unidas al LS_{H205A/E206A}. Se añadió Etanercept asialilado (50 µg; 100 µl en PBS) a 50 µl de resina LS_{H205A/E206A} equilibrada con PBS. Se dejó que las proteínas se unieran a la resina durante 2 horas a temperatura ambiente con rotación de extremo a extremo. La resina se lavó tres veces con PBS (350 µl) y después se eluyó con la adición de (A) 1-8 M de urea o (B) 1.25-10% de SDS (50 µl, 5 min de incubación; 2 repeticiones). Todas las muestras se separaron en SDS-PAGE para su análisis.

35 **Figura 16A:** Resultados del experimento que muestra la elución enzimática de O-glicoproteínas unidas al LS_{H205A/E206A} utilizando LS. Se añadieron Abatacept asialilado (10 µg, 100 µl en PBS) y Etanercept (50 µg; 100 µl en PBS) a 50 µl de resina LS_{H205A/E206A} equilibrada con PBS. Se dejó que las proteínas se unieran a la resina durante 2 horas a temperatura ambiente con rotación de extremo a extremo. La resina se lavó tres veces con tampones de unión (350 µl) antes de añadir 50 unidades del LS en un volumen total de 100 µl de PBS. Las muestras se incubaron durante 6-24 horas más a 37°C con agitación (450 rpm). Las O-glicoproteínas/glicopéptidos liberados por el LS se recogieron mediante centrifugación (1000 g, 1 min) antes de que la columna se eluyera finalmente con la adición de 8 M de Urea (50 µl, 5 min de incubación; 2 repeticiones). Todas las muestras se separaron en SDS-PAGE para su análisis.

40 **Figura 16B:** Resultados del análisis por espectrometría de masas (LC/MS y MS/MS) de Etanercept eluido con LS. Los péptidos identificados (Fig 16B.1) coincidían con los generados en una digestión del LS de

etanercept (Fig 16B.2). Los péptidos identificados (recuadros blancos) se ajustaron a la secuencia de Etanercept basándose en los valores m/z y los datos MS/MS, con los iones y' y b' marcados como pequeños recuadros grises. Todos los recuadros blancos (p. ej. péptidos) comienzan directamente en una T o una S, donde se unen los O-glicanos.

Figura 17: Resultados que demuestran la purificación por afinidad y el enriquecimiento de proteínas séricas O-glicosiladas. (A) Se añadió suero asialilado (20 µl; 100 µl en PBS) a 50 µl de resina LS_{H205A/E206A} equilibrada con PBS. Se dejó que las proteínas se unieran a la resina durante 2 horas a temperatura ambiente con rotación de extremo a extremo. La resina se lavó tres veces con tampón de unión (350 µl) y después se eluyó con la adición de 8 M de Urea. (B) Para investigar el impacto de los glicanos en la interacción, se pretrataron las muestras con una mezcla de sialidasa +/- O-glicosidasa. La purificación posterior se realizó como se ha descrito anteriormente. (C) El suero (40 µl) se mezcló con PBS (hasta 100 µl) y la mezcla de sialidasa (50-500 Unidades) y se añadió a una columna equilibrada con PBS, incubándose durante 2 horas a temperatura ambiente con rotación de extremo a extremo, tras lo cual las muestras se lavaron y eluyeron como se ha descrito anteriormente. Todas las muestras se separaron en SDS-PAGE para su análisis.

Figura 18: Resultados que muestran el enriquecimiento de O-glicoproteínas del suero humano. Se aplicó suero humano diluido 2.5 veces en PBS hasta 100 µl sobre 50 µl de resina LS_{H205A/E206A} equilibrada con PBS en una columna de centrifugación. Se añadió una mezcla de 50-500 unidades de Sialidasa y se coincubó en la resina a temperatura ambiente durante 2 horas. Se recogió el flujo y la resina se lavó 5-10 veces con PBS. Las proteínas unidas se eluyeron en 8 M de urea seguida de desnaturización y reducción con la adición de 5 mM de DTT y una incubación de 60 minutos a 37°C. Las cisteínas reducidas se alquilaron con 15 mM de yodoacetamida a temperatura ambiente y en oscuridad durante 30 minutos. Las muestras se intercambiaron en tampón a 50 mM Tris pH 8.0 en una columna de desalinización por centrifugación. Se añadió tripsina (2.5 µg) a la solución y la digestión se realizó durante toda la noche a 37°C. Los péptidos se separaron y analizaron mediante RP-LC MS/MS en una columna C18 en un 0.1%FA en MQ : gradiente de 0.1%FA en 95% ACN a 45°C y un flujo de 0.2 ml/min. La detección se realizó en un ESI-Q-TOF. Los datos se convirtieron en archivos de formato mgf y se buscaron en la base de datos Swiss Prot (A) Péptidos identificados procedentes de proteínas anotadas como proteínas O-glicosiladas o no O-glicosiladas. Sólo se incluyeron las proteínas con >6 péptidos coincidentes y una puntuación MASCOT >200. (B) Diferentes pasos de lavado produjeron cambios en los péptidos identificados, así como cambios en la proporción de proteínas O-glicosiladas y no O-glicosiladas (C). Sia = Sialidasa tratada; Sia Pre = Sialidasa pretratada.

Figura 19: Resultados de experimentos que demuestran que el doble mutante inmovilizado también se une a O-glicopéptidos más cortos. **A** muestra resultados representativos del análisis LC/MS de la unión a una mezcla preparada de un péptido O-glicosilado (glicodrosocina (GD)) y varios péptidos no glicosilados (H2686, H4062 H8390 y cadena beta oxidada de insulina (IOB)). **B** muestra un diagrama esquemático de IgA que ilustra que la digestión tríptica producirá un único péptido O-glicosilado. **C** muestra resultados representativos del análisis LC/MS de la unión al digerido tríptico de IgA.

Figura 20: Resultados de experimentos que demuestran que el doble mutante inmovilizado se compara favorablemente con otras matrices de unión de O-glicoproteínas disponibles en el mercado. **A** muestra geles SDS-PAGE representativos que comparan la presencia de etanercept o etanercept asialilado (etanercept(S)) en flujo (FT) o eluido (E) tras incubación con diferentes lectinas inmovilizadas o LS doble mutante como se muestra. **B** muestra el análisis de densitometría de los geles, en relación con el control positivo de 1.5 µg de sustrato cargado directamente.

Breve descripción de las secuencias

La SEQ ID NO: 1 es la secuencia de aminoácidos de un polipéptido con actividad endoproteasa específica de O-glicoproteína.

La SEQ ID NO: 2 es una secuencia de aminoácidos de un polipéptido ejemplar con actividad endoproteasa específica de O-glicoproteína. Relativo a la SEQ ID NO: 1 incluye una Metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆. El polipéptido que consiste en esta secuencia puede denominarse en la presente LS.

La SEQ ID NO: 3 es una secuencia de nucleótidos que codifica un polipéptido cuya secuencia es la SEQ ID NO: 2.

La SEQ ID NO: 4 es la secuencia de aminoácidos de tipo salvaje de un polipéptido aislado de *Akkermansia muciniphila* con actividad endoproteasa específica de O-glicoproteína. Relativo a la SEQ ID NO: 1 incluye un motivo de señal en el extremo N.

- La SEQ ID NO: 5 es la secuencia de aminoácidos de un polipéptido capaz de unirse a O-glicanos pero que carece de actividad endoproteasa específica de O-glicoproteína o la tiene reducida.
- 5 La SEQ ID NO: 6 es una secuencia de aminoácidos de un polipéptido ejemplar capaz de unirse a O-glicanos pero que carece de actividad endoproteasa específica de O-glicoproteína o la tiene reducida. Relativo a la SEQ ID NO: 5 incluye una Metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆. El polipéptido consistente en esta secuencia puede denominarse en el presente documento LS_{E206A}.
- 10 La SEQ ID NO: 7 es una secuencia de nucleótidos que codifica un polipéptido cuya secuencia es SEQ ID NO: 6.
- La SEQ ID NO: 8 es el motivo de dominio metaloproteasa de un polipéptido ejemplar que tiene actividad endoproteasa específica de O-glicoproteína.
- 15 La SEQ ID NO: 9 es la secuencia de aminoácidos de tipo salvaje de una sialidasa, Am1757, aislada de *Akkermansia muciniphila*. Incluye un motivo de señal en el extremo N.
- 20 La SEQ ID NO: 10 es la secuencia de aminoácidos de tipo salvaje de una sialidasa, Am1757, que carece de un motivo de señal en el extremo N en relación con la SEQ ID NO: 9.
- 25 La SEQ ID NO: 11 es una secuencia de aminoácidos de una sialidasa ejemplar, Am1757. Relativo a la SEQ ID NO: 10 incluye una Metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆.
- La SEQ ID NO: 12 es la secuencia de aminoácidos de tipo salvaje de una sialidasa, Am0707, aislada de *Akkermansia muciniphila*. Incluye un motivo de señal en el extremo N.
- 30 La SEQ ID NO: 13 es la secuencia de aminoácidos de tipo salvaje de una sialidasa, Am0707, que carece de un motivo de señal en el extremo N en relación con la SEQ ID NO: 12.
- La SEQ ID NO: 14 es una secuencia de aminoácidos de una sialidasa ejemplar, Am0707. Relativo a la SEQ ID NO: 13 incluye una Metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆.
- 35 La SEQ ID NO: 15 es la secuencia de aminoácidos de una O-glicosidasa aislada de *S. oralis*.
- Las SEQ ID NO: 16 y 17 son secuencias de cebadores.
- La SEQ ID NO: 18 muestra la secuencia de aminoácidos de la EPO.
- 40 La SEQ ID NO: 20 es la secuencia de aminoácidos de un polipéptido capaz de unirse a O-glicanos pero que carece de actividad endoproteasa específica de O-glicoproteína o la tiene reducida.
- 45 La SEQ ID NO: 21 es una secuencia de aminoácidos de un polipéptido ejemplar capaz de unirse a O-glicanos pero que carece de actividad endoproteasa específica de O-glicoproteína o la tiene reducida. Relativo a la SEQ ID NO: 20 incluye una Metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆. El polipéptido consistente en esta secuencia puede denominarse en el presente documento LS_{HE206AA} o LS_{H205A/E206A}.
- 50 La SEQ ID NO: 22 es una secuencia de nucleótidos que codifica un polipéptido cuya secuencia es la SEQ ID NO: 21.
- Las SEQ ID NO: 23, 24 y 25 son secuencias de motivos de dominio de metaloproteasa alterados, cada una de ellas de un polipéptido de la invención que es capaz de unirse a O-glicanos pero carece de actividad endoproteasa específica de O-glicoproteína o la tiene reducida.
- 55 Las SEQ ID NO: 26, 27 y 28 son las secuencias de aminoácidos de polipéptidos que tienen actividad endoproteasa específica de O-glicoproteína.
- 60 Las SEQ ID NO: 29, 30 y 31 son las secuencias de aminoácidos de polipéptidos ejemplares que tienen actividad endoproteasa específica de O-glicoproteína. Relativo a las SEQ ID NOs: 26, 27 y 28, respectivamente, SEQ ID NOs: 29, 30 y 31 incluyen cada uno una metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆.
- 65 Las SEQ ID NO: 32, 33 y 34 son secuencias de aminoácidos de tipo salvaje de polipéptidos con actividad endoproteasa específica de O-glicoproteína, aislados de *Pseudomonas aeruginosa* PAO1, *Bacteroides thetaiotaomicron* VPI-5482 y *Clostridium perfringens*, respectivamente. Relativo a las SEQ ID NOs: 26, 27 y 28, respectivamente, cada uno incluye un motivo de señal en el extremo N.

Las SEQ ID NO: 35, 36 y 37 son las secuencias de aminoácidos de polipéptidos capaces de unirse a O-glicanos pero que carecen de actividad endoproteasa específica de O-glicoproteína o la tienen reducida.

- 5 La SEQ ID NO: 38, 39 y 40 son las secuencias de aminoácidos de polipéptidos ejemplares capaces de unirse a O-glicanos pero que carecen de actividad endoproteasa específica de O-glicoproteína o la tienen reducida. Relativo a las SEQ ID NOs: 35, 36 y 37, respectivamente, SEQ ID NOs: 38, 39 y 40 incluyen cada uno una metionina N terminal adicional y un enlazador C-terminal + etiqueta His₆.
- 10 Las SEQ ID NO: 41 - 43 son las secuencias de aminoácidos de motivos representativos de metaloproteasas de endoproteasas específicas de O-glicoproteínas.
- 15 Las SEQ ID NO: 44 - 46 son las secuencias de aminoácidos de motivos representativos de metaloproteasas alteradas de polipéptidos capaces de unirse a O-glicanos pero que carecen de actividad endoproteasa específica de O-glicoproteína o la tienen reducida.
- 20 La SEQ ID NO: 47 es la secuencia de aminoácidos del péptido glicodrosocina. Hay un sitio de O-glicosilación en el residuo T.
- 25 Las SEQ ID NO: 48 a 50 son las secuencias de aminoácidos de los péptidos que no están O-glicosilados.

Descripción detallada de la invención

- 25 Debe entenderse que las diferentes aplicaciones de los productos y métodos descritos pueden adaptarse a las necesidades específicas de la técnica. También debe entenderse que la terminología utilizada en la presente tiene el propósito de describir realizaciones particulares de la invención únicamente, y no pretende ser limitativa.
- 30 Tal como se usa en la presente memoria descriptiva y las reivindicaciones adjuntas, las formas singulares "un", "una" y "el/la" incluyen referentes plurales a menos que el contenido indique claramente lo contrario. Por lo tanto, por ejemplo, la referencia a "un polipéptido" incluye "polipéptidos" y similares.

Características generales del polipéptido

- 35 Un "polipéptido" se usa en la presente en su sentido más amplio para referirse a un compuesto de dos o más subunidades de aminoácidos, análogos de aminoácidos u otros peptidomiméticos. Por lo tanto, el término "polipéptido" incluye secuencias peptídicas cortas y también polipéptidos y proteínas más largos. Los términos "proteína", "péptido" y "polipéptido" se pueden usar indistintamente. Tal como se usa en la presente, el término "aminoácido" se refiere a aminoácidos naturales y/o no naturales o sintéticos, que incluyen isómeros ópticos D o L, y análogos de aminoácidos y peptidomiméticos.
- 40 Un polipéptido puede producirse mediante un método adecuado, que incluye métodos recombinantes o sintéticos. Por ejemplo, el polipéptido se puede sintetizar directamente usando técnicas estándar conocidas en la técnica, tales como química de fase sólida Fmoc, química de fase sólida Boc o mediante síntesis de péptidos en fase de solución. Alternativamente, un polipéptido puede producirse transformando una célula, típicamente una célula bacteriana, con una molécula de ácido nucleico o vector que codifica dicho polipéptido. La producción de polipéptidos mediante expresión en células hospedadoras bacterianas se describe a continuación y se ejemplifica en los Ejemplos. En el presente documento se describen moléculas de ácido nucleico y vectores que codifican un polipéptido. También se describe en el presente documento una célula huésped que comprende dicho ácido nucleico o vector. Las moléculas polinucleotídicas ejemplares que codifican los polipéptidos aquí divulgados se proporcionan como SEQ ID NOs: 3 y 7. Cada una de estas secuencias incluye en el extremo 5' un codón para la metionina terminal N (ATG) y, antes del codón de parada (TAA) en el extremo 3', codones para un enlazador Gly-Ser-Gly y una etiqueta 6x His, que opcionalmente pueden excluirse. La inclusión opcional de una metionina adicional y una etiqueta se analizan con más detalle a continuación.
- 45 55 Los términos "molécula de ácido nucleico" y "polinucleótido" se usan indistintamente en la presente y se refieren a una forma polimérica de nucleótidos de cualquier longitud, ya sea desoxirribonucleótidos o ribonucleótidos, o análogos de estos. Los ejemplos no limitativos de polinucleótidos incluyen un gen, un fragmento de gen, ARN mensajero (ARNm), ADNc, polinucleótidos recombinantes, plásmidos, vectores, ADN aislado de cualquier secuencia, ARN aislado de cualquier secuencia, sondas de ácido nucleico y cebadores. Un polinucleótido codifica un polipéptido y puede proporcionarse en forma aislada o sustancialmente aislada. Por sustancialmente aislada, se entiende que puede haber un aislamiento sustancial, pero no total, del polipéptido de cualquier medio circundante. Los polinucleótidos se pueden mezclar con portadores o diluyentes que no interferirán con su uso previsto y aún se considerarán sustancialmente aislados. Una secuencia de ácido nucleico que "codifica" un polipéptido seleccionado es una molécula de ácido nucleico que se transcribe (en el caso del ADN) y se traduce (en el caso del ARNm) en un polipéptido *in vivo* cuando se coloca bajo el control de secuencias reguladoras apropiadas, por ejemplo, en un vector de expresión. Los límites de la secuencia codificante están determinados
- 60 65

por un codón de inicio en el extremo 5' (amino) y un codón de terminación de la traducción en el extremo 3' (carboxi). Tales secuencias de ácido nucleico pueden incluir, pero sin limitarse a ADNc de ARNm viral, procariota o eucariota, secuencias genómicas de ADN o ARN viral o procariota e incluso secuencias de ADN sintéticas. Una secuencia de terminación de la transcripción puede localizarse en 3' con respecto a la secuencia codificante.

5 Los polinucleótidos se pueden sintetizar de acuerdo con métodos bien conocidos en la técnica, como se describe a modo de ejemplo en Sambrook et al (1989, Molecular Cloning - a laboratory manual; Cold Spring Harbor Press). Puede proporcionarse una molécula de ácido nucleico en forma de casete de expresión que incluya secuencias de control enlazadas operablemente a la secuencia insertada, permitiendo así la expresión del polipéptido in vivo (por ejemplo, en sistemas de expresión procarióticos o eucarióticos). Estos cassetes de expresión, a su vez, se proporcionan típicamente dentro de vectores (por ejemplo, plásmidos o vectores virales recombinantes). Tal casete de expresión se puede administrar directamente a un sujeto hospedador. Alternativamente, un vector que comprende un polinucleótido puede ser administrado a un sujeto huésped. Preferentemente, el polinucleótido se prepara y/o administra utilizando un vector genético. Un vector adecuado puede ser cualquier vector que sea capaz de portar una cantidad suficiente de información genética y permitir la expresión de un polipéptido.

10 20 En la presente se describen vectores de expresión que comprenden tales secuencias de polinucleótidos. Tales vectores de expresión se construyen rutinariamente en el arte de la biología molecular y pueden, por ejemplo, implicar el uso de ADN plasmídico e iniciadores apropiados, promotores, potenciadores y otros elementos, tales como, por ejemplo, señales de poliadenilación que pueden ser necesarias, y que se colocan en la orientación correcta, con el fin de permitir la expresión de un péptido de la invención. Otros vectores adecuados serían evidentes para los expertos en la técnica. A modo de ejemplo adicional en este sentido, nos remitimos a Sambrook et al.

25 30 También se describen en la presente células que se han modificado para expresar un polipéptido. Tales células típicamente incluyen células procariotas tales como células bacterianas, por ejemplo, *E. coli*. Tales células se pueden cultivar usando métodos rutinarios para producir un polipéptido.

35 40 Un polipéptido puede derivatizarse o modificarse para ayudar con su producción, aislamiento o purificación. Por ejemplo, cuando se produce un polipéptido mediante expresión recombinante en una célula hospedadora bacteriana, la secuencia del polipéptido puede incluir un residuo de metionina (M) adicional en el extremo N para mejorar la expresión. Como otro ejemplo, el polipéptido puede derivatizarse o modificarse mediante la adición de un ligando que es capaz de unirse directa y específicamente a un medio de separación. Alternativamente, el polipéptido puede derivatizarse o modificarse mediante la adición de un miembro de un par de unión y el medio de separación comprende un reactivo que se deriva o modifica mediante la adición del otro miembro de un par de unión. Se puede utilizar cualquier par de unión adecuado. Cuando el polipéptido se derivatiza o modifica mediante la adición de un miembro de un par de unión, el polipéptido es preferiblemente marcado con histidina o biotina. Típicamente, la secuencia codificante de aminoácidos de la etiqueta de histidina o biotina se incluye a nivel génico y el polipéptido se expresa de forma recombinante en *E. coli*. La etiqueta de histidina o biotina está típicamente presente en cualquier extremo del polipéptido, preferentemente en el extremo C. Puede unirse directamente al polipéptido o indirectamente mediante cualquier secuencia enlazadora adecuada, como 3, 4 o 5 residuos de glicina, o una mezcla de residuos de glicina y serina. La etiqueta de histidina típicamente consiste en seis residuos de histidina, aunque puede ser más larga que esto, típicamente hasta 7, 8, 9, 10 o 20 aminoácidos o más corta, por ejemplo 5, 4, 3, 2 o 1 aminoácidos.

45 50 55 Un polipéptido puede proporcionarse en una forma sustancialmente aislada o purificada. Es decir, aislado de la mayoría de los otros componentes presentes en un extracto celular de una célula en la que se expresó el polipéptido. Por sustancialmente purificado, se entenderá que el polipéptido se purifica hasta al menos 50%, 60%, 70%, 80% o preferentemente al menos 90% de homogeneidad. El nivel de pureza se puede evaluar por cualquier medio adecuado, pero normalmente implica el análisis SDS-PAGE de una muestra, seguido de la detección con azul de Coomassie. Un polipéptido puede mezclarse con portadores, diluyentes o conservantes que no interfieran con el propósito previsto del polipéptido y que aún se consideren sustancialmente aislados o purificados. Cuando se proporciona un polipéptido en una composición con un componente activo adicional, tal como otro polipéptido, cada uno de dichos polipéptidos se purificó individualmente hasta un alto nivel de homogeneidad antes de mezclarse a una relación apropiada para el propósito previsto de cada uno. Por ejemplo, cada uno de dos polipéptidos puede purificarse hasta al menos 90% de homogeneidad antes de combinarse en una relación 1:1.

60 65 Se puede proporcionar un polipéptido (o mezcla de estos) en forma liofilizada, adecuado para la reconstitución en solución acuosa antes de su uso. La composición liofilizada tiene una estabilidad mejorada que permite un almacenamiento más prolongado del polipéptido. En la presente se proporciona un método para preparar un polipéptido (o mezcla de estos) en forma liofilizada, que comprende liofilizar dicho polipéptido (o mezcla) en un tampón adecuado, tal como solución salina tamponada con Tris (TBS). Típicamente, un polipéptido se purifica sustancialmente antes de la liofilización. El polipéptido (o mezcla) resultante en forma liofilizada también se proporciona. Un método para preparar una solución de un polipéptido (o mezcla), que comprende proporcionar el polipéptido (o mezcla) en forma liofilizada y reconstituir con un portador o diluyente adecuado, tal como agua, también se proporciona.

5 Un polipéptido puede inmovilizarse utilizando métodos conocidos en la técnica, por ejemplo como se describe en Datta S et al., Enzyme immobilization: an overview on techniques and support materials, 3 Biotech, 3(1): 1-9 (2013). Por ejemplo, el polipéptido puede inmovilizarse por adsorción, unión covalente, inmovilización por afinidad o atrapamiento. Los materiales que se pueden usar como soportes incluyen, pero no se limitan a, por ejemplo, soportes naturales tales como agarosa, colágeno, gelatina, celulosa, pectina, sefarosa, materiales inorgánicos tales como cerámica, sílice, vidrio, carbón activado o carbón vegetal, o polímeros sintéticos. Por ejemplo, el polipéptido puede inmovilizarse en sefarosa o agarosa, opcionalmente suministrada como resina.

10 **Polipéptidos con actividad endoproteasa**

Características funcionales de un polipéptido con actividad endoproteasa

15 En el presente documento se describe un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas. En otras palabras, el polipéptido tiene actividad endoproteasa específica de O-glicoproteína. El polipéptido escinde cualquier glicoproteína O-ligada, preferiblemente cualquier glicoproteína O-ligada humana. Los ejemplos de glicoproteínas O-ligadas incluyen cualquier proteína que comprenda o consista total o parcialmente en una inmunoglobulina, incluidos los anticuerpos de longitud completa, los fragmentos Fc y las proteínas de fusión Fc, en particular las de los isotipos IgA, IgD e IgG3. Otro ejemplo de glicoproteína O-ligada es el Etanercept, que es una proteína de fusión del dominio de unión al ligando del receptor 2 del TNF α humano unido a la porción Fc de la IgG1, con numerosos sitios de O-glicosilación. Otros ejemplos de glicoproteínas O-ligadas son la eritropoyetina (EPO), los receptores TNF α , la fetuina y el plasminógeno.

20 La hidrólisis (*es decir*, la escisión) de la glicoproteína sustrato suele producirse con gran especificidad en un enlace peptídico N-terminal y muy cerca de una serina o treonina O-glicosilada, y depende del O-glicano. Un polipéptido como el descrito en el presente documento, es preferiblemente capaz de escindir dicho enlace peptídico en estrecha proximidad a cada sitio O-glicosilado en la glicoproteína sustrato. La reacción preferentemente no muestra ninguna especificidad o limitación de aminoácidos, y en particular no requiere que ningún aminoácido(s) específico(s) esté(n) presente(s) en el extremo N-terminal de la serina o treonina O-glicosilada. Cuando se evalúa 25 utilizando parámetros estándar de espectrometría de masas, generalmente se observa que el sitio de escisión se encuentra en el enlace peptídico inmediatamente N terminal a cada residuo O-glicosilado.

30 La actividad endoproteasa y la especificidad de un polipéptido dado pueden determinarse mediante un ensayo adecuado. Por ejemplo, un sustrato de O-glicoproteína estándar, como una molécula de IgA o eritropoyetina (EPO), 35 puede incubarse con un polipéptido de prueba. Los materiales de partida y los productos de reacción pueden analizarse mediante SDS-PAGE y/o espectrometría de masas para determinar la presencia de productos de escisión (si los hay) y, si es necesario, también para caracterizar aún más esos productos. Un sustrato de glucoproteína que no está O-glicosilado, tal como una molécula de IgG1, se puede utilizar como control negativo. Los resultados pueden compararse con los obtenidos en el mismo ensayo cuando el sustrato se pone en contacto 40 con un polipéptido ejemplar según se describe en el presente documento, como un polipéptido que consiste en la secuencia de aminoácidos de SEQ ID NO: 2. Una unidad del polipéptido de la SEQ ID NO: 2 se define como la cantidad necesaria para digerir > 90% de 1 μ g de Eritropoyetina (EPO) en combinación con una unidad de una mezcla de sialidasa en tampón Tris 20 mM pH 6.8, durante la noche a 37°C, según se controla por SDS-PAGE (las mezclas de sialidasa preferidas se describen más adelante). Un polipéptido de prueba preferiblemente logra un 45 nivel similar de actividad cuando está presente en la misma cantidad. Los ensayos de ejemplo también se describen en los Ejemplos.

Características estructurales de un polipéptido con actividad endoproteasa

50 En esta sección se exponen las características estructurales de un polipéptido, que se aplican, además de las características funcionales descritas en la sección anterior.

55 El polipéptido suele tener al menos 150, 200, 250, 275, 280, 290, 300, 310, 320, 330, 340, 350 o 360 aminoácidos de longitud. El polipéptido no suele tener más de 400, 395, 390, 385, 380, 375, 370 o 365 aminoácidos de longitud. Se apreciará que cualquiera de los límites inferiores enumerados anteriormente puede combinarse con cualquiera de los límites superiores enumerados anteriormente para proporcionar un rango para la longitud del polipéptido. Por ejemplo, el polipéptido puede tener una longitud de 150 a 400 aminoácidos, o de 280 a 380 aminoácidos. El polipéptido tiene preferiblemente una longitud de 340 a 380 aminoácidos, y más preferiblemente de 360 a 375 aminoácidos.

60 La estructura primaria (secuencia de aminoácidos) del polipéptido se basa en la estructura primaria del polipéptido codificado por el gen *Amuc1119* de *Akkermansia muciniphila*. La secuencia completa de este polipéptido se muestra en la SEQ ID NO: 4, que incluye un motivo señal en las posiciones 1-24. La secuencia con el motivo de señal eliminado se muestra en la SEQ ID NO: 1.

65 El polipéptido puede comprender, consistir esencialmente o estar formado por la secuencia de SEQ ID NO: 1.

Alternativamente, el polipéptido puede comprender, consistir esencialmente o consistir en una variante de la secuencia de aminoácidos de SEQ ID NO: 1 que es al menos un 50% idéntica a la secuencia de aminoácidos de SEQ ID NO: 1. La secuencia variante puede ser al menos un 60%, al menos un 70%, al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 98% o al menos un 99% idéntica a la secuencia de SEQ ID NO: 1. El nivel de identidad es preferentemente de al menos el 85% o superior. Identidad relativa a la secuencia de SEQ ID NO: 1 puede medirse sobre una región de al menos 100, al menos 200, al menos 300 o al menos 350 o más aminoácidos contiguos de la secuencia mostrada en la SEQ ID NO: 1, o más preferiblemente en toda la longitud de la SEQ ID NO: 1. Una variante es típicamente de una longitud que no es más de 50 aminoácidos más larga o más corta que la secuencia de referencia, y es preferiblemente de aproximadamente (o exactamente) la misma longitud que la secuencia de referencia.

La identidad de aminoácidos se puede calcular usando cualquier algoritmo adecuado. Por ejemplo, los algoritmos PILEUP y BLAST se pueden utilizar para calcular secuencias de identidad o alineación (tales como la identificación de secuencias equivalentes o correspondientes (típicamente en sus configuraciones predeterminadas), por ejemplo, como se describe en Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S. F et al (1990) J Mol Biol 215:403-10. El software para realizar análisis BLAST está disponible públicamente a través del Centro Nacional de Información Biotecnológica (<http://www.ncbi.nlm.nih.gov/>). Este algoritmo implica identificar primero el par de secuencias de alta puntuación (HSP) mediante la identificación de palabras cortas de longitud W en la secuencia de consulta que coinciden o satisfacen alguna puntuación umbral T de valor positivo cuando se alinean con una palabra de la misma longitud en una secuencia de base de datos. T se conoce como el umbral de puntuación de la palabra vecina (Altschul *et al*, supra). Estos aciertos iniciales de palabras vecinas actúan como semillas para iniciar búsquedas para encontrar HSP que los contengan. Los aciertos de palabras se extienden en ambas direcciones a lo largo de cada secuencia en la medida en que se puede aumentar la puntuación de alineación acumulativa. Las extensiones de las palabras coincidentes en cada dirección se detienen cuando: la puntuación acumulada de la alineación cae en la cantidad X desde su valor máximo alcanzado; la puntuación acumulada llega a cero o menos, debido a la acumulación de una o más alineaciones de residuos con puntuación negativa; o se alcanza el final de cualquiera de las secuencias. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y la velocidad de la alineación. El programa BLAST utiliza por defecto una longitud de palabra (W) de 11, la matriz de puntuación BLOSUM62 (véase Henikoff and Henikoff (1992) Proc. Natl. Acad. Sci. USA 89: 10915-10919) alineaciones (B) de 50, expectativa (E) de 10, M=5, N=4 y una comparación de ambas cadenas.

El algoritmo BLAST realiza un análisis estadístico de la similitud entre dos secuencias; véase, por ejemplo, Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5787. Una medida de similitud proporcionada por el algoritmo BLAST es la probabilidad de suma más pequeña (P(N)), que proporciona una indicación de la probabilidad por la cual una coincidencia entre dos secuencias de polinucleótidos o aminoácidos ocurriría por casualidad. Por ejemplo, una secuencia se considera similar a otra secuencia si la probabilidad de suma más pequeña en comparación de la primera secuencia con la segunda secuencia es menor que aproximadamente 1, preferentemente menor que aproximadamente 0.1, más preferentemente menor que aproximadamente 0.01 y lo más preferentemente menor que aproximadamente 0.001. Alternativamente, el paquete UWGCG proporciona el programa BESTFIT que se puede utilizar para calcular la identidad (por ejemplo, se utiliza en su configuración predeterminada) (Devereux *et al* (1984) Nucleic Acids Research 12, 387-395).

La secuencia de un polipéptido puede comprender una variante de la secuencia de aminoácidos de SEQ ID NO: 1 en la que se realizan modificaciones, como adiciones, supresiones o sustituciones de aminoácidos, en relación con la secuencia de SEQ ID NO: 1. A menos que se especifique lo contrario, las modificaciones son preferentemente sustituciones de aminoácidos conservativas. Las sustituciones conservadoras reemplazan los aminoácidos con otros aminoácidos de estructura química similar, propiedades químicas similares o volumen de cadena lateral similar. Los aminoácidos introducidos pueden tener polaridad, hidrofilia, hidrofobicidad, basicidad, acidez, neutralidad o carga similares a los aminoácidos que reemplazan. Alternativamente, la sustitución conservadora puede introducir otro aminoácido que es aromático o alifático en lugar de un aminoácido aromático o alifático preexistente. Los cambios conservadores de aminoácidos son bien conocidos en la técnica y se pueden seleccionar de acuerdo con las propiedades de los 20 aminoácidos principales como se define a continuación en la Tabla A1. Cuando los aminoácidos tienen una polaridad similar, esto se puede determinar por referencia a la escala de hidropatía para las cadenas laterales de aminoácidos en la Tabla A2. Una secuencia de un polipéptido descrito en el presente documento puede comprender una variante de la secuencia de aminoácidos de SEQ ID NO: 1 en la que se realizan hasta 10, 20, 30, 40, 50 o 60 sustituciones conservadoras.

Tabla A1 - Propiedades químicas de los aminoácidos

Ala (A)	alifático, hidrófobo, neutro	Met (M)	hidrófobo, neutro
Cys (C)	polar, hidrófobo, neutro	Asn (N)	polar, hidrófilo, neutro
Asp (D)	polar, hidrófilo, cargado (-)	Pro (P)	hidrófobo, neutro

Glu (E)	polar, hidrófilo, cargado (-)	Gln (Q)	polar, hidrófilo, neutro
Phe (F)	aromático, hidrófobo, neutro	Arg (R)	polar, hidrófilo, cargado (+)
Gly (G)	alifático, neutro	Ser (S)	polar, hidrófilo, neutro
His (H)	aromático, polar, hidrófilo, cargado (+)	Thr (T)	polar, hidrófilo, neutro
Ile (I)	alifático, hidrófobo, neutro	Val (V)	alifático, hidrófobo, neutro
Lys (K)	polar, hidrófilo, cargado(+)	Trp (W)	aromático, hidrófobo, neutro
Leu (L)	alifático, hidrófobo, neutro	Tyr (Y)	aromático, polar, hidrófobo

Tabla A2 - Escala de hidropatía

Cadena lateral	Hidropatía
Ile	4.5
Val	4.2
Leu	3.8
Phe	2.8
Cys	2.5
Met	1.9
Ala	1.8
Gly	-0.4
Thr	-0.7
Ser	-0.8
Tip	-0.9
Tyr	-1.3
Pro	-1.6
His	-3.2
Glu	-3.5
Gln	-3.5
Asp	-3.5
Asn	-3.5
Lys	-3.9
Arg	-4.5

- 5 La secuencia de aminoácidos de un polipéptido descrito en el presente documento puede comprender una variante de la secuencia de aminoácidos de SEQ ID NO: 1 como se ha descrito anteriormente. Sin embargo, algunos residuos de la secuencia de aminoácidos de SEQ ID NO: 1 se mantienen preferentemente dentro de dicha secuencia variante. Por ejemplo, dicha secuencia variante suele conservar ciertos residuos que se sabe que son necesarios para la actividad endoproteasa. Así, el glutamato en la posición 182 de la SEQ ID NO: 1 (corresponde a la posición 206 de la SEQ ID NO: 4) se mantiene preferiblemente en la secuencia de aminoácidos de un polipéptido descrito en el presente documento. Se cree que este residuo es necesario para la transferencia de electrones en el sitio activo. Así, un polipéptido descrito en el presente documento comprende típicamente una variante de la secuencia de aminoácidos de SEQ ID NO: 1 que tiene un glutamato (E) en la posición de dicha secuencia variante que corresponde a la posición 182 de la SEQ ID NO: 1. Del mismo modo, la histidina en la posición 181 de la SEQ ID NO: 1 (corresponde a la posición 205 de la SEQ ID NO: 4) se mantiene preferiblemente en la secuencia de aminoácidos de un polipéptido descrito en el presente documento. Se cree que este residuo es necesario para la unión a un cofactor de ion de zinc.
- 10 15 20 Dicho glutamato y dicho residuo de histidina están ambos típicamente comprendidos dentro de un dominio metaloproteasa que tiene el motivo HEbbH, donde b es un aminoácido sin carga, como el aminoácido A, C, F, G,

- I, L, M, N, P, Q, S, T, V o W. Un ejemplo preferido de dicho dominio tiene la secuencia HELGH (SEQ ID NO: 41), que corresponde a las posiciones 181 a 185 de la SEQ ID NO: 1 (posiciones 205 a 209 en la SEQ ID NO: 4). Así, un polipéptido descrito en el presente documento comprende típicamente una variante de la secuencia de aminoácidos de SEQ ID NO: 1 que comprende el motivo HEbbH (tal como HEIGH (SEQ ID NO: 42) o HELGH, preferiblemente HELGH), en las posiciones correspondientes a las posiciones 181 a 185 de la SEQ ID NO: 1. Un polipéptido descrito en el presente documento comprende típicamente un dominio de unión específico de O-glicano situado en el extremo C-terminal del dominio metaloproteasa.
- 5 El motivo HEbbH puede estar comprendido dentro de un dominio metaloproteasa mayor que tenga el motivo abxHEbbHbc, donde a es el aminoácido V, T o G, b es un aminoácido sin carga, como el aminoácido A, C, F, G, I, L, M, N, P, Q, S, T, V o W, x es cualquier aminoácido, y c es un aminoácido hidrófobo como A, C, F, I, L, M, P, V, W o Y. Un ejemplo preferido de dicho dominio tiene la secuencia GMAHELGHGL (SEQ ID NO: 8), que corresponde a las posiciones 178 a 187 de la SEQ ID NO: 1 (posiciones 202 a 211 en la SEQ ID NO: 4). Otros ejemplos incluyen GVAHELGHNF (SEQ ID NO: 43). Así, un polipéptido descrito en el presente documento comprende preferiblemente una variante de la secuencia de aminoácidos de SEQ ID NO: 1 que comprende el motivo abxHEbbHbc, (como GMAHELGHGL o GVAHELGHNF, preferiblemente GMAHELGHGL), en las posiciones correspondientes a las posiciones 178 a 187 de la SEQ ID NO: 1. Un polipéptido descrito en el presente documento comprende típicamente un dominio de unión específico de O-glicano situado en el extremo C-terminal del dominio metaloproteasa.
- 10 Alternativamente, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente o consistir en un fragmento más corto de la SEQ ID NO: 1 o de una variante de la misma como se describió anteriormente. Los fragmentos pueden describirse como una forma truncada de la SEQ ID NO: 1 que conserva la actividad endoproteasa específica de la O-glicoproteína. Tales fragmentos son más cortos que la SEQ ID NO: 1 y suelen tener al menos 100, 150 o 200 aminoácidos de longitud. Los fragmentos comprenden típicamente un dominio metaloproteasa en posiciones correspondientes a las posiciones 178 a 187 de la SEQ ID NO: 1, incluyendo un residuo de ácido glutámico (E) en una posición que corresponde a la posición 182 de la SEQ ID NO: 1 y un residuo de histidina (H) en una posición que corresponde a la posición 181 de la SEQ ID NO: 1, y un dominio de unión específico de O-glicano situado en el extremo C-terminal del dominio metaloproteasa.
- 15 20 Alternativamente, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente o consistir en un fragmento más corto de la SEQ ID NO: 1 o de una variante de la misma como se describió anteriormente. Los fragmentos pueden describirse como una forma truncada de la SEQ ID NO: 1 que conserva la actividad endoproteasa específica de la O-glicoproteína. Tales fragmentos son más cortos que la SEQ ID NO: 1 y suelen tener al menos 100, 150 o 200 aminoácidos de longitud. Los fragmentos comprenden típicamente un dominio metaloproteasa en posiciones correspondientes a las posiciones 178 a 187 de la SEQ ID NO: 1, incluyendo un residuo de ácido glutámico (E) en una posición que corresponde a la posición 182 de la SEQ ID NO: 1 y un residuo de histidina (H) en una posición que corresponde a la posición 181 de la SEQ ID NO: 1, y un dominio de unión específico de O-glicano situado en el extremo C-terminal del dominio metaloproteasa.
- 25 30 Alternativamente, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente o consistir en un fragmento más corto de la SEQ ID NO: 1 o de una variante de la misma como se describió anteriormente. Los fragmentos pueden describirse como una forma truncada de la SEQ ID NO: 1 que conserva la actividad endoproteasa específica de la O-glicoproteína. Tales fragmentos son más cortos que la SEQ ID NO: 1 y suelen tener al menos 100, 150 o 200 aminoácidos de longitud. Los fragmentos comprenden típicamente un dominio metaloproteasa en posiciones correspondientes a las posiciones 178 a 187 de la SEQ ID NO: 1, incluyendo un residuo de ácido glutámico (E) en una posición que corresponde a la posición 182 de la SEQ ID NO: 1 y un residuo de histidina (H) en una posición que corresponde a la posición 181 de la SEQ ID NO: 1, y un dominio de unión específico de O-glicano situado en el extremo C-terminal del dominio metaloproteasa.
- 35 40 En resumen, por lo tanto, un polipéptido descrito en el presente documento es un polipéptido que tiene actividad endoproteasa específica de O-glicoproteína que comprende:
- (a) una secuencia de aminoácidos de SEQ ID NO: 1;
 - (b) una secuencia de aminoácidos que sea idéntica en al menos un 85% a la secuencia de aminoácidos de SEQ ID NO: 1 o
 - 45 (c) una secuencia de aminoácidos que es un fragmento de la secuencia de SEQ ID NO: 1 o un fragmento de un aminoácido que es 85% idéntico a la secuencia de aminoácidos de SEQ ID NO: 1;
- 50 55 60 65 70 75 80 85 90 95 100
- La secuencia de un polipéptido exemplar se proporciona como SEQ ID NO: 2. El polipéptido puede comprender o consistir en la secuencia de aminoácidos de SEQ ID NO: 2. Una secuencia polinucleotídica exemplar que codifica este polipéptido se muestra en la SEQ ID NO: 3.
- Se han identificado polipéptidos alternativos que tienen actividad endoproteasa específica de O-glicoproteína en *Pseudomonas aeruginosa* PAO1, *Bacteroides thetaiotaomicron* VPI-5482 y *Clostridium perfringens* (véanse tres peptidasas descritas en Noach et al; PNAS 2017, pE679-688 y apéndices de apoyo, específicamente Materiales y Métodos de Clonación, Expresión y Purificación de Proteínas). Las secuencias completas de estos polipéptidos se proporcionan como SEQ ID NOs: 32, 33 y 34. Cada una de estas secuencias incluye un dominio metaloproteasa que tiene el motivo HEbbH descrito anteriormente. La secuencia de *Clostridium perfringens* también tiene el dominio metaloproteasa más largo con el motivo abxHEbbHbc descrito anteriormente. Cada una de estas secuencias puede modificarse opcionalmente para eliminar cualquier secuencia señal o secuencias proenzimáticas que puedan estar presentes y/o para incluir una metionina adicional en el extremo N y/o una histidina u otra etiqueta en el extremo C. Estas secuencias adicionales pueden facilitar la expresión (por ejemplo,

en *E. coli*) y/o la purificación. Las secuencias correspondientes con la señal y otras secuencias inmaduras eliminadas se proporcionan como SEQ ID NOs: 26, 27 y 28. Las versiones de estas secuencias optimizadas para la expresión en *E. coli* y la posterior purificación (mediante la inclusión de una metionina adicional en el extremo N y una etiqueta de histidina en el extremo C) se proporcionan como SEQ ID NOs: 29, 30 y 31. En los métodos descritos en el presente documento para el uso de un polipéptido con actividad endoproteasa específica de O-glicoproteína, el polipéptido de la invención puede sustituirse opcionalmente por uno de estos polipéptidos. Por lo tanto, los polipéptidos preferidos para su uso en tales métodos comprenden, consisten esencialmente o consisten en cualquiera de las SEQ ID NOs: 26 a 31.

10 **Métodos que utilizan la actividad endoproteasa del polipéptido**

La presente invención también proporciona un método de hidrólisis de una O-glicoproteína, en el que el método comprende poner en contacto una muestra de dicha O-glicoproteína con una composición de la invención que tiene actividad endoproteasa específica de O-glicoproteína y opcionalmente comprende además la detección de los productos de hidrólisis.

La presente invención también puede incluir un método para evaluar el estado de glicosilación de una proteína, que comprende poner en contacto una muestra de dicha proteína con una composición de la invención que tiene actividad endoproteasa específica de O-glicoproteína y analizar los productos producidos. La presencia de productos de escisión indica que la proteína de dicha muestra está O-glicosilada, por lo que el método también puede utilizarse para la detección de O-glicoproteínas. Opcionalmente, los productos de escisión pueden analizarse posteriormente para identificar la cadena de glicanos y su posición de unión a la proteína.

25 En tales métodos, una muestra se pone en contacto con una composición de la invención, en condiciones adecuadas para que el polipéptido que contiene interactúe con cualquier proteína de la muestra y para que se produzcan reacciones de hidrólisis/limpieza (actividad endoproteasa). Las condiciones adecuadas incluyen la incubación con un polipéptido de la invención durante al menos 20 minutos, 30 minutos, 40 minutos, 50 minutos, 60 minutos, 70 minutos, 80 minutos, 90 minutos o 120 minutos, 3 horas, 5 horas, 10 horas o toda la noche. La incubación tiene lugar preferentemente a temperatura ambiente, más preferentemente a aproximadamente 20°C, 30 25°C, 30°C, 35°C, 40°C o 45°C, y más preferentemente a aproximadamente 37°C. Los métodos descritos anteriormente pueden llevarse a cabo con cualquier pH adecuado. Los valores de pH adecuados incluyen, por ejemplo, un pH de aproximadamente 3.0, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9 o 9.5. El pH preferido para la actividad de un polipéptido de la invención está en el intervalo de 5.6 a 6.8. El método puede realizarse en cualquier tampón adecuado, como solución salina tamponada con tris (TBS) o solución salina tamponada con fosfato (PBS).

35 La relación aproximada entre el polipéptido de la invención y el contenido proteico de la muestra (enzima:sustrato) puede ser de 1:1, 2:1, 4:1, 6:1, 10:1, 15:1, 20:1, 1:2, 1:4, o 1:6, 1:10, 1:15, 1:20, 1:40, 1:100, 1:200 o 1:400. La proporción preferida es de 1:20. Una mayor proporción de enzima en relación con el sustrato puede ser beneficiosa si se requiere un tiempo de reacción más corto, o si la O-glicoproteína está muy sialilada. El sustrato suele estar presente a una concentración de 0.1mg/ml a 10mg/ml, preferiblemente de 0.1 a 2mg/ml.

40 La detección o el análisis de los productos producidos puede evaluarse mediante cualquier método analítico adecuado, como por ejemplo, aunque no exclusivamente, espectrometría de masas, HPLC, cromatografía de afinidad, electroforesis en gel, SDS-PAGE, ELISA, transferencia de lectina, espectrometría, electroforesis capilar y otras técnicas de laboratorio estándar para el análisis de proteínas.

45 La muestra en cualquiera de los métodos anteriores puede ser una muestra tomada de un paciente, preferiblemente un paciente humano. Los resultados obtenidos pueden utilizarse con fines de diagnóstico, por ejemplo para detectar la presencia de cánceres en los que interviene la glicosilación O-ligada. Dicho uso puede implicar la comparación de los resultados obtenidos a partir de la muestra del paciente, con los obtenidos utilizando una muestra obtenida de un control sano.

50 Como se describe aquí como referencia, el polipéptido puede utilizarse en combinación con otra enzima, como una proteasa o una glicosidasa. La proteasa o glicosidasa adicional normalmente digerirá más las proteínas del sustrato, lo que puede mejorar la actividad del polipéptido de la invención y/o permitir un análisis más fácil o detallado de los productos.

55 Por ejemplo, los presentes inventores han determinado que un polipéptido descrito en el presente documento demuestra una actividad endoproteasa mejorada si los O-glicanos de una proteína sustrato se modifican primero para eliminar el ácido siálico. Así, en un método preferido, la muestra se pone en contacto con un agente para eliminar el ácido siálico. Dicho agente puede ser preferiblemente una enzima sialidasa o una mezcla de dichas enzimas, que pueden estar presentes en un tampón adecuado como TBS o PBS. El tampón comprende preferentemente una baja concentración de NaCl, típicamente hasta 300mM, 250mM, 200mM, o 150mM. La concentración de NaCl es preferentemente de aproximadamente 150 mM, tal como entre 125 mM y 175 mM. Las sialidasas (o neuraminidasas) catalizan la escisión de los ácidos siálicos terminales de los carbohidratos complejos de las glicoproteínas y muestran un alto grado de especificidad. Estas enzimas se dirigen a tres enlaces distintos de ácido siálico que se encuentran habitualmente en las O-glicoproteínas, a saber, los enlaces α2-3, α2-6 y α2-8.

Las sialidasas que son adecuadas para su uso en los métodos descritos incluyen sialidasas de amplio espectro que se dirigen a todos los enlaces α2-3, α2-6 o α2-8, así como sialidasas de espectro estrecho que normalmente se dirigen a un solo tipo de enlace. El enlace α2-3 es el más común en las glicoproteínas humanas, por lo que si se utiliza una sialidasa de espectro estrecho es preferible que se dirija a este enlace. Las sialidasas adecuadas pueden incluir sialidasas víricas o de mamíferos, pero son preferiblemente sialidasas aisladas de bacterias, incluidas, entre otras, cepas de *Clostridium perfringens*, *Arthrobacter ureafaciens*, *Streptococcus pneumoniae*, *Vibrio cholerae* and *Akkermansia muciniphila*.

- 5 Una sialidasa de espectro estrecho preferida es Am1757 aislada de *Akkermansia muciniphila*. Am1757 tiene actividad específica contra los enlaces α2-3. La secuencia de tipo salvaje de Am1757 se proporciona como SEQ ID NO: 9, que incluye una secuencia de señales. La secuencia de tipo salvaje de Am1757 que carece de la secuencia señal se proporciona como SEQ ID NO: 10. Estas secuencias pueden modificarse opcionalmente para incluir una metionina adicional en el extremo N y/o una histidina u otra etiqueta en el extremo C. Estas secuencias adicionales pueden facilitar la expresión (por ejemplo, en *E. coli*) y/o la purificación. Una etiqueta de histidina consiste preferentemente en seis residuos de histidina. La etiqueta de histidina se une preferentemente al extremo C mediante un enlazador, que es típicamente una secuencia corta de aminoácidos, tal como 3 - 5 aminoácidos. El enlazador típicamente consiste predominantemente en residuos de glicina y serina, y puede incluir preferentemente la secuencia GSG. Por ejemplo, GSG y GSGLÉ son enlazadores adecuados. Una secuencia Am1757 exemplar, con una metionina adicional en el extremo N y un enlazador GSGLÉ y una etiqueta His₆ en el extremo C se proporciona como SEQ ID NO: 11. Cualquier referencia a Am1757 en la presente divulgación puede significar cualquiera de las SEQ ID NOs: 9, 10 u 11, pero preferentemente se refiere a un polipéptido que comprende o consiste en la secuencia de aminoácidos de SEQ ID NO: 10. Lo más preferido es un polipéptido que consiste en la secuencia de aminoácidos de la SEQ ID NO: 11.
- 10 Una sialidasa de amplio espectro preferida es Am0707 aislada de *Akkermansia muciniphila*. Am0707 tiene actividad contra los enlaces α2-3, α2-6 y α2-8. La secuencia de tipo salvaje de Am0707 se proporciona como SEQ ID NO: 12, que incluye una secuencia de señales. La secuencia de tipo salvaje de Am0707 que carece de la secuencia señal se proporciona como SEQ ID NO: 13. Estas secuencias pueden modificarse opcionalmente para incluir una metionina adicional en el extremo N y/o una histidina u otra etiqueta en el extremo C. Tales secuencias adicionales pueden ayudar con la expresión y/o purificación. Una etiqueta de histidina consiste preferentemente en seis residuos de histidina. La etiqueta de histidina se une preferentemente al extremo C mediante un enlazador, que es típicamente una secuencia corta de aminoácidos, tal como 3 - 5 aminoácidos. El enlazador típicamente consiste predominantemente en residuos de glicina y serina, y puede incluir preferentemente la secuencia GSG. Por ejemplo, GSG y GSGLÉ son enlazadores adecuados. Una secuencia Am0707 exemplar, con una metionina adicional en el extremo N y un enlazador GSGLÉ y una etiqueta His₆ en el extremo C se proporciona como SEQ ID NO: 14. Cualquier referencia a Am0707 en la presente divulgación puede significar cualquiera de las SEQ ID NOs: 12, 13 ó 14, pero preferentemente se refiere a un polipéptido que comprende o consiste en la secuencia de aminoácidos de SEQ ID NO: 13. Lo más preferido es un polipéptido que consiste en la secuencia de aminoácidos de la SEQ ID NO: 14.
- 15 Una mezcla de sialidasas preferida, capaz de hidrolizar todos los enlaces de ácido siálico, comprende Am1757 y Am0707 aislados de *Akkermansia muciniphila*. La mezcla de Am1757 y Am0707 suele tener una proporción de 1:1. Una mezcla particularmente preferida puede comprender un polipéptido consistente en la secuencia de aminoácidos de SEQ ID NO: 11 y un polipéptido consistente en la secuencia de aminoácidos de SEQ ID NO: 14.
- 20 Los métodos descritos en el presente documento pueden comprender preferiblemente la incubación de una muestra con Am 1757 o con una mezcla de Am1757 y Am0707 antes o simultáneamente con el polipéptido descrito en el presente documento, en condiciones adecuadas para la actividad de las sialidasas. La invención también proporciona una composición (en forma liofilizada o de solución) que comprende un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas que comprende:
- 25 (a) una secuencia de aminoácidos de SEQ ID NO: 1;
- 30 (b) una secuencia de aminoácidos que sea idéntica en al menos un 85% a la secuencia de aminoácidos de SEQ ID NO: 1 o
- 35 (c) una secuencia de aminoácidos que es un fragmento de la secuencia de SEQ ID NO: 1 o un fragmento de una secuencia de aminoácidos que sea idéntica en un 85% a la secuencia de aminoácidos de SEQ ID NO: 1,
- 40 Una mezcla de sialidasas preferida, capaz de hidrolizar todos los enlaces de ácido siálico, comprende Am1757 y Am0707 aislados de *Akkermansia muciniphila*. La mezcla de Am1757 y Am0707 suele tener una proporción de 1:1. Una mezcla particularmente preferida puede comprender un polipéptido consistente en la secuencia de aminoácidos de SEQ ID NO: 11 y un polipéptido consistente en la secuencia de aminoácidos de SEQ ID NO: 14.
- 45 Los métodos descritos en el presente documento pueden comprender preferiblemente la incubación de una muestra con Am 1757 o con una mezcla de Am1757 y Am0707 antes o simultáneamente con el polipéptido descrito en el presente documento, en condiciones adecuadas para la actividad de las sialidasas. La invención también proporciona una composición (en forma liofilizada o de solución) que comprende un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas que comprende:
- 50 (a) una secuencia de aminoácidos de SEQ ID NO: 1;
- 55 (b) una secuencia de aminoácidos que sea idéntica en al menos un 85% a la secuencia de aminoácidos de SEQ ID NO: 1 o
- 60 (c) una secuencia de aminoácidos que es un fragmento de la secuencia de SEQ ID NO: 1 o un fragmento de una secuencia de aminoácidos que sea idéntica en un 85% a la secuencia de aminoácidos de SEQ ID NO: 1,
- 65 en el que la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en el que Am1757 es un polipéptido de cualquiera de los SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de los SEQ ID NOs: 12 a 14.
- Dicha composición puede preferiblemente liofilizarse en solución salina tamponada con tris, que puede estar en torno a un pH 7.6. En dicha composición, Am1757 y Am0707 estarán preferiblemente presentes en una proporción 1:1 entre sí, estando también presente el contenido total de sialidasa (Am1757+Am0707) en una proporción 1:1

- con respecto al polipéptido de la invención. Por ejemplo, si una composición incluye 2000 unidades del polipéptido de la invención, también incluirá 2000 unidades de sialidasa, en la que dichas 2000 unidades de sialidasa comprenden 1000 unidades Am1757 y 1000 unidades Am0707. Una unidad de mezcla de sialidasa es típicamente la cantidad requerida para hidrolizar los ácidos siálicos de ≥ 90 % de 1 µg de glicoproteína (fetuina) cuando se 5 incuba en 20 mM de Tris, pH 6.8 a 37 °C durante 2 h a 37 °C según se monitorea por SDS-PAGE. Una unidad del polipéptido de la invención es típicamente la cantidad requerida para digerir > 90% de 1 µg de Eritropoyetina (EPO) cuando se incuba en tampón Tris 20 mM, pH 6.8, durante la noche con una unidad de mezcla de sialidasa a 37°C como se monitorea por SDS-PAGE.
- 10 También se describe en el presente documento un kit que comprende un polipéptido en un recipiente separado de Am1757 o una mezcla de Am1757 y Am0707, con instrucciones para el uso combinado de las diferentes enzimas.
- Como otro ejemplo, en cualquiera de los métodos descritos en el presente documento, la muestra puede incubarse 15 con una N-glicosidasa antes, al mismo tiempo o después de poner en contacto la muestra con el polipéptido descrito en el presente documento, para eliminar los N-glicanos de las proteínas diana. Una N-glicosidasa ejemplar es la PNGasaF. Otras N-glicosidasas que pueden utilizarse cuando la muestra incluye inmunoglobulinas, son EndoS (véase SEQ ID NO: 1 de WO2008071418) o EndoS2 (puede denominarse EndoS49 - véase SEQ ID NO: 1 de la publicación WO2013037824). Cada una de estas enzimas elimina la glicoproteína ligada al N del Asn-297 20 de la IgG1. Por ejemplo, la muestra puede ponerse en contacto con una N-glicosidasa y una sialidasa (o una mezcla de ambas) además del polipéptido descrito en el presente documento. En dicho método, la sialidasa (o mezcla) puede aplicarse primero, antes de la adición simultánea de la N-glicosidasa y el polipéptido. Este método es especialmente adecuado para la evaluación posterior de los sitios de O-glicosilación, que normalmente se consigue mediante la separación de los productos, por ejemplo, utilizando RPLC y el posterior análisis de las 25 diferentes fracciones, por ejemplo, utilizando espectrometría de masas.
- Como otro ejemplo, en cualquiera de los métodos descritos en el presente documento, la muestra puede incubarse 30 con una proteasa antes, al mismo tiempo o después del contacto de la muestra con el polipéptido, para digerir aún más la proteína diana. Las proteasas generales adecuadas incluyen tripsina, quimotripsina, Lys-C, Asp-N, Glu-C, Arg-C o endoproteasas similares, o Arg-gingipaína (RgpB) de *Porphyromonas gingivalis*.
- Si la muestra incluye inmunoglobulinas, pueden utilizarse inmunoglobulina proteasas como SpeB (véase la secuencia 35 en el documento WO2015040125), enzima degradadora de inmunoglobulina G de *S. pyogenes* (IdeS - véase la secuencia en el documento WO2015040125), enzima degradadora de inmunoglobulina G de *S. equi* subespecie *zooepidemicus* (IdeZ), Lys-gingipaína (Kgp) de *Porphyromonas gingivalis*, y enzima degradadora de inmunoglobulina G de *S. agalactiae* (IgdE_{agalactiae} - véase SEQ ID NO: 3 del PCT/EP2017/052463). El uso de cualquier combinación de estas proteasas en un método descrito en el presente documento puede ayudar a 40 determinar los sitios de O-glicosilación en anticuerpos monoclonales y subunidades de los mismos, por ejemplo mediante espectrometría de masas (enfoque middle down).
- Como otro ejemplo, en cualquiera de los métodos descritos en el presente documento, la muestra puede incubarse 45 con una O-glicosidasa después de poner en contacto la muestra con un polipéptido descrito en el presente documento. Por ejemplo, para simplificar el análisis de los productos obtenidos, éstos se someten a digestión por una O-glicosidasa para eliminar los O-glicanos antes de su posterior análisis por cualquier método adecuado. Las O-glicosidasas adecuadas pueden obtenerse a partir de una cepa de *Enterococcus faecalis*, *Streptococcus oralis*, o *Bifidobacterium bifidum*, preferiblemente *Enterococcus faecalis* o *Streptococcus oralis*, más preferiblemente *Streptococcus oralis*. La secuencia de una O-glicosidasa ejemplar de *Streptococcus oralis* se proporciona como 50 SEQ ID NO: 15.
- 50 Polipéptidos que se unen a glicoproteínas O-ligadas pero que carecen o tienen una actividad endoproteasa reducida.**
- 55 Características funcionales de un polipéptido carente de actividad endoproteasa**
- En una realización, la presente invención se refiere a una composición que comprende un polipéptido que 60 comprende una secuencia de aminoácidos de la SEQ ID NO: 5 o SEQ ID NO: 20, que carecen de actividad endoproteasa o la tienen reducida, pero conservan la capacidad de unirse a O-glicanos, y una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en la que Am1757 es un polipéptido de cualquiera de las SEQ ID NOS: 9 a 11 y Am0707 es un polipéptido de cualquiera de las SEQ ID NOS: 12 a 14. En otras palabras, el polipéptido puede describirse como un agente de unión O-glicano-específico que no hidroliza significativamente una glicoproteína a la que está unido dicho glicano.
- La actividad endoproteasa de la O-glicoproteína puede determinarse utilizando cualquier método adecuado, pero 65 típicamente puede emplear el mismo ensayo descrito anteriormente para los polipéptidos de la invención que poseen dicha actividad. La falta de actividad de un polipéptido de prueba se indicará por la ausencia de productos de escisión tras la incubación con un sustrato de O-glicoproteína. Escisión del mismo sustrato por un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO: 1 puede utilizarse como control positivo. La reducción

5 de la actividad del polipéptido de prueba puede determinarse por comparación con el mismo control. El polipéptido tiene típicamente una actividad de endoproteasa de O-glicoproteína que se reduce en relación con la actividad de un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO: 1. El polipéptido suele tener una actividad de endoproteasa de O-glicoproteína inferior al 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 35%, 30%, 25%, 20%, 15%, 10% o 5% en comparación con la actividad de un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO: 1.

10 La capacidad de un polipéptido para unirse a un O-glicano o a una O-glicoproteína también puede evaluarse mediante cualquier método adecuado. Uno de estos métodos consiste en inmovilizar un polipéptido de prueba, por ejemplo en sefarosa en una columna de centrifugación, seguido de la incubación con una muestra que contenga O-glicoproteínas y/o O-glicanos. Si el polipéptido de prueba tiene capacidad de unión a O-glicanos y/o O-glicoproteínas, las O-glicoproteínas y/o O-glicanos serán detectables unidos a la columna o en un eluyente posterior. Preferiblemente, el polipéptido es capaz de unirse a todas las O-glicoproteínas que son hidrolizables por un polipéptido que comprende la secuencia de aminoácidos de SEQ ID NO: 1.

15 En los ejemplos se describen ensayos ejemplares de este tipo.

Características estructurales de un polipéptido carente de actividad endoproteasa

20 Esta sección establece las características estructurales de un polipéptido descrito en el presente documento, que se aplican además de las características funcionales descritas en la sección anterior. Un polipéptido puede poseer las mismas características estructurales que las descritas anteriormente en relación con un polipéptido que tenga actividad endoproteasa, con la excepción de que la secuencia de aminoácidos se modifique mediante una o más adiciones, supresiones o sustituciones de aminoácidos de manera que se reduzca o elimine dicha actividad.

25 Típicamente, un polipéptido descrito en el presente documento no incluirá un motivo metaloproteasa intacto de HEbbH o abxHEbbHbc. Dicho motivo puede estar alterado por adición, delección o sustitución, pero preferiblemente está alterado por al menos una sustitución de aminoácidos. Preferiblemente, la sustitución implica la sustitución del residuo de ácido glutámico (E) en dicho motivo por un aminoácido alternativo y/o la sustitución del residuo de histidina (H) en la posición correspondiente a la 1^a posición del motivo más corto (la 4^a posición del motivo más largo) y/o la sustitución del residuo de histidina (H) en la posición correspondiente a la 5^a posición del motivo más corto (la 8^a posición del motivo más largo). Preferiblemente, alguna de dichas sustituciones, ambas o las tres, es no conservativa. La sustitución del residuo E debería reducir o eliminar la transferencia de electrones. La sustitución de cualquiera de los residuos H debería reducir o eliminar la unión del cofactor ion de Zinc. Por lo tanto, el residuo E se sustituye preferiblemente con un aminoácido no polar o no cargado, como A, C, F, G, I, L, M, N, P, Q, S, T, V o W, pero más preferiblemente se sustituye con Alanina (A) o Glicina (G). Cada uno de los residuos H puede sustituirse individualmente por cualquier aminoácido no H, pero se prefieren de nuevo los aminoácidos no polares como A y G.

30 Así, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente o consistir de la secuencia de aminoácidos de SEQ ID NO: 1 en el que se altera el motivo metaloproteasa de HEbbH o abxHEbbHbc, preferiblemente mediante la sustitución del residuo de ácido glutámico en la posición correspondiente a la posición 182 de la SEQ ID NO: 1 y/o la sustitución del residuo de histidina correspondiente a la posición 181 de la SEQ ID NO: 1 con un aminoácido alternativo y/o la sustitución del residuo de histidina correspondiente a la posición 185 de la SEQ ID NO: 1 con un aminoácido alternativo. En otras palabras, el polipéptido puede describirse como no comprendiendo el motivo de metaloproteasa HEbbH y preferiblemente comprendiendo una versión alterada de dicho motivo, tal que:

- 40 (a) el H de la primera posición se sustituye por un aminoácido alternativo, preferiblemente A o G, y/o
- 45 (b) E en la segunda posición se sustituye por un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W, preferiblemente A o G; y/o
- 50 (c) el H de la quinta posición se sustituye por un aminoácido alternativo, preferiblemente A o G
- 55 donde b en dicho motivo es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W.

Por lo tanto, dicho polipéptido puede describirse como que comprende el motivo xbbb, en el que:

- 60 (a) x es preferiblemente cualquier aminoácido excepto H, y es preferiblemente A o G; y/o
 - 65 (b) b es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W, preferiblemente A o G;
- 65 opcionalmente, en el que dicho motivo está presente en dicho polipéptido en las posiciones correspondientes a las posiciones 181 a 185 de la SEQ ID NO: 1.

Por consiguiente, dicho polipéptido puede comprender un motivo de metaloproteasa alterado, por ejemplo con una cualquiera de las secuencias siguientes: HALGH (SEQ ID NO: 44), AELGH (SEQ ID NO: 45) o más preferiblemente AALGH (SEQ ID NO: 46). Secuencias que comprenden este tipo de cambio específico de la SEQ ID NO: 1 se muestran en la SEQ ID NO: 5 y la SEQ ID NO: 20. En otras palabras, por lo tanto, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente, o consistir en la secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20.

Dicho polipéptido puede describirse alternativamente como que comprende el motivo abxxbbbxbc, en el que:

- (a) a es el aminoácido V, T o G;
 - (b) b es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W, preferiblemente A o G;
 - (c) x es cualquier aminoácido excepto que el aminoácido en la 4^a y/o 8^a posición del motivo preferiblemente no es H, y es preferiblemente A o G; y
 - (d) c es un aminoácido hidrófobo, opcionalmente A, C, F, I, L, M, P, V, W o Y;

opcionalmente, donde dicho motivo está presente en dicho polipéptido en las posiciones correspondientes a las posiciones 178 a 187 de la SEQ ID NO: 1.

Por consiguiente, dicho polipéptido puede comprender un motivo de metaloproteasa alterado, por ejemplo con una cualquiera de las secuencias siguientes: GMAHALGHGL (SEQ ID NO: 23), GMAAELGHGL (SEQ ID NO: 24) o más preferiblemente GMAAALGHGL (SEQ ID NO: 25). Secuencias que comprenden este tipo de cambio específico de la SEQ ID NO: 1 se muestran en la SEQ ID NO: 5 y la SEQ ID NO: 20. En otras palabras, por lo tanto, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente, o consistir en la secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20.

Alternativamente, el polipéptido aquí descrito puede comprender, consistir esencialmente o consistir en una variante de la secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20 que es al menos un 50% idéntica a la secuencia de aminoácidos de la SEQ ID NO: 5, siempre que no se introduzca un residuo de ácido glutámico en la posición correspondiente a la posición 182 de la SEQ ID NO: 1 y/o no se introduce un residuo de histidina en la posición correspondiente a la posición 181 de la SEQ ID NO: 1 y/o no se introduce un residuo de histidina en la posición correspondiente a la posición 185 de la SEQ ID NO: 1.

La secuencia variante puede ser al menos un 60%, al menos un 70%, al menos un 80%, al menos un 85%, al menos un 90%, al menos un 95%, al menos un 98% o al menos un 99% idéntica a la secuencia de SEQ ID NO: 5.

menos un 90%, al menos un 95%, al menos un 98% o al menos un 99% identica a la secuencia de SEQ ID NO: 5. El nivel de identidad es preferentemente de al menos el 85% o superior. Identidad relativa a la secuencia de SEQ ID NO: 5 o SEQ ID NO: 20 puede medirse sobre una region de al menos 100, al menos 200, al menos 300 o al menos 350 o más aminoácidos contiguos de la secuencia mostrada en la SEQ ID NO: 5 o SEQ ID NO: 20, o más preferiblemente en toda la longitud de la SEQ ID NO: 5 o SEQ ID NO: 20. Una secuencia de un polipéptido descrito en el presente documento puede comprender una variante de la secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20 en la que se realizan hasta 10, 20, 30, 40, 50 o 60 sustituciones conservadoras. Las determinaciones de la identidad de secuencia y una explicación de las sustituciones conservativas y no conservativas se proporcionan en la sección relativa a los polipéptidos con actividad endoproteasa y se aplican igualmente en este caso.

Alternativamente, un polipéptido descrito en el presente documento puede comprender, consistir esencialmente o consistir en un fragmento más corto de la SEQ ID NO: 5 o SEQ ID NO: 20, o de una variante del mismo como se ha descrito anteriormente. Los fragmentos pueden describirse como una forma truncada de la SEQ ID NO: 5 o SEQ ID NO: 20 que conserva la actividad de unión a la O-glicoproteína. Tales fragmentos son más cortos que la SEQ ID NO: 1 y suelen tener al menos 100, 150 o 200 aminoácidos de longitud.

Cualquier polipéptido que comprenda la SEQ ID NO: 5 o la SEQ ID NO :20, o una variante de la misma, o un fragmento de cualquiera de ellos, puede incluir opcionalmente una metionina adicional en el extremo N y/o una histidina u otra etiqueta en el extremo C. Tales secuencias adicionales pueden ayudar con la expresión y/o purificación. Una etiqueta de histidina consiste preferentemente en seis residuos de histidina. La etiqueta de histidina se une preferentemente al extremo C mediante un enlazador, que es típicamente una secuencia corta de aminoácidos, tal como 3 - 5 aminoácidos. El enlazador típicamente consiste predominantemente en residuos de glicina y serina, y puede incluir preferentemente la secuencia GSG. Por ejemplo, GSG y GSGLÉ son enlazadores adecuados.

65 En resumen, por lo tanto, una composición de la invención comprende un polipéptido que es capaz de unirse a un O-glicano o a una O-clícoproteína y que carece o tiene una actividad endoproteasa reducida específica para

5 proteínas O-glicosiladas que comprenden una secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20, en donde la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en donde Am1757 es un polipéptido de cualquiera de las SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de las SEQ ID NOs: 12 a 14 opcionalmente, en el que dicho polipéptido incluye una metionina adicional en el extremo N y/o una etiqueta de histidina en el extremo C, etiqueta que puede unirse al extremo C mediante un enlazador.

10 La secuencia de un polipéptido ejemplar descrito en el presente documento se proporciona como SEQ ID NO: 6. El polipéptido puede comprender o consistir en la secuencia de aminoácidos de SEQ ID NO: 6. Una secuencia polinucleotídica ejemplar que codifica este polipéptido se muestra en la SEQ ID NO: 7. La secuencia de otro polipéptido ejemplar se proporciona como SEQ ID NO: 21. El polipéptido puede comprender o consistir en la secuencia de aminoácidos de SEQ ID NO: 21. Una secuencia polinucleotídica ejemplar que codifica este polipéptido se muestra en la SEQ ID NO: 22.

15 Dicho polipéptido se suministra preferiblemente en forma inmovilizada, por ejemplo sobre agarosa o sefarosa, opcionalmente en forma de resina.

20 Pueden producirse polipéptidos adicionales que tengan actividad de unión a O-glicoproteína pero que carezcan de actividad endoproteasa específica de O-glicoproteína o que la tengan reducida, alterando el motivo HEbbH o abxHEbbHbc del dominio metaloproteasa en cualquier otro polipéptido que tenga actividad endoproteasa específica de O-glicoproteína que comprenda dicho motivo. En los métodos descritos en el presente documento para el uso de un polipéptido que carece de actividad endoproteasa o que la tiene reducida, las referencias al polipéptido incluyen dichos polipéptidos. La alteración de dicho motivo se consigue preferiblemente como se ha descrito anteriormente de forma que:

- 25 (a) el H de la primera posición se sustituye por un aminoácido alternativo, preferiblemente A o G, y/o
 (b) E en la segunda posición se sustituye por un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W, preferiblemente A o G; y/o
 30 (c) el H de la quinta posición se sustituye por un aminoácido alternativo, preferiblemente A o G

donde b en dicho motivo es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W.

35 Por lo tanto, dicho polipéptido puede describirse como que comprende el motivo xbbb, en el que:

- 35 (a) x es preferiblemente cualquier aminoácido excepto H, y es preferiblemente A o G; y/o
 (b) b es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W, preferiblemente A o G;

40 Otros polipéptidos que tienen actividad endoproteasa específica de O-glicoproteína y que pueden alterarse de esta manera se han descrito anteriormente como identificados en *Pseudomonas aeruginosa* PAO1, *Bacteroides thetaiotaomicron* VPI-5482 y *Clostridium perfringens* (véanse tres peptidasas descritas en Noach et al; PNAS 2017, pE679-688 y apéndices de apoyo, específicamente Materiales y Métodos de Clonación, Expresión y Purificación de Proteínas). Las secuencias completas de estos polipéptidos se proporcionan como SEQ ID NOs: 32, 33 y 34. Las secuencias maduras correspondientes (por ejemplo, con la señal y otras secuencias eliminadas) se proporcionan como SEQ ID NOs: 26, 27 y 28. Las versiones de estas secuencias optimizadas para la expresión en *E. coli* y la posterior purificación (mediante la inclusión de una metionina adicional en el extremo N y una etiqueta de histidina en el extremo C) se proporcionan como SEQ ID NO: 29, 30 y 31. Cada una de las SEQ ID NOs: 26 a 34 incluye, por tanto, un dominio metaloproteasa que tiene el motivo HEbbH que puede alterarse para producir un motivo xbbb como se ha descrito anteriormente. Versiones de la SEQ ID NOs: 26, 27 y 28 en los que el motivo HEbbH ha sido alterado, se proporcionan como SEQ ID NOs: 35, 36 y 37. Las versiones de estas secuencias optimizadas para la expresión en *E. coli* y la posterior purificación (mediante la inclusión de una metionina adicional en el extremo N y una etiqueta de histidina en el extremo C) se proporcionan como SEQ ID NO: 38, 39 y 40. Los polipéptidos que carecen o tienen reducida la actividad endoproteasa específica de la O-glicoproteína pueden comprender, consistir esencialmente o consistir en cualquiera de las SEQ ID NOs: 35, 36, 37, 38, 39 o 40.

Métodos que utilizan mutantes LS que carecen o tienen una actividad endoproteasa reducida

60 La presente invención también proporciona un método de unión a un O-glicano, en el que el método comprende poner en contacto una muestra que comprende el O-glicano con una composición de la invención que comprende un polipéptido capaz de unirse a un O-glicano y que carece o tiene actividad endoproteasa reducida específica para proteínas O-glicosiladas, en el que el polipéptido comprende una secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20, y en donde la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en donde Am1757 es un polipéptido de cualquiera de las SEQ ID NOs: 9 a 11 y Am0707 es un

polipéptido de cualquiera de las SEQ ID NOs: 12 a 14. El método opcionalmente incluye además determinar si se ha unido o no un O-glicano y/o separar el O-glicano y cualquier glicoproteína unida de la mezcla resultante.

5 La presente invención también puede incluir un método para evaluar el estado de glicosilación de una proteína, que comprende poner en contacto una muestra de dicha proteína con un polipéptido de la invención, que comprende un polipéptido capaz de unirse a un O-glicano y que carece o tiene una actividad endoproteasa reducida específica para proteínas O-glicosiladas, y determinar si la proteína está unida o no por dicho polipéptido; en el que el polipéptido comprende una secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20, y en donde la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en donde Am1757 es un polipéptido de cualquiera de las SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de las SEQ ID NOs: 12 a 14.

10 La presente invención también puede incluir un método para detectar glicoproteínas O-ligadas en una muestra, en el que el método comprende poner en contacto dicha muestra con una composición de la invención, que comprende un polipéptido capaz de unirse a un O-glicano y que carece o tiene una actividad endoproteasa reducida específica para proteínas O-glicosiladas, para permitir así la formación de un complejo glicoproteína-polipéptido O-ligado; en el que el polipéptido comprende una secuencia de aminoácidos de SEQ ID NO: 5 o SEQ ID NO: 20, y en donde la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en donde Am1757 es un polipéptido de cualquiera de las SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de las SEQ ID NOs: 12 a 14. El método puede incluir opcionalmente separar dicho polipéptido de la muestra contactada y determinar si el polipéptido separado está unido a las glicoproteínas O-ligadas, con lo que se puede determinar la presencia o ausencia de glicoproteínas O-ligadas en la muestra. El método también puede utilizarse para aislar un O-glicano o una glicoproteína O-ligada a partir de una muestra que contenga O-glicanos o glicoproteínas O-ligadas.

15 25 En tales métodos, una muestra se pone en contacto con una composición de la invención en condiciones adecuadas para que el polipéptido interactúe con cualquier O-glicano o proteínas de la muestra y para que se produzca la unión. Las condiciones adecuadas incluyen la incubación con un polipéptido durante al menos 20 minutos, 30 minutos, 40 minutos, 50 minutos, 60 minutos, 70 minutos, 80 minutos, 90 minutos o 120 minutos, 3 horas, 5 horas, 10 horas o toda la noche, normalmente con mezcla, por ejemplo, mezcla de extremo a extremo. La incubación tiene lugar preferentemente a temperatura ambiente, más preferentemente a aproximadamente 20°C, 25°C, 30°C, 35°C, 40°C o 45°C, y más preferentemente a aproximadamente 37°C. Los métodos descritos anteriormente pueden llevarse a cabo con cualquier pH adecuado. Los valores de pH adecuados incluyen, por ejemplo, un pH de aproximadamente 3.0, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9 o 9.5. El pH preferido para la actividad de un polipéptido de la invención está en el intervalo de 5.6 a 6.8. El método puede realizarse en cualquier tampón adecuado, como solución salina tamponada con tris (TBS) o solución salina tamponada con fosfato (PBS). La relación aproximada entre el polipéptido y el contenido proteínico de la muestra puede ser de 1:1, 2:1, 4:1, 6:1, 10:1, 15:1, 20:1, 1:2, 1:4, o 1:6, 1:10, 1:15, 1:20, 1:40, 1:100, 1: 200 o 1:400 (peso:peso). Una proporción preferible es 1:1 (peso:peso). Una mayor proporción de polipéptido respecto al sustrato puede ser beneficiosa si se requiere un tiempo de reacción más corto, o si la O-glicoproteína está muy sialilada. Alternativamente, puede utilizarse una etapa de incubación de sialidasa anterior o simultánea para reducir el contenido de ácido siálico, como se explica con más detalle a continuación. El sustrato suele estar presente a una concentración de alrededor de 0.01mg/ml a 10mg/ml, preferiblemente de alrededor de 0.1mg/ml a 10mg/ml, de alrededor de 0.01mg/ml a 2mg/ml, o de alrededor de 0.1mg/ml a 2mg/ml.

40 45 50 La detección o el análisis de la muestra para determinar si se ha unido un O-glicano o una glicoproteína O-ligada puede evaluarse mediante cualquier método analítico adecuado, como por ejemplo, entre otros, espectrometría de masas, HPLC, cromatografía de afinidad, electroforesis en gel, SDS-PAGE, ELISA, transferencia de lectina, espectrometría, electroforesis capilar y otras técnicas de laboratorio estándar para el análisis de proteínas. Por ejemplo, puede analizarse el peso molecular del polipéptido. El polipéptido aquí descrito unido a un O-glicano o a una glicoproteína O-ligada tendrá un peso molecular superior al de un polipéptido no unido a un O-glicano o a una glicoproteína O-ligada.

55 60 65 La separación del O-glicano unido o de la glicoproteína O-ligada y del polipéptido descrito en el presente documento puede llevarse a cabo por cualquier medio de separación adecuado. Por ejemplo, los medios de separación pueden comprender una población de nanopartículas magnéticas. Éstos pueden separarse de una muestra mediante separación por campo magnético, preferiblemente separación por campo magnético de alto gradiente. Ejemplos de reactivos o medios de separación son poblaciones de partículas magnéticas capaces de unirse al polipéptido de la invención. Por ejemplo, cuando el polipéptido está derivatizado con una etiqueta de histidina, las partículas magnéticas contienen en su superficie grupos quelantes que transportan un ion de níquel, cobre ^o zinc. Alternativamente, cuando el polipéptido se derivatiza con una etiqueta de biotina, las partículas magnéticas contienen en su superficie estreptavidina.

65 Los medios de separación también pueden comprender un soporte sólido al que se inmoviliza el polipéptido. Los ejemplos de soportes sólidos incluyen los descritos en secciones anteriores, y pueden incluir resinas de agarosa

o sefarosa, perlas de agarosa reticuladas o similares. El soporte puede utilizarse como matriz en una columna de cromatografía de afinidad. Alternativamente, el soporte sólido puede comprender un material adecuado a base de sílice o poliestireno, o un recipiente de plástico como una placa de microtitulación o equivalente, en el que pueda adsorberse directamente el polipéptido de la invención.

5 Los medios de separación alternativos incluyen reactivos que comprenden anticuerpos específicos para el polipéptido descrito en el presente documento, que pueden generarse por métodos habituales en la técnica. Los anticuerpos en este sentido incluyen un anticuerpo monoclonal, un anticuerpo policlonal, un anticuerpo de cadena simple, un anticuerpo quimérico, un anticuerpo injertado con CDR o un anticuerpo humanizado. El anticuerpo puede ser una molécula de inmunoglobulina intacta o un fragmento de la misma, como un fragmento Fab, F(ab')2 o Fv. Si hay más de un anticuerpo presente, los anticuerpos tienen preferiblemente determinantes diferentes que no se solapan, de modo que puedan unirse simultáneamente al polipéptido descrito en el presente documento. El anticuerpo puede estar unido a un soporte sólido o puede estar marcado o conjugado con otro grupo químico o molécula para ayudar a su separación o aislamiento. Por ejemplo, los grupos químicos típicos incluyen etiquetas fluorescentes como la fluoresceína (FITC) o la ficoeritrina (PE), o etiquetas como la biotina.

10 Otros medios adecuados de separación incluyen la elución de la proteína del polipéptido (típicamente inmovilizado) poniendo en contacto el polipéptido de la muestra contactada con un tampón de elución adecuado. La elección del tampón de elución puede depender de la sensibilidad al ácido de la proteína. Los tampones de elución preferidos pueden comprender altas concentraciones molares de urea (típicamente al menos 5, 6, 7 o más preferiblemente al menos 8M) o altas concentraciones de un detergente (típicamente al menos alrededor del 1%, 5% o 10%). Los detergentes adecuados incluyen Nonidet P40, Triton X-100, Tween 20, CHAPS, desoxicolato sódico y surfactante RapiGest SF, pero se prefiere el dodecyl sulfato sódico (SDS). Se prefiere la urea de alto molar al detergente, ya que es más probable que los procedimientos posteriores sean más sensibles a la presencia de detergente.

15 20 25 Otro tampón de elución preferido comprende una concentración adecuada de un polipéptido descrito en el presente documento que tenga actividad endoproteasa de O-glicoproteína, por ejemplo un polipéptido de la SEQ ID NO: 1. La ruptura del O-glicano por este polipéptido liberará las O-glicoproteínas unidas, evitando así la necesidad de elución con urea o detergentes.

30 35 40 Los métodos preferidos de elución de O-glicoproteínas a partir de polipéptidos inmovilizados se demuestran en los Ejemplos.

45 La muestra en cualquiera de los métodos anteriores puede ser una muestra tomada de un paciente, preferiblemente un paciente humano. Los resultados obtenidos pueden utilizarse con fines de diagnóstico, por ejemplo para detectar la presencia de cánceres en los que interviene la glicosilación O-ligada. Dicho uso puede implicar la comparación de los resultados obtenidos a partir de la muestra del paciente, con los obtenidos utilizando una muestra obtenida de un control sano.

50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740 9745 9750 9755 9760 9765 9770 9775 9780 9785 9790 9795 9800 9805 9810 9815 9820 9825 9830 9835 9840 9845 9850 9855 9860 9865 9870 9875 9880 9885 9890 9895 9900 9905 9910 9915 9920 9925 9930 9935 9940 9945 9950 9955 9960 9965 9970 9975 9980 9985 9990 9995 10000 10005 10010 10015 10020 10025 10030 10035 10040 10045 10050 10055 10060 10065 10070 10075 10080 10085 10090 10095 10100 10105 10110 10115 10120 10125 10130 10135 10140 10145 10150 10155 10160 10165 10170 10175 10180

más la proteína diana. Las proteasas generales adecuadas incluyen tripsina, quimotripsina, Lys-C, Asp-N, Glu-C, Arg-C o endoproteasas similares, o Arg-gingipaína (RgpB) de *Porphyromonas gingivalis*.

Si la muestra incluye inmunoglobulinas, pueden utilizarse inmunoglobulina proteasas como SpeB (véase la secuencia en el documento WO2015040125), enzima degradadora de inmunoglobulina G de *S. pyogenes* (IdeS - véase la secuencia en el documento WO2015040125), enzima degradadora de inmunoglobulina G de *S. equi* subespecie *zooepidemicus* (IdeZ), Lys-gingipaína (Kgp) de *Porphyromonas gingivalis*, y enzima degradadora de inmunoglobulina G de *S. agalactiae* (IgdE_{agalactiae} - véase SEQ ID NO: 3 del PCT/EP2017/052463). El uso de cualquier combinación de estas proteasas en un método descrito en el presente documento puede ayudar con el análisis de la proteína sustrato o glicano, por ejemplo mediante espectrometría de masas.

Como otro ejemplo, en cualquiera de los métodos aquí descritos, la glicoproteína O-ligada aislada puede incubarse con una O-glicosidasa para eliminar los O-glicanos antes de su posterior análisis por cualquier método adecuado. Las O-glicosidasas adecuadas pueden obtenerse a partir de una cepa de *Enterococcus faecalis*, *Streptococcus oralis*, o *Bifidobacterium bifidum*, preferiblemente *Enterococcus faecalis* o *Streptococcus oralis*, más preferiblemente *Streptococcus oralis*. La secuencia de una O-glicosidasa ejemplar de *Streptococcus oralis* se proporciona como SEQ ID NO: 15.

Los siguientes ejemplos ilustran la invención:

Ejemplo 1

MATERIALES Y MÉTODOS

Mutagénesis del LS

La mutagénesis dirigida al sitio utilizando Q5 (NEB) se realizó siguiendo las instrucciones del fabricante (temperatura de recocido 68°C, 3 min de elongación) utilizando los cebadores E206A_fwd 5'-ATGGCGCACGC GCTGGGCCACG-3' y 5'-GCCACCGTAC CATTTCGTC-3' (rev); cambiando así un ácido glutámico por una alanina en un gen Amuc1119 de *Akkermansia muciniphila* para crear el mutante, Amuc1119_{E206A} (LS_{E206A}). La construcción se transformó en DH5α *E. coli*, se aisló y se verificó mediante secuenciación (GATC Biotech).

Expresión recombinante del LS y LS_{E206A}

El gen Amuc1119 de *Akkermansia muciniphila* ATCC BAA-835, y un mutante, Amuc1119_{E206A} (Amuc1119 - LS; Amuc1119_{E206A} - LS_{E206A}), se optimizaron en codones para su expresión en *E. coli* (DNA 2.0) y se clonaron en un vector de expresión con una etiqueta 6xHis C-terminal como parte de la proteína de fusión.

Los genes con codones optimizados se transformaron en células BL21(DE3) Star. *E. coli* se cultivó rutinariamente en LB a 37°C, 180 rpm. En presencia del plásmido, se añadieron 50 µg/ml de kanamicina. Tras una incubación de una noche, los cultivos se diluyeron 1:20 en LB(kana) fresco, y se cultivaron hasta que alcanzaron una DO₆₂₀ ~ 0.7-0.8, tras lo cual se indujo la expresión de la proteína recombinante mediante la adición de 1 mM de IPTG, y la expresión continuó durante 6 horas antes de que las células se recogieran y congelaran. Las células congeladas se descongelaron y se resolvieron en tampón de unión a His (NaP 20 mM pH 7.4, NaCl 500 mM, imidazol 20 mM) y se sometieron a ultrasonido para la liberación de proteínas intracelulares. Los residuos celulares se eliminaron por centrifugación. El sobrenadante filtrado estéril se purificó por afinidad en una columna de níquel y se volvió a tamponar a Tris-HCl 20 mM pH 8.0 en una columna PD-25. La concentración de las proteínas se determinó utilizando el Nanodrop, y la pureza se estimó a través de SDS-PAGE.

Evaluación de la actividad utilizando sustrato proteico

El TNFαR se mezcló con el LS en una proporción de 2:1 y se incubó durante 15-60 minutos a 37°C, tras lo cual las proteínas se separaron en un SDS-PAGE de gradiente Novex al 4-20%. Se investigó el impacto del NaCl (0-1 M), los cationes divalentes, el EDTA y el pH en la actividad del LS, y se midieron las diferencias en los fragmentos hidrolíticos generados mediante análisis densitométrico utilizando Gel Doc EZ (BioRad).

Dependencia del tiempo y la dosis para la actividad

El TNFαR (0.5 µg) se incubó con dosis variables del LS durante 15 o 60 minutos a 37°C en PBS, tras lo cual las proteínas se separaron en un SDS-PAGE de gradiente Novex al 4-20%. La intensidad de los fragmentos generados (densitometría) se utilizó para determinar la dosis y el tiempo óptimos para unas condiciones de incubación eficaces.

Especificidad del sustrato

El LS se incubó con una variedad de sustratos N- y O-ligados durante la noche a 37°C en una proporción de 2:1 (sustrato:enzima). El LS se incubó con EPO (0.3mg/ml) en una proporción de 50:1 (sustrato:enzima). Las proteínas se separaron y analizaron en geles SDS-PAGE de gradiente Novex al 4-20%.

5 Inmovilización del LS_{E206A}

El LS_{E206A} se resuspendió en un tampón de acoplamiento (0.2 M NaHCO₃, 0.5 M NaCl pH 8.3), y se concentró a 20 mg/ml. La sefarosa 4 Fast Flow activada con NHS (GE Healthcare) se preparó para el acoplamiento siguiendo las instrucciones del fabricante (por ejemplo, lavado con HCl y equilibrado en tampón de acoplamiento). El LS_{E206A} se inmovilizó incubándolo durante una noche con la sefarosa a 4°C, agitándola lentamente para conseguir una mezcla constante. La sefarosa se bloqueó añadiendo 0.1 M Tris pH 8.5, se lavó con 3 repeticiones de 0.1 M Tris pH 8.5 / 0.1 M NaAc, 0.5 M NaCl pH 5.0, y se almacenó en EtOH hasta su uso.

15 Afinidad de unión del LS_{E206A}

Las columnas de centrifugación con 50 µl del LS_{E206A} inmovilizada (por ejemplo, 50 µg de proteína) equilibradas en PBS se incubaron con 10 µg de glicoproteína pretratada con una mezcla de sialidasas (Am0707:Am1757) o con una combinación de sialidasas y una Endo-α-N-acetil-galactosaminidasa de *Streptococcus oralis* (por ejemplo, una O-glicosidasa). Las muestras se dejaron incubar durante 2 h a 37°C, tras lo cual las columnas se lavaron con PBS (10 volúmenes; 100 g, 30 s) y se eluyeron con glicina 0,1 M, pH 3.0. Las fracciones se analizaron en SDS-PAGE.

20 Análisis por espectrometría de masas

El etanercept (Enbrel®) es una proteína de fusión Fc clínicamente aprobada que se une al TNF α . El etanercept contiene varios O-glicanos. Para comprobar las especificidades de escisión enzimática, la endoproteasa se incubó junto con etanercept durante toda la noche a 37°C. Para simplificar el análisis por espectrometría de masas, se realizó una segunda ronda de tratamiento enzimático para eliminar los O-glicanos restantes utilizando sialidasa y O-glicosidasa (durante la noche, en PBS, proporción 1:40 de todas las enzimas individuales). Los péptidos generados se analizaron por MS/MS tras separarlos mediante cromatografía líquida de fase inversa C18.

RESULTADOS

25 El LS es una metaloproteasa putativa

Basándose en la similitud de secuencias y dominios, el LS comparte homología con varias metaloproteasas, conteniendo la secuencia putativa del sitio activo GMAHELGHGL, que comparte similitud con la secuencia general de metaloproteasas abxHEbbHbc (a = V/T, b = sin carga, c = hidrofóbica). En general, las histidinas intervienen en la unión del sustrato y la afinidad por el Zn²⁺, mientras que el ácido glutámico, junto con las histidinas, media en la transferencia de electrones y, por tanto, en el efecto hidrolítico. Para poder caracterizar aún más la enzima, construimos un mutante LS_{E206A}, capaz de unir los sustratos, pero que carece de capacidad hidrolítica o la tiene reducida mediante la alteración de la E por una A. Pueden ser necesarias más modificaciones (por ejemplo, alterar la H por una A) para la inactividad total. Ambas construcciones se expresaron bien y se purificaron fácilmente mediante cromatografía de afinidad basada en sus etiquetas His (Fig. 1).

45 El LS hidroliza específicamente las glicoproteínas con O-glicanos

Para investigar la especificidad del sustrato del LS, se incubó la proteasa con una diversidad de proteínas. Como se muestra en la Fig. 2, el LS se incubó con IgA y Herceptin (trastuzumab). El LS sólo pudo actuar sobre proteínas con glicanos O-ligados, como la IgA. Mientras que la presencia de ácidos siálicos terminales parece inhibir parcialmente la actividad del LS, la ausencia de ácidos siálicos no es un requisito previo para la hidrólisis (Fig. 4).

55 El LS puede actuar sobre glicoproteínas O-ligadas en diversas condiciones

Se realizaron análisis densitométricos de geles SDS-PAGE para evaluar las propiedades enzimáticas del LS. El LS es activo en la mayoría de las condiciones, con preferencia por un pH ligeramente ácido y una baja concentración de NaCl (Fig. 3A-B). Mientras que los iones Mg²⁺ y Ca²⁺ afectaban positivamente a la actividad hidrolítica del LS, la presencia de Zn²⁺ disminuía significativamente la actividad, y el EDTA la abolía por completo (Fig. 3C-D).

60 Los residuos de galactosidasa O-ligados son críticos para la actividad del LS

A pesar de tener una mayor actividad en ausencia de ácidos siálicos terminales, la importancia de los otros carbohidratos en los O-glicanos para la actividad del LS no se comprendía completamente. Mientras que la actividad del LS aumenta significativamente en ausencia de ácidos siálicos terminales, la eliminación de galactosas inhibe completamente la actividad del LS (Fig. 4A). Además, la menor actividad de LS sobre proteínas sialiladas,

no se debe a una incapacidad para hidrolizar el enlace en presencia de ácidos siálicos, como demuestra la hidrólisis completa tras la incubación de una noche (Fig. 4B). La actividad del LS depende totalmente de los O-glicanos, ya que la eliminación de los N-glicanos no afectó a la hidrólisis por LS (Fig. 4C).

5 **Los glicanos O-ligados dirigen el sitio de escisión del LS**

Una vez demostrado que el O-glicano es crítico para la actividad, intentamos investigar el lugar específico de escisión del LS. Mediante espectrometría de masas, pudimos demostrar que el LS hidroliza el enlace amino entre el Ser/Thr O-glicosilado y su aminoácido N-terminal, independientemente de su tipo (*por ejemplo*, la prolina no parece inhibir la hidrólisis) (Fig. 5).

Utilizando etanercept como proteína modelo debido a su gran abundancia de glicanos O-ligados, la glicoproteína se trató con LS, tras lo cual se trató posteriormente con O-glicosidasas para facilitar el análisis de espectrometría de masas. Los valores m/z generados a partir del análisis de espectrometría de masas, en combinación con los datos MS/MS, se ajustaron a etanercept. Todos los péptidos identificados tenían una serina o treonina N-terminal, lo que concuerda con la escisión del LS justo en el N-terminal de los O-glicanos (Fig. 5). El análisis identificó péptidos tanto en una búsqueda dirigida (definiendo la hidrólisis S/T en los parámetros; Fig. 5A), como en un enfoque imparcial (Fig. 5B).

20 **Una variante inactiva hidrolítica del LS se une específicamente a proteínas que contienen O-glicanos**

Dada la capacidad del LS para unirse a O-glicanos e hidrolizar específicamente el enlace aminoácido próximo al glicano (*por ejemplo*, junto a Ser/Thr), planteamos la hipótesis de que un mutante E_{206A} del LS carecería de actividad hidrolítica, pero conservaría la capacidad de unión. Una herramienta de este tipo sería valiosa, entre otras cosas, para a) identificar glicoproteínas O-ligadas, b) purificar por afinidad glicopeptídos O-ligados para eliminarlos o estudiarlos, y c) purificar por afinidad O-glicanos.

La Fig. 6A muestra que el mutante LS no tenía ninguna actividad hidrolítica detectable. Mientras que el LS era capaz de hidrolizar etanercept en presencia de sialidasa, el LSmut no podía hidrolizar etanercept, confirmando que la alteración genética inactivaba efectivamente la O-glicoproteasa en las condiciones probadas.

El LSE_{206A} se inmovilizó en sefarosa y se añadió a columnas de centrifugación para facilitar su manipulación. Es importante destacar que la unión del LSE_{206A} a diferentes sustratos se correlacionó perfectamente con la actividad hidrolítica del LS (Fig. 6B). El LSE_{206A} (etiquetado como LSmut) demostró una afinidad específica por las glicoproteínas O-ligadas. Inmovilizando el LSmut en sepharose pudimos purificar IgA por afinidad. Sin embargo, no pudimos eluir la proteína, probablemente debido a una fuerte afinidad. El Herceptin (trastuzumab), que carece de O-glicanos, así como la IgA tratada con O-glicosidasa, no se unió a la columna, pero pudo detectarse en el flujo (FT).

40 **Es importante eliminar 2-3 enlaces siálicos para una actividad del LS completa**

Recientemente determinamos que la actividad endoproteasa dependía de enlaces de ácido siálico específicos, lo que requería la eliminación tanto de los ácidos siálicos enlazados 2-3 como de los enlazados 2-6 para un efecto completo. Para determinar el papel individual de enlaces específicos de ácido siálico para la actividad del LS, incubamos Enbrel con diferentes sialidasas en combinación con el LS durante 30 min - 20 h. La eliminación de los enlaces 2-3 parecía suficiente para la hidrólisis por LS (Fig. 7).

El LS escinde la eritropoyetina (EPO)

50 La EPO se trató con PNGasaF, una sialidasa (Smix, comprende Am0707 y Am1757) y/o una O-glicosidasa y se incubó con LS.

A continuación, los productos resultantes se analizaron mediante SDS-PAGE y tinción con azul de Coomassie, así como mediante RPLC y espectrometría de masas. Los resultados del SDS-PAGE se muestran en la Figura 9A, que muestra que el LS escinde la EPO tanto cuando se han eliminado los ácidos siálicos como cuando están intactos. Además, el LS también digiere la EPO en la que se han eliminado los N-glicanos con PNGasaF, lo que confirma que la actividad del LS no se ve afectada por la eliminación de los N-glicanos. Sin embargo, el LS no escindió la EPO cuando se eliminaron los O-glicanos con O-glicosidasa, lo que demuestra que el O-glicano es necesario para que el LS escinda una proteína. Se observaron resultados equivalentes en proporciones de 10:1, 5:1 y 2:1 (sustrato:enzima) (datos no mostrados).

Las mezclas de muestras tras la incubación con PNGasaF, Smix y LS se separaron mediante cromatografía líquida de fase inversa y se analizaron por espectrometría de masas ESI para la identificación de los productos de reacción tras el tratamiento enzimático.

La Figura 9B muestra un cromatograma UV de la RPLC. Como era de esperar, dado que la EPO sólo tiene una posición sugerida de O-glicano (véase la posición predicha en SEQ ID NO: 14 abajo), el cromatograma muestra 2 picos que corresponden a los 2 fragmentos resultantes de la escisión por LS.

5 Estos fragmentos se analizaron posteriormente por EM (véanse las figuras 9C y D) y se identificaron como sigue:

10 SAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGD (Masa = 4900.5868Da - corresponde a la secuencia C terminal al punto de escisión y por lo tanto incluye el O-glicano todavía unido a la serina N terminal); y

15 APPRLICDSRVLERYLLEAKEAEDITTGCAEHCSLDENITVPDTKVDFYAWKRMEVGQQAVEVWQGLALL SEAVLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTLLR ALGAQKEAISPPDAA (Masa = 13714.1199Da, corresponde a la secuencia N terminal al punto de escisión).

15 Por lo tanto, el uso combinado de PNGasaF, mezcla de sialidasa y LS permitió el aislamiento y la identificación precisa de la serina portadora de O-glicanos en la EPO. Los métodos de este tipo son aplicables a cualquier O-glicoproteína y permiten la rápida identificación de las posiciones de unión de los O-glicanos.

Ejemplo 2

Introducción

El mutante LSE206A descrito en el Ejemplo 1 incorpora una mutación dirigida al sitio del sitio activo de LS (abx~~H~~bbHbc a abx~~A~~bbHbc), eliminando la capacidad de transferencia de electrones de la hendidura enzimática.

25 Como se explica más adelante, al realizar más pruebas de estrés se descubrió que, aunque este cambio reducía la actividad de la O-glicoproteasa en relación con la secuencia de tipo salvaje, no la eliminaba por completo. En consecuencia, los inventores han desarrollado y caracterizado otro mutante que incorpora una sustitución adicional en la hendidura enzimática. En concreto, se sustituyó un residuo His importante en la orientación del cofactor de ión de zinc por un Ala. El doble mutante resultante se denomina H205A/E206A (abx~~H~~bbHbc a abx~~A~~bbHbc).

2.1 Producción del doble mutante

Se utilizó mutagénesis dirigida al sitio utilizando protocolos estándar (por ejemplo, como en el Ejemplo 1) para cambiar tanto una histidina como un ácido glutámico por alanina en relación con el gen Amuc1119 de *Akkermansia muciniphila*, para crear el doble mutante, Amuc1119H205A/E206A (LS_{H205A/E206A}). La construcción se transformó en *E. coli*, se aisló y se verificó mediante secuenciación como en el Ejemplo 1. La expresión en *E. coli* se llevó a cabo como se describe en el Ejemplo 1. La secuencia de la proteína expresada se proporciona como SEQ ID NO: 21.

2.2 Caracterización del doble mutante

2.2.1 El doble mutante inactiva completamente la actividad del LS

Como se muestra en el Ejemplo 1, se observó que el único mutante LSE206A era inactivo dada su incapacidad para hidrolizar una O-glicoproteína en 2 horas. Sin embargo, en una prueba de estrés se comprobó que la actividad de la O-glicoproteasa no quedaba completamente abolida, sino que sólo se reducía en el sentido de que se observaba cierta actividad en proporciones más elevadas de la enzima : O-glicoproteína y tiempos de incubación más largos.

50 La incubación durante 24 horas en una proporción 1:1 (peso:peso) de LSE206A : O-glicoproteína asialilada, dio lugar a una hidrólisis significativa del sustrato, aunque no en la misma medida que el LS de tipo salvaje (Fig 10A). Por el contrario, el doble mutante LSH205A/E206A no produjo ninguna evidencia de hidrólisis incluso a proporciones 15:1 (peso:peso) para la enzima : O-glicoproteína, después de la incubación durante la noche (Fig 10B), lo que sugiere que la enzima estaba completamente inactiva con la adición de la segunda mutación.

2.2.2 El doble mutante se une específicamente a las O-glicoproteínas

Para evaluar la unión a diferentes proteínas, el LS_{H205A/E206A} inmovilizado (50 µl de resina) (preparado siguiendo el mismo protocolo que en el Ejemplo 1) se equilibró en PBS, tras lo cual se añadieron 50 µg de diferentes muestras de proteínas, en una concentración de 0.5 mg/ml y se incubaron con rotación de extremo a extremo durante 2 horas a temperatura ambiente. El flujo se recogió mediante centrifugación (200 g, 1 min) y la resina se lavó 3 veces con 350 µl de PBS. Las proteínas unidas se eluyeron mediante dos incubaciones secuenciales de 5 minutos con 50 µl 8 M de urea, seguidas de centrifugación (1000 g, 1 min). Todas las muestras se cargaron en volúmenes iguales. El material de partida/carga, el flujo y el eluido se analizaron mediante SDS-PAGE.

En el primer experimento (véase la Figura 11A), las proteínas glicosiladas o no glicosiladas se pretrataron con una mezcla de sialidasas (Am0707:Am1757) o con una combinación de la mezcla de sialidasas y una Endo- α -N-acetilgalactosaminidasa de *Streptococcus oralis* (por ejemplo, una O-glicosidasa) antes de incubarlas con la resina, lavarlas y eluir las. El pretratamiento de las muestras (mezcla de sialidasa +/- O-glicosidasa) se realizó de acuerdo con las instrucciones del fabricante. Sólo las proteínas con O-glicanos se unieron a la resina, con una mayor afinidad en ausencia de ácidos sílicos. La presencia de O-glicanos era imprescindible para que se produjera cualquier unión, como demuestra la falta de interacción tras el tratamiento con O-glicosidasa.

En el segundo experimento (véase la Figura 11B), se incubó una mezcla de proteínas N-glicosiladas, O-glicosiladas y no glicosiladas con resina LS doble mutante. Sólo las proteínas O-glicosiladas (TNF α R y ApoE) se unieron a la matriz y eluyeron con 8 M de urea. Los N-glicosilados (afiblertcept, AGP (glicoproteína alfa-1-ácido), el dominio Fc de IgG (IgG Fc) y los no glicosilados (BSA) no se unieron a la resina LS doble mutante y se encontraron en el flujo de paso. Así, la resina doble mutante se une específicamente sólo a proteínas O-glicosiladas cuando una muestra contiene una mezcla de proteínas N-, O- y no glicosiladas.

En el tercer experimento (véase la Figura 11C), se incubó una mezcla de proteínas N-glicosiladas y no glicosiladas con resina LS doble mutante. No se observó ninguna unión inespecífica, ni siquiera en ausencia de la posible competencia de las O-glicoproteínas (no hay ninguna). No se encontraron proteínas en el eluido. Así pues, la resina doble mutante no se une a las proteínas que carecen de O-glicanos.

2.2.3 El doble mutante puede inmovilizarse en la resina a diferentes concentraciones para mejorar la capacidad

Para investigar la capacidad de mejorar la capacidad de la resina de doble mutante inmovilizada, para unir más proteínas O-glicosiladas, se utilizaron diferentes concentraciones de doble mutante (5-15 mg/ml) durante la inmovilización en la resina. En la Figura 12A se muestra un gel representativo. El % mostrado es el nivel de unión relativo al control positivo y se determinó mediante análisis de densitometría del gel. Los resultados se muestran en el gráfico de la Figura 12B. Se observó un aumento de la capacidad dependiente de la dosis, con una mayor capacidad de unión a la O-glicoproteína, cuando se utilizó una mayor concentración de doble mutante durante la inmovilización. Los experimentos posteriores continuaron utilizando 15 mg/ml del doble mutante inmovilizado. Además, se mantuvo un alto grado de unión a la O-glicoproteína incluso en presencia de 1 M de urea y 1 M de GHI, aunque este último redujo significativamente la eficacia de la unión.

2.2.4 La capacidad de purificación por afinidad del doble mutante es de ~3 mg de glicoproteína/ml de resina

Con el fin de investigar específicamente la capacidad de la resina de doble mutante para purificar por afinidad las O-glicoproteínas, así como el impacto de la concentración de la muestra en esta capacidad, se añadieron a la resina diferentes cantidades y concentraciones de etanercept asialilado. Una columna individual (que contenía 50 μ l de resina doble mutante) tenía una capacidad de unión de unos 150 μ g de O-glicoproteína, es decir, 3 mg de O-glicoproteína / ml de resina). La figura 13 muestra un gel representativo.

2.2.5 La unión de O-glicoproteínas al doble mutante no se ve afectada significativamente por la fuerza iónica o el volumen/tipo de tampón, y funciona en un amplio rango de pH

Se dejó que las proteínas de la muestra se unieran a la resina de doble mutante durante 2 horas a temperatura ambiente, con rotación de extremo a extremo en una serie de condiciones diferentes, para comprobar el efecto de la fuerza iónica, el volumen/tipo de tampón y el pH en la capacidad de unión de la resina. En cada caso, la resina se lavó tres veces con su tampón de unión respectivo (350 μ l) y después se eluyó con la adición de 8 M de urea (50 μ l, 5 min de incubación; 2 repeticiones). A continuación, todas las muestras se analizaron mediante SDS-PAGE.

En un primer experimento (véase la Figura 14A), para investigar la estabilidad de la interacción en tampones con distinta fuerza iónica, la muestra consistió en etanercept asialilado, que se incubó con resina doble-mutante en NaCl 0-4 M, además de realizar todos los pasos de lavado con la concentración respectiva de NaCl. La adición de NaCl no afectó significativamente a la unión del etanercept asialilado.

En un segundo experimento (véase la Figura 14B), la muestra consistió en etanercept asialilado en una gama de diferentes volúmenes de PBS. En los pasos de lavado se utilizó PBS. La variación del volumen de sustrato entre 100-300 μ l no afectó significativamente a la eficacia.

En un tercer experimento (véanse las figuras 14C y D), la muestra consistió en etanercept asialilado y BSA en diferentes tampones (100 mM de acetato sódico, 50 mM de fosfato sódico y 50 mM de Tris) a diferentes pH (pH 4-9). Se comprobó que el pH 6-8 funcionaba mejor, mientras que el pH 4 no funcionaba en absoluto y el pH 9 era ligeramente menos eficaz que el pH 8. La BSA, que no contiene O-glicanos, no se unió a la resina en ninguna de las condiciones de unión.

2.2.6 La urea y el SDS pueden eluir O-glicoproteínas unidas por afinidad

Basándose en la alta afinidad entre el doble mutante y el sustrato O-glicoproteína, los inventores investigaron diferentes medios para eluir la proteína unida de la resina, no basados en la fuerza iónica. La urea tuvo una elución dependiente de la dosis, con una elución cercana al 100% con el uso de 8 M de urea (Fig. 15A). Altas 5 concentraciones de SDS (por ejemplo, 5-10%) también eluyeron la mayor parte de la proteína unida (Fig. 15B). Sin embargo, dado que muchas aplicaciones posteriores son sensibles a la presencia de detergentes, es probable que el uso de altos niveles de urea sea de mayor utilidad práctica para la liberación no enzimática de proteínas/péptidos unidos.

10 **2.2.7 Los LS de tipo salvaje pueden utilizarse para eluir O-glicoproteínas unidas a doble mutante**

Los inventores especularon con la posibilidad de que la adición de LS a las proteínas unidas a doble mutante pudiera dar lugar a una liberación de las mismas y, por tanto, no fuera necesaria la adición de urea para la elución. Tanto abatacept como etanercept pudieron ser hidrolizados y eluidos de la resina doble-mutante por LS en 6 h 15 pero tuvieron una elución ligeramente más completa después de 24 h (Fig. 16A). La adición posterior de urea mostró que quedaba muy poca O-glicoproteína adherida a la matriz de afinidad, lo que demuestra que la estrategia de elución del LS era muy eficaz.

20 El etanercept eluido con el LS también se sometió a análisis por espectrometría de masas (LC/MS y MS/MS). Los péptidos identificados (Fig 16B.1) coincidían con los generados en una digestión del LS de etanercept (Fig 16B.2). En la tabla siguiente se muestran otros datos de EM de este experimento:

Fila	Ork	Cópia	m/z medida	Mz cál.	Z	A m/z lopos	RMS90 (ppm)
1	VERDADERO	204	592.18579	582.1545447	2	3.18935727	39.5628229
2	VERDADERO	1240	480.24064	459.2541484	1	3.873732048	6.970140398
3	VERDADERO	2393	593.93513	578.59071	3	-4.012654897	37.85883109
4	VERDADERO	1613	598.28991	586.418529	5	-8.916891727	23.18053663
5	VERDADERO	1368	642.65438	1924.956428	3	7.665331795	12.36198202
6	VERDADERO	2169	672.31437	1543.621461	2	-5.366321116	7.28732308
7	VERDADERO	2167	672.31437	1542.621461	2	-5.366321116	7.28732308
8	VERDADERO	1666	695.83359	1389.70547	7	-2.245927789	13.08730439
9	VERDADERO	1887	745.4213	544.4170258	1	-4.021055897	40.39538533
10	VERDADERO	2308	846.8813	1689.758682	2	-8.278836744	10.77131192
11	VERDADERO	1636	878.52107	1734.839667	7	-6.875687232	8.158438723
12	VERDADERO	2798	895.09553	3685.161633	3	-1.614552583	8.779152853
13	VERDADERO	2648	998.39499	1794.785823	2	-5.875354572	10.636492108
14	VERDADERO	2319	924.06736	2769.59361	3	-4.877599176	9.79889835
15	ENRODADO	2269	1072.20596	7143.92580	2	-1.87227591	10.42121345
16	VERDADERO	1315	1189.55981	2377.320982	2	-6.589738429	9.551830784
17	VERDADERO	1325	1189.55981	2377.3206552	2	-6.540738429	9.551830784
18	VERDADERO	2375	1219.18607	3654.352682	3	-4.091374291	9.325771178
Fila	tot.		Resultados			Resumen	
1	160010		20.3 (D.mezzscore: 20.3) D.bScore: 16.0 DfragCov: 20.0 D.int.Cov: 21.3			431.269	
2	29833		23.0 (D.mezzscore: 23.0) D.bScore: 16.0 DfragCov: 22.3 D.int.Cov: 22.3			182.138	
3	30476		31.2 (D.mezzscore: 31.2) D.bScore: 157.0 DfragCov: 40.0 D.int.Cov: 28.3			216.325	
4	219206		26.2 (D.mezzscore: 26.2) D.bScore: 265.0 DfragCov: 53.2 D.int.Cov: 21.1			384.268	

5	657586	23.9 (D.mezzscore: 23.9) D.bScore: 136.0 DfragCov: 46.7 D.int.Cov: 12.3	184.198
6	374698	36.2 (D.mezzscore: 36.2) D.bScore: 61.0 DfragCov: 33.3 D.int.Cov: 39.2	217.225
7	374698	36.2 (D.mezzscore: 36.2) D.bScore: 61.0 DfragCov: 33.3 D.int.Cov: 39.2	217.225
8	394898	26.9 (D.mezzscore: 26.9) D.bScore: 12.0 DfragCov: 31.7 D.int.Cov: 20.3	196.198
9	18890	68.8 (D.mezzscore: 68.8) D.bScore: 191.0 DfragCov: 92.0 D.int.Cov: 48.1	1.7
10	188910	27.4 (D.mezzscore: 27.4) D.bScore: 138.0 DfragCov: 27.3 D.int.Cov: 22.3	217.225
11	275854	29.6 (D.mezzscore: 29.6) D.bScore: 260.0 DfragCov: 59.8 D.int.Cov: 22.3	186.199
12	365866	24.2 (D.mezzscore: 24.2) D.bScore: 569.0 DfragCov: 28.1 D.int.Cov: 20.8	216.231
13	48878	35.0 (D.mezzscore: 35.0) D.bScore: 153.0 DfragCov: 35.0 D.int.Cov: 35.0	216.225
14	38885	23.7 (D.mezzscore: 23.7) D.bScore: 25.0 DfragCov: 33.5 D.int.Cov: 3.8	216.225
15	29176	20.1 (D.mezzscore: 20.1) D.bScore: 28.6 DfragCov: 25.0 D.int.Cov: 16.2	216.225
16	2495838	23.1 (D.mezzscore: 23.1) D.bScore: 52.8 DfragCov: 59.4 D.int.Cov: 16.1	184.199
17	187236	22.9 (D.mezzscore: 22.9) D.bScore: 40.9 DfragCov: 43.8 D.int.Cov: 12.0	184.199
18	380318	24.1 (D.mezzscore: 24.1) D.bScore: 33.6 DfragCov: 22.2 D.int.Cov: 20.0	208.225

Fila	#Cpds.	P	Secuencia	Modificaciones	Proteína
1	0	Y.TCK.S		Acetyl, 1-HexylhexNAc(1,1)	Elasorecept
2	1	TSPTK.S			Elasorecept
3	5	2	P.TYSFLPQMGPS	Beta(1,3)Beta(NAc(1,3),2)	Elasorecept
4	13	4	P.TRSMAANGAVH.PQPVSTRSCHI	Beta(1,3)Beta(NAc(1,3),16,17)	Elasorecept
5	6	1	P.TRSMAANGAVH.PQPV.S	Beta(1,3)Beta(NAc(1,3))	Elasorecept
6	4	3	S.TSFLLPMGCP.S	Oxidación, 3-Beta(1,3)Beta(NAc(1,3))	Elasorecept
7	4	1	S.TSFLLPMGCP.S	Oxidación, 3-Beta(1,3)Beta(NAc(1,3))	Elasorecept
8	2	1	P.SMASAAGVHLPCPV.S		Elasorecept
9	5	0	-LPAQVAKT		Elasorecept
10	2	1	S.TSFLLPMGCP.S	Beta(1,3)Beta(NAc(1,3),2)	Elasorecept
11	3	1	R.SMASAAGVHLPCPV.S	Beta(1,3)Beta(NAc(1,3),14)	Elasorecept
12	2	3	P.TSFLLPMGCPSSPPAHGS	Beta(1,3)Beta(NAc(1,3),2,3)	Elasorecept
13	10	2	P.TSFLLPMGCP.S	Oxidación, 3-Beta(1,3)Beta(NAc(1,3),2)	Elasorecept
14	2	4	P.TSFLLPMGCPFAEGST	Beta(1,3)Beta(NAc(1,3),2,3)	Elasorecept
15	5	2	P.TSFLLPMGCP.S	Beta(1,3)Beta(NAc(1,3),2,3)	Elasorecept
16	19	2	P.TBSMAGAVHLPCPV.S	Beta(1,3)Beta(NAc(1,3),16)	Elasorecept
17	19	2	P.TBSMAGAVHLPCPV.S	Beta(1,3)Beta(NAc(1,3),3)	Elasorecept
18	8	5	P.TDPSTAPSTNSL.LPAGK.S	Beta(1,3)Beta(NAc(1,3),5,8,9,10,11)	Elasorecept

2.2.8 El doble mutante puede utilizarse para purificar por afinidad O-glicoproteínas de muestras complejas

- Como prueba del concepto de que el sistema puede funcionar como matriz de afinidad general para proteínas O-glicosiladas, no sólo en sistemas simplificados sino en medios complejos, los inventores investigaron la capacidad del doble mutante para purificar O-glicoproteínas a partir de suero humano. El suero humano se compone principalmente de proteínas no glicosiladas (BSA) y N-glicosiladas (IgG), y sólo una pequeña fracción del proteoma sérico total está O-glicosilado.
- La aplicación de 20 µl de suero tratado con sialidasa (aproximadamente 1.2 mg de proteína) a una columna de resina doble mutante inmovilizada de 50 µl permitió eliminar casi todas las proteínas no glicosiladas y N-glicosiladas, eluyendo al mismo tiempo unas pocas proteínas seleccionadas (Fig. 17A). Añadiendo mayores cantidades de suero (por ejemplo, 2.5 mg de proteína) con o sin pretratamiento de sialidasas y O-glicosidasas se demostró que la interacción depende de los O-glicanos y de la eliminación de los ácidos siálicos terminales (Fig. 17B). Además, se concluyó que el pretratamiento con sialidasas aumentaba significativamente la cantidad de O-glicoproteínas unidas en comparación con las muestras no tratadas con sialidasas. La adición de 50 U de mezcla de sialidasa (Am0707:Am1757) fue suficiente para mejorar la cantidad de O-glicoproteínas purificadas por afinidad (Fig. 17C).
- Por análisis en espectrometría de masas, la gran mayoría de las proteínas séricas purificadas por afinidad, pueden ser anotadas como proteínas O-glicosiladas (véase la Fig. 18A, y los nombres en **negrita-cursiva** en la tabla siguiente). El número de péptidos O-glicoproteínicos identificados en relación con los péptidos no O-

ES 2 986 910 T3

5

glicoproteínicos pudo verse afectado por las diferentes rigurosidades en los pasos de lavado, tanto en términos del número total de péptidos identificados (Fig. 18B) como en la proporción de péptidos O-glicoproteínicos frente a péptidos no O-glicoproteínicos (Fig. 18C). Así pues, está claro que la resina de afinidad es muy eficaz en su capacidad de purificar y enriquecer específica y selectivamente las O-glicoproteínas por afinidad. En la tabla siguiente se muestran otros datos de EM de este experimento:

Fila	OR	Proteína	MW kDa
1	XXXXXX0000	<i>Ig cadena alfa-1 región C</i> OS-Homo sapiens Gln>Rala; Phe>1 SV>2	37.6
2	XXXX000000	<i>Apolipoproteína B-199</i> OS-Homo sapiens Gln>Asp; Phe>1 SV>2	515.3
3	XXXX000000	<i>Kininaseta-1</i> OS-Homo sapiens Gln>Eng; Phe>1 SV>2	71.9
4	XXXX000000	<i>Complemento C4-B</i> OS-Homo sapiens Gln>Cys; Phe>1 SV>2	197.8
5	XXXX000000	<i>Complemento C4-A</i> OS-Homo sapiens Gln>Cys; Phe>1 SV>2	192.7
6	XXXX000000	<i>Complemento C3</i> OS-Homo sapiens Gln>S; Phe>1 SV>2	97
7	XXXX000000	<i>Inhibidor inter-alfa-tripsina cadena pesada H4</i> OS-Homo sapiens Gln>Ile(G; Phe>1 SV>4	102.3
8	XXXX000000	<i>Inhibidor inter-alfa-tripsina cadena pesada H2</i> OS-3Homo sapiens Gln>Utb2; Phe>1 SV>2	106.4
9	XXXX000000	<i>Plasminogeno</i> OS-Homo sapiens Gln>P; G; Phe>1 SV>2	90.5
10	XXXX000000	<i>Inhibidor inter-alfa-tripsina cadena pesada H3</i> OS-3Homo sapiens Gln>Utb3; Phe>1 SV>2	101.3
11	XXXX000000	<i>Albumina cadena OS-Homo sapiens Gln>A; B Phe>1 SV>2</i>	69.5
12	XXXX000000	<i>Alfa 2-macroglobulina</i> OS-Homo sapiens Gln>A2M Phe>1 SV>3	163.2
13	XXXX000000	<i>Lipoproteína</i> OS-Homo sapiens Gln>I; Phe>1 SV>4	262.5
14	XXXX000000	<i>Complemento C1s subcomponente</i> OS-Homo sapiens Gln>C1S Phe>1 SV>1	6.5
15	XXXX000000	<i>N-aciltransferasa-L-diamina oxidasa</i> OS-Homo sapiens Gln>Vgl; Vrp2; Phe>1 SV>1	62.2
16	XXXX000000	<i>Inhibidor de la proteasa plasmática C1</i> OS-Homo sapiens Gln>NRPIN1; Phe>1 SV>2	53.1
17	XXXX000000	<i>Proteomina</i> OS-Homo sapiens Gln>F; Phe>1 SV>2	70
18	XXXX000000	Cadena alfa de la proteasa de moco a C30 OS-Homo sapiens Gln>C30P; Phe>1 SV>2	67
19	XXXX000000	<i>Factor de coagulación XII</i> OS-Homo sapiens Gln>F12 Phe>1 SV>2	67.7
20	XXXX000000	<i>Complemento C1r subcomponente</i> OS-Homo sapiens Gln>C1R; Phe>1 SV>2	80.4
21	XXXX000000	<i>Ig cadena beta 2 región C</i> OS-Homo sapiens Gln>WHD; Phe>1 SV>2	52.7
22	XXXX000000	<i>Apolipoproteína E</i> OS-Homo sapiens Gln>A; Phe>1 SV>2	38.5
23	XXXX000000	<i>Complemento componente C7</i> OS-Homo sapiens Gln>C7 Phe>1 SV>2	93.5
24	XXXX000000	<i>Ig cadena alfa-2 región C</i> OS-Homo sapiens Gln>IGMN; Phe>1 SV>3	49.3

			Individuo: otoño-alta-tríptica; cadena pesada H3: OS-His-Glycine-Glycine-Arginine-Phe-Tyr	
	NOMBRE	SVs		
23	NC005000	SVs2	Glycoprotein rich in histidine OS->Lysine sequence GN-NH2 Phe-Tyr-SVs1	99.8
26	NC005000		Glycoprotein rich in histidine OS->Lysine sequence GN-NH2 Phe-Tyr-SVs1	99.5
37	NC005000		Capsid protein of bacteriophage C3 OS->Histidine sequence GH-Cys-Phe-Tyr-SVs2	63.1
28	NC005000		Af9-4-HS glycoprotein OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	99.3
29	NC005000		Protein SMH OS->Lysine sequence GN-NH2 Phe-Tyr-SVs1	98
30	NC005000		The endotoxin gamma Urease C OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	96.1
31	NC005000		Alpha-1 antitrypsin OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	96.7
32	NC005000		Cofactor phosphatase OS->Fatty acids GN-NH2 K3 Phe-Tyr-SVs2	71.3
25	NC005000		Yersinia enterocolitica OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	54.5
34	NC005000		Apolipoprotein A-I OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	30.8
35	NC005000		Protein S-deoxyribose N-oxide OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	75.1
36	NC005000		Serotransferrin OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	77
37	NC005000		Apolipoprotein E OS-Histidine sequence GN-NH2 Phe-Tyr-SVs1	96.1
38	NC005000		Factor 10a Wilberand OS-Histidine sequence SVs1 SVs2 SVs3	99.9
39	NC005000		Proteoglycan 4 OS-Histidine sequence GN-NH2 Phe-Tyr-SVs2	100

Filo	ph	# Peptidos	SG (%)	Resultados	KMSW0 (ppm)
1	6.2	132	75.1	1239.4 (M+expext:0.0 M score:1239.4 M sigleut:38.0)	5.36
2	6.0	132	41.2	3043.5 (M+expext:0.0 M score:3043.5 M sigleut:38.0)	5.62
3	5.8	132	55.7	1036.5 (M+expext:0.0 M score:1036.5 M sigleut:38.0)	5.77
4	6.9	78	61.3	4951.1 (M+expext:0.0 M score:4951.1 M sigleut:38.0)	6.36
5	6.7	76	61.3	4488.7 (M+expext:0.0 M score:4488.7 M sigleut:38.0)	6.12
6	6	75	84.1	4599.7 (M+expext:0.0 M score:4599.7 M sigleut:38.0)	5.13
7	6.5	82	56.9	2389.7 (M+expext:0.0 M score:2389.7 M sigleut:38.0)	6.29
8	6.8	49	50.4	2175.2 (M+expext:0.0 M score:2175.2 M sigleut:38.0)	7.02
9	7	85	67.1	2752.2 (M+expext:0.0 M score:2752.2 M sigleut:38.0)	5.78
10	6.3	39	43.8	1878.3 (M+expext:0.0 M score:1878.3 M sigleut:38.0)	5.16
11	5.9	37	63.7	2350.1 (M+expext:0.0 M score:2350.1 M sigleut:38.0)	5.59
12	6	36	39.1	1979.1 (M+expext:0.0 M score:1979.1 M sigleut:38.0)	5.07
13	5.5	32	28.3	1671.8 (M+expext:0.0 M score:1671.8 M sigleut:38.0)	5.38
14	4.8	24	46.1	1454.6 (M+expext:0.0 M score:1454.6 M sigleut:38.0)	6.01
15	5.2	22	36.4	1289.8 (M+expext:0.0 M score:1289.8 M sigleut:38.0)	5.72
16	6.1	23	38.2	1586.8 (M+expext:0.0 M score:1586.8 M sigleut:38.0)	6.15
17	5.6	20	43.4	1158.1 (M+expext:0.0 M score:1158.1 M sigleut:38.0)	6.13
18	7.2	20	51.4	1111.2 (M+expext:0.0 M score:1111.2 M sigleut:38.0)	5.13
19	8	20	59.7	601.9 (M+expext:0.0 M score:601.9 M sigleut:38.0)	5.05
20	5.8	19	39.6	1097.7 (M+expext:0.0 M score:1097.7 M sigleut:38.0)	6.07
21	8.2	59	36.9	677.5 (M+expext:0.0 M score:677.5 M sigleut:38.0)	5.97
22	5.4	39	43.3	1444.0 (M+expext:0.0 M score:1444.0 M sigleut:38.0)	5.73
23	6.6	38	36.1	1100.7 (M+expext:0.0 M score:1100.7 M sigleut:38.0)	5.73
24	6.2	17	30.9	11530.0 (M+expext:0.0 M score:11530.0 M sigleut:38.0)	5.15
25	5.5	36	747.2 (M+expext:0.0 M score:747.2 M sigleut:38.0)	5.8	
26	7.1	35	35.4	904.5 (M+expext:0.0 M score:904.5 M sigleut:38.0)	5.33
27	5.2	15	34.3	876.5 (M+expext:0.0 M score:876.5 M sigleut:38.0)	5.88
28	5.5	14	46.1	941.9 (M+expext:0.0 M score:941.9 M sigleut:38.0)	5.72
29	5.3	54	35.4	886.3 (M+expext:0.0 M score:886.3 M sigleut:38.0)	5.98
30	5.5	54	60.3	825.2 (M+expext:0.0 M score:825.2 M sigleut:38.0)	6.08
31	5.7	34	780.9 (M+expext:0.0 M score:780.9 M sigleut:38.0)	6.02	
32	8.6	54	26	647.7 (M+expext:0.0 M score:647.7 M sigleut:38.0)	5.82
33	5.6	25	57	752.2 (M+expext:0.0 M score:752.2 M sigleut:38.0)	5.1
34	5.6	32	46.8	1211.0 (M+expext:0.0 M score:1211.0 M sigleut:38.0)	5.77
35	5.8	32	27.7	676.0 (M+expext:0.0 M score:676.0 M sigleut:38.0)	5.18
36	6.8	33	23.9	584.9 (M+expext:0.0 M score:584.9 M sigleut:38.0)	6.02
37	5.6	38	45.1	515.3 (M+expext:0.0 M score:515.3 M sigleut:38.0)	5.75
38	5.3	53	6.3	505.8 (M+expext:0.0 M score:505.8 M sigleut:38.0)	5.82

[39 : 8.5 : 37 : 5.6 : 314.2 M+expext:0.0 M score:2132.8 M sigleut:38.0 : 4.7]

2.2.9 El doble mutante inmovilizado también se une a O-glicopéptidos más cortos

- 5 Se realizaron una serie de experimentos para demostrar la especificidad del doble mutante del LS también para los O-glicopéptidos. En el primer experimento, se mezcló un péptido O-glicosilado (glicodrosocina (GD) = GKPRPYSPRPTSHPRPIRV (SEQ ID NO: 47) con un núcleo 1 O-glicano en la treonina) y varios péptidos no glicosilados (H2686, H4062 H8390 y cadena beta oxidada de insulina (IOB)), se incubó con resina LS doble mutante. (H2686 = YIYGSFK (SEQ ID NO: 48), H4062 = KKLVFFA (SEQ ID NO: 49), H8390 = FLPLILGKLVKGLL (SEQ ID NO: 50)).
- Se dejó que la mezcla de péptidos se uniera a 50 µl de resina doble mutante inmovilizada durante 2 horas a temperatura ambiente con rotación de extremo a extremo. La resina se lavó cinco veces con tampón de unión (300 µl) y después se eluyó con la adición de 8 M de urea. Los péptidos de la carga, el flujo y el eluido se analizaron con LC/MS. La separación se realizó en una columna RP-LC C18 (Advance BioPeptide Map 2.1x100 2.7µm de Agilent) y se detectó con ESI-Q-TOF Bruker Impact II. Los resultados se muestran en la Figura 19A. La glicodrosocina, el único péptido de la mezcla que contiene un O-GalNAcGal, se encontró predominantemente en la fracción eluida y los péptidos no glicosilados en la fracción de flujo.
- 20 En el segundo experimento, se investigó si el doble mutante del LS podía enriquecer péptidos O-glicosilados a partir de una digestión tríptica de proteínas (por ejemplo, un tipo diferente de mezcla de péptidos). Se eligió la IgA como diana para la digestión. Basándose en los sitios de tripsina y los sitios de O-glicosilación reportados en la IgA, una digestión con tripsina debería resultar en un solo péptido O-glicosilado correspondiente a las posiciones 89-126 de la IgA (ver diagrama esquemático en la Figura 19B). Para crear la digestión tríptica, la IgA se mezcló con urea a 6M y DTT a 5 mM seguido de incubación a 37°C durante 1h. Se añadió IAM a 15 mM y se incubó a

temperatura ambiente durante 30 minutos en la oscuridad. A continuación, la muestra se intercambió de tampón a 50 mM Tris, pH 8.0 en la columna 7000 K de centrifugación Zeba. A continuación, se añadió tripsina en una proporción de 1:20 y se incubó a 37 °C durante toda la noche. Se añadió un inhibidor de tripsina de 1 mg/mg y se incubó a temperatura ambiente durante 20 minutos. Se añadieron la mezcla de sialidasa y NaCl al digerido tríptico resultante. Se dejó que la mezcla se uniera a la resina durante 2 h a temperatura ambiente con rotación de extremo a extremo. La resina se lavó diez veces con tampón PBS (300 µl) y después se eluyó con la adición de 8 M de urea (50 µl, 2 min, 2 repeticiones).

Los péptidos de la carga, el flujo y el eluido se separaron y analizaron mediante RP-LC MSMS en una columna C18 (Advance BioPeptide Plus 2.1x150mm 2.7µm de Agilent Technologies) en un 0.1%FA en MQ: gradiente de 0.1%FA en 95% ACN a 45°C y un flujo de 0.2 ml/min. La detección se realizó en un instrumento ESI-Q-TOF Bruker Impact II. Los resultados se muestran en la Figura 19C. El péptido O-glicosilado 89-126 se enriqueció significativamente en el eluido y el O-glicopéptido 89-126 específico se identificó con masa intacta.

2.2.10 El doble mutante inmovilizado se compara favorablemente con otras matrices de unión de O-glicoproteínas

Los inventores evaluaron la capacidad del doble mutante para purificar por afinidad la O-glicoproteína en comparación con otras matrices de unión de O-glicoproteína disponibles en el mercado, en concreto las lectinas aglutinina de Cacahuete (PNA), y lectina de Vicia villosa (VVA). El etanercept y el etanercept asialilado se utilizaron como sustratos modelo.

Se añadieron 50 µg de sustrato en PBS (PNA y LS doble mutante) o tampón de unión a lectina (VVA) a 50 µl de volumen de diferentes resinas de lectina inmovilizada o de LS doble mutante preequilibradas en los respectivos tampones (total 100 µl). (El tampón de unión a la lectina es 20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM de MgCl₂, CaCl₂, ZnCl₂ y MnCl₂). Se dejó que el sustrato interactuara con las resinas con mezcla de extremo a extremo durante 2 h a temperatura ambiente. Las proteínas no unidas se lavaron (100 g, 1 min; 3 x) con PBS o tampón de unión a lectina, respectivamente. Las resinas se secaron por centrifugación (1000 g, 1 min). Las proteínas unidas se eluyeron con la adición de 8 M de urea (para las resinas PNA y LS doble mutante), o el tampón de elución VVA, según el fabricante (para la resina VVA) (50 µl, 5 min de tratamiento antes de la centrifugación 1000 g 1 min; 2 x) y tanto el flujo (FT) como el eluido (E) se analizaron en SDS-PAGE. Se añadieron 1.5 µg de sustrato a cada gel (por ejemplo, 3 µl) como control positivo y se realizó un análisis densitométrico para evaluar la eficacia de las resinas, en relación con el sustrato cargado con 1.5 µg, lo que supone una eficacia del 100%. En la Figura 20A se muestran geles representativos para etanercept y etanercept asialilado (Etanercept S). Los resultados del análisis densitométrico se muestran en la Figura 20B. El doble mutante LSH205A/E206A funciona al menos tan bien como la lectina comercial de mejor rendimiento en cuanto a eficacia de purificación del sustrato asialilado.

Secuencias

	SEQ	ID	NO:	1
	EVTPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYKGKEMQRHGYGARSFGLDIKSPGRVNIIIEYKAKNPAAHYPYENGGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDPNGGVPFYGMGRNCFALDYPAFDIKHLGQKTREREGLLTWKWYGGMAHELGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSPFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNDYDAVSFSRRLGKSGKFSMKIDKKELEGNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKS			
45	MEVTPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYKGKEMQRHGYGARSFGLDIKSPGRVNIIIEYKAKNPAAHYPYENGGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDPNGGVPFYGMGRNCFALDYPAFDIKHLGQKTREREGLLTWKWYGGMAHELGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSPFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQDYDAVSFSRRLGKSGKFSMKIDKKELEGNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKS	NO: ID: HHHHHH	NO: ID: HHHHHH	2

ES 2 986 910 T3

SEQ	ID	NO:
ATGGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCCTGAAGAAAACCCTCGTCAGCTGAATATCG TCTACTTCCGGTTCTGATACCGAACCGGTTCCGGACTACAGAGCGCCGCTGAGCGAGCTGCTGGTGTATCT GCAGCAATTCTATGGTAAAGAAATGCAGGCCATGGCTATGGCGCACCGAGCTTGGTCTGGACATTAAGTC CCGGGTCTGTGAACATTATCGAGTACAAGCGAAGAACCCGGCAGCGCAATTACCGTATGAGAATGGTGGCG GCTGGAAGCTGCACAAGAACTGGACGAATTTTCAGGGCCATCCAGACCGAAGAAAAGCCACACCCCT GATCATCATGCCTACTGGAATGATGAGAAAATGGTCTGACAATCCGGTGGCTTCGTTCTATGGTATG GGTCGTAATTGTTTGCCTGGACTACCCGGCTTGTATACAGCACCTGGTCAGAAAACGCGTGAGGGTC GTCTGCTGACGAAATGGTACGGTGGCAACTGGCAGCGCAACTGGGCAACCTGGCTGAATCTGCCGCACAATCACCA GACCGCGAGCGATGGCAAGAAAATATGGCACCCCTGATGGTGGCGCAACTACAGCTCGTACCCAGCG ACGTTCTGACCCCCGGCAGCTGTCGCTGCTGGATGCTGCAAGTGTTCAGCGTTACCCCGAGCCAACAGT TTTATGAGGTTAACCGAGTCAGGTTGGTGTATGTTGCAATTCTCTAACGGGTGATCAAATCTGGTCA CGGTAACATACAAGAGCCCGAACCGTGAAGCTCTGAACTGTTAACGGATCCGGCTACCGCTGAC CAAGACTACGATGCGAGCTTAGCCGTCGCTGGCAAAAGTCGGTAAGTGTAGCATGAAGATTGACA AAAAAGAAGTGGAGGCCTGAATAACACGAACTTCCGTATTCCTGATGTTCATTCAGGCAACGGCTTACA CATGCAGAAGCACTTACGTTACTGGGATGCGCTGCAAGACTACCGTGACGGTAGCAAATCTGGTCCGGT CATCATCACCACCATCACTGAA		

SEQ **ID** **NO:**
MLKRLLSAFFSLFFLGAASGTSFAEVTPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRSELLLYIQ
QFYGKEMQRHGYGARSFGLDIKSPGRVNIIIEYKAKNPAAHYPYENGGWKAQELDEFFKAHPDRKKSQHTLI
IMPTWNDEKNGPDNPAGVFPYGMGRNCFALDYPAFDIKHGQKTRERGLLTWKWYGGMAHELGHGLNLPHNHQT
ASDGKKYGTALMGSGNYTFGTSPTFLTPASCALLDACEVSVTPSQFYEGKPEVEVGDAISFKGDQIILVSG
NYKSPQTVKALNVYIQDPPYAVQNQDYDAVSFSRRLGKKSGKFSMKIDKKELEGNNNEFRISLMFI LANGLHM
QKHFTFWDALQDYYRDGSKS (signal sequence underlined)

5 QKHFTFHWDALQDYRDGSKS (signal sequence underlined)

SEQ ID NO:
EVTVPDAKDRIALKKTARQLNIVFLGSDTEPVPDYERRLSELLLYLQQFYKGEMQRHGYGARSFGLDIKSP
GRVNIIEYKAKNPAAHYPYENGGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPAGVPFYGMG
RNCFALDYPAFDIKHGQKTRERGLLTWKYGGMAHALGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSP
FLTPASCALLDACEVSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQ
DYDAVSFSRLGKKGKFSMKIDKELEGLNNNEFRISLMFI LANGLHMQKHFTRHWDALQDYRDGSKS

10 SEQ ID NO:
MEVTVPDALKDRIALKKTARQLNIVYFLGSDEPVPDYERRLSELLLYLQQFYKGEMQRHGYGARSFGLDI
KSPGRVNIIYEYAKAKNPAAHYPYENGGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNP
PGGVVFYGMGRNCFALDYPAFDFDIKHLGQKTCERGRLLTKWYGGMAHALGHGLNLPHNHQTASDG
KYYGTALMGSGNYTFGTSP
TFLTPASCALLDACEVFSVTPSQFYEGKPEVEVGDAISFKGDIILVSGNYKSPQTVKALNVYIQD
PPYAVN
QDYDAVFSRSLRGKKSFKSMKIDKKELEGLNNNEFRISLMFI
LANGLHMQKHFTFHWDALQDYRDGSKSGSG
HHHHHH

SEQ	ID	NO:
ATGGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAATATCG TCTACTTCCGGGTTCTGATACCGAACCGGTTCCGGACTACGAGCGCCGCTGAGCGAGCTGCTGGTGTATCT GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACCGAGCTTGGTCTGGACATTAAGTC CCGGGTCGTGTGAACATTATCGAGTACAAGCGAAGAACCCGGCAGCCATTACCGTATGAGAATGGTGGCG GCTGGAAGCTGCACAAGAACTGGACGAATTTCAGGGCCATCCAGACCGAAGAAAAGCCAGCACACCC GATCATCATGCCTACCTGGAATGATGAGAAAAATGGTCTGACAATCCGGTGGCGTCTATGGTATG GGTCGTAATTGTTTGCCTTGGACTACCCGGCTTGATATCAAGCACCTGGTCAAGAACCGCTGAGGGTC GTCTGCTGACGAAATGGTACGGTGGCATGGCGCACCGCTGGGCCACCGCTGAATCTGCCGCACAATCACC GACCGCGAGCGATGGCAAGAAAATAGGGCACCCCCCTGATGGGTAGCGCAACTACACCTTCGGTACCGAGCCCC ACGTTCTGACCCCCGGCAGCTGTGCGCTGCTGGATGCGCAAGTGGTCTACGGCTTACCCCGAGCCAACAGT TTATGAGGGTAAGCCAGAAGTCAGGGTTGGTATGTTGCAATTCTCTCAAGGGTATCAAAATCTGGTCAG CGGTAACTACAAGAGCCCCGGCAACCGTGAAGCTCTGAACTACATCAGGATCCCCCTACGCCGTGAAC CAAGACTACGATGCACTGAGCTTAGCCGTCGCTGGGCAAAAGTCGGTAAGTTAGCATGAAGATTGACA AAAAAGAAGTGGAGGCCTGAATAACAAACGAATTCCGTATTCCTGATGTCATTCTGGCAAACGGCTTACA CATCGAGAAGCACTTACGTTCACTGGGATGCGCTGCAAGACTACCGTGACGGTAGCAAATCTGGTCCGGT CATCATCACCAACCATCACTGAA		

15 SEQ ID NO: 8GMAHELGHGL (motivo de metaloproteasa)

ES 2 986 910 T3

	SEQ	ID	NO:	9
	MKNLLFALLTGSFCCCYAQOKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVS RDNGKTWE MKAAPAGYPPNYGGIPKESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKN LKKLGGIMRTPVFNKGRRVIVPFHNMGGGTKFHISDDGGLTWHVSRNGVTSPRHEARPPHQGVRFNNAVEA TVLEMKDGTI WALARTSQDQAWQAFSKDYGETWSKEPEPSRFFGTLMTLGRLLDDGTIVSLWTNTMALPENAT AGNGTWEDVFTNRDSHHIAMSGDEGKTYWGFRERII DEHRNHPGYATLDGPEDRGKHQSEMVLQDKNRLISL GQHKNHRRRLVIVDRRWVGAKTRATQTGKDLSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSSGGTKVQLQIKR LDDPELVNEKSNDYRNGGATWNFPNGTTGLVKFRFVVGDGEQADDSGLQVSLDRLFNACDSTTKDYALFTFP PIRLKPAPHLLGMKKVPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFISTGSQPDAG ILLDTVNARVK (signal sequence underlined)			
	SEQ	ID	NO:	10
5	QQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVS RDNGKTWE MKAAPAGYPPNYGGIPK ESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFNKG RRVIVPFHNMGGGTKFHISDDGGLTWHVSRNGVTSPRHEARPPHQGVRFNNAVEATVLEMKDGTI WALARTS DQAWQAFSKDYGETWSKEPEPSRFFGTLMTLGRLLDDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSHH IAMSGDEGKTYWGFRERII DEHRNHPGYATLDGPEDRGKHQSEMVLQDKNRLISLQGHKNHRRRLVIVDRRWVG AKTRATQTGKDLSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSSGGTKVQLQIKRLLDDPELVNEKSNDYRNG GATWNFPNGTTGLVKFRFVVGDGEQADDSGLQVSLDRLFNACDSTTKDYALFTFP PIRLKPAPHLLGMKKVPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFISTGSQPDAG ILLDTVNARVK			
	SEQ	ID	NO:	11
10	MQQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVS RDNGKTWE MKAAPAGYPPNYGGIPK ESPAIVRNPLTREFIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFNKG RRVIVPFHNMGGGTKFHISDDGGLTWHVSRNGVTSPRHEARPPHQGVRFNNAVEATVLEMKDGTI WALARTS DQAWQAFSKDYGETWSKEPEPSRFFGTLMTLGRLLDDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSHH IAMSGDEGKTYWGFRERII DEHRNHPGYATLDGPEDRGKHQSEMVLQDKNRLISLQGHKNHRRRLVIVDRRWVG AKTRATQTGKDLSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSSGGTKVQLQIKRLLDDPELVNEKSNDYRNG GATWNFPNGTTGLVKFRFVVGDGEQADDSGLQVSLDRLFNACDSTTKDYALFTFP PIRLKPAPHLLGMKKVPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFISTGSQPDAG ILLDTVNARVK HHHH			
	SEQ	ID	NO:	12
15	MTWLLCGRGKWNKVKRMMNSVFCKLMSAVCAVALPAGFQEKEKTFPTDRAVTVFSAGEGNPYASIRIPALLSI GKGQLLAFAEGRYKNTDQGENDIIMS VSKNGGKTSR PRAIAKAHGATFNNPCPVYDAKTRTVVFCRQYPAGV KERQPNIPDGWDEKCIRNFMIQSRNGGSSWTKPQEITKTTKRPSPGVDIMASGPNAQTQLSGAHKGRLVIP MNEGPGFWKVISCIYSSDDGKSWKLQPTANMGMVNETSIAETDNGGVMVARHWGAGNCRRIAWSQDG GETWQVEDAPELFCDS TQNSLMTYSLSDQPAYGGKSRLIFSGPSAGRR IKGQVAMS YDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEE NQEHIKKLKFVPITMEWLTDGEDTGLAPGKKAPVLK (signal sequence underlined)			
	SEQ	ID	NO:	13
15	QEEKTGFPTDRAVTVFSAGEGNPYASIRIPALLSIGKGQLLAFAEGRYKNTDQGENDIIMS VSKNGGKTSR PRAIAKAHGATFNNPCPVYDAKTRTVVFCRQYPAGVKERQPNIPDGWDEKCIRNFMIQSRNGGSSWTKPQE ITKTTKRPSPGVDIMASGPNAQTQLSGAHKGRLVIPMNEGPGFWKVISCIYSSDDGKSWKLQPTANMGMV NETSIAETDNGGVMVARHWGAGNCRRIAWSQDG GETWQVEDAPELFCDS TQNSLMTYSLSDQPAYGGKSRLIFSGPSAGRR IKGQVAMS YDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEE NQEHIKKLKFVPITMEWLTDGEDTGLAPGKKAPVLK			
	SEQ	ID	NO:	14
	MQE EKTGFPTDRAVTVFSAGEGNPYASIRIPALLSIGKGQLLAFAEGRYKNTDQGENDIIMS VSKNGGKTSR PRAIAKAHGATFNNPCPVYDAKTRTVVFCRQYPAGVKERQPNIPDGWDEKCIRNFMIQSRNGGSSWTKPQE ITKTTKRPSPGVDIMASGPNAQTQLSGAHKGRLVIPMNEGPGFWKVISCIYSSDDGKSWKLQPTANMGMV NETSIAETDNGGVMVARHWGAGNCRRIAWSQDG GETWQVEDAPELFCDS TQNSLMTYSLSDQPAYGGKSRLIFSGPSAGRR IKGQVAMS YDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEE NQEHIKKLKFVPITMEWLTDGEDTGLAPGKKAPVLKGSGLEHHHHHH			

SEQ ID NO: 15 - O-glicosidasa de S. oralis
MDKRFKEKRCKFSIRKFTLGVASVMIGATFFAASPVLADQARVGSTDNLPSLEADLDKKASDEGHDFDKEAAA
QNPGSAETTEGPQTEEELLAQEKEKSEKPSNLPKELEDKLEKAEDNGREVDKDQLAQDTGKLVPEDVAKTTNG
ELNYGATVKIKTPSGEGSGIVVAKDLVLTSHNFIDKSQEGNIRKVVNDQGDGDIYSISYPGLPDVKFSKKD
IIHWDRGYLGKFKNDLALVRLRTVLENTPVEVTKKPVVKIGDKLHVFGYPEGKLNPIVNTTVDAEPYGE
VQGIGYQGGKPGASGGIFDTEGKLGVHQNGVVGKRSGGILFSPAPQLKWIQDHMQGISSVKPADLEEKEKPA
EEKPKEDKPAAKPTPKAVTPEWQTVANKEQQGTVTIREEKGVRYNQLSSTAQNNDGKPALFEQKQGLTVDA
NGNATVDTLFKDDSEKGSKSRGFLKFKDTKNNVFVGYDQGGWFWEYKTPGNSTWYKGNRVAAPEPGSVNRLS
ITLKSQDGQLNASNNNDVLFDTVTLPGAVNENIKNNEKKILLKAGTYSNDRTVSVKTDNQEGVKADDTPAQET
GPAVDDSKVTVYDTIQSKVLUKAVIDQAFPRVKEYTLNGHTLPGQVQOFNQVFINNHRITPEVTVKINETTAEY
LMKLRDDAHLLINAEMTVRLQVVDNQLHFDVTKIVHNQVTPGQKIDDERKLSTISFLGNALVSNSDQAGAK
FDGATMSNNTHVSGDDHDVTPNPKDLMQYGMVQFVSTDKLAACGVWSNSQNSYGGGSNDWTRLTAYKETVGN
NYVGIHSSEWQWEKAYKGIVFPEYTKELPSAKVITEDANADNVWQDGAIAYRSIMNNPQGWKEVKDITAY
RIAMNFGSQAQNPFLMTLDGKIKINLHTDGLGQGVLLKGYGSSEGHDGSHLNYADIGKRIGGVEDFKTLIEKAK
KYGAHLGIHVNASETYPESKYFNENILRKNPDSGSYSYGNWLDQGINIDAAYDLAHGRLARWEDLKKKLGEGL
DFIYVDWGNQGSGDNGAWATHVIAKEINKQGWRFAIEWGHGGEYDSTFQHWAADLTYYGGYTNKGINSAITRF
IRNHQKDWSVGDYRSYGGANYPLGGYSMKDFEGWQGRSDYNGYVTNLFADHDMTKYFQHFTVSKWENGTPV
TMTDNGSTYKWTPEMKVELVDAAGNKVVTRKSNDVNSPQYRERTVTLNGRVIQDG SAYLTPWNWDANGKKLP
TEKEKMYFNTQAGATTWTLPSDWANSKVLYKLTDQGKTEEQELTVTDGKITLDDLLANQPYVLYRSKQTNP
MSWSEGMIYDQGFNSGTLKHWTISGDAKAEIVKSQGANEMLRIQGNKSKVSLTQKLTGLKPNTKYAVYVG
DNRNSAKASITVNTGEKEVTTYTNAKSLALNYIKAYAHNNRRENATVDDTSYFQNMYAFFTTGSDVSNVTLTLS
REAGDEATYFDEIRTEENNNSMMYGDKHDGTQGQTFKQDFENVAQGIFPFVVGVEGVEDNRTHLSEKHDPTYQR
GWNGKKVDDVIEGNWSLTKNTLIVSRRNLVYQTIPQNFRFEACKTYRVTFEYEAEGSDNTYAFVVGKGEFQSGRR
GTQASNLMEHLPNTWTDSDKAKKVTFLVTGAETGDTWVGIVYSTGNASNTRGDAGGNANFRGYNDFMMDNLQI
EEITLTGKMLTENALKNLYLPTVAMNTYKESMDALKEAVFNLSQLQADDISVEEARAETAKIEALKNALVQKKT
ALVADEFESLDAPAQPQGEGLNAFDGNVSLWHTSWNGGDVGPATMVLKEPIETGLRYVPRASDSNGNLRD
VKLVVTDESGKEHTFNVTDWPNNPKDIDFGKTIKAKKIVLTTGKTYGDGGDKYQSAEELIFTRPQVAETPL
DLSGYEALAKAQKLTDKDNQEEVASVQASMKYATDNHLLTERMVAYFADYLNLQKDSATKPDAPTSSKGEEQ
PPVLDVPEFKGGVNAMEAAVHEVPEFKGGVNAVQALVHELPEYKGGANAVLAAANEVPEYKGGANAVEALVNE
KPAYTGVLATAGDQAAPTVEKPEYPLTPSPVADTKTPGAKDEEKLPATGEHSSEVALFLASVSI
ALSAAVLAT
KRKEEGSGLEHHHHHH

5 SEQ ID NO: 16 - E206A _cebador directoATGGCGCACGC GCTGGGCCACG

SEQ ID NO: 17 - E206A_cebador inversoGCCACCGTAC CATTTCGTC

SEQ ID NO: 18 - EPO
 APPRLICDSRVLERYLLEAKEAEDITTGCAEHCSLDENITVPDTKVDFYAWKRMEVGQQAVEVWQGLALLSEA
 VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLLRALGAQKEAISPPDAASAAPLRTITADTFRKLFRVYS
 NFLRGKLKLYTGEACRTGDR (Notas:
 el O-glicano con serina está subrayado; la arginina C terminal suele truncarse durante la expresión)

SEQ EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRSELLLYLQQFYGKEMQRHGYGARSFGLDIKSP
ID GRVNIIIEYKAKNPAAHYPYENGGWKAQAELEDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPCCVFPYGMG
NO: RNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAALGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSPT
20
FLTPASCALLDACEVSVTSPSQFYEGKPEVEVGDVIAISFKGDQIILVSGNYKSPQTVKALNVYIQDPPYAVNQ
DYDAVSFSRLLGKSKGFSMKIDKKELEGNNNEFRISLMFI LANGLHMQKHFTFHWDALQDYRDGSKS

SEQ	ID	NO:
MEVTVPDALKDRIALKKTARQINIVYFLGSDTEPVPDYERRLSELLLYLQQFYKGEMQRHGYGARSFGLDIKS PGRVNIIYEYAKKNPAAHYPYENGGWKAQELDEFFKAHPDRKKSQHTLIMPTWNDEKNGPDNPAGVPFYGM GRNCFALDYPAFDFIKHLGQKTRREGLLTKWYGGMAAALGHGLNLPHNHQTASDGKKYGTALMGSGNYTFGTSP TFLTPASCALLDACEVFSVTPSQFYEGKPEVEVDVAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVN QDYDAVSFSRRLGKSGKFSMKIDKKELEGNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKSGSG HHHHHH		21

SEQ	ID	NO:	22
	ATGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCCTGAAGAAAACCGCTCGTCAGCTGAATATCG TCTACTTCCTGGTTCTGATACCGAACCGGTTCCGACTACGAGCCCGTCTGAGCAGCTGCTGTTGATCT GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCCACGCAGCTTGGCTGGACATTAAGTCA CCGGGTCGTGAACATTATCGAGTACAAGGAAAGAACCCGGCAGCGCATTACCGTATGAGAATGGTGGCG GCTGGAAAGCTGCACAAGAACTGGACGAATTTCAGGGCCATCCAGACCGCAAGAAAAGCCAGCACACCCT GATCATCATGCTACCTGAAATGATGAGAAAAATGGCCTGACAATCCGGTGGCGTCCGTTCTATGGTATG GGTCGTAATTGTTGCTGGACTACCCGGCTTGATATCAAGCACCTGGCTAGAAAACCGCTGAGGGTC GTCTGCTGACGAAATGGTACGGTGGATGGCGCCGCTGGGCACGGCTGAATCTGCCGACAATCACCA GACCGCGAGCGATGGCAAGAAATATGGCACCGCCCTGATGGTAGCGGCAACTACACGTTGGTACCGCCCG ACGTTCTGACCCCCGGCAGCTGCGCTGCTGGATGCGCAACTGTTACCCGAGCCAACAGT TTTATGAGGGTAAGCCAGAAGTCGAGGTTGGTGAATGTTGCAATTCTCTCAAGGGTGTCAAATCTTGGTCAG CGGTAACTACAAGAGCCCGCAACCGTGAAGCTCTGAACGTTACATTCAAGGATCCGGCTACGCCGTGAAC CAAGACTACATGAGCTGAGCTTGTAGCGCTGCTGGCAAAAGTCCGTAAGTTAGCATGAAGATTGACA AAAAAGAACTGGAAGGCCTGAATAACAAGAATTCCGTATTCCTGATGTTCATCTGGCAAACGGCTTACA CATGAGAAGCACTTACGTTACTGGATGCGCTGCAAGACTACCGTAGCAAAATCTGGTICGGGT CATCATCACCACCATCACTGA		
5	SEQ ID NO: 23GMAHALGHGL (motivo de metaloproteasa alterado)		
	SEQ ID NO: 24GMAELGHGL (motivo de metaloproteasa alterado)		
	SEQ ID NO: 25GMAALGHGL (motivo de metaloproteasa alterado)		
10	SEQ ID 26Pseudomonas aeruginosa PAO1 (secuencia nativa con secuencia de señal eliminada) ATQEEILDAALVSGDSSQLTDHSHLVALRLQQQVERIRQTRTQLLDGLYQNLSQLQAYDPGAASMWVLPA PFLIGDKGRVLASLSLEAGGRGLAYGTVNLTQLSGTNAAHAPLLKRAVQWLNVNGDPAATAKDFKVSVGV DKTAALNGLKSAGLQPADAACNALTDASCSTSLLVNGASAASLSATVRARLQAGLPILFVHTNGWNQSSTG QQILAGLGLQEGPYGGNYWDKDRVPSSRTRTRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTS YVGRITCDDVPGLSDFSFRDVLDKALDAYNQKAQNLFALPGTTSLRLWLWADAVRQNIRYPMDKA ADTARFQETFVADAIIVYVREAGAAQKELGSYAGQRQOSMPVSGSEETILTLPSAQGFT AIGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRWLNTQYDPRFLKSPDIIKQLQANQVALVSPY CGLLQLVYSGATPGQT VTVKTGAASQPFLDIQPGEDSSQAIA DFIQLADAKDADKADLEIRSGSVEVHAKVEKVR GSIIDKDYGGDVQRFIRELNEVFIDAYTL AGFAIPNQAKTPAIQQECAARGWCDSETIHKLPGTQHINVDQY AQC GGGCSGNPYDQTWGLNPRGWGE SHEL GHNLQVNRLKVYGRSGEISNQIFPLHKDWRVLR FEGQNLDDTRV NAYRNAYN LIVAGRAEADPLAGVYKRLWE DPGTYALNGERMAFY TQWVHYWADLKDPLQGDW DI VLLYLHQRQVDKSDW DANKAALGYGTYA QRP GN GDA SSTDGNDNLL LGLSWLTQRDQRPTF ALWGIRTSAAA QVAAYGFAEQ PAFFY ANNRT NEY STV KLL MSQGS PAWPFP	Motivo	
	de metaloproteasa subrayado.		
	SEQID 27Bacteroides thetaiotaomicron VPI-5482 (secuencia nativa con secuencia de señal eliminada) DKWEKEFRIRSYEPYSNIAEWADKLMKKYSDLNPTGISVKAGDIIIVLVGDTYQQNISMQCIWETGTEYKQ TASSGDVYMLNPVGVNLTMKGEQLFVMYNTELTNSATAKPIKIHIPLGSGTVNGFFDLKEHKTDEKYAE LLKK STHKYFCIRGEKIMFYFHRNKLL EYV P NNILSAIHLWDNIVGWQ QELMGIDDVRPSQVNNHLFA ISPEGSYMW ASDYQIGFVYT Y LGN N I LLED NV M A E D N A W G P A E I GH V H Q A A I N W A S T E S S N N I L F S N F I I Y K L G K Y K S R G N GL G S V A T A R Y A N Q A W Y N M G D A T H Q N E D T E T H M R M N W L I Y Y H R C E Y K T D F W Q T L F K L M R V N M T E G E D P G K K Q L E F A K M A K Q N L T D F E M W G F F E P V N T T I E Q Y G T Y K Y V S D A M I R E A K Y M A Q F P A K H A F Q Y I E D R K K S E F P S N D Y R Y S A V G D V G Y Y T Q F K E N Q K I T K A I T A E L A G R K V S I Q N G D E A V A F E L R E N D E N G K L L Y F S T F T F E I P S S I L M V N A K L Y A V Q A D G K R I L L E I P T P N I F A V S S N K S L E F V Q A T Y A L D W Y K K N N K T P K Y T A E Q F W D Q Y I A D A M G F W G F D N S K D V N S D F N F R I M P M V K N L S G G A F M N A G N G V I G R P G N Q D A I L A A N K G W G V A H E L G H N F D T G G R T I V E T V N N M M P L F F E S K Y K T K T R I T D Q N I W E N N T P K V G L L D Y S N N E L Y N K A D S T H I A Q L A P L W Q L Y D N T F Y G K F E R Q F R E R D F G N K R E D I Y K S W V V A A S D A M E L D L T E F F A R H G I R V D K V K E D L A K Y P K P D K I Y Y L N D L A M N Y K G D G F T E N A K V S V S T S G S G N N I K L S F S V D D E N K D N I L G Y E I R R D G K Y V G F T S N D F V D T K S N L D E G V Y V V T P Y D R K L N T L N P I E V N	Motivo	
15	de metaloproteasa subrayado.		
	SEQ ID NO: 28Clostridium perfringens (secuencia nativa con secuencia de señal eliminada) VLELEMRGDSI SEAKKRKVWNFQDWQITGLSARADKITVYVDVAEGDPTTLLYKQSLTQHGGATSFQLKPG KNEITIPEINYESNGI PKDVIQGGDIFTNYKSDSQKRAPKVRTEGASKYPVFILGKSDENEV MKELEAYVEK IKAEPKTTPNIA FAVSSNKSLE FVQATYALD WYKKNNK TPK YTA F Q W D Q Y I A D A M G F W G F D N S K D V N S D F N F R I M P M V K N L S G G A F M N A G N G V I G R P G N Q D A I L A A N K G W G V A H E L G H N F D T G G R T I V E T V N N M M P L F F E S K Y K T K T R I T D Q N I W E N N T P K V G L L D Y S N N E L Y N K A D S T H I A Q L A P L W Q L Y D N T F Y G K F E R Q F R E R D F G N K R E D I Y K S W V V A A S D A M E L D L T E F F A R H G I R V D K V K E D L A K Y P K P D K I Y Y L N D L A M N Y K G D G F T E N A K V S V S T S G S G N N I K L S F S V D D E N K D N I L G Y E I R R D G K Y V G F T S N D F V D T K S N L D E G V Y V V T P Y D R K L N T L N P I E V N	Motivo	
20	de metaloproteasa subrayado.		
	SED ID NO: 29Pseudomonas aeruginosa PAO1 (N-term Met, C-term enlace/etiqueta) MATQEEILDAALVSGDSSQLTDHSHLVALRLQQQVERIRQTRTQLLDGLYQNLSQLQAYDPGAASMWVL PANPDNTLPFLIGDKGRVLASLSLEAGGRGLAYGTVNLTQLSGTNAAHAPLLKRAVQWLNVNGDPAATA KDFKVSVGV		

ES 2 986 910 T3

KTAALNGLKSAGLQPADAACNLTDASCASTSKLLVLGNAGASAASLSATVRARLQAGLPILFVHTNGWNQSST
GQQILAGLQLQECPYGGNYWDKDRVPSSRTRSVELGGAYQGDPAVQQIVDGSWRDYDWSKCTSIVGRRT
CDDVPGLSDFSKRVDVLKGALDAYNQKAQNLFALPGTTSRLWLWADAVRNIRYPMDKAADTARFQETFVA
DAIVGYVREAGAQKELGSYAGQRQQSMPVSGSEETLTTLPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLA
VGLNTQRTGSTRILWNTFRQYDRPRFLKSPDIKLQANQSVALSPYGGLLQJVYSGATPGQTVTVKVTGAASQPF
LDIOPGDESSQAIADEFIQALDADKADWLEIRSGBVEHAKVEKVRGSIKDYGGDVQRFIRELNEVFIDDAYT
LAGFAIPNQAKTPAIQQECAARGWCDSETLHKLPGTQHINVQYACQCGGGCSGNPYDQTWGLNPRGWGESHE
LGHNLQVNRLKVGYYGRSGEISNQIFPLHKDWRLVREFGQNLDDTRVNRYRNAYNLIVAGRAEADPLAGVYKRLW
EDPFGTYALINGERMAFYTOVWVHYWADLKNDPLQGWDITWLLYLHQRVQDKSDWDANKAALGYGTYAQRPGNSGD
ASSTDGNDNLLLQLSWLWTQDQRPTFALWGIRTSAAAQAQVAAYGFABQPAFFYANNRTNEYSTVKLLDMSQG
SPAWPFPGSGHHHHHHHH

Motivo

de metaloproteasa subrayado.

SEQ ID NO: 30Bacteroides thetaiotaomicron VPI-5482 (N-term Met, C-term enlace/etiqueta)
 MDKWEKEFRIRSYEPYSNIAEWADKLMKTYKSYSDLNDPTGISVKAGDDIVLVGDTYQONISMQCWIETGTEYK
 QTASSGDVYMLNPGVNKLTMKGEGQLFVMYNTELTSNTAKPIKIHIPLGSVTNGFDFLDKEHKTDEKYAELLK
 KSTHKYFCIRGEKIMFYFHRNLLEYVPNNILSAIHLWDNIVGWQQELMGIDDVRPSQVNHHLFAlSPEGSYM
 WASDYQIGFVYTLYGLNILLDEDNMAEDNAWGAPEAHEIGHVHQAAIWASSTESSNNLFSNFIIYKLGKYKSRG
 NGLGSVATARYANGQAWYNMGDATHQNEDTETHMRMNWQWLWYHRCEYKTDWFQTLFKLMLREVNMTGEDEPG
 KKQLEFKAMASKAANQNLTDFFEMWGFFEPVNTEQYGTYKYYVS DAMIREAKEYMAQFPAPKHAFOQYIEDR
 KKSEFSPSDNYRSAVGDVGYYTQFKENQKITAKITAEELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTT
 EFLPSSILMVNKAQYAOADGKRILLGCSHHHHHHH Motivo

Motivo

de metaloproteasa subrayado

Motivo

de metaloproteasa subrayado.

SEQ ID NO: 32 Pseudomonas aeruginosa PAO1 (secuencia nativa completa, incluida la secuencia de **MSLSTTAFPSIQLGENMSRSPIPRHALLAGFCLAGALSQA**ATQEEILDAALVSGDSSQLTDSDLVALRLQQQVERIRQTRTQLLDGLYQNLISQAYDPGAASMWLPANPDNTLPFLIGDKGRVLASLSLEAGGRLAYGTNVLTQLSGTNAAHAPLLKRAVQWLNVNGDPGAATAKDFKVSVGVVDKTAALNGLKSAGLQPADAACNALTDASCASTSKLLVLGNNGASAASLSATVVRARLQAGLPILFVHTNGWNQSSTGQQILAGLGLQEGPYGGNYWDKDRVPSSRTTRTSVLEGGAYGQDPALVQQJLVDGWSRTDYDWSKCTSIVGRTTCDDVPGLSDFSFRVDSLKGALDAYNQKAQNLFALPGTTSLRLWLLWADAVRQNIRYPMDKAADTARFOETFVADAIVGYVREAGAAQKELGSYAGQRQOSMPVSGSEETTLTLPSPAQGFTAIGRMAAPGKRLSIRTEDAQASLAVGLNTQREGTRSTRLWNTRQYDRPRFLKSPDIKLQANQSVALVSPYGGILLQIVYSGATPGQTVTVKVTGAASQPFLDIOPQGEDDSQIAIDFIQALDADKADWLEIRSGSVEVHAKVEVKVRGSIDKDYGGDVQRFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQCEAARGWDCDSETLHKLPGTQHINVQYAQCGGGCSGNPDTWGLNPRGWGESELGHNLQVNRLKVYGGRSGEISNQIFPLHKDWVRVLRFGQNLLDDTRVNYRNAYNLLIVAGRAEADPLAGVYKRLWEDPGTYALNGERMAFYTQWVHYWADLKNDPLQGWDIWTLLYLHQRVQDKSDWDANKAALGYGTYAQRPGNNSGDASSTDGNDNLLGLSWLTQRDQRPFTALWGIRTseñal) SAAAQAVAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGS PAWPFP

Acceso Uniprot: Q9I5W4.1

Motivo de metaloproteasa subrayado.

Secuencia de señales en negrita, subrayada.

SEQ ID NO: 33Bacteroides thetaiotaomicron VPI-5482 (secuencia nativa completa, incluidas la señal y otras secuencias))

MTIKRIFTNLLALFTLFTVSLACKDETEKSIINSSFSISEEYLIQNLDSSTSVOIPIINTSMELAQWSVSYEAN
WLQCSKQKTAEGTFLRITVNVNENTGETKRTANIKVTSTTATYTTITVNQYAKEGEVIVEGDIKVTPGGKASEHQ
EGQDIENTYDGKFSTDGAAPFHTPWQSAKFPTVLEYYFKGDTEIDYLIVYTRSGNGNFGKVKVYTTTNPDRS
DYTLLQGEYDFKEQNAPSFKVSEGIKATGIKFEVLISGLGDFVSCDEMZFYKTNTDKTLKQLLTVFTDITCTE
TKNNVTNEQIQALPDYFVRIAEAVRDNTYDKWEKEFRIRSYEPYSNTAERADKLMRKYSLDNPTGTGKAG
DDIIVLVGDTTYGQN1SMQCIWETGTEYKQTASSGCVYMLNPGVNKLTMKGEGLFVPMYNTELTSNTANPIKH
IPLGSSTVNGFEDLKERKTDKYEALIINKSTHKYFCIRGRKIMFYFHRNKLLLEYVPNNITLSAIHLWDNTVGWQ
QELMGIDDVVRPSQVNNHFLFAISPEGSYNNWASDYQ1GFVITYLGNIILEDNVMAAEADNAWCPAHEIGHVHQAAI
NWASSTESSNNLFSNFIIYKLGKYKSRGNGLGSVATARVANGQAWYTRNGDATHQNETEHNMRMNWQLWIYYR
RCEYKTDFWQTLFKLNREVNMTGEDEPGKKQLEPAKMASKAANQNLTDFFEMWGFFEPVNNTIEQYCTYKYYV
SDAMIREAKEYMAQFPAPKHAFCQYIEDRKKSETPSNDYRYSAVGDVGYYTCQPKENQKITKAITAELAGRKVI
QNGDEAVAFELRENDENGKLLYFSTFTFEIPSSILMVNAKLYAVQADGKRII

5 Acceso Uniprot: Q89ZX7.1

Motivo de metaloproteasa subrayado.

Secuencia de señales en negrita, subrayada.

10 Otras secuencias eliminadas en la proteína madura en negrita y cursiva.

SEQ ID NO: 34Clostridium perfringens (secuencia nativa completa, incluidas la señal y otras secuencias)
MNKRKIAIILATMITNLATTIDVLAQELNTKNNSKVEVSHDDESHQARVSKFDLYNSDQLDAYNOEQVSR
SNIKSINNNNGGKYNSSIDKAIDGNLETHWETGKPNANDFTNEVVFVTFNEITNIDRIVYSARRDSARGKGF
EFEIYASLKDEGDDFNLVSSGEYTESTRDLVEIKFNPTDFKRLKFKKADQNWasAAEFMFYKEDKLNKFN
GLFTDSSMNKVSEEFNTLEKLNafENEELDHPIYDLYKEGLNNARAILTETSENPTKATLGQITYNLNDDYNN
QYRMPYKNIKAIKNNGRHYAAQNIKEAIDNDVNTYWETGTLNSSFNNNEVEVFENDLVTLDRI
VYGSRQSDLK GFAEEVYIYASRTSKGDTYKLVATGAHEATRGLVEAKFEPTEFKRVFKFRSKQNSATLNELMFYK
PDEVYS SIPKLFTDGTMSELSEEFNSLEKINA
FEKAKNHPLYNDFNETIELAESLISNPRKD
DVIELENRGD
SI
EAK KRKVWNFQDWQITG1SAPAGDKITVYDVARGDPTPHLYFQSLTQROGATSPQLKPGKNE
IPKDVIQGGDLLEFTNNKSIQSKRAFKVRIEGASKYVFI
LGSDEMEVMEKELEYAVVEKIAEPKTTTPNIFAVS
SNKSLEFVQATYALDVKNNKTPKTYAEQWDQYIADAMCFWGFNLSKDVNSDFNFKIMPVMVKNLSSGAFMNA
GNGVIGIRPGNQDA
ILAANKGNGVAHELGHNFTDGRRTIVEV
TNNMMP1FFEESKYKTKTRITDQNTIWERNTYP
KVGLDDOYNNELYNKADSTHLAQLAPIWQIYLYDNTFYGKFERQFREDFGKRNPREDIYKSWVVAASDAMELD
LTEFFARHGIRVDDKVKE
DLAK7PKFDDK1IYIYLNDA
MINYKGDGFTENIAKVSVTSGSNCNIKLSFSV
DENK DN
ILGYEIRR.DGKVVGFTSND
FVDTKSNLIDEDGVV
VV/PYDRLNLTLP1EVNALOPTLSVNPVITIALGEE
FN
EEYIYAKDIKGNSLSESVVKVKSNNNTSKVGEYEVL
YLED
SKGN
NEY
YTKTSKVNVVSRKEYMSDLTPKQS
SNGWGTVRKDKSISGGVIGLTRDGDFVDYKGLGLHSNAEYVYDLEGKDYDYFESYVCVDKAMSSRPASSVIF
KVLVDGEEKFN
SGVMRSTTPQK
YV
KVDVKNAELKLIV
NDA
GDGDS
DHASF
GDAK
LATLSSKP
I
KGENLAY SMDEKV
DLMKGITATDIEDGNITSKVQIKSSDFVEGKSGIFTVVYSV
TDS
DGLTSEC
RTIAVTDKETQLSDL
NWKSATIGSGSVRKDRAVSGNQIRLLNEDNSVETFAKGIGTHSYSEIV
YNSEGYD
IFD
TWVGIDRH
VADKRV
SVFKFVYV
DGELKAETDVMR
IDTPKKRLLV
DVRNSRE
I
KLVVD
A
DNGNNWD
HADWADAKFRN
LA
YDASELN
KAI
EEARKL
DNNY
TEES
SEAL
RNAISK
GE
ALL
SKDK
KET
INS
AEEL
NKEM
NSL
VKVD
LN
NAV
IN
IPDRYLLK
SI
QNLNK
GT
ITLGDMYSL
TTL
LSGVED
LTGLEN
AKNLET
LNMDY
NEV
VKDLRFLSK
KKL
INTL
NAQE
QFTIA
AGE
KLPSNGK
VIGDSK
VYNREG
KNVAKT
IRVVD
KGNTILE
QDAK
DEF
INT
KDL
SSGLY
GVHVL
FEDEG
FSG
VMFYLFNV

15 Acceso Uniprot: A0A0H2YN38.1

Motivo de metaloproteasa subrayado.

20 Secuencia de señales en negrita, subrayada.

Otras secuencias eliminadas en la proteína madura en negrita y cursiva.

- SEQ ID NO: 35Pseudomonas aeruginosa PAO1 (doble mutante con secuencia de señal eliminada) ATQEEILDAALVSGDSSQLTDShLVALRLQQVERIQRTRTQLLDGLYQNLQSQAYDPGAASMWVLPAAPDNTL PFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQSLGTNAAHAPLLKRAVQWLNGDPAATAKDFKVSVGVVDK TAALNGLKSAGLQPADAACNALTDCASTSKLLVLNGASAASLSATVRARLQAGLPILFVHTNGWNQSSTG QQILAGLGLQEGPYGGNYWDKDRVPSSRTRTSVEILGGAYQDPAVQQIVDGSWRDWDWSKCTSYVGRTTC DDVPGLSDFSkrDVlKGAlDAYNQKAQNLFALPGTTSLRLWLLWADAVRQNIRYPMDKAADTARFQETFVAD AIVGYVREAGAAQKELGSYAGQRQQSMPVSGSEETLTLPQAQGFTAIGRMAAPGKRLSIRIEDAGQASLA GLNTQRIGSTRlwNTRQYDRPRFLKSPDIKLQANQVALVSPYGLLQLVYSGATPGQTVTVKVTGAASQPFL DIQPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSIDKDYGGDVRQFIRELNEVFIDAYTL AGFAIPNQAKTPAIQQECAARGWDCDSETLHKLPGTQHINVDQYAQCAGGCGSNNPQDTWGLNPRGWGES**AAL** GHNL**QVNRLK**VYGGRSGEISNQIFPLHKDWRLREFGQNLLDDTRVNRNAYNLIVAGRAEADPLAGVYKRLWE DPGTYALNGERMAFYTQWVHYWADLKNDPLQGWDIWTLLYLHQROVDKSDWDANKAALGYGTYAQRPGNSGDA SSTDGNDNLLLGLSWLTQRDQRPTFALWGIRTSAAAQQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGS SPAWPFP Motivo
de metaloproteasa alterado subrayado
- 5 SEQ ID NO: 36Bacteroides thetaiotaomicron VPI-5482 (doble mutante con señal eliminada y otras secuencias inmaduras) DKWEKEFRIRSYPEPSNIAEWADKLMKKYSDLDNPTGISVKAGDDIIVLVGDTYGQNI SMQCIWETGTEYKQ TASGDVYMLNPVGVNKLTMKGEQQLFVMYNTELTNTAKPIKIHIPLGSGTVNGFFDLKEHKTDEKYAELLKK STHKYFCIRGEKIMFYFHRNKLLEYVPNNILSAIHLWDNIVGWQOELMGIDDVRPSQVNHLFAISPEGSYM WASDYQIGFVYTLYGNILLEDNVMAAEDNAW**GPA****A**IGHVHQAAINWASSTESSNNLFSNFIIYKLGKYKSRGN GLGSVATARYANGQAWYNMAGDATHQNEDTETHMRMNWQLWIYYHRCEYKTDFWQTLFKLMREVNMTEGEDPGK KQLEFAKMASKAANQNLTDFFEMWGFFEPVNTTIEQYGTYKYYVS DAMIREAKEYMAQFPAPKHAFOYIEDRK KSEFSPSNDYRYSAVGDVGYYTQFKENQKITKAITELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTT EIPSSILMVNAKLYAVQADGKRILL Motivo
de metaloproteasa alterado subrayado
- 10 SEQ ID NO: 37Clostridium perfringens (doble mutante con señal eliminada y otras secuencias inmaduras) VLELEMRGDSI SEAKKRKVWNFQDWQITG L SARAGDKITVYVVDVAEGDPTPTLLYKQSLTQHGGATSFQLKPG KNEITIPEINYESNGTPKDVIQGGDLFFTNYKSDS**Q**KRAPKVRIEGASKYPVFI LGKSDENEVMKELEAYVEK IKAEPKTTPNIAFVSSNKSLEFVQATYALDWYKKNNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRI MPMVKNLSSGGAFMAGNGVIGIRPGNQDAI LAANKGNGVA**A****A**LGHNFDTGGRTIVEVTNNMMPLFFESKYKTK TRITDQNIWENNTYPKVGLDDYSNNELYNKADSTHIAQLAQPLWQLYLYDNTFYGKFERQFRERDFGNKNREDI YKSWVVAASDAMELDLTEFFARHGIRVDDVKEDLAKYPKPDKKIYYLNDLAMNYKGDGFTENAKVSVSTSGS NGNIKLSFSVDDENKDNLG YEIRRDGKVYGF TSNSDFVDTKS NLD EGDVYVWT PYDRKINTLNPIEVN Motivo
de metaloproteasa alterado subrayado
- 15 SEQ ID NO: 38Pseudomonas aeruginosa PAO1 (doble mutante con secuencia señal eliminada, con Met en el N-terminal, enlazador/etiqueta en el C-terminal) MATQEEILDAALVSGDSSQLTDShLVALRLQQVERIQRTRTQLLDGLYQNLQSQAYDPGAASMWVLPAAPDNTL LPFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQSLGTNAAHAPLLKRAVQWLNGDPAATAKDFKVSVGVVDK KTAALNGLKSAGLQPADAACNALTDCASTSKLLVLNGASAASLSATVRARLQAGLPILFVHTNGWNQSST GQQILAGLGLQEGPYGGNYWDKDRVPSSRTRTSVEILGGAYQDPAVQQIVDGSWRDWDWSKCTSYVGRTTC CDDVPGLSDFSkrDVlKGAlDAYNQKAQNLFALPGTTSLRLWLLWADAVRQNIRYPMDKAADTARFQETFVAD AIVGYVREAGAAQKELGSYAGQRQQSMPVSGSEETLTLPQAQGFTAIGRMAAPGKRLSIRIEDAGQASLA VGLNTQRIGSTRlwNTRQYDRPRFLKSPDIKLQANQVALVSPYGLLQLVYSGATPGQTVTVKVTGAASQPFL LDIOPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSIDKDYGGDVRQFIRELNEVFIDAYTL LAGFAIPNQAKTPAIQQECAARGWDCDSETLHKLPGTQHINVDQYAQCAGGCGSNNPQDTWGLNPRGWGES**AAL** GHNL**QVNRLK**VYGGRSGEISNQIFPLHKDWRLREFGQNLLDDTRVNRNAYNLIVAGRAEADPLAGVYKRLWE DPGTYALNGERMAFYTQWVHYWADLKNDPLQGWDIWTLLYLHQROVDKSDWDANKAALGYGTYAQRPGNSGDA ASSTDGNDNLLLGLSWLTQRDQRPTFALWGIRTSAAAQQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGS SPAWPFPGSHHHHHH Motivo
de metaloproteasa alterado subrayado
- 20 SEQ ID NO: 39Bacteroides thetaiotaomicron VPI-5482 (doble mutante con señal eliminada y otras secuencias de la proteína inmadura, con Met en el N-terminal, enlazador/etiqueta en el C-terminal) MDKWEKEFRIRSYPEPSNIAEWADKLMKKYSDLDNPTGISVKAGDDIIVLVGDTYGQNI SMQCIWETGTEYKQ QTASSGDVYMLNPVGVNKLTMKGEQQLFVMYNTELTNTAKPIKIHIPLGSGTVNGFFDLKEHKTDEKYAELLKK STHKYFCIRGEKIMFYFHRNKLLEYVPNNILSAIHLWDNIVGWQOELMGIDDVRPSQVNHLFAISPEGSYM WASDYQIGFVYTLYGNILLEDNVMAAEDNAW**GPA****A**IGHVHQAAINWASSTESSNNLFSNFIIYKLGKYKSRGN NGGSVATARYANGQAWYNMAGDATHQNEDTETHMRMNWQLWIYYHRCEYKTDFWQTLFKLMREVNMTEGEDPGK KQLEFAKMASKAANQNLTDFFEMWGFFEPVNTTIEQYGTYKYYVS DAMIREAKEYMAQFPAPKHAFOYIEDRK KSEFSPSNDYRYSAVGDVGYYTQFKENQKITKAITELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTT EIPSSILMVNAKLYAVQADGKRILLGGHHHHHH Motivo
de metaloproteasa alterado subrayado

SEQ ID NO: 40Clostridium perfringens (doble mutante con señal eliminada y otras secuencias de la proteína inmadura, con Met en el N-terminal, enlazador/etiqueta en el C-terminal)
 MVLELEMRGDSI SEAKKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQLKP

GKNEITIPEINYESNGIPKDVIQGGDLFFFNYKSDSQKRAPKVRLEGASKYPVFILGKSDENEVMKELEAYVE
 KIKAEPKTTPNIFA VSSNKSLEFVQATYALDWYKKNNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFR
 IMPMVKNLSGGAFMNAGNGVGIRPGNQDAILAANKGWGVAAALGHNFDTGGRTIVEVTNNMMPLFFESKYKT
 KTRIDQNIWENNNTPKVGLDDYSNNELYNKADSTHQLAPLWQLYLYDNTFYGKFERQFRERDFGNKNRED
 IYKSWVVAASDAMELDLTEFFARHGIRVDDKVKE DIAKYPKPDKKIYYLNLDIAMNYKGDGFTENAKVSSTSG
 SNGNIKLSFSVDDENKDNILGYEIRRDGKYVGFTSNDSFVDTKSNLDEDGVVVTPYDRKLNTLNPIEVNGSG
 HHHHHH

Motivo

de metaloproteasa alterado subrayado

- 5 SEQ ID NO: 41HELGH (motivo de metaloproteasa)
- 10 SEQ ID NO: 42HEIGH (motivo de metaloproteasa)
- 15 SEQ ID NO: 43GVAHELGHNF (motivo de metaloproteasa)
- 20 SEQ ID NO: 44HALGH (motivo de metaloproteasa alterado)
- 25 SEQ ID NO: 45AELGH (motivo de metaloproteasa alterado)
- 30 SEQ ID NO: 46AALGH (motivo de metaloproteasa alterado)
- 35 SEQ ID NO: 47GKPRPYSPRPTSHPRPIRV (péptido de glicodrosocina con sitio O-gly en la T)
- 40 SEQ ID NO: 48YIYGSFK (Péptido no O-glicosilado)
- 45 SEQ ID NO: 49KKLVFFA (Péptido no O-glicosilado)
- 50 SEQ ID NO: 50FLPLILGKLVKGLL (Péptido no O-glicosilado)

REIVINDICACIONES

1. Composición que comprende un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas que comprende:
 5 (a) una secuencia de aminoácidos de SEQ ID NO: 1;
 (b) una secuencia de aminoácidos que sea idéntica en al menos un 85% a la secuencia de aminoácidos de SEQ ID NO: 1 o
 (c) una secuencia de aminoácidos que es un fragmento de la secuencia de SEQ ID NO: 1 o un fragmento de una secuencia de aminoácidos que sea idéntica en un 85% a la secuencia de aminoácidos de SEQ ID NO: 1,
 10 en el que la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en el que Am1757 es un polipéptido de cualquiera de los SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de las SEQ ID NOs: 12 a 14.
2. La composición según la reivindicación 1, en la que el polipéptido tiene una secuencia de aminoácidos que es al menos un 85% idéntica a la secuencia de aminoácidos de la SEQ ID NO: 1, o el fragmento del mismo, comprende el motivo HEbbH, en el que b es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W, y opcionalmente en el que dicho motivo está presente en dicho polipéptido en las posiciones correspondientes a las posiciones 181 a 185 de la SEQ ID NO: 1.
 15 20 3. La composición según la reivindicación 2, en la que dicho motivo comprende la secuencia HEIGH o HELGH, preferiblemente HELGH.
4. La composición según una cualquiera de las reivindicaciones precedentes, en la que el polipéptido comprende el motivo abxHEbbHbc, en el que:
 25 (a) a es el aminoácido V, T o G;
 (b) b es un aminoácido no cargado, opcionalmente A, C, F, G, I, L, M, N, P, Q, S, T, V o W;
 (c) x es cualquier aminoácido; y
 (d) c es un aminoácido hidrófobo, opcionalmente A, C, F, I, L, M, P, V, W o Y;
 30 opcionalmente, en el que dicho motivo comprende la secuencia GMAHELGHGL o GVAHELGHNF, preferiblemente GMAHELGHGL.
5. La composición según una cualquiera de las reivindicaciones precedentes, en la que el polipéptido,
 35 (a) incluye una metionina adicional en el extremo N y/o una etiqueta His en el extremo C, etiqueta que puede estar unida al extremo C por un enlazador, opcionalmente en el que dicho polipéptido comprende o consiste en la secuencia de aminoácidos de la SEQ ID NO: 2 o
 (b) se suministra en solución, liofilizado o inmovilizado, opcionalmente en el que Am 1757 es un polipéptido consistente de la SEQ ID NO: 11 y/o en el que Am0707 es un polipéptido consistente de la SEQ ID NO: 14.
6. Un método de hidrólisis de una O-glicoproteína, en el que el método comprende poner en contacto una muestra que comprende la O-glicoproteína, con una composición según una cualquiera de las reivindicaciones 1 a 5 y de ese modo hidrolizar la O-glicoproteína, comprendiendo opcionalmente además la detección o análisis de los productos de la hidrólisis.
 40 45 50
7. Un método para evaluar el estado de glicosilación de una proteína, que comprende poner en contacto una muestra, que comprende la proteína con una composición según una cualquiera de las reivindicaciones 1 a 5 y detectar y/o analizar los productos producidos, opcionalmente en el que la presencia o ausencia de productos de escisión se utiliza para determinar la presencia o ausencia de una O-glicoproteína en la muestra, y/o en el que dicho análisis se lleva a cabo para identificar el tipo de una cadena de O-glicano y/o su posición de unión a una O-glicoproteína.
 55 60
8. El método según la reivindicación 6 o 7, en el que el análisis o la detección se lleva a cabo mediante cromatografía de afinidad, SDS-PAGE, HPLC o espectrometría de masas.
9. Una composición que comprende un polipéptido capaz de unirse a un O-glicano o a una O-glicoproteína y que carece o tiene una actividad endoproteasa reducida específica para proteínas O-glicosiladas que comprenden una secuencia de aminoácidos de la SEQ ID NO: 5 o SEQ ID NO: 20, en donde la composición comprende además una sialidasa Am1757 o una mezcla de Am1757 y Am0707, en donde Am1757 es un polipéptido de cualquiera de las SEQ ID NOs: 9 a 11 y Am0707 es un polipéptido de cualquiera de las SEQ ID NOs: 12 a 14.
 65 66 67
10. La composición según la reivindicación 9, en la que el polipéptido:
 (a) incluye una metionina adicional en el extremo N y/o una etiqueta His en el extremo C, etiqueta que puede estar unida al extremo C por un enlazador, opcionalmente en el que dicho polipéptido comprende o consiste en la secuencia de aminoácidos de la SEQ ID NO: 6 o 21; y/o
 (b) se suministra en solución, liofilizado o inmovilizado.

11. Un método de unión a un O-glicano, O-glicopéptido y/o O-glicoproteína, en el que el método comprende poner en contacto una muestra que comprende el O-glicano, O-glicopéptido y/o O-glicoproteína con una composición según la reivindicación 9 o 10, y opcionalmente determinar si se ha unido o no un O-glicano, O-glicopéptido y/o O-glicoproteína y/o separar el O-glicano y cualquier glicoproteína enlazada, el O-glicopéptido o la O-glicoproteína de la mezcla resultante; opcionalmente, en el que el método tiene por objeto aislar un O-glicano o una glicoproteína enlazada, un O-glicopéptido o una O-glicoproteína de la muestra.
- 5
12. Un método para evaluar el estado de glicosilación de una proteína, que comprende poner en contacto una muestra que comprende la proteína con una composición según una cualquiera de las reivindicaciones 9 o 10 y determinar si la proteína está unida o no al polipéptido.
- 10
13. Un método para detectar O-glicopéptidos y/o O-glicoproteínas en una muestra, en el que el método comprende:
(a) poner en contacto dicha muestra con una composición según la reivindicación 9 o 10 para permitir así la formación de un complejo entre un O-glicopéptido y/o O-glicoproteína y el polipéptido;
15
(b) opcionalmente, separar el polipéptido de la muestra contactada; y
(c) determinar si el polipéptido separado está unido a una glicoproteína o glicopéptido O-ligado, determinando así la presencia o ausencia de glicopéptidos o glicoproteínas O-ligados en la muestra.
14. El método según cualquiera de las reivindicaciones 11 a 13,
20
(a) en la que dicha determinación y/o separación se lleva a cabo mediante cromatografía de afinidad, SDS-PAGE, HPLC, transferencia de lectina, ELISA o espectrometría de masas; y/o
(b) que comprende además una etapa de elución del material ligado del polipéptido de la reivindicación 10 u 11 con un tampón que comprende:
25
(a) Urea de alta concentración molar;
(b) Detergente de alta concentración; o
(c) Un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas según cualquiera de las reivindicaciones 1 a 5.

DIBUJOS

Fig. 1

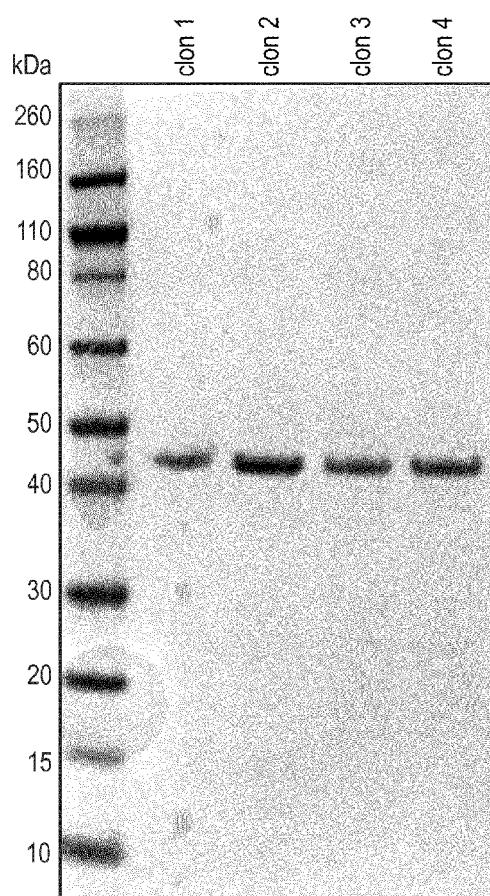


Fig. 2

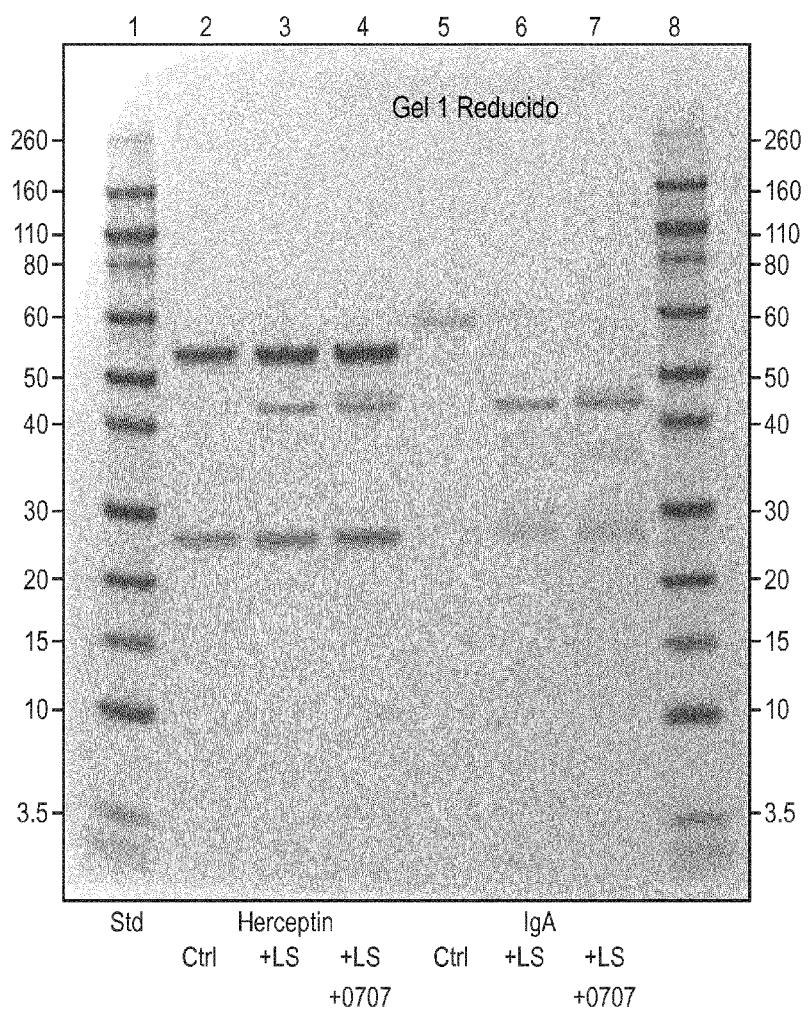


Fig. 3A

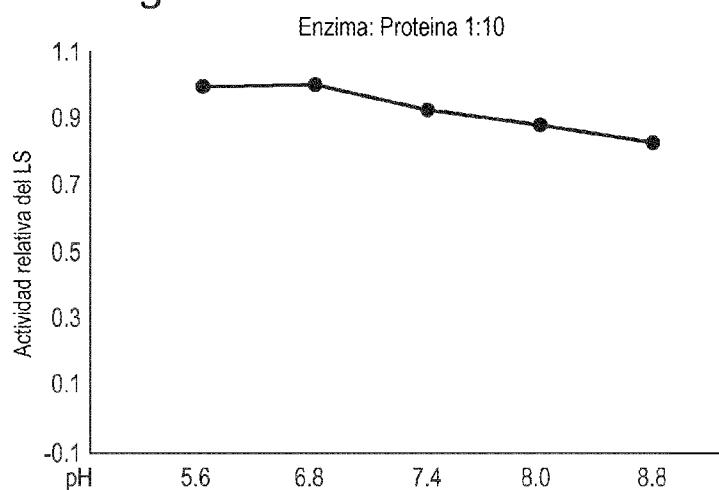


Fig. 3B

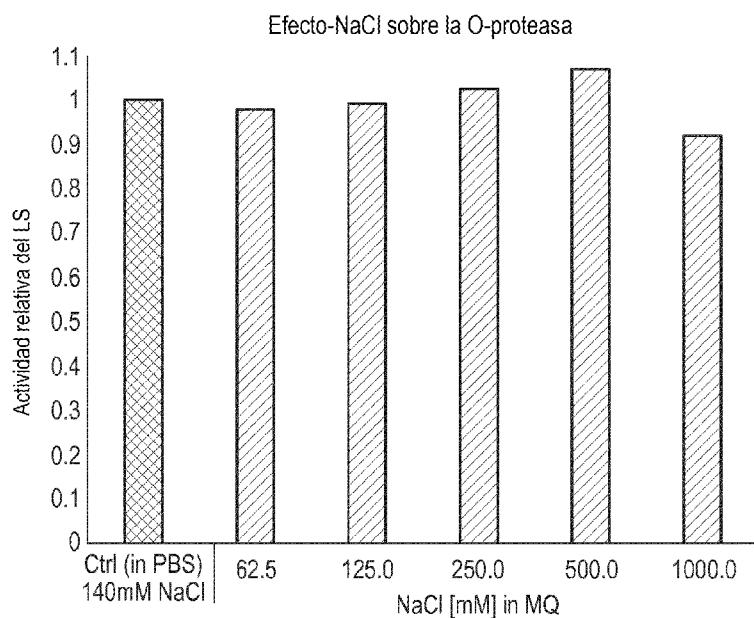


Fig. 3C

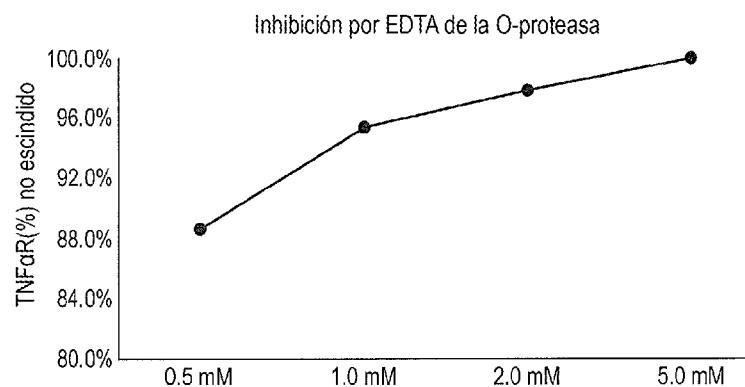


Fig. 3D

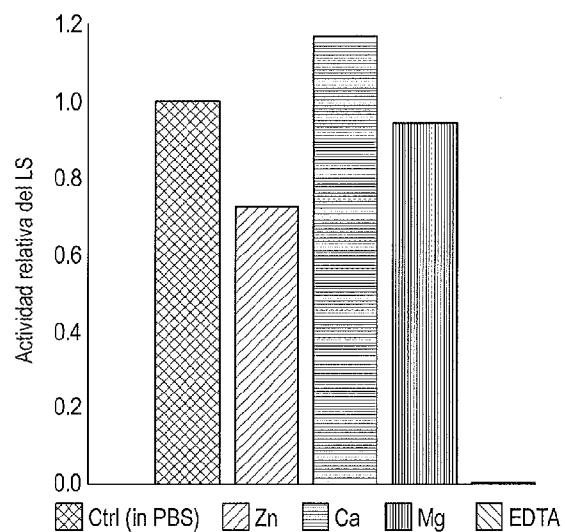


Fig. 4A

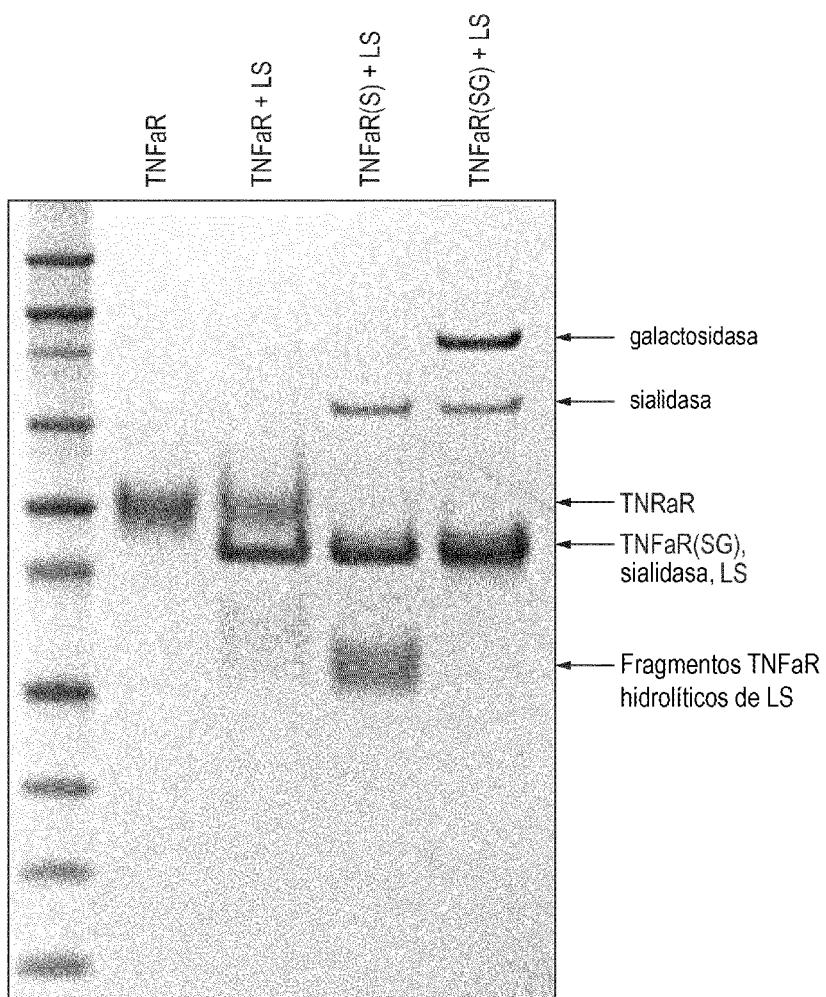


Fig. 4B

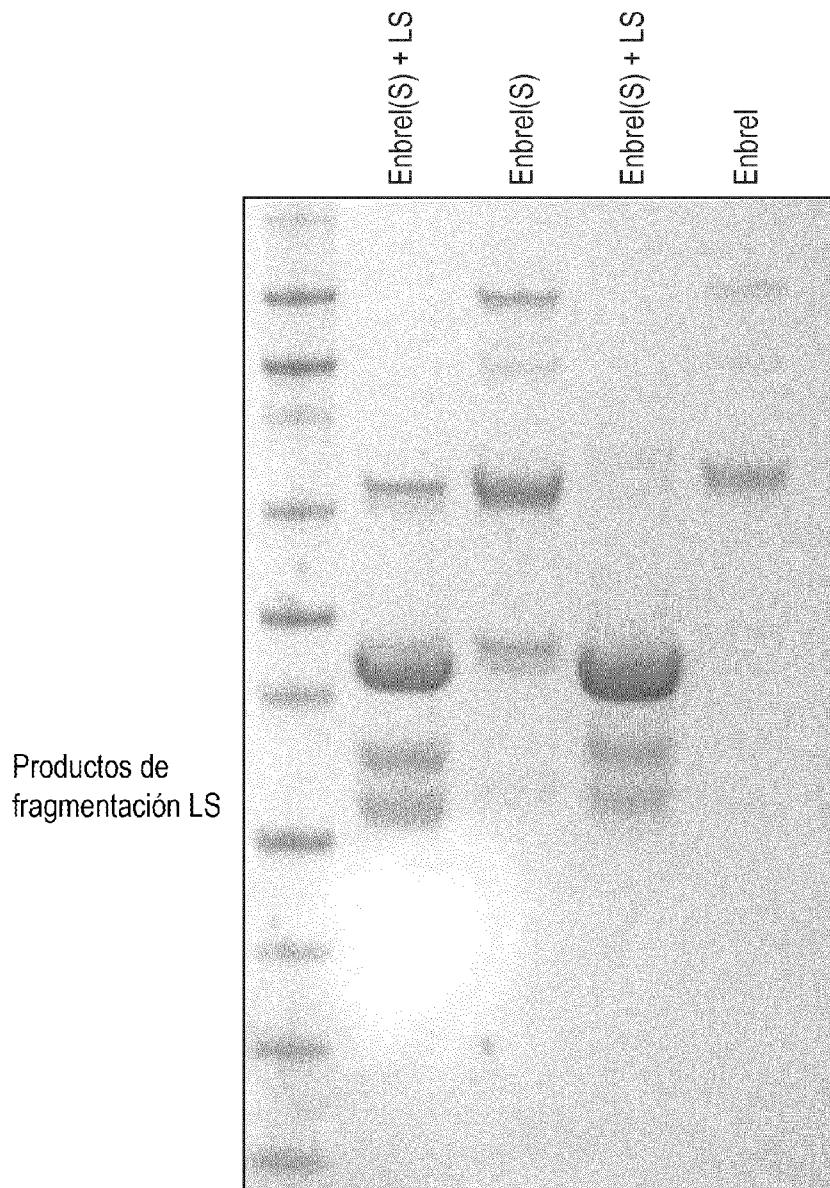


Fig. 4C

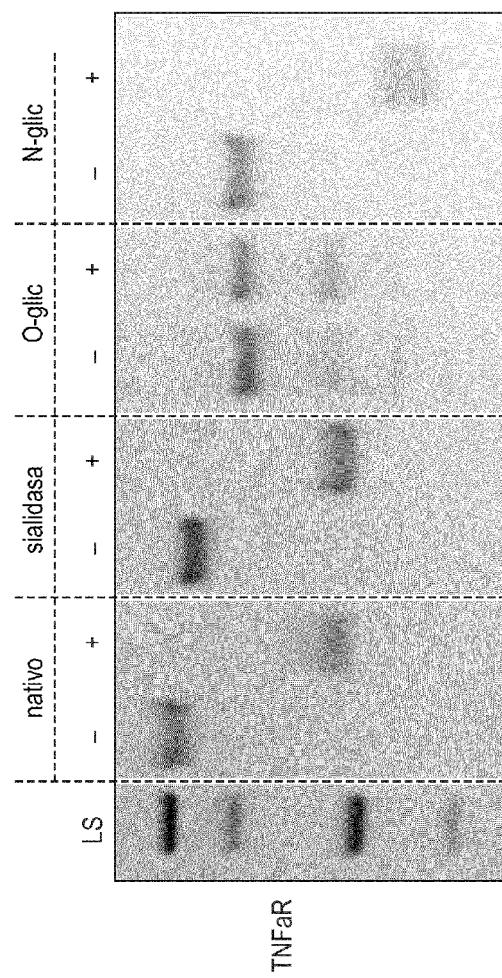


Fig. 5A

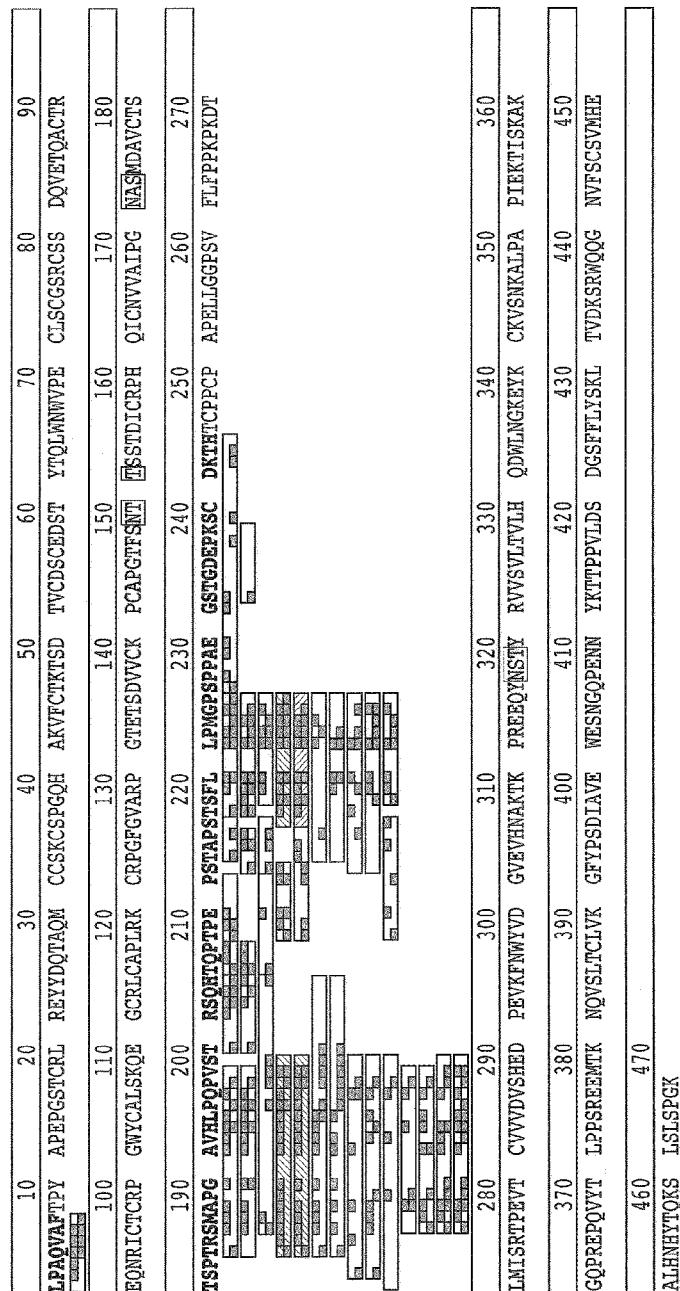


Fig. 5B

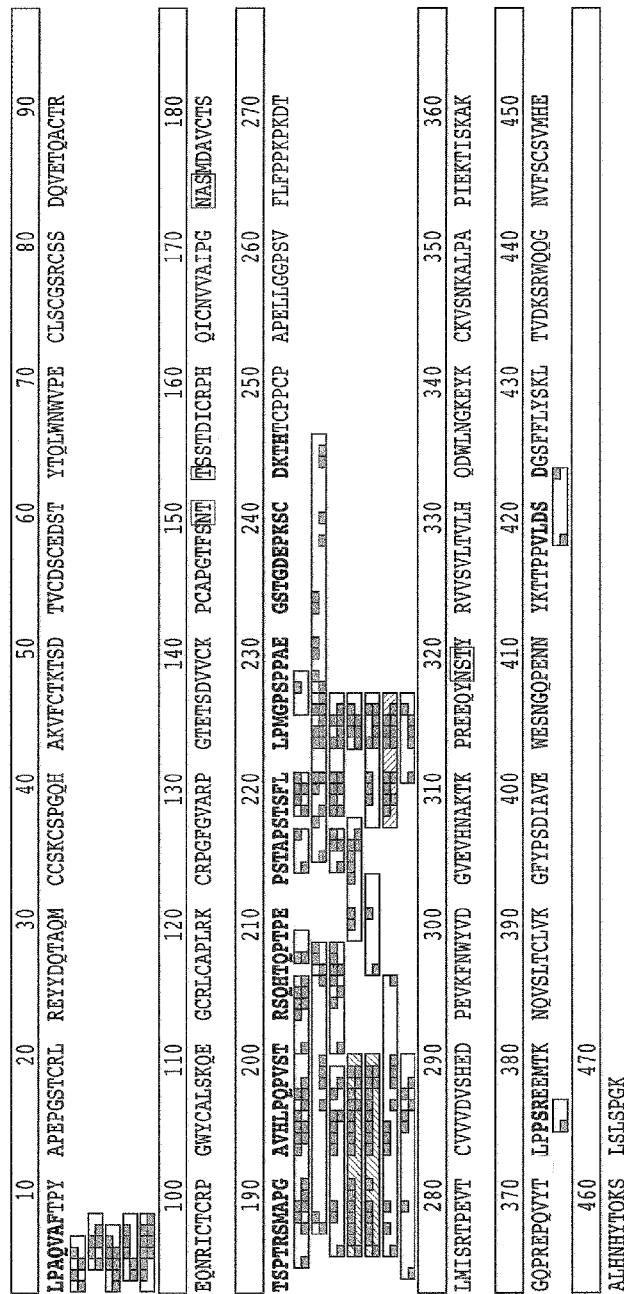


Fig. 6A

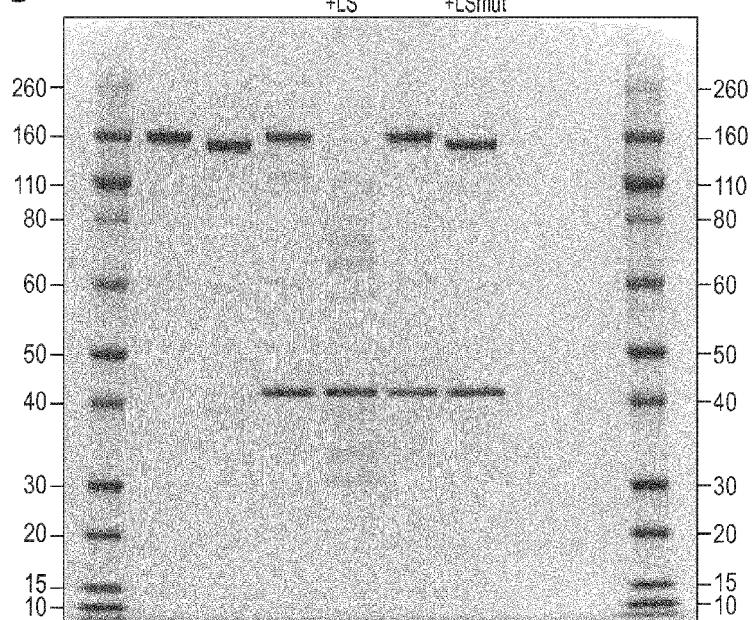


Fig. 6B

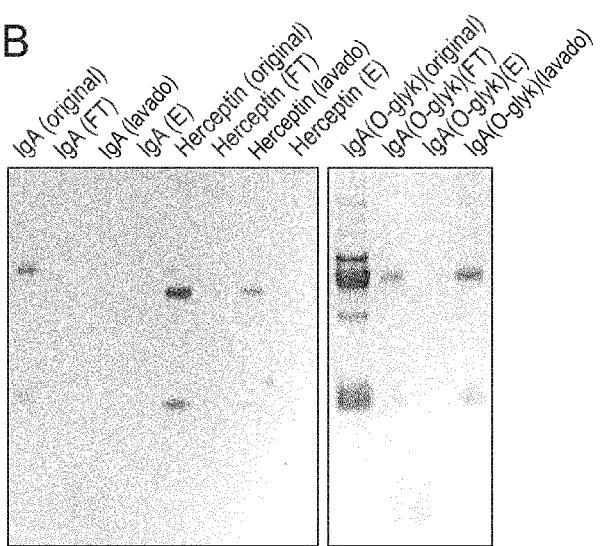


Fig. 7

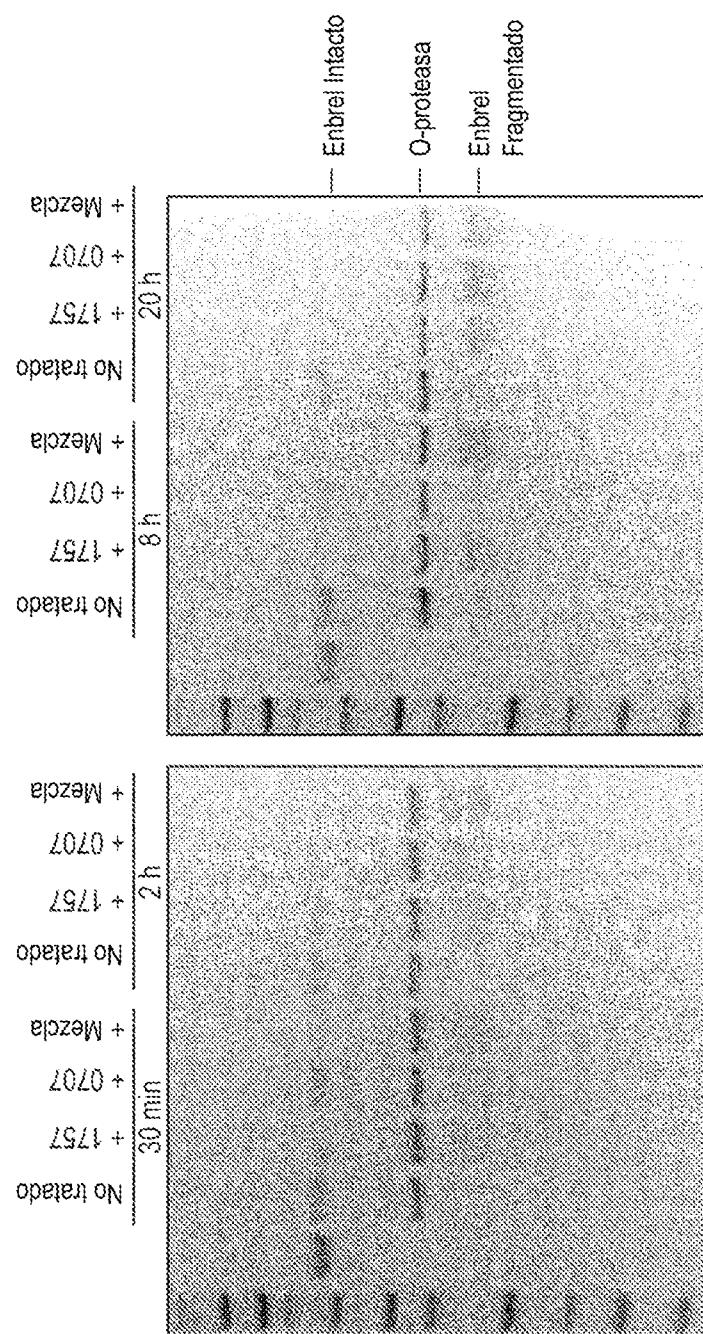


Fig. 8

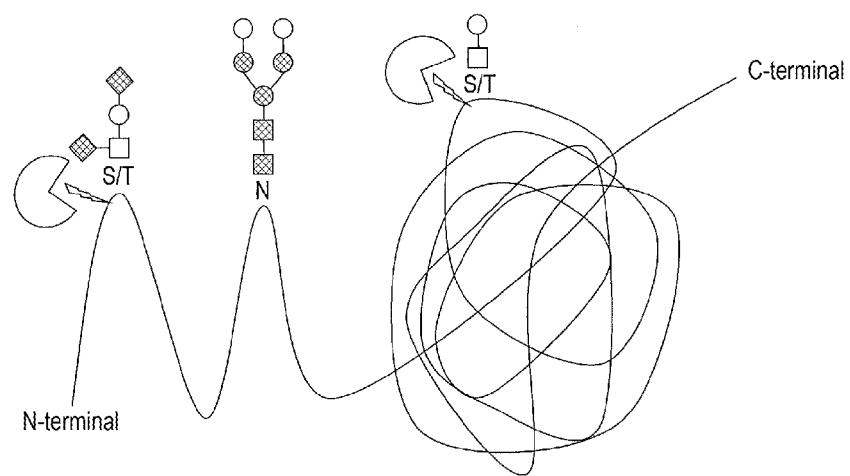


Fig. 9A

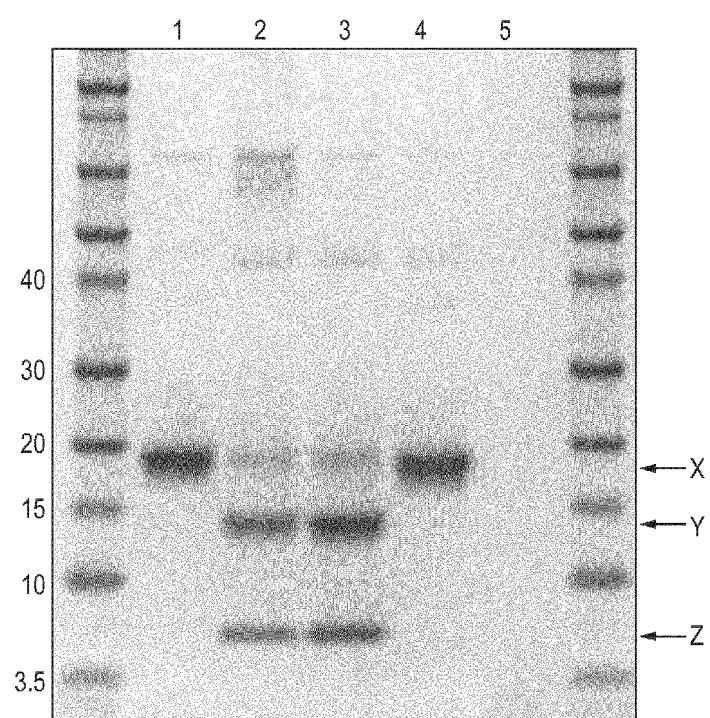


Fig. 9B

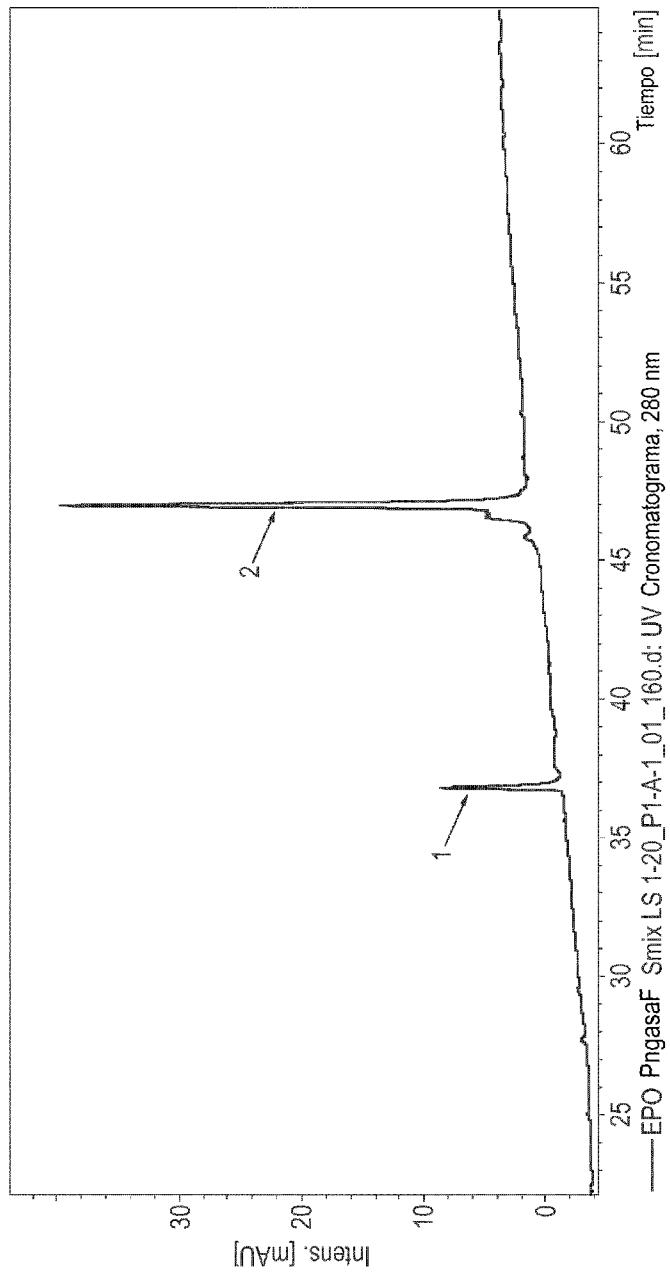


Fig. 9C

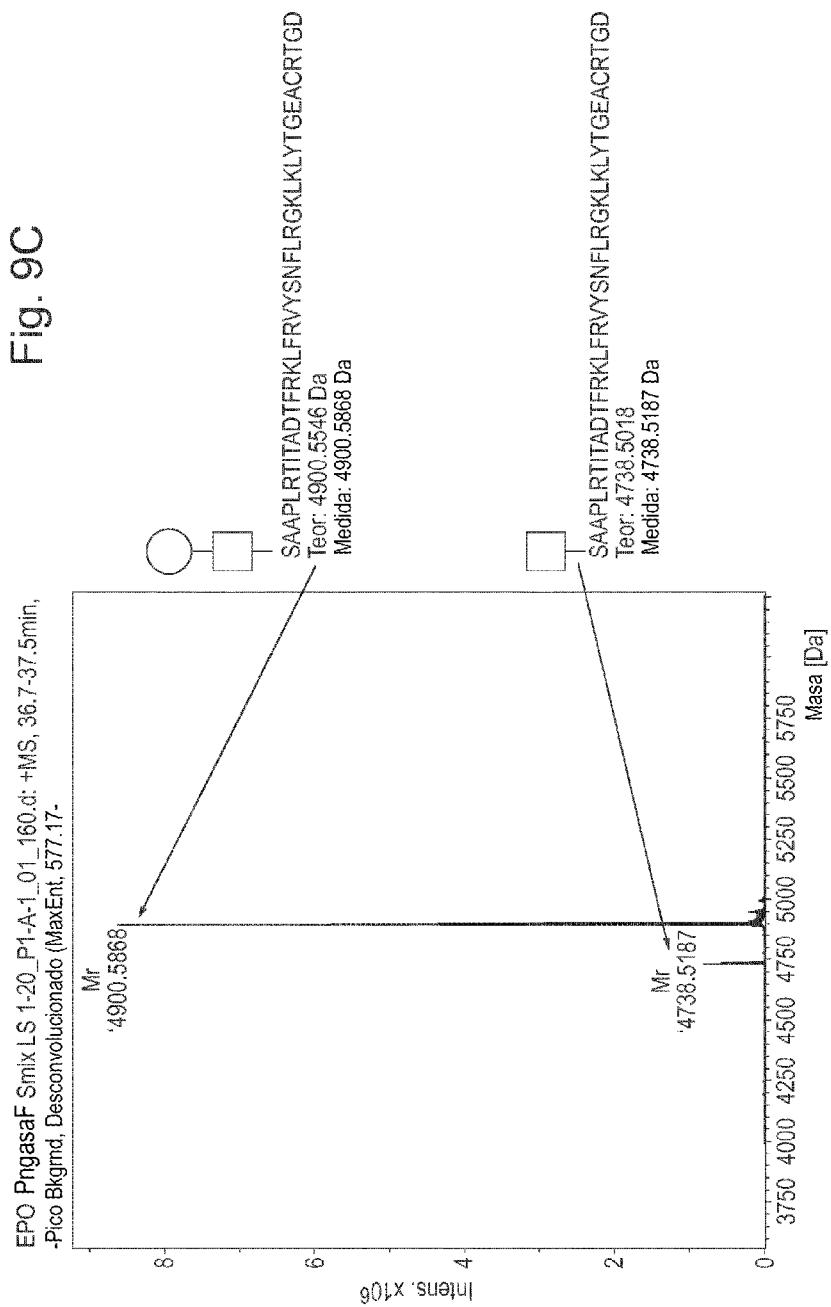
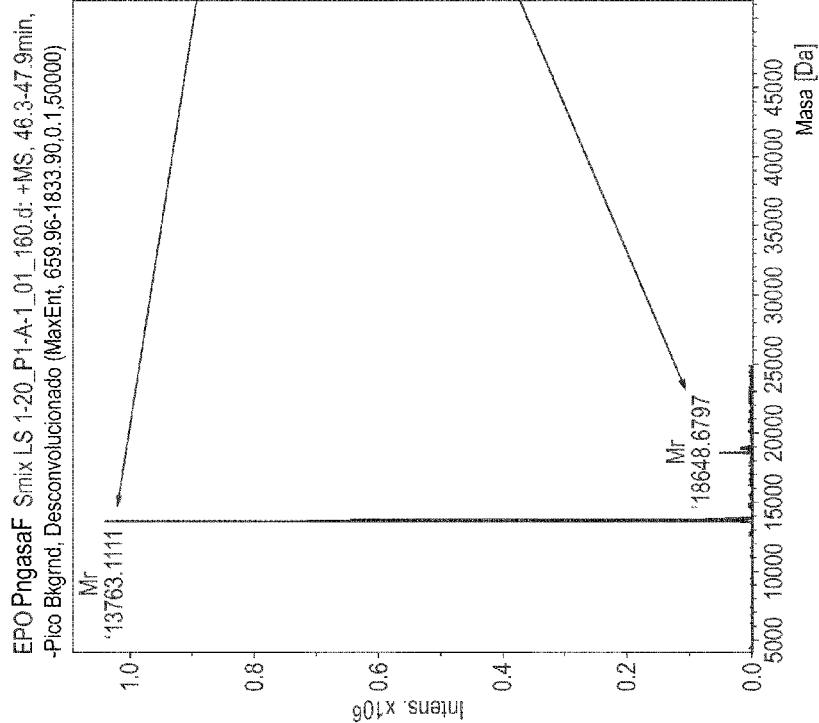


Fig. 9D



APRILICDSRVILERYLEAKEAEDITGCAEHCSLD
ENITVPTKVDYAWKRMEVGQQAVENWQGL
ALLSEANLRGQALLVNSQWPWERQLHDKAVS
GLRSITLRLGAQKEAISPPDAA
Teor: 13714.0957 Da
Medida: 13714.1199 Da

EPN no sugerida tratada con PgngasaF
APRILICDSRVILERYLEAKEAEDITGCAEHCSLD
MEVGQQAVEWQGLALLSEANLRGQALLVNSQWPWERQLHDKAVS
RSLTLLRLGAQKEAISPPDAAASAFLRTTAOTFRKLFRVYSNFRGKLYT
GEACRTGD
Teor: 18598.6397 Da
Medida: 18648.6797 Da

Fig. 10B

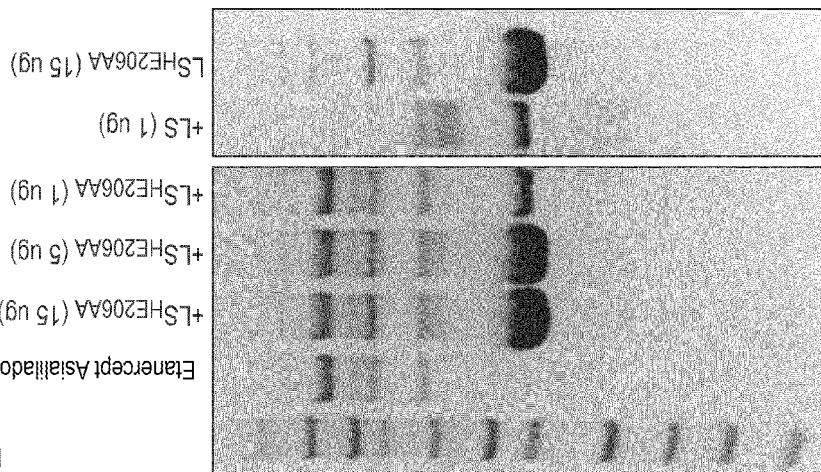


Fig. 10A

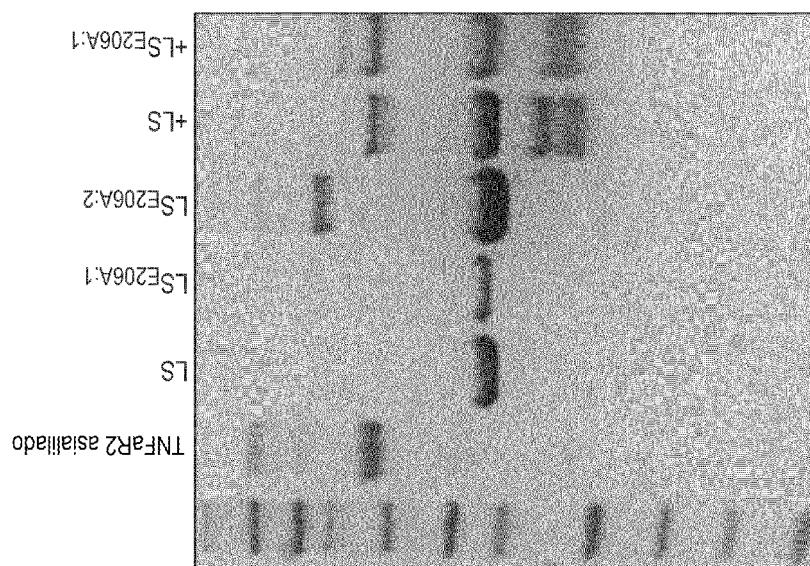


Fig. 11A

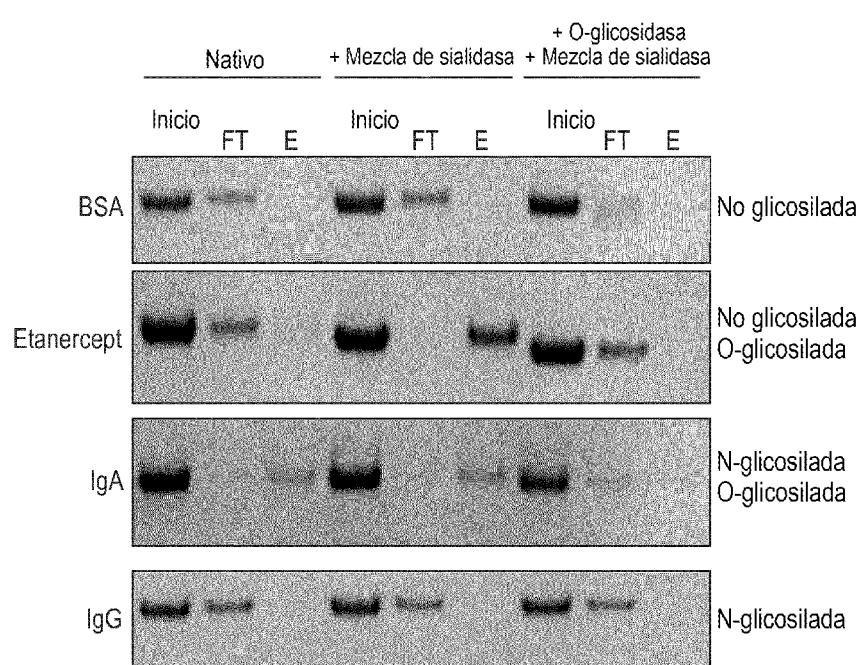


Fig. 11B
Fig. 11C

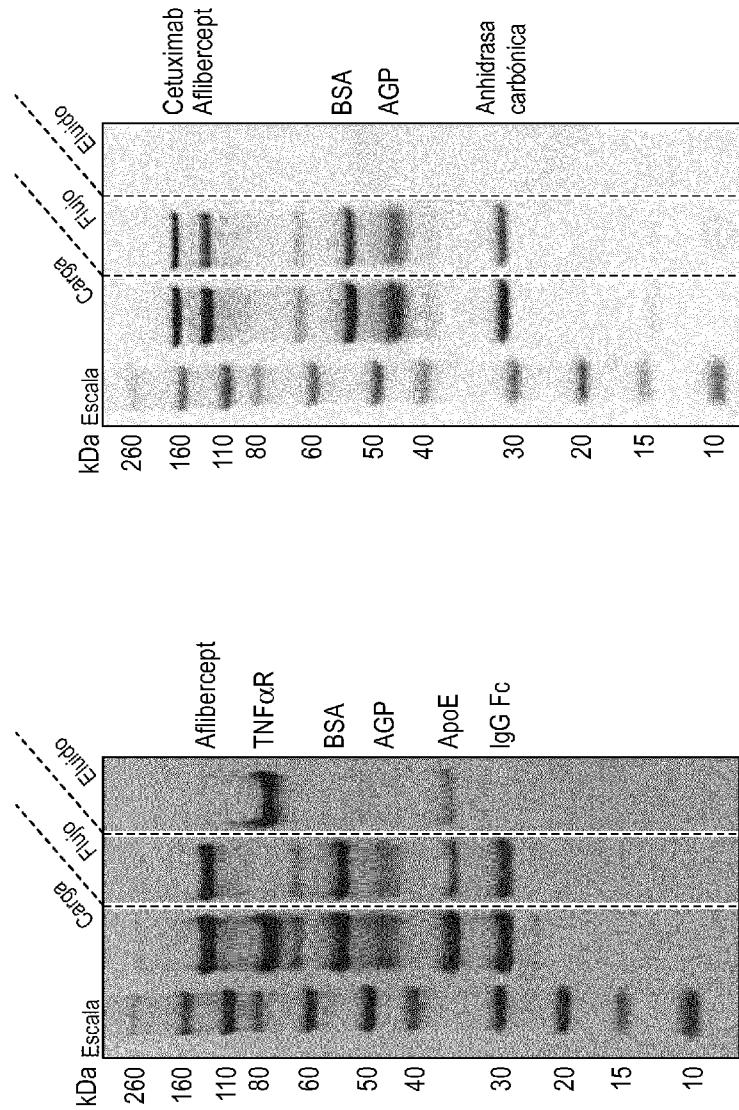


Fig. 12A

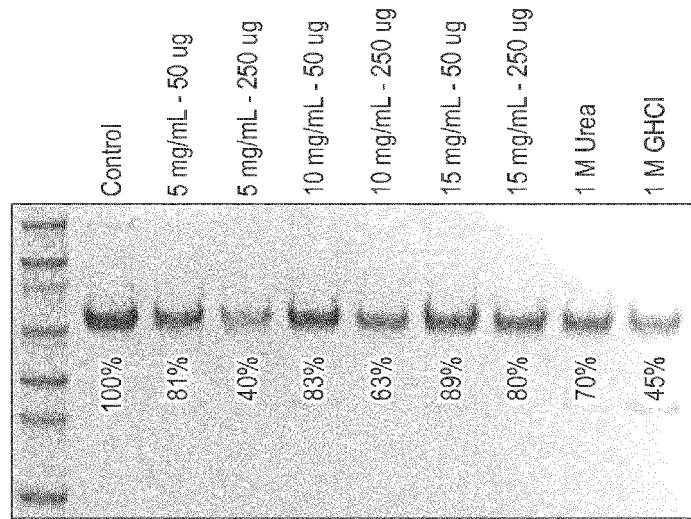


Fig. 12B

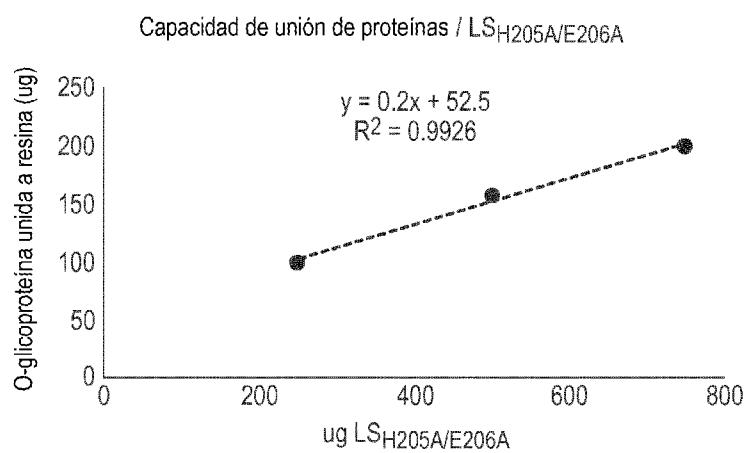


Fig. 13

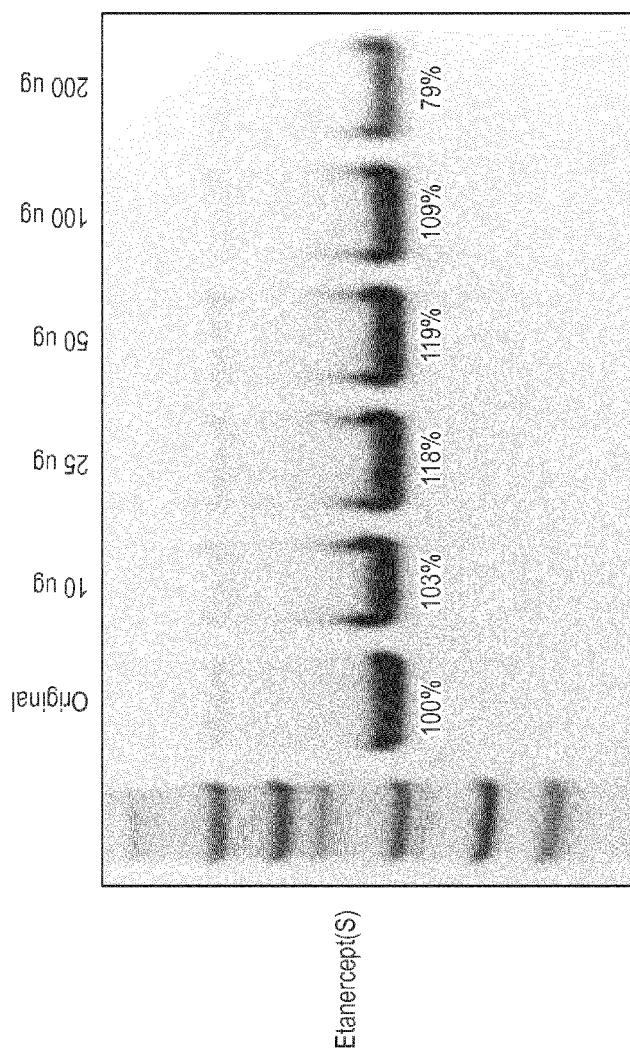


Fig. 14A
Fig. 14B

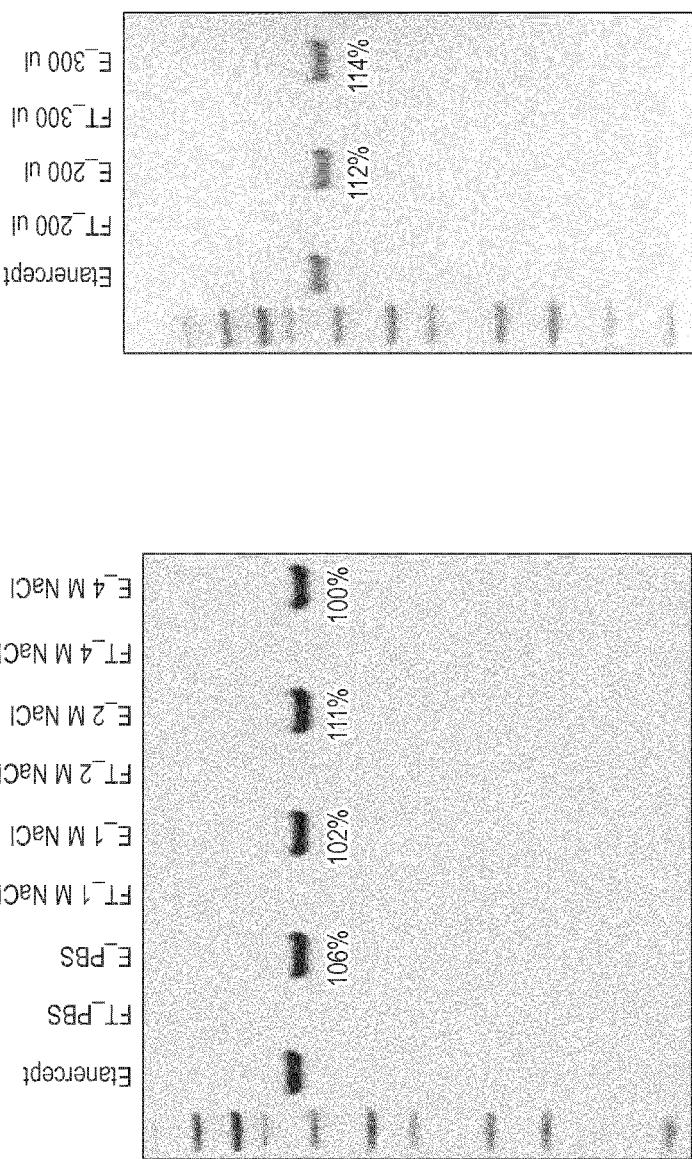


Fig. 14C

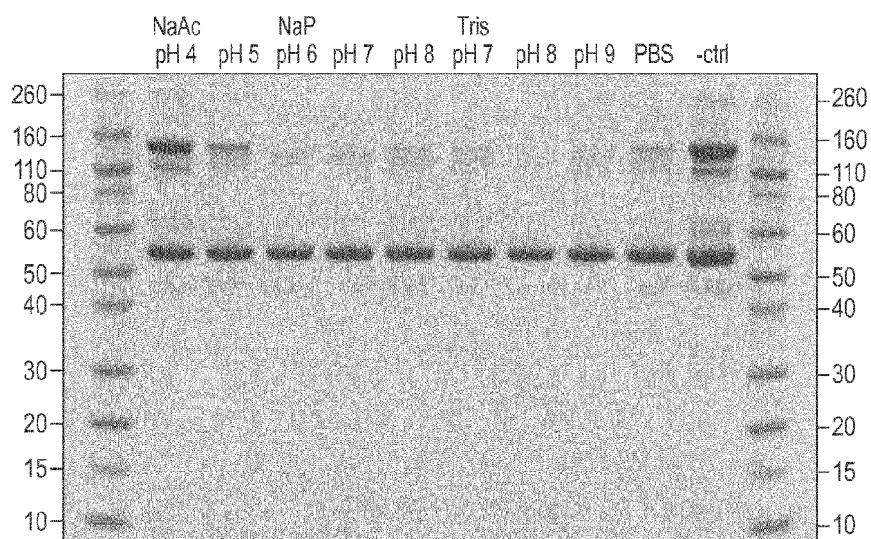


Fig. 14D

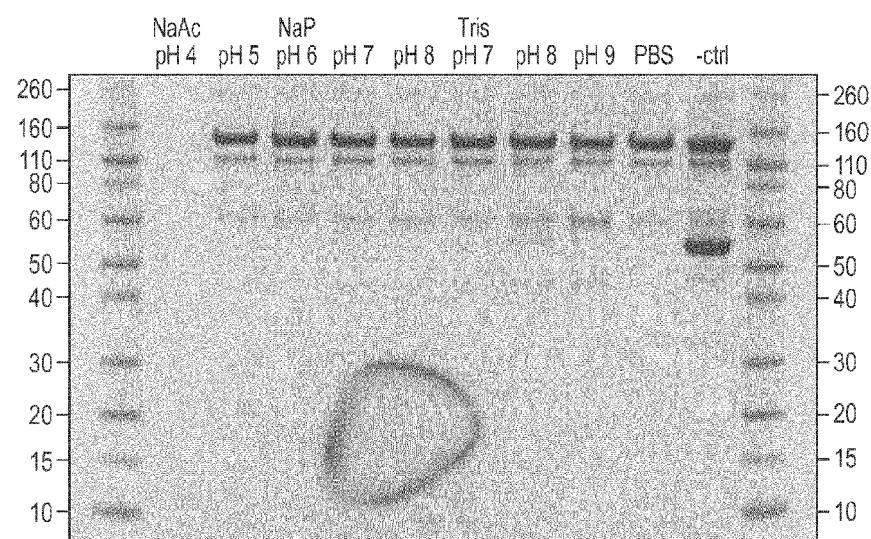


Fig. 15A

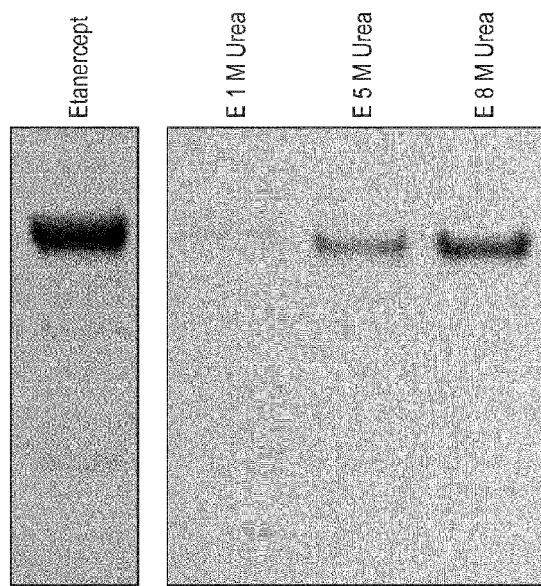


Fig. 15B

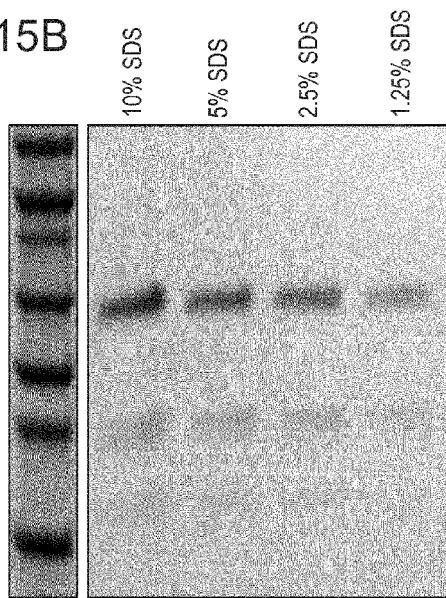


Fig. 16A

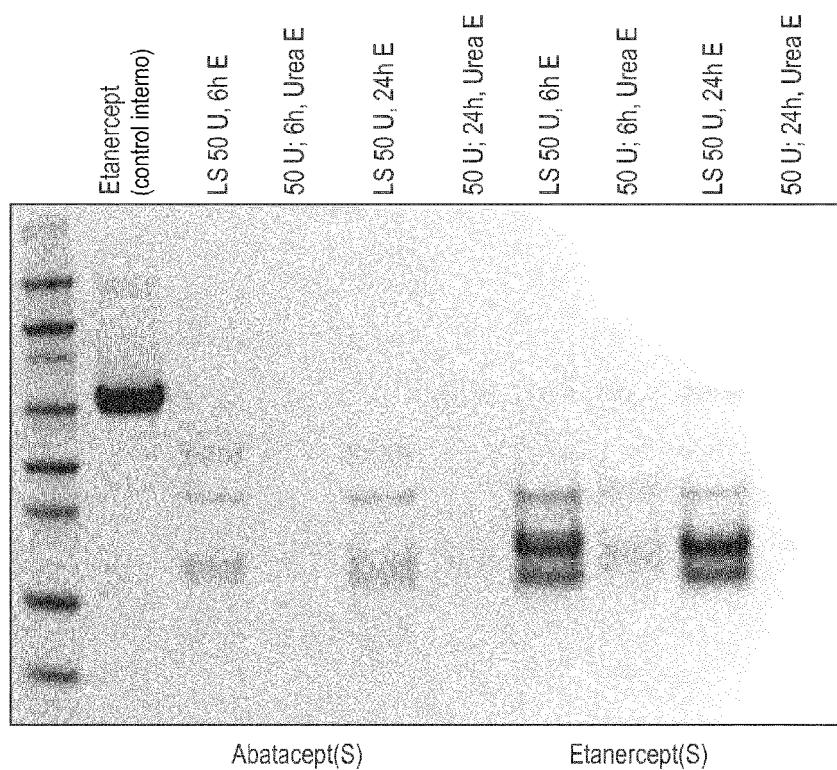


Fig. 16B.1

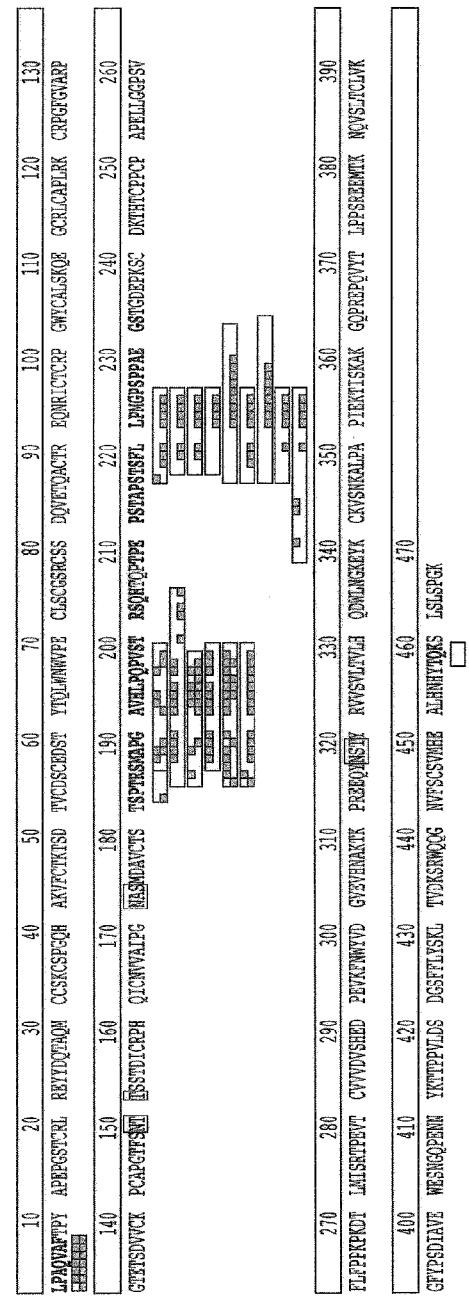
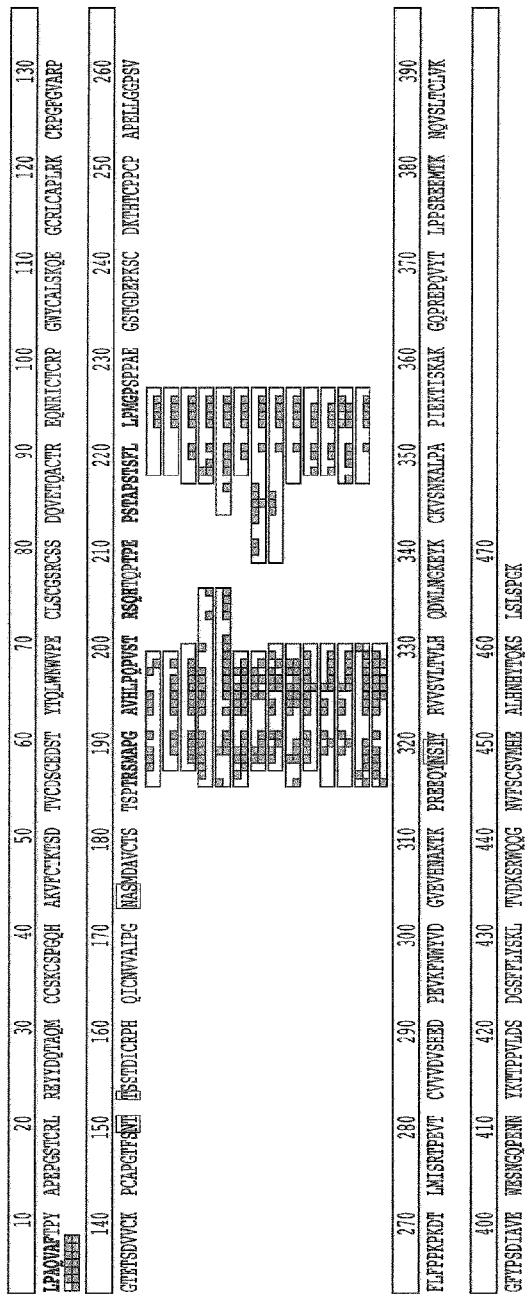
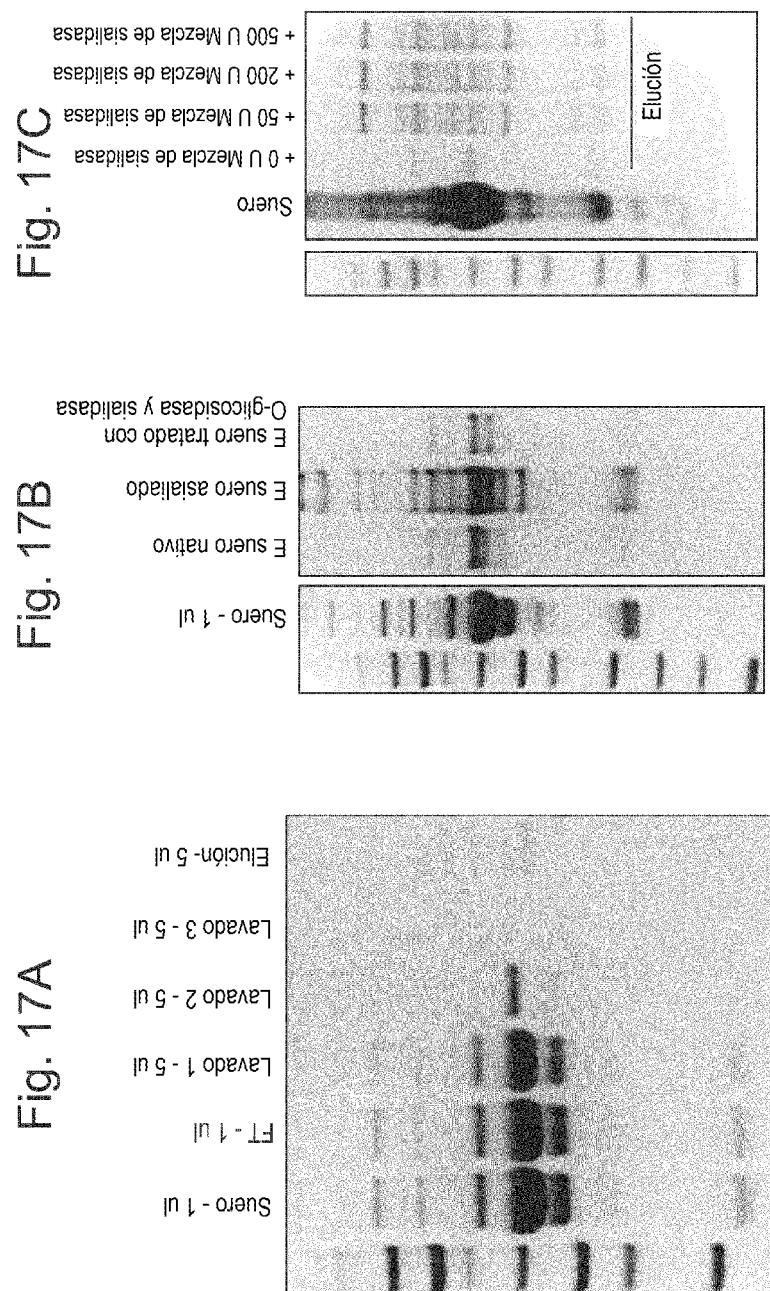


Fig. 16B.2





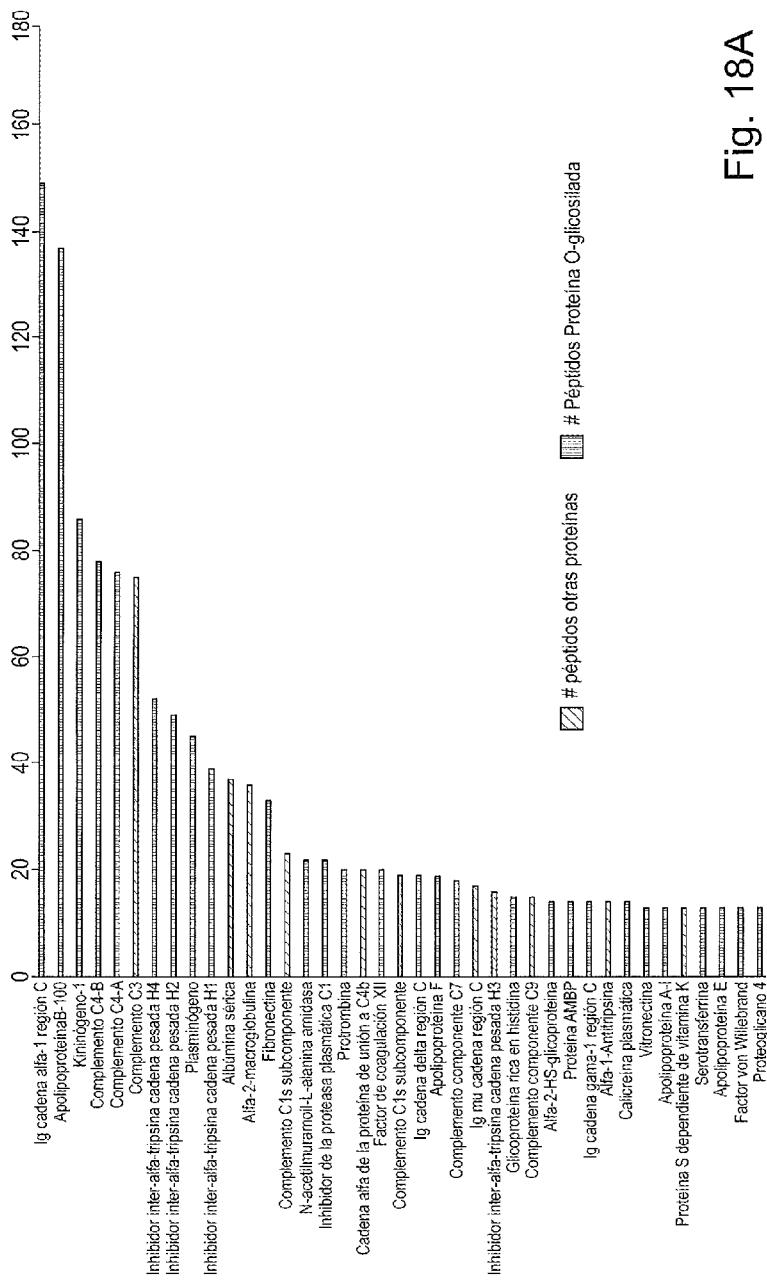
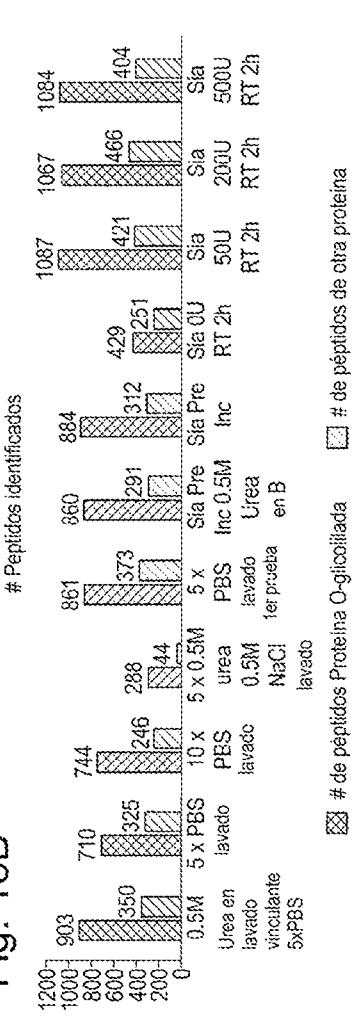


Fig. 18A

Fig. 18B**Fig. 18C**

Relación # péptidos O-glicoproteína identificados / otros péptidos proteicos

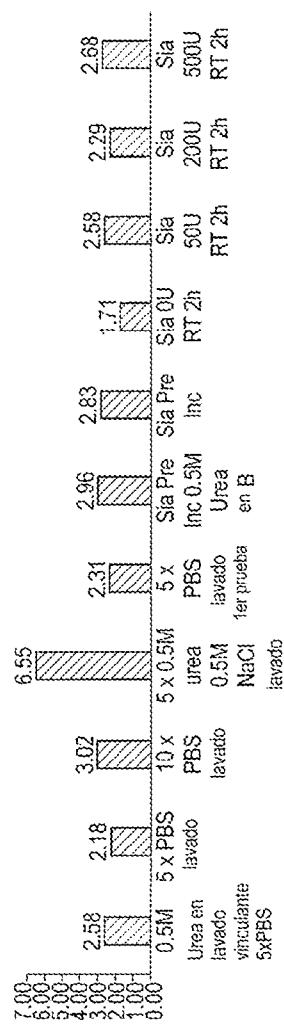


Fig. 19A

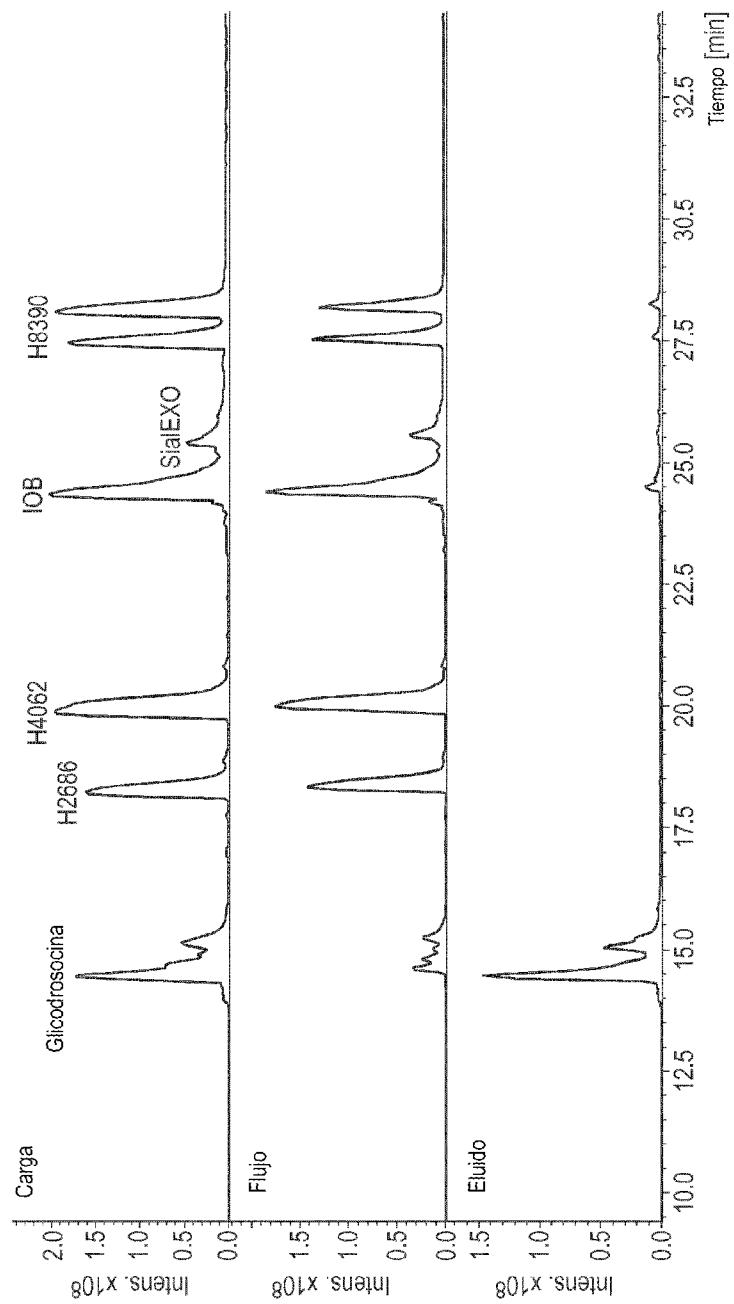


Fig. 19B

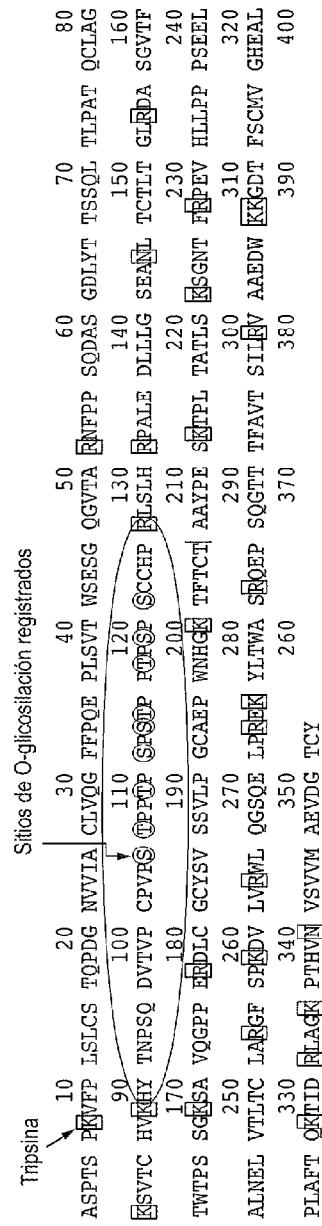


Fig. 19C

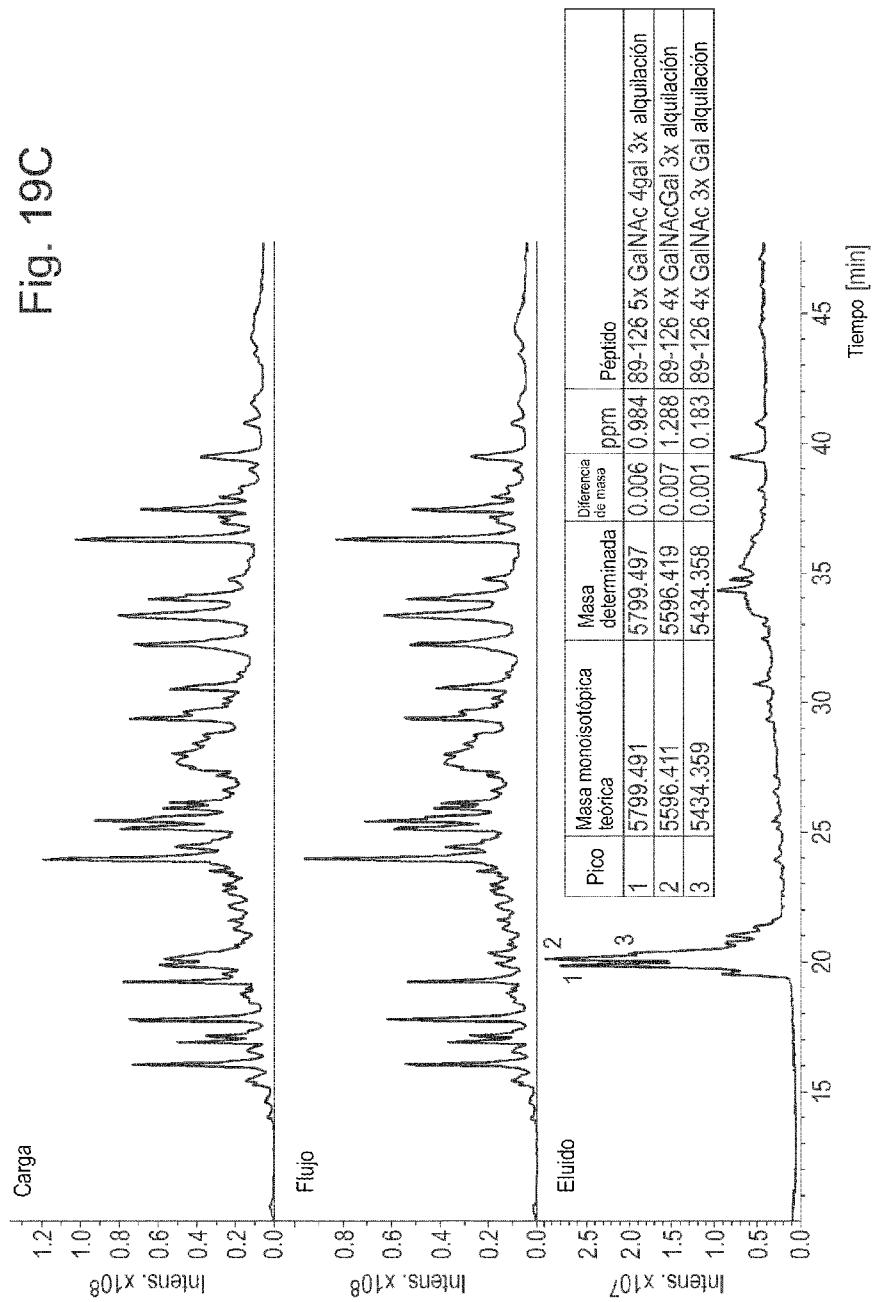


Fig. 20A

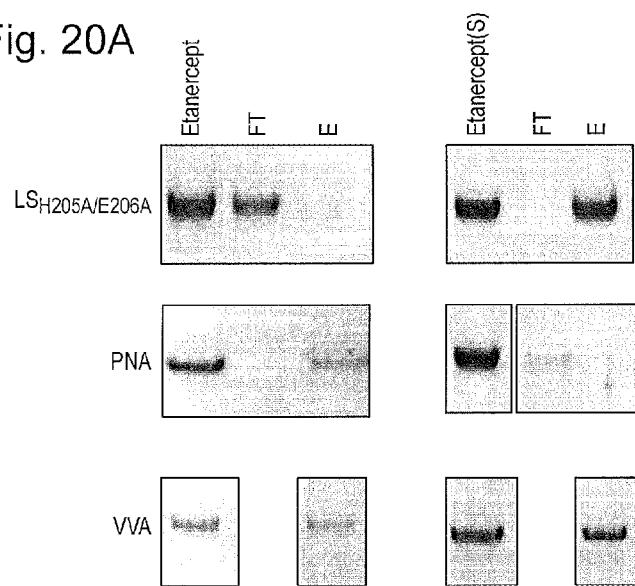


Fig. 20B

