

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 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 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, 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 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.

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.

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:

(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;

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.

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.

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.

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:

(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);

(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.

Breve descripción de las figuras

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.

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.

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).

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/n) 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 TNF α del etanercept (TNF α R) se pretrató con sialidasa ("sialidasa"), O-glicosidasa/sialidasa ("O-glic"), o con PNGaseF ("N-glic"), para eliminar los ácidos siálicos, los O-glicanos y los N-glicanos, respectivamente. Se añadió el LS a las muestras y se dejó que la incubación continuara o/n 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 clon "LS_{mut}" (también denominado LS_{E206A}). (A) Mientras que el LS fue capaz de hidrolizar Enbrel en presencia de sialidasa, el LS_{mut} 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 LS_{mut} 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 α2-3 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 α2-3 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 α2-6 (y α2-8) 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 TNFα del etanercept (TNFαR2; también puede denominarse en el presente documento TNFαR) y el propio etanercept (Etanercept), en comparación con la enzima LS de tipo salvaje. Se añadieron diferentes concentraciones de los mutantes LS a 1 µg de sustrato (1:1 - 15:1, enzima: sustrato), incubando en PBS a 37°C durante la noche antes del análisis en SDS-PAGE.

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).

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}

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 (aflibercept, 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, aflibercept, AGP) y proteínas no glicosiladas (BSA, anhidrasa carbónica), pretratadas con la mezcla Sialidasa.

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 (GHCi) 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.

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.

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.

Figura 15: Resultados del experimento que demuestra que la desnaturalización 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.

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.

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 desnaturalización 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 triptica producirá un único péptido O-glicosilado. **C** muestra resultados representativos del análisis LC/MS de la unión al digerido triptico 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.

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}.

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.

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.

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.

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.

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.

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.

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.

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}.

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.

Las SEQ ID NO: 26, 27 y 28 son las secuencias de aminoácidos de polipéptidos que tienen actividad endoproteasa específica de O-glicoproteína.

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.

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.

La SEQ ID NO: 47 es la secuencia de aminoácidos del péptido glicodrosocina. Hay un sitio de O-glicosilación en el residuo T.

20 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.

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

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.

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 casetes 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.

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.

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.

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.

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.

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.

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.

Polipéptidos con actividad endoproteasa

Características funcionales de un polipéptido con actividad endoproteasa

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.

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 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.

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), 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 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 preferidas se describen más adelante). Un polipéptido de prueba preferiblemente logra un 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

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

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.

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.

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.

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.

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.

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.

Cualquier polipéptido que comprenda la SEQ ID NO: 1 o una variante de la misma, o un fragmento de cualquiera de ellas, 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 GSGLE son enlazadores adecuados.

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

(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;

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.

La secuencia de un polipéptido ejemplar 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 ejemplar 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.

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.

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, 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 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.

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.

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.

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.

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 α 2-3, α 2-6 y α 2-8.

Las sialidasas que son adecuadas para su uso en los métodos descritos incluyen sialidasas de amplio espectro que se dirigen a todos los enlaces α 2-3, α 2-6 o α 2-8, así como sialidasas de espectro estrecho que normalmente se dirigen a un solo tipo de enlace. El enlace α 2-3 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*.

Una sialidasa de espectro estrecho preferida es Am1757 aislada de *Akkermansia muciniphila*. Am1757 tiene actividad específica contra los enlaces α 2-3. 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 GSGLE son enlazadores adecuados. Una secuencia Am1757 ejemplar, con una metionina adicional en el extremo N y un enlazador GSGLE 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.

Una sialidasa de amplio espectro preferida es Am0707 aislada de *Akkermansia muciniphila*. Am0707 tiene actividad contra los enlaces α 2-3, α 2-6 y α 2-8. 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 GSGLE son enlazadores adecuados. Una secuencia Am0707 ejemplar, con una metionina adicional en el extremo N y un enlazador GSGLE 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.

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.

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:

(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,

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 $\geq 90\%$ de 1 μg de glicoproteína (fetuina) cuando se incubaba 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 incubaba 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.

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 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 PNGaseF. 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 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 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 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-gingipaina (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-gingipaina (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 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 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 SEQ ID NO: 15.

Polipéptidos que se unen a glicoproteínas O-ligadas pero que carecen o tienen una actividad endoproteasa reducida.

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 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 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

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.

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 sefarsa 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.

En los ejemplos se describen ensayos ejemplares de este tipo.

Características estructurales de un polipéptido carente de actividad endoproteasa

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. 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 1ª posición del motivo más corto (la 4ª posición del motivo más largo) y/o la sustitución del residuo de histidina (H) en la posición correspondiente a la 5ª posición del motivo más corto (la 8ª 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.

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:

(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

(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.

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

(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;

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 abxxbbxbx, 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 4ª y/o 8ª 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. 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 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: 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 GSGLE son enlazadores adecuados.

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-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, 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.

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.

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

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:

(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

(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, VorW.

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

(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;

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 xbbbx 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

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.

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.

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.

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.

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.

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.

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.

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.

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 dodecil 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.

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.

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

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.

El polipéptido aquí descrito puede utilizarse en combinación con otra enzima, como una proteasa o una glucosidasa. La proteasa o glicosidasa adicional normalmente digerirá más las proteínas o glicanos del sustrato, lo que puede permitir un análisis más fácil o detallado de los productos.

Por ejemplo, el polipéptido aquí descrito puede utilizarse en combinación con un agente para eliminar el ácido siálico. Dicho agente puede ser preferiblemente una enzima sialidasa o una mezcla de dichas enzimas como se describe en la sección anterior. La invención proporciona una composición (en forma liofilizada o de solución) que comprende un polipéptido, en el que el polipéptido comprende una secuencia de aminoácidos de la 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. Se describe en el presente documento un kit que comprende un polipéptido como el descrito en el presente documento en un envase separado de Am1757 o una mezcla de Am1757 y Am0707, con instrucciones para el uso combinado de las diferentes enzimas.

Como otro ejemplo, sólo como referencia, la muestra puede incubarse 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 de la IgG1. La muestra puede ponerse en contacto con una N-glicosidasa y una sialidasa (o una mezcla de ambas) además del polipéptido. En dicho método, la sialidasa (o mezcla) puede aplicarse primero, antes de la adición simultánea de la N-glicosidasa y el polipéptido.

Como otro ejemplo, sólo como referencia, la muestra puede incubarse con una proteasa antes, al mismo tiempo o después de poner en contacto la muestra con el polipéptido descrito en el presente documento, 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-gingipaina (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-gingipaina (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.

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.

Análisis por espectrometría de masas

El etanercept (Enbrel®) es una proteína de fusión Fc clínicamente aprobada que se une al TNFα. 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

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).

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).

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).

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 glicopeptidos 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 LS_{mut} no podía hidrolizar etanercept, confirmando que la alteración genética inactivaba efectivamente la O-glicoproteasa en las condiciones probadas.

El LS_{E206A} se inmovilizó en sefarosa y se añadió a columnas de centrifugación para facilitar su manipulación. Es importante destacar que la unión del LS_{E206A} a diferentes sustratos se correlacionó perfectamente con la actividad hidrolítica del LS (Fig. 6B). El LS_{E206A} (etiquetado como LS_{mut}) demostró una afinidad específica por las glicoproteínas O-ligadas. Inmovilizando el LS_{mut} 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)

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:

SAAPLRITADTFRKLFRVYSNFLRGKCLKLYTGACRTGD (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

10 APPRLICDSRVLERYLLEAKEAEEDITTGCAEHCSLDENITVPDTKVDFYAWKRMEVGQQAVEVWQGLALL
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.

20 **Ejemplo 2**

Introducción

El mutante LSE206A descrito en el Ejemplo 1 incorpora una mutación dirigida al sitio del sitio activo de LS (abxHEbbHbc a abxHAAbbHbc), eliminando la capacidad de transferencia de electrones de la hendidura enzimática. 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 (abxHEbbHbc a abxAAAbbHbc).

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 eluirlas. 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 siálicos. 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 (aflibercept, 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 GHCl, 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 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.

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 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.

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	OK	Cmpd.	m/z medida	Mr calc.	z	A m/z [ppm]	RMS90 [ppm]
1	VERDADERO	1901	592.18576	582.3545447	2	3.130952727	39.5628229
2	VERDADERO	1240	460.24964	459.2441464	1	-3.873732048	6.970140996
3	VERDADERO	2393	592.93513	1778.79071	3	-4.012614897	37.83863199
4	VERDADERO	1019	596.28693	2986.410592	5	-6.916951722	23.18052633
5	VERDADERO	1368	642.63449	1954.956428	3	-7.065331795	12.36796202
6	VERDADERO	2169	672.31437	1342.621401	2	-5.364792216	7.287732308
7	VERDADERO	2160	672.51437	1342.621401	2	-5.364972116	7.267752308
8	VERDADERO	1666	695.83597	1389.70747	2	-7.245935789	13.98730439
9	VERDADERO	1987	745.4213	744.4170254	1	-4.027105599	40.39538537
10	VERDADERO	2308	846.5813	1691.756882	3	-6.278836544	10.77151727
11	VERDADERO	1636	878.42107	1754.839667	2	-6.875687276	9.154478723
12	VERDADERO	2298	895.05953	2682.161633	3	-1.614552481	8.730132859
13	VERDADERO	2644	896.59499	1794.785623	2	-5.675814672	10.69492018
14	VERDADERO	2319	954.06736	2769.193661	3	-4.817709076	9.79889834
15	VERDADERO	2209	1077.36596	7143.957806	2	-7.56127791	10.42175433
16	VERDADERO	1313	1189.55961	2377.120852	2	-6.550138429	9.551830744
17	VERDADERO	1325	1189.55981	2377.120852	2	-6.550738429	9.55102786
18	VERDADERO	2377	1219.18407	5654.557682	3	-6.093774791	9.275746116
Fila	tot.	Resultados					Rango
1	100050	20.3 (D.metascore:20.3,D.bScore:3.6,D.fragCov:50.0,D.mCov:8.2)					157 - 459
2	29654	23.0 (D.metascore:23.0,D.bScore:0.8,D.fragCov:12.3,D.mCov:92.3)					182 - 185
3	50476	31.2 (D.metascore:31.2,D.bScore:1.75,D.fragCov:40.0,D.mCov:24.5)					216 - 225
4	219206	26.5 (D.metascore:26.5,D.bScore:26.5,D.fragCov:33.3,D.mCov:21.1)					184 - 204

5	657526	23.9 (D.metascore:23.9,D.bScore:136.0,D.fragCov:46.7,D.mCov:12.3)	184 - 198
6	374688	36.2 (D.metascore:36.2,D.bScore:61.0,D.fragCov:33.3,D.mCov:30.1)	217 - 225
7	374698	36.2 (D.metascore:36.2,D.bScore:61.0,D.fragCov:33.3,D.mCov:39.2)	217 - 225
8	36489	26.9 (D.metascore:26.9,D.bScore:12.0,D.fragCov:35.7,D.mCov:20.3)	156 - 196
9	18900	66.8 (D.metascore:66.8,D.bScore:191.0,D.fragCov:82.9,D.mCov:48.1)	1 - 7
10	198010	27.4 (D.metascore:27.4,D.bScore:13.0,D.fragCov:22.7,D.mCov:22.5)	217 - 225
11	275854	29.6 (D.metascore:29.6,D.bScore:260.0,D.fragCov:29.3,D.mCov:22.1)	186 - 199
12	30556	24.2 (D.metascore:24.2,D.bScore:569.0,D.fragCov:28.1,D.mCov:20.8)	216 - 231
13	48878	35.0 (D.metascore:35.0,D.bScore:125.0,D.fragCov:35.0,D.mCov:35.0)	216 - 225
14	38966	23.7 (D.metascore:23.7,D.bScore:253.0,D.fragCov:23.5,D.mCov:23.8)	216 - 232
15	26176	20.1 (D.metascore:20.1,D.bScore:28.0,D.fragCov:25.0,D.mCov:16.3)	216 - 225
16	249538	23.1 (D.metascore:23.1,D.bScore:235.0,D.fragCov:59.4,D.mCov:9.0)	184 - 199
17	182236	22.9 (D.metascore:22.9,D.bScore:4036.0,D.fragCov:43.6,D.mCov:12.0)	184 - 199
18	360318	21.1 (D.metascore:21.1,D.bScore:33.0,D.fragCov:22.3,D.mCov:20.0)	208 - 223

Fila	#Cmpds.	P	Secuencia	Modificaciones	Proteína
1	1	0	Y T K K S	Acetil. 1, Hexa (H)HexNAc (F) 1	Eisnercept
2	1	1	T S P I R S		Eisnercept
3	5	2	P S T S F L P M G P S	Hex (H)HexNAc (F) 1, 2	Eisnercept
4	1	4	P T R S M A P G A V H L P Q P V S T R S G H T	Hex (H)HexNAc (F) 1, 16, 17	Eisnercept
5	6	1	P T R S M A P G A V H L P Q P V S	Hex (H)HexNAc (F) 3	Eisnercept
6	4	1	S T S F L P M G P S	Oxidación 7, Hex (H)HexNAc (F) 2	Eisnercept
7	4	1	S T S F L P M G P S	Oxidación 7, Hex (H)HexNAc (F) 1	Eisnercept
8	2	1	R S M A P G A V H L P Q P V S I		Eisnercept
9	2	0	- L P A Q V A F T		Eisnercept
10	2	1	S T S F L P M G P S	Hex (H)HexNAc (F) 1, 2	Eisnercept
11	5	1	R S M A P G A V H L P Q P V S T	Hex (H)HexNAc (F) 1, 14	Eisnercept
12	2	3	P S T S F L P M G P S P P A L G S	Hex (H)HexNAc (F) 1, 2, 3	Eisnercept
13	10	2	P S T S F L P M G P S	Oxidación 8, Hex (H)HexNAc (F) 1, 2	Eisnercept
14	3	4	P S T S F L P M G P S P P A L G S T	Hex (H)HexNAc (F) 1, 2, 3	Eisnercept
15	5	2	P S T S F L P M G P S	Hex (H)HexNAc (F) 1, 2, 3	Eisnercept
16	19	2	P T R S M A P G A V H L P Q P V S T	Hex (H)HexNAc (F) 3, 16	Eisnercept
17	19	2	P T R S M A P G A V H L P Q P V S T	Hex (H)HexNAc (F) 1, 3	Eisnercept
18	8	5	P T P S T S F S T S F L P M G P S	Hex (H)HexNAc (F) 5, 8, 9, 10, 11	Eisnercept

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-

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:

5

Fila	OK	Proteína	MW [kDa]
1	922000930	Ig cadena alfa-1 región C OS=Homo sapiens GN=IGHA1 PE=1 SV=2	77.6
2	000000000	Apolipoproteína B-100 OS=Homo sapiens GN=APOB PE=1 SV=2	515.3
3	933000930	Kininogeno-1 OS=Homo sapiens GN=KNG1 PE=1 SV=2	71.9
4	922000930	Complemento C4-B OS=Homo sapiens GN=C4B PE=1 SV=2	197.6
5	000000000	Complemento C4-A OS=Homo sapiens GN=C4A PE=1 SV=2	192.7
6	933000930	Complemento C3 OS=Homo sapiens GN=C3 PE=1 SV=2	187
7	000000000	Inhibidor inter-alfa-tripsina cadena pesada H4 OS=Homo sapiens GN=ITIH4 PE=1 SV=4	183.3
8	922000930	Inhibidor inter-alfa-tripsina cadena pesada H2 OS=Homo sapiens GN=ITIH2 PE=1 SV=1	180.4
9	933000930	Plasminogen OS=Homo sapiens GN=PLG PE=1 SV=2	90.5
10	922000930	Inhibidor inter-alfa-tripsina cadena pesada H1 OS=Homo sapiens GN=ITIH1 PE=1 SV=3	180.3
11	933000930	Albúmina sérica OS=Homo sapiens GN=ALB PE=1 SV=2	66.5
12	922000930	Alfa-2-macroglobulina OS=Homo sapiens GN=ALM PE=1 SV=3	163.2
13	000000000	Fibrinógeno OS=Homo sapiens GN=F1 PE=1 SV=4	362.5
14	933000930	Complemento C1s subcomponente OS=Homo sapiens GN=C1S PE=1 SV=3	76.6
15	000000000	N-acetilglucosaminil-L-alutina amidasa OS=Homo sapiens GN=EGH VRP2 PE=1 SV=3	62.2
16	933000930	Inhibidor de la proteasa plasminolítica OS=Homo sapiens GN=SERPIN1 PE=1 SV=2	53.1
17	922000930	Protrombina OS=Homo sapiens GN=F2 PE=1 SV=2	70
18	933000930	Cadena alfa de la proteasa de rodón a C9 OS=Homo sapiens GN=C9RFA PE=1 SV=2	67
19	922000930	Factor de coagulación XII OS=Homo sapiens GN=F12 PE=1 SV=2	67.7
20	000000000	Complemento C1r subcomponente OS=Homo sapiens GN=C1R PE=1 SV=2	60.1
21	933000930	Ig cadena delta región C OS=Homo sapiens GN=IGHD PE=1 SV=2	45.2
22	000000000	Apolipoproteína F OS=Homo sapiens GN=APOF PE=1 SV=2	35.4
23	000000000	Complemento componente C7 OS=Homo sapiens GN=C7 PE=1 SV=2	93.5
24	933000930	Ig cadena que región C OS=Homo sapiens GN=IGHM PE=1 SV=3	49.3

23					Inhibidor inter-difosforato: cadena pesada H3 OS-Homo sapiens (GN-ITR13 PE=1 SV=2)	99.8
26					<i>Glycoproteína rica en histidina</i> (G6-Homo sapiens (GN-13H1) PE=1 SV=1)	99.5
27					Componente del coagulo: C9 OS-Homo sapiens (GN-C9 PE=1 SV=2)	63.1
29					<i>Alfa-2-HS-glycoproteína</i> (OS-Homo sapiens (GN-AHS3) PE=1 SV=1)	59.5
29					<i>Proteína AMBP</i> (OS-Homo sapiens (GN-AMBP) PE=1 SV=1)	59
30					<i>Le cadavre spina-1 region C</i> (OS-Homo sapiens (GN-BRC1) PE=1 SV=1)	56.1
31					<i>Alfa-1-antitripsina</i> (OS-Homo sapiens (GN-SLRPINA) PE=1 SV=2)	56.7
32					<i>Cadherina plasmocita</i> (OS-Homo sapiens (GN-KLR1) PE=1 SV=1)	51.3
33					<i>Vitelonutina</i> (OS-Homo sapiens (GN-VTN) PE=1 SV=1)	54.5
34					<i>Apolipoproteína A-I</i> (OS-Homo sapiens (GN-APAI) PE=1 SV=1)	50.6
35					<i>Proteína S de plasma de la serpiente N</i> (OS-Homo sapiens (GN-PRON) PE=1 SV=1)	55.1
36					<i>Serotoninaferina</i> (OS-Homo sapiens (GN-TS) PE=1 SV=2)	57
37					<i>Apolipoproteína E</i> (OS-Homo sapiens (GN-APV) PE=1 SV=1)	56.1
38					<i>Factor von Willebrand</i> (OS-Homo sapiens (GN-VWF) PE=1 SV=4)	502.1
39					<i>Proteoglicano 4</i> (OS-Homo sapiens (GN-PR4) PE=1 SV=2)	151

Fila	pl	# Péptidos	% (%)	Resultados	RMSD [ppm]
1	6.3	132	75.3	1239.4 (M.expect:0.0,M.score:1239.4,M.sig:0.0)	5.36
2	6.0	137	74.2	7043.5 (M.expect:0.0,M.score:7043.5,M.sig:0.0)	6.62
3	6.5	76	52.7	1636.5 (M.expect:0.0,M.score:1636.5,M.sig:0.0)	5.59
4	6.9	78	61.3	4921.1 (M.expect:0.0,M.score:4921.1,M.sig:0.0)	6.46
5	6.7	76	61.3	4468.7 (M.expect:0.0,M.score:4468.7,M.sig:0.0)	6.12
6	6	75	64.1	4596.5 (M.expect:0.0,M.score:4596.5,M.sig:0.0)	6.13
7	6.3	82	56.9	2329.1 (M.expect:0.0,M.score:2329.1,M.sig:0.0)	6.29
8	6.4	49	58.4	2175.2 (M.expect:0.0,M.score:2175.2,M.sig:0.0)	7.02
9	7	45	67.1	2773.2 (M.expect:0.0,M.score:2773.2,M.sig:0.0)	5.59
10	6.3	39	43.8	1838.3 (M.expect:0.0,M.score:1838.3,M.sig:0.0)	5.16
11	5.9	37	63.7	2340.1 (M.expect:0.0,M.score:2340.1,M.sig:0.0)	5.59
12	6	36	29.1	1930.1 (M.expect:0.0,M.score:1930.1,M.sig:0.0)	5.67
13	5.5	33	25.2	1672.8 (M.expect:0.0,M.score:1672.8,M.sig:0.0)	5.38
14	4.8	21	46.1	1454.6 (M.expect:0.0,M.score:1454.6,M.sig:0.0)	6.01
15	7.3	22	26.4	1289.8 (M.expect:0.0,M.score:1289.8,M.sig:0.0)	5.72
16	6.1	22	46.2	1786.8 (M.expect:0.0,M.score:1786.8,M.sig:0.0)	6.15
17	5.6	20	43.4	1128.1 (M.expect:0.0,M.score:1128.1,M.sig:0.0)	6.13
18	7.2	20	51.4	1111.9 (M.expect:0.0,M.score:1111.9,M.sig:0.0)	3.13
19	6	20	39.2	691.9 (M.expect:0.0,M.score:691.9,M.sig:0.0)	5.03
20	5.8	19	39.6	1097.7 (M.expect:0.0,M.score:1097.7,M.sig:0.0)	6.07
21	8.1	19	46.9	667.5 (M.expect:0.0,M.score:667.5,M.sig:0.0)	4.57
22	5.4	19	43.2	144.0 (M.expect:0.0,M.score:144.0,M.sig:0.0)	7.73
23	6.1	18	39.1	1100.7 (M.expect:0.0,M.score:1100.7,M.sig:0.0)	5.73
24	6.3	17	30.9	1199.0 (M.expect:0.0,M.score:1199.0,M.sig:0.0)	5.25
25	5.5	16	28.3	747.2 (M.expect:0.0,M.score:747.2,M.sig:0.0)	5.8
26	7.1	15	25.4	964.5 (M.expect:0.0,M.score:964.5,M.sig:0.0)	5.71
27	5.4	15	34.3	876.5 (M.expect:0.0,M.score:876.5,M.sig:0.0)	5.86
28	5.2	14	46	941.9 (M.expect:0.0,M.score:941.9,M.sig:0.0)	5.79
29	5.9	14	55.4	886.7 (M.expect:0.0,M.score:886.7,M.sig:0.0)	6.06
30	6.5	14	60.3	825.9 (M.expect:0.0,M.score:825.9,M.sig:0.0)	6.04
31	5.4	14	39	740.9 (M.expect:0.0,M.score:740.9,M.sig:0.0)	6.02
32	8.0	14	26	647.7 (M.expect:0.0,M.score:647.7,M.sig:0.0)	3.62
33	5.6	13	37	752.2 (M.expect:0.0,M.score:752.2,M.sig:0.0)	5.1
34	5.6	12	46.6	721.0 (M.expect:0.0,M.score:721.0,M.sig:0.0)	5.77
35	5.5	12	27.7	676.0 (M.expect:0.0,M.score:676.0,M.sig:0.0)	6.18
36	6.8	11	22.9	544.9 (M.expect:0.0,M.score:544.9,M.sig:0.0)	6.02
37	5.6	11	45.1	525.1 (M.expect:0.0,M.score:525.1,M.sig:0.0)	5.75
38	5.3	11	6.3	505.5 (M.expect:0.0,M.score:505.5,M.sig:0.0)	5.52
39	9.5	11	5.6	714.7 (M.expect:0.0,M.score:714.7,M.sig:0.0)	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.

En el segundo experimento, se investigó si el doble mutante del LS podía enriquecer péptidos O-glicosilados a partir de una digestión triptica 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 triptica, 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 intercambi6 de tamp6n a 50 mM Tris, pH 8.0 en la columna 7000 K de centrifugaci6n Zeba. A continuaci6n, se a6adi6 tripsina en una proporci6n de 1:20 y se incub6 a 37 °C durante toda la noche. Se a6adi6 un inhibidor de tripsina de 1 mg/mg y se incub6 a temperatura ambiente durante 20 minutos. Se a6adieron la mezcla de sialidasa y NaCl al digerido triptico resultante. Se dej6 que la mezcla se uniera a la resina durante 2 h a temperatura ambiente con rotaci6n de extremo a extremo. La resina se lav6 diez veces con tamp6n PBS (300 µl) y despu6s se eluy6 con la adici6n de 8 M de urea (50 µl, 2 min, 2 repeticiones).

Los p6ptidos 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 detecci6n se realiz6 en un instrumento ESI-Q-TOF Bruker Impact II. Los resultados se muestran en la Figura 19C. El p6ptido O-glicosilado 89-126 se enriqueci6 significativamente en el eluido y el O-glicop6ptido 89-126 espec6fico se identific6 con masa intacta.

2.2.10 El doble mutante inmovilizado se compara favorablemente con otras matrices de uni6n de O-glicoprote6nas

Los inventores evaluaron la capacidad del doble mutante para purificar por afinidad la O-glicoprote6na en comparaci6n con otras matrices de uni6n de O-glicoprote6na 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 a6adieron 50 µg de sustrato en PBS (PNA y LS doble mutante) o tamp6n de uni6n 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 tamp6n de uni6n a la lectina es 20 mM Tris-HCl pH 7.4, 150 mM NaCl, 1 mM de MgCl₂, CaCl₂, ZnCl₂ y MnCl₂). Se dej6 que el sustrato interactuara con las resinas con mezcla de extremo a extremo durante 2 h a temperatura ambiente. Las prote6nas no unidas se lavaron (100 g, 1 min; 3 x) con PBS o tamp6n de uni6n a lectina, respectivamente. Las resinas se secaron por centrifugaci6n (1000 g, 1 min). Las prote6nas unidas se eluyeron con la adici6n de 8 M de urea (para las resinas PNA y LS doble mutante), o el tamp6n de eluci6n VVA, seg6n el fabricante (para la resina VVA) (50 µl, 5 min de tratamiento antes de la centrifugaci6n 1000 g 1 min; 2 x) y tanto el flujo (FT) como el eluido (E) se analizaron en SDS-PAGE. Se a6adieron 1.5 µg de sustrato a cada gel (por ejemplo, 3 µl) como control positivo y se realiz6 un an6lisis densitom6trico para evaluar la eficacia de las resinas, en relaci6n 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 an6lisis densitom6trico 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 purificaci6n del sustrato asialilado.

Secuencias

SEQ	ID	NO:	
EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLDIKSP			1
GRVNIIEYKAKNPAAHYPYENGSGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVFFYGMG			
RNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHELGHGLNLPNNHQTASDGKKYGTALMGSGNYTFTGTSP			
FLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQ			
DYDAVSFSRRLGKKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDAEQDYRDGSKS			

SEQ	ID	NO:	
MEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLDIKS			2
PGRVNIIEYKAKNPAAHYPYENGSGWKAQELDEFFKAHPDRKKSQHTLIIMPTWNDEKNGPDNPGGVFFYGM			
GRNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHELGHGLNLPNNHQTASDGKKYGTALMGSGNYTFTGTSP			
TFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVN			
QDYDAVSFSRRLGKKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDAEQDYRDGSKSGSG			
HHHHHH			

SEQ	ID	NO:	3
<p>ATGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAATATCG TCTACTTCCTGGGTCTGTATACCGAACCGGTTCCGGACTACGAGCGCCGCTGTGAGCGAGCTGCTGTTGTATCT GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACGCAGCTTTGGTCTGGACATTAAGTCA CCGGGTCGTGTGAACATTATCGAGTACAAAGCGAAGAACCCGGCAGCGCATTACCCGTATGAGAATGGTGGCG GCTGGAAAGCTGCACAAGAACTGGACGAATTTTTCAAGGCCCATCCAGACCGCAAGAAAAGCCAGCACACCCT GATCATCATGCCTACCTGGAATGATGAGAAAAATGGTCCTGACAATCCGGGTGGCGTTCCGTTCTATGGTATG GGTCGTAATTGTTTTGCGTTGGACTACCCGGCGTTTGATATCAAGCACCTGGGTGAGAAAACGCGTGAGGGTC GTCTGCTGACGAAATGGTACGGTGGCATGGCGCACGAACTGGGCCACGGCCTGAATCTGCCGCACAATCACCA GACCCGAGCGATGGCAAGAAATATGGCACCGCCCTGATGGGTAGCGGCAACTACACGTTCCGGTACCAGCCCCG ACGTTCCCTGACCCCGGCGAGCTGTGCGCTGCTGGATGCCGAGTGTTACAGCGTTACCCCGAGCCAACAGT TTTATGAGGGTAAGCCAGAAGTCGAGGTTGGTGATGTTGCAATTTCTTCAAGGGTGATCAAATCTTGGTCAG CGGTAACACAAAGAGCCCCGAAACCGTGAAAGCTCTGAACGTTTACATTACAGGATCCGCCGTACGCCGTGAAC CAAGACTACGATGACGTGAGCTTTAGCCGTCGTCTGGGCAAAAAGTCCGGTAAGTTTACGATGAAGATTGACA AAAAAGAACTGGAAGGCCTGAATAACAACGAATTCGGTATTTCTTGTATGTTTATTCTGGCAAACGGCTTACA CATGCAGAAGCACTTTACGTTTCACTGGGATGCGCTGCAAGACTACCGTGACGGTAGCAAATCTGGTTCCGGGT CATCATCACCACCATCACTGA</p>			

SEQ	ID	NO:	4
<p>MLKRLLSAFFSLFLGAASGTSFAEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQ QFYGKEMQRHGYGARSFGLDIKSPGRVNI IEYKAKNPAAHYPYENGCGWKAQELDEFFKAHPDRKKSQHTLI IMPTWNDEKNGPDNPGGVPFYGMGRNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHELGHGLNLPNNHQT ASDGKKYGTALMGSGNYTFGTSPFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSG NYKSPQTVKALNVYIQDPPYAVNQDYDAVSFSRRLGKKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHM QKHFTFHWDAHQDYRDGSKS (signal sequence underlined)</p>			

SEQ	ID	NO:	5
<p>EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQFYGKEMQRHGYGARSFGLDIKSP GRVNI IEYKAKNPAAHYPYENGCGWKAQELDEFFKAHPDRKKSQHTLI IMPTWNDEKNGPDNPGGVPFYGMG RNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHALGHGLNLPNNHQTASDGKKYGTALMGSGNYTFGTSP FLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQ DYDAVSFSRRLGKKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDAHQDYRDGSKSGS</p>			

SEQ	ID	NO:	6
<p>MEVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQFYGKEMQRHGYGARSFGLDIKS PGRVNI IEYKAKNPAAHYPYENGCGWKAQELDEFFKAHPDRKKSQHTLI IMPTWNDEKNGPDNPGGVPFYGM GRNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAHALGHGLNLPNNHQTASDGKKYGTALMGSGNYTFGTSP TFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVN QDYDAVSFSRRLGKKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDAHQDYRDGSKSGSG HHHHH</p>			

SEQ	ID	NO:	7
<p>ATGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAATATCG TCTACTTCCTGGGTCTGTATACCGAACCGGTTCCGGACTACGAGCGCCGCTGTGAGCGAGCTGCTGTTGTATCT GCAGCAATTCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACGCAGCTTTGGTCTGGACATTAAGTCA CCGGGTCGTGTGAACATTATCGAGTACAAAGCGAAGAACCCGGCAGCGCATTACCCGTATGAGAATGGTGGCG GCTGGAAAGCTGCACAAGAACTGGACGAATTTTTCAAGGCCCATCCAGACCGCAAGAAAAGCCAGCACACCCT GATCATCATGCCTACCTGGAATGATGAGAAAAATGGTCCTGACAATCCGGGTGGCGTTCCGTTCTATGGTATG GGTCGTAATTGTTTTGCGTTGGACTACCCGGCGTTTGATATCAAGCACCTGGGTGAGAAAACGCGTGAGGGTC GTCTGCTGACGAAATGGTACGGTGGCATGGCGCACGCGCTGGGCCACGGCCTGAATCTGCCGCACAATCACCA GACCCGAGCGATGGCAAGAAATATGGCACCGCCCTGATGGGTAGCGGCAACTACACGTTCCGGTACCAGCCCCG ACGTTCCCTGACCCCGGCGAGCTGTGCGCTGCTGGATGCCGAGTGTTACAGCGTTACCCCGAGCCAACAGT TTTATGAGGGTAAGCCAGAAGTCGAGGTTGGTGATGTTGCAATTTCTTCAAGGGTGATCAAATCTTGGTCAG CGGTAACACAAAGAGCCCCGAAACCGTGAAAGCTCTGAACGTTTACATTACAGGATCCGCCGTACGCCGTGAAC CAAGACTACGATGACGTGAGCTTTAGCCGTCGTCTGGGCAAAAAGTCCGGTAAGTTTACGATGAAGATTGACA AAAAAGAACTGGAAGGCCTGAATAACAACGAATTCGGTATTTCTTGTATGTTTATTCTGGCAAACGGCTTACA CATGCAGAAGCACTTTACGTTTCACTGGGATGCGCTGCAAGACTACCGTGACGGTAGCAAATCTGGTTCCGGGT CATCATCACCACCATCACTGA</p>			

SEQ ID NO: 8GMAHELGHGL (motivo de metaloproteasa)

	SEQ	ID	NO:	9
	<p><u>MKNLLFALLTG</u>SFCCCYAQQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVSRDNGKTWE MKAAPAGYPPNYGGIPKESPAIVRNPLTREIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKN LKKLGGIMRTPVFVNKGRRVIVPFHNMGGGTFKFIISDDGGLTWHVSRNGVTSRHEARPPHQGVRWFNNAVEA TVLEMKDGTLWALARTSQDQAWQAFSKDYGETWSKPEPSRFFGTLTMNTLGRLLDDGTIVSLWTNTMALPENAT AGNGTWEDVFTNRDSSHIIAMSGDEGKTWYGFREIILDEHRNHHPGYATLDGPEDRGKHQSEMVLQDKNRILISL GQHKHNRRLLVIVDRRWVGAKTATQTGKDLDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKR LDDPELVNEKSNVDYRNGGATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTF PIRLKPAPHLLLGMKKVPFPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFIISTGSQPDAG ILLDTVNARVK (signal sequence underlined)</p>			
	SEQ	ID	NO:	10
5	<p>QQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVSRDNGKTWEMKAAPAGYPPNYGGIPKE SPAIVRNPLTREIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFVNKG RVIVPFHNMGGGTFKFIISDDGGLTWHVSRNGVTSRHEARPPHQGVRWFNNAVEATVLEMKDGTLWALARTSQ DQAWQAFSKDYGETWSKPEPSRFFGTLTMNTLGRLLDDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSSHII AMSGDEGKTWYGFREIILDEHRNHHPGYATLDGPEDRGKHQSEMVLQDKNRILISLQHKHNRRLLVIVDRRWV AKTRATQTGKDLDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEKSNVDYRNG GATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLLGMKKVP FPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFIISTGSQPDAGILLDTVNARVK</p>			
	SEQ	ID	NO:	11
10	<p>MQQKAAPVPEPEVVATPPADAGRGLIRVDSREIRHYSGTRKEPDYLVSRDNGKTWEMKAAPAGYPPNYGGIPK ESPAIVRNPLTREIRVQPIGGFVFLSRGGLDGKWLAVTNDGKLEEDWKDPEKRKNLKKLGGIMRTPVFVNKG RRVIVPFHNMGGGTFKFIISDDGGLTWHVSRNGVTSRHEARPPHQGVRWFNNAVEATVLEMKDGTLWALARTSQ DQAWQAFSKDYGETWSKPEPSRFFGTLTMNTLGRLLDDGTIVSLWTNTMALPENATAGNGTWEDVFTNRDSSHII AMSGDEGKTWYGFREIILDEHRNHHPGYATLDGPEDRGKHQSEMVLQDKNRILISLQHKHNRRLLVIVDRRWV GAKTATQTGKDLDSQWTIHTYIPQKKGHCSYNRKPSAELVQDPSGGTKKVLQIKRLDDPELVNEKSNVDYRNG GGATWNFPNGTTGLVKFRFRVVDGEQADDSGLQVSLTDRLFNACDSTTKDYALFTFPIRLKPAPHLLLGMKKVP FPTPGAWHEISLLWQGGQAVVSLDGKKAGTLKMANKSPNGASYIHFIISTGSQPDAGILLDTVNARVKGSGLEH HHHHH</p>			
	SEQ	ID	NO:	12
	<p>MTWLLCGRGKWNKVKRMNSVFKCLMSAVCAVALPAFGQEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSI GKGQLLAFAGRYKNNDQGENDIIMSVSKNGGKTWSRPAIAKAHGATFNNPCPVYDAKTRTVTVVFQRYPA VKERQPNIPDGWDDEKCIIRNFMIQSRNGGSSWTKPQEIITKTTKRPSGVDIMASGPNAGTQLKSGAHKGR LVIPMNEGPFPGKWIISCIYSDDGGKSWKLGOPTANMKGMVNETSIAETDNGGVVMVARHWGAGNCRRI AWSQDGGGETWGQVEDAPELFCDSQNSLMTYSLSDQPAYGGKSRIILFSGPSAGRRIGQVAMS YDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEENQEHIKKLFVPITMEWLTG EDTGLAPGKKAPVLK (signal sequence underlined)</p>			
	SEQ	ID	NO:	13
15	<p>QEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKGQLLAFAGRYKNNDQGENDIIMSVSKNGGKTWSR PAIAKAHGATFNNPCPVYDAKTRTVTVVFQRYPAVKERQPNIPDGWDDEKCIIRNFMIQSRNGGSSWTKPQ EITKTTKRPSGVDIMASGPNAGTQLKSGAHKGRLVIPMNEGPFPGKWIISCIYSDDGGKSWKLGOPTANMKGMVNE TSIAETDNGGVVMVARHWGAGNCRRIAWSQDGGGETWGQVEDAPELFCDSQNSLMTYSLSDQPAYGGKSRIILF SGPSAGRRIGQVAMS YDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEENQEHIKKLFVPITMEWLTG EDTGLAPGKKAPVLK</p>			
	SEQ	ID	NO:	14
	<p>MQEEKTGFPDRAVTVFSAGEGNPYASIRIPALLSIGKGQLLAFAGRYKNNDQGENDIIMSVSKNGGKTWSR PAIAKAHGATFNNPCPVYDAKTRTVTVVFQRYPAVKERQPNIPDGWDDEKCIIRNFMIQSRNGGSSWTKPQ EITKTTKRPSGVDIMASGPNAGTQLKSGAHKGRLVIPMNEGPFPGKWIISCIYSDDGGKSWKLGOPTANMKGMVNE TSIAETDNGGVVMVARHWGAGNCRRIAWSQDGGGETWGQVEDAPELFCDSQNSLMTYSLSDQPAYGGKSRIILF SGPSAGRRIGQVAMS YDNGKTWPVKLLGEGGFAYSSLAMVEPGIVGVLYEENQEHIKKLFVPITMEWLTG EDTGLAPGKKAPVLKGSGLHHHHHHH</p>			

SEQ	ID	NO:	15	-	O-glicosidasa	de	S.	oralis
MDKRFFEKRCCKFSIRKFTLGVASVMIGATFFAASPVADQARVGSTDNLPSELADLDKKASDEGHDFDKEAAA QNPGSATTEGPQTEEBELLAQEKEKSEKPSNLPKELEDKLEKAEDNGREVVDKDLAQDTGKLVPEVDVAKTTNG ELNYGATVKIKTPSGEGSGIIVAKDLVLTVSHNFIKDSQEGNIRKVVNDNDQGDGIYSISYPGLPDVKFSKKD IIHWDREGYLKGFKNLALVRLRTVLENTPVEVTKKPVVKKIGDKLHVFGYPEGKLNPIVNTTVDFAEPYGEG VQGIQYQGGKPGASGGGIFDTEGKLVGVHQNGVVGKRSGGILFSPAQLKWIQDHMQGISSVKPADLEEKEKPA EEKPKEDKPAAPKTPKAVTPEWQTVANKEQQGTVTIREEKGVRYNQLSSTAQNNDNGKPALFEKQGLTVDA NGNATVDLTFKDDSEKGSRFVFLKFKDTKNNVFGYDQGGWFWEYKTPGNSTWYKGNRVAAPPGSVNRLS ITLKSQGLNANNDVNLFDVTPLPGAVNENLKNEKKILLKAGTYSNDRTVSVKTDNQEGVKADTPAQKET CPAVDDSKVTYDTIQSKVLKAVIDQAFPRVKEYTLNGHTLPGQVQGFNQVFINNHRITPEVTYKKINETTAEY LMKLRDDAHLINAEMTVRLQVVDNQLHFDVTKIIVNHQVTPGQKIDDERKLLSTISFLGNALVSVSSDQAGAK FDGATMSNNTHVSGDDHIDVTNPMKDLAKGYMGFVSTDKLAAGVWSNSQNSYGGGSNDWTRLTAYKETVGNA NYVGIHSSSEWQWEKAYKGI VPEYTKELPSAKVITEDANADNKVQWQDGAIAYSIMNPNQGWKVKDITAY RIAMNFGSQAQNPFLMTLDGKIKINLHTDGLGQGVLLKGYGSEGHDSGHLNYADIGKRIGGVEDFKTLIEKAK KYGAHLGIHVNASSETYPESKYFNENILRKNPDGSYSYGNWLDQGINIDAAYDLAHLARWEDLKKKLGEG DFIYVDVWNGQSGDNGAWATHVLAKEINKQGWRAIEWGHGGEYDSTFQHWAADLTYYGTYNKGINSAITRF IRNHQKDSWVGDRSYGGAANYPLGGYSMKDFEGWQGRSDYNGYVTNLFAHDVMTKYFQHFTVSKWENGTPV TMTDNGSTYKWTPEMKVELVDAAGNKVVVTRKSNVDNSPQYRERTVTNLGRVIQDGSAYLTPWNWDANGKKLP TEKEKMYFNTQAGATTWTPLPSDWANSKYLYKLTDQOGKTEEQLTVDGKITLDLLANQPYVLYRSKQTNPE MSWSEGMIHYDQGFNSGTLKHWITISGDASKAEIVKSGQANEMLRIQGNKSKVSLTKLTLGLKPNTRYAVYGV DNRSNAKASITVNTGEKEVTYTNKSLALNYIKAYAHNNRRENATVDDTSYFQNMAYFTTGSQVSNVTLLTS REAGDEATYFDEIRTFENNSSMYGDKHDTGQGTFFKQDFENVAQGIFFPVVGGVEGVEDNRTHLSEKHDPTQR GWNGKKVDDVIEGNWSLKTNGLVSRRLVYQTIQPNRFEAGKTYRVTFEYEAGSDNTYAFVVGKGEFQSGRR GTQASNLEMHELPTWTDSKKAKKVTFLVTGAETGDTWVGIIYSTGNASNTRGDAGGNANFRGYNDMMNDLQI EETITLTKMLTENALKNYLPTVAMTNYTKESMDALKEAVFNLSQAQDDISVEEARAEIAKIEALKNALVQKKT ALVAEDFESLDAPAQPGEGLENAFDGNVSSLWHTSWNGGDVGKPMVTLKEPTEITGLRYVPRASDSNGNLDR VKLVVTDESGKEHTFNVDWPNNNKPKDIDFGKTIKAKKIVLTGKTYGDDGDKYQSAELIFTRPQVAETPL DLSGYEALAKAQKLTDKDNQEEVASVQASMKYATDNHLLTERMVAYFADYLNQLKDSATKPDAPTSSKGEEQ PPVLDVPEFKGGVNATEAAVHEVPEFKGGVNAVQALVHELPEYKGGANAVLAAANEVPEYKGGANAVEALVNE KPAYTVGLATAGDQAAPTVEKPEYPLTPSPVADTKTPGAKDEEKLPATGEHSSEVALFLASVSIALSAAVLAT KRKEEGSGLEHHHHHH								

5 SEQ ID NO: 16 - E206A _cebador directoATGGCGCACGC GCTGGGCCACG

SEQ ID NO: 17 - E206A _cebador inversoGCCACCGTAC CATTTCGTC

SEQ	ID	NO:	18	-	EPO
APPRLICDSRVLERYLEAKEAEDITTGCAEHCSLDENITVPDTKVDFYAWKRMEVGQQAVEVWQGLALLSEA VLRGQALLVNSSQPWEPLQLHVDAVSGRLSLTTLRALGAQKEAISPPDAAS <u>SA</u> APLRTITADTFRKLFVYS NFLRGKLLKLYTGEACRTGDR					

(Notas:
el O-glicano con serina está subrayado; la arginina C terminal suele truncarse durante la expresión)

SEQ	ID	NO:	20
EVTVPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLDIKSP GRVNIIEYKAKNPAAHYPYENGCGWKAQELDEFFKAHPDRKKSQHTLIIMPTWDEKNGPDNPGGVPPFYGMG RNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAAALGHGLNLPNNHQTASDGKKYGTALMGSGNYTFGTSP FLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVNQ DYDAVGSFRRRLGKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKS			

SEQ	ID	NO:	21
MEVTPDALKDRIALKKTARQLNIVYFLGSDTEPVPDYERRLSELLLYLQQFYGKEMQRHGYGARSFGLDIKS PGRVNIIEYKAKNPAAHYPYENGCGWKAQELDEFFKAHPDRKKSQHTLIIMPTWDEKNGPDNPGGVPPFYGM GRNCFALDYPAFDIKHLGQKTREGRLLTWKYGGMAAALGHGLNLPNNHQTASDGKKYGTALMGSGNYTFGTSP TFLTPASCALLDACEVFSVTPSQQFYEGKPEVEVGDAISFKGDQILVSGNYKSPQTVKALNVYIQDPPYAVN QDYDAVGSFRRRLGKSGKFSMKIDKKELEGLNNNEFRISLMFILANGLHMQKHFTFHWDALQDYRDGSKSGSG HHHHHH			

SEQ	ID	NO:	22
ATGGAAGTCACTGTGCCGGACGCCCTGAAAGATCGCATCGCGCTGAAGAAAACCGCTCGTCAGCTGAATATCG TCTACTTCCTGGGTCTGTATACCGAACCAGTTCGGGACTACGAGCGCCGCTGAGCGAGCTGCTGTGTATCT GCAGCAATCTATGGTAAAGAAATGCAGCGCCATGGCTATGGCGCACGCAGCTTTGGTCTGGACATTAAGTCA CCGGGTCGTGTGAACATTATCGAGTACAAGCGAAGAACCCGGCAGCGCATTACCCGATGAGAATGGTGGCG GCTGGAAAGCTGCACAAGAACTGGACGAATTTTTCAGGCCCATCCAGACCGCAAGAAAAGCCAGCACACCCT GATCATCATGCCTACCTGGAATGATGAGAAAAATGGTCTGACAATCCGGGTGGCGTTCCGTTCTATGGTATG GGTCTGAATTTGTTTGGCTTGGACTACCCGGCGTTTATATCAAGCACCTGGGTGAGAAAACCGCTGAGGGTC GTCTGCTGACGAAATGGTACGGTGGCATGGCGGCCGCGCTGGGCCACGGCCTGAATCTGCCGCACATCACCA ACCCGCAGCGATGGCAAGAAATATGGCACCGCCCTGATGGGTAGCGGCAACTACACGTTCCGGTACCAGCCCCG ACGTTCTGACCCCGCGAGCTGTGCGCTGCTGGATGCCTGCGAAGTGTTACAGCGTTACCCCGAGCCAACAGT TTTATGAGGGTAAGCCAGAAGTCGAGGTTGGTGATGTTGCAATTTCTTCAAGGGTGATCAAATCTTGGTTCAG CGGTAACACAGAGCCCCGAAACCGTGAAAGCTCTGAACGTTTACATTAGGATCCGCCGTACGCCGTGAAC CAAGACTACGATGAGTGAAGCTTTAGCCGCTGCTGCGCAAAAAGTCCGGTAAAGTTTAGCATGAAGATTGACA AAAAAGAACTGGAAGGCCTGAATAACAACGAATTCGGTATTTCTTGTATGTTTATTCTGGCAAACGGCTTACA CATGCAGAAGCACTTTACGTTTCACTGGGATGCGCTGCAAGACTACCGTGACGGTAGCAAATCTGGTTCGGGT CATCATCACCACCATCACTGA			

SEQ ID NO: 23GMAHALGHGL (motivo de metaloproteasa alterado)

SEQ ID NO: 24GMAAELGHGL (motivo de metaloproteasa alterado)

SEQ ID NO: 25GMAAALGHGL (motivo de metaloproteasa alterado)

SEQ	ID	NO:	22
10	26Pseudomonas aeruginosa PAO1 (secuencia nativa con secuencia de señal eliminada)		
ATQEEILDAAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLQAYDPGAASMWVLPANPDNTL PFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWLNVNDPGAATAKDFKVSVVGVVDK TAALNGLKSAGLQPADAAACNALTDASCSTSKLLVLCNGASAAASLSATVRARLQAGLPILFVHTNGWNQSSSTG QQILAGLGLQEGPYGGNYWDKDRVPSSRTTRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRITC DDVPLGSLDFSKRVDVLKALDAYNQKAQNLFALPGTTSRLWLLWADAVRQNIYRPMDKAADTARFQETFFVAD AIVGYVREAGAAQKELGSYAGQRQQSMPVSGSEETLTTLPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLAV GLNTQRIGSTRLWNTRQYDRPRFLKSPDIKLQANQSVLSPYGLLQLVYSGATPGQTVTVKVTGAASQPFLL DIQPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSIDKDYGGDVQRFIRELNEVFIDDAYTL AGFAIPNQAKTPAIQEACAARGWDCDSETLHKLPGTQHINVDQYACCGGCSGNPYDQTWGLNPRGWGESHEL GHNLQVNRLLKVGGRSGEISNQIFPLHKDWRVLRFEQGNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLWE DPGTYALNGERMAFYTQWVHYWADLKNLPLQGWDIWTLTLLYHQRQVDKSDWDANKAALGYGTYAQRPGNSGDA SSTDGNLNLGLSWLTQRDQRPFTALWGIRTSAAAQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGS PAWPF			

de metaloproteasa subrayado.

Motivo

SEQ	ID	NO:	22
15	27Bacteroides thetaiotaomicron VPI-5482 (secuencia nativa con secuencia de señal eliminada)		
DKWEKEFRIRSYEPYSNIAEWADKLMKKYSDDLNTGTSVKAGDJIIVLVGDTYQGNISMQCIWETGTEYKQ TASSGDVYMLNPGVKNLTMKGGQLFVMYNTELTSTAKPIKIHPLSGTVNGFFDLKEHKTDEKYAELLKK STHKYFCIRGEKIMFYFHRNKLEYPNNILSAIHLWDNIVGWQELMGIDVDRPSQVNNHLFAISPEGSYMW ASDYQIGFVYTYLGNILLEDNVMAAEDNAWGPAHEIGHVHQAANWASSTESSNNLFSNFIYKLGKYKSRGN GLGSVATARYANGQAWYNGDATHQNETETHMRMNWLWIYYHRCYKTDFTWQTLFKLMREVNMTGEDPDK KQLEFAKMASKAANLTDFFEMWGFEPVNTTIEQGTYYKYYVSDAMIREAKYMAQFPAPKHAFQYIEDRK KSEFPNSNDYRYSAGVDVGYTQFKENQKITKAITAELAGRKVSIQNGDEAVAFELRENDENGKLLYFTFTTF EIPSSILMVNAKLYAVQADGKRILL			

de metaloproteasa subrayado.

Motivo

SEQ	ID	NO:	22
20	28Clostridium perfringens (secuencia nativa con secuencia de señal eliminada)		
VLELEMRGDSISEAKKRKVVNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGAATSFQKPG KNEITTEPEINYESNGIPKDVIGGGDLFFTNYSKSDSKRAPKVRLEGASKYPVILGKSDENEVMKELEYVEK IKAEPKTPNIFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRI MPMVKNLSGGAFMNAGNGVIGIRPGNQDAILAANKGWVAHELGHNFDTGGRTIVEVTNNMMLPFESKYKTK TRITDQNIWENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLYDNTFYGKFERQFRERDFGNKNREDI YKSWVAASDAMELDLTEFFARHGIRVDDKVKEDLAKYKPKDKKIYYLNDLAMNYKGDGTENAKVSVSTSGS NGNIKLSFSVDENKDNILGYEIRRDGKYVGFTSNDSSFVDTKSNLJEDGVYVVTYPDRKLNLTNPIEVN			

de metaloproteasa subrayado.

Motivo

SED	ID	NO:	22
	29Pseudomonas aeruginosa PAO1 (N-term Met, C-term enlace/etiqueta)		
MATQEEILDAAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLQAYDPGAASMWVLPANPDNT LPFLIGDKGRVLASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWLNVNDPGAATAKDFKVSVVGVVD			

KTAALNGLKSAGLQPADAACNALTDASCASTSKLLVLGNGASASAASLSATVRARLQAGLPILFVHTNGWNQSSST
 GQQILAGLGLQEGPYGGNYWDKDRVPSSRTRRSVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRRT
 CDDVPGLSDFS KRVDVLK GALDAYNQKAQNL FALPGTTSRLRLWLLWADAVRQNI RYPMDKAADTARFQET FVA
 DAIVGYVREAGAAQKELG SYAGQRQQSMPVSGSEETLT LTLPSAQGF TAIGRMAAPGKRLSIRIEDAGQASLA
 VGLNTQRIGSTRLWNTRQYDRPRFLKSPDIKLQANQSVALVSPYGGLLQLVYSGATPGQTVTVKVTGAASQPF
 LDIQPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSDIKDYGGDVQRFIRELNEVFIDDAYT
 LAGFAIPNQAKTPAIQQECAARGWDCDSETHLKLPGTQHINVDQYAQCGGGCSGNPYDQTWGLNPRGWGESHE
 LGHNLQVNRLKVYGGRSGEISNQIFPLHKDWRVLRFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLW
 EDPGTYALNGERMAFYTQWVHYWADLKNLPLQGWDIWTL LYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGD
 ASSTDGNDNLLGLSWLTQRDQRPTFALWGIRTSAAAQAQVAAYGF AEQPAFFYANNRTNEYSTVKLLDMSQG
 SPAWPFPGSGHHHHHH

Motivo

de metaloproteasa subrayado.

SEQ ID NO: 30Bacteroides thetaiotaomicron VPI-5482 (N-term Met, C-term enlace/etiqueta)
 MDKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDDLNP TGISVKAGDDIIVLVGDTY GQNI SMQC IWETGTEYK
 QTASSGDVYMLNPGVKNKLTMKGEGQLFV MYNTELTSTAKPIKIH IPLGSGTVNGFFDLKEHKTDEKYAELLK
 KSTHKYFCIRGEKIMFYFHRNKLLEYV PNNILSAIHLWDNIVGWQQLMGIDDV RPSQVNNH LFAISPEGSYM
 WASDYQIGFVYTYLGNILLEDNVMAAEDNAWGP AHEIGHVHQAAINWASSTESSNNLFSNFI IYKLGKYKSRG
 NGLGSVATARYANGQAWYNMGDATHQNEDETE THMRMNWQLWIYYHRCEYKTD FWTFLKLMREVNMTEGEDPG
 KKQLEFAKMASKAANQNLTDFFEMWGFEPVNTTIEQYGT YKYVSDAMIREAKEYMAQFPAPKHAFQYIEDR
 KKSEFPNDYRYSAGVDVGYTQFKENQKITKAITAELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTT
 FEIPSSILMVNAKLYAVQADGKRILLGSGHHHHHH

Motivo

de metaloproteasa subrayado.

SEQ ID NO: 31Clostridium perfringens (N-term Met, C-term enlace/etiqueta)
 MVLELEMRGDSISEAKKRKVNWFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQLKP
 GKNEITIP EINYENSGIPKDV IQGGDLFTNYKSDSQKRAPKVRIEGASKYPVFILGKSDENEVMKELEAYVE
 KIKAEPKTTPNIFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFR
 IMPMVKNLSGGAFMNAAGNVGIRPGNQDAILAANKGWGVAHELGHNFDTGGRITVEVTNNMPLFFESKYKT
 KTRITDQNIWENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLYDNTFYGKFERQFRERDFGNKNRED
 IYKSWVVAASDAMELDLTEFFARHGIRVDDKVKEDLAKYKPKDKKIYYLNDLAMNYKGDGFTENAKVSVSTSG
 SNGNIKLSFSVDDENKDNILGYEIRRDGKYVGFTSND SFVDTKSNLDEDDGVYVVTPTYDRKLNTLNPIEVNGSG
 HHHHHH

Motivo

de metaloproteasa subrayado.

SEQ ID NO: 32 Pseudomonas aeruginosa PAO1 (secuencia nativa completa, incluida la secuencia de
MSLSTTAFFSLQGENMSRSPIPRHRALLAGFCLAGALSAQAATQEEILDAAALVSGDSSQLTDSHLVALRLQQQ
 VERIRQTRTQLLDGLYQNL SQAYDPGAASMVLPANPDNTLPFLIGDKGRVLASLSLEAGGRGLAYGTNVL TQ
 LSGTNAAHAPLLKRAVQWL VNGDPGAATAKDFKVS VVGVDKTAALNGLKSAGLQPADAACNALTDASCASTSK
 LLVLGNGASASAASLSATVRARLQAGLPILFVHTNGWNQSSSTGQQILAGLGLQEGPYGGNYWDKDRVPSSRTRTR
 SVELGGAYGQDPALVQQIVDGSWRTDYDWSKCTSYVGRRTT CDDVPGLSDFS KRVDVLK GALDAYNQKAQNLFA
 LPGTTSLRLWLLWADAVRQNI RYPMDKAADTARFQET FVADAIVGYVREAGAAQKELG SYAGQRQQSMPVSGS
 EETLT LTLPSAQGF TAIGRMAAPGKRLSIRIEDAGQASLAVGLNTQRIGSTRLWNTRQYDRPRFLKSPDIKLQ
 ANQSVALVSPYGGLLQLVYSGATPGQTVTVKVTGAASQPF LDIQPGEDSSQAIADFIQALDADKADWLEIRSG
 SVEVHAKVEKVRGSDIKDYGGDVQRFIRELNEVFIDDAYTLAGFAIPNQAKTPAIQQECAARGWDCDSETHLKL
 PGTQHINVDQYAQCGGGCSGNPYDQTWGLNPRGWGESHELGHNLQVNRLKVYGGRSGEISNQIFPLHKDWRV
 LREFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLWEDPGTYALNGERMAFYTQWVHYWADLKNLPLQ
 WDIWTL LYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGDASSTDGNDNLLGLSWLTQRDQRPTFALWGIRT
 seña) SAAAQAQVAAYGF AEQPAFFYANNRTNEYSTVKLLDMSQGSPA WPF

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))

MTIKRFITNLLALFTLFTVSLACKDTEKSIINSSFSISEEYLIQNLDKSSTSVQIPINTSMELAQWSVSYEAN
WLQCSKQKTAAGTFLRITVNENTGETKRTANIKVTSTTATYITITVNQYAKGEVIVEGDIKVTPTGGKASEHQ
EGQDIENYTDGKFSTDGAAPFHTFWGQSAKFPVTLEYFKGDTEIDYLIYYTRSGNGNFGKVKVYTTTNPDRS
DYTLQGEYDFKEQNAPSKVSFSEGIKATGIKFEVLGSLGDFVSCDEMEFYKTNITDKTLDRQLLTVFTDITCTE
IKNNVTNEQIQALPDYFVRIAEAVRDNTYDKWEKEFRIRSYEPYSNIAEWADKIMTKKYSDLNFTGISVKAG
DDIIVLVGDTYQGNISMQCIWETGTEYKQTASSGDEVYMLNPGVKNLTMKGEQQLFVVMYNTELTSNTANFIKIH
IPLGSGTVNGFFDLKEKTEDEKYAELLKKSTHRYFCIRGEKIMFYFHRNKLELYVPNNILSAIHLWDNIVGWQ
QELMGIDDVRFPSQVNNHLFAISPEGSYMWASDYQIGFVYTYLGNILLEDNVMAEDNAWGPAHEIGHVHQAAI
NWASSTESSNNLFSNFIIYKLGKYKSRGNGLGSVATARYANGQAWYTMGDATHQNEDETHMRMNWQLWIYYH
RCEYKTFWFQTLFKLMREVNMTGEDPGKKQLEFAKMASKAANQNLTDFENWGFFEPVNTTIEQYCTYKYYV
SDAMIREAKYMAQFPAPKHAFCYIEDRKSEFPNDYRYSAGVDVGYTQPKENQKITKAITAELAGKRVSI
QNGDEAVAFELRENDENGKLLYFSTFTTFFILPSSILMVNAKLYAVQADGKRILL

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)

MNKRKIAAIIILATMITNLSTATTIDVLAQELNNTKNNKVEVSHDDESHQARVSKFDLYNSDKLDAYNQEFQVSR
SNIRKINNNGGKYNSSTIDKAIIDGNLETHWETGKPNANFTNEVVVTFNEITNIDRIVSARRDSARGKFAK
EFEIYASLKDEGDDFNLVSSGEYTESTRDLVEIKFNPTDFKRLKFKFKKADQNWASAAEFMFYKEDKLEKFN
GLFTDSSMNKVSEEFNTLEKLNAFENELKDHPIDYLYKEGLNNARAILTETSENPTKATLGQITYNLNDYNN
QYRMPYKNIKAIKNNGRHYAAQNIKAIIDNDVNTYWETGTLNSSFNNEVEVEFNDLVTLDRIVYGSQSDLK
GFAEEVYIYASRTSGDITYKLVAATGAHEATKGLVEAKFEPTTEFKRVKFKFKKSKQNSATLNELMFYKPEDEVYS
SIPKLFDTGTMSELSEEFNSLEKINAFKEKARNHPLYNDFNETIELAESLISNPKEDVLELEMRGDSISEAK
KRKVNFQDWQITGLSAPAGDKITVYVDVAREGDPTPTLLYKQSLTQGGATSFQLKPGKNEITITPEINYESNG
IPKDVIOGGDLFFETNYKSDSQKRAKPVRIEGASKYEPVILGKSDENEVMKELEAYVEKIKAEPKTTPNIFAVS
SNKSLFVQATYALDWYKNNKTPKYTAEQWQYIADAMGFWGFDNSKDVNSDFNFRIMPVKNLSGGAFMNA
GNGVIGTRPGNQDAIILANEGWQVARELGHNFDTGGRTIVEVTNNMMPLFFESKYKTKTRIITQNIWENNTYP
KVGLDDYSNNELYNKADSTHLAQIAPLWQILYDNTFYGKEERQFRERDFGNKNREDIYKSWVVAASDAMELD
LTEFFARHGIRVDKVKEDLAKYKPKDKKIYYINDLAMNYKGDGFTEAKVSVSTSGSNGNKLKLSFSVDDENK
DNILGYEIRRDGKYVGFTSNDSSFVDTKSNLDEDDGVYVTPYDRKLTNLPLEVNALQPTLSVNPVITLALGEE
FNEEYIYAKDIKGNLSLESVKVKSNNVNTSKVGEYEVLYSLEDKNGNEYTKTSKVNVS SRKEYMSDLTPKQS
SNGWGTVRKDKSISGGVIGLTRDGDVFDYNNKGLGLHSNAEYVYDLEGKDYYFESYVGVDKAMSSRPASSVIF
KVLVDGEEKFNSGVMRSTTPQKYVKVDVKNKELKLI VNDAGDGDSSDHASFGDAKLATLSSKPIIKGENLAY
SMDEKVDLMKGITATDIEDGNITSKVQIKSSDFVEGKSGIFTVVYSVTDSGLTSECRTIAVTDKETQLSDL
NWKSATIGSGSVRKDRVSGNQIRLLNEDNSVETFAKGIGTHSYSEIVYNSEGYDIFDTWVGIDRHVADKKVS
SVKFKVYVDGELKAEITDVMRIDTPKKRLVVDVRNSKEIKLVVDVADMGNNWDHADWADAKFRNLAEYDASELN
KAIEEARKLDLNNYTEESSEALKNAISKGEEALLSKDKETINSALAEELNKEMNSLVKVDLNAVINIPDRYLLK
SIQNQLNKTGDTITLGDMSLTTLTSLGVEDLTGLENAKNETLNMDYNEVKDLRFLSKLKLINTLNAEQEFTIA
AGELKPSNGKVIKDSKVYNREGKNVAKTIRVVDKNGNTILEQDAKDEFTINTRDLSSGLYGVHVLFEDEGFSG
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.

- 5
- SEQ ID NO: 35Pseudomonas aeruginosa PAO1 (doble mutante con secuencia de señal eliminada)
- ATQEEILDAAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLSSQAYDPGAASMWVLPANPDNTL
PFLIGDKGRVLAASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWLNVGDPGAATAKDFKVSVVGVVDK
TAALNGLKSAGLQPADAAACNALTDASCASSTKLLVLGNGASAAASLSATVRARLQAGLPILFVHTNGWNQSSSTG
QQILAGLGLQEGPYGGNYWDKDRVPSSRTRRSVELGGAYGQDPALVQQIVDGSWRDIDYDWSKCTSYVGRRTT
DDVPGLSDFS KRVDVLKGALDAYNQKAQNLFALPGTTSRLRLWLLWADAVRQNI RYPMDKAADTARFQETTFVAD
AIVGYVREAGAAQKELGSYAGQRQQSMPVSGSEETLTTLPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLAV
GLNTQRIGSTRLWNTRQYDRPRFLKSPDIKLANQSVLVSPYGGLLQLVYSGATPGQT VTVKVTGAASQPF
DIQPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSI DKDYGGDVQRFIRELNEVFIDDAYTL
AGFAIPNQAKTPAIQQECAARGWDCDSETLHKLPGTQHINVDQYACGGGCSGNPYDQ TWGLNPRGWGESAA
GHNLQVNRLKVGGRSGEISNQIFPLHKDWRVLRREFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLWE
DPGTYALNGERMAFYTQWVHYWADLKNLPLQGWDIWTL LLYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGDA
SSTDGNDNLLGLSWLTQRDQRPTFALWGIRTSAQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGS
PAWPF
- Motivo
de metaloproteasa alterado subrayado
- 10
- SEQ ID NO: 36Bacteroides thetaiotaomicron VPI-5482 (doble mutante con señal eliminada y otras secuencias inmaduras)
- DKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDDLNP TGISVKAGDDIIVLVGDTYQGNISMQCIWETGTEYKQ
TASSGDVYMLNPGVKNLTKMGEGQLFVMYNTELTSTAKPIKIHIPLGSGTVNGFFDLKEHKTDEKYAELLK
STHKYFCIRGEKIMFYFHRNKLL EYVPNNILSAIHLWDNIVGWQQELMGIDVDRPSQVNNHLFAISPEGSYMW
ASDYQIGFVYTYLGNILLEDNVMAAEDNAWGPA~~AA~~IGHVHQAAINWASSTESSNNLFSNFI IYKLGKYKSRGN
GLGSVATARYANGQAWYNMGDATHQNEDETETHMRMNWQLWIYYHRCEYKTD FWTFLFKLMREVNMTGEDPGK
KQLEFAKMASKAANQNLTDFEMWGFEPVNTTIEQYGTYYVSDAMIREAKEYMAQFPAPKHAFQYIEDRK
KSEFPSNDYRYSAVGDVGYYTQFKENQKITKAITAELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTTF
EIPSSILMVNAKLYAVQADGKRILL
- Motivo
de metaloproteasa alterado subrayado
- 15
- SEQ ID NO: 37Clostridium perfringens (doble mutante con señal eliminada y otras secuencias inmaduras)
- VLELEMRGDSISEAKKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQKLPK
KNEITIP EINYESNGIPKDVITQGGDLFFTNYSKSDSKRAPKVRIEGASKYPVILGKSDENEVMKELEAYVEK
IKAEPKTPNIFAVSSNKSLEFVQATYALDWYKKNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFRI
MPMVKNLSGGAFMNAGNGVIGIRPGNQDAILAANKGW~~GVA~~AAALGHNFDTGGRTIVEVTNNMPLFFESKYKTK
TRITDQNIWENNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLYDNTFYGKFERQFRERDFGNKNREDI
YKSWVVAASDAMELDLTEFFARHGIRVDDKVKEDLAKYKPKDKKIYYLNDLAMNYKGDGFTENAKVSVSTSGS
NGNIKLSFSVDENKDNILGYEIRRDGKYVGFTSNDSFVDTKSNLDEDGVYVVTYPYDRKLNLTLPNIEVN
- Motivo
de metaloproteasa alterado subrayado
- 20
- SEQ ID NO: 38Pseudomonas aeruginosa PAO1 (doble mutante con secuencia señal eliminada, con Met en el N-terminal, enlazador/etiqueta en el C-terminal)
- MATQEEILDAAALVSGDSSQLTDSHLVALRLQQQVERIRQTRTQLLDGLYQNLSSQAYDPGAASMWVLPANPDNT
LPFLIGDKGRVLAASLSLEAGGRGLAYGTNVLTQLSGTNAAHAPLLKRAVQWLNVGDPGAATAKDFKVSVVGVVD
KTAALNGLKSAGLQPADAAACNALTDASCASSTKLLVLGNGASAAASLSATVRARLQAGLPILFVHTNGWNQSSST
GQQILAGLGLQEGPYGGNYWDKDRVPSSRTRRSVELGGAYGQDPALVQQIVDGSWRDIDYDWSKCTSYVGRRTT
CDDVPGLSDFS KRVDVLKGALDAYNQKAQNLFALPGTTSRLRLWLLWADAVRQNI RYPMDKAADTARFQETTFVA
DAIVGYVREAGAAQKELGSYAGQRQQSMPVSGSEETLTTLPSAQGFTAIGRMAAPGKRLSIRIEDAGQASLA
VGLNTQRIGSTRLWNTRQYDRPRFLKSPDIKLANQSVLVSPYGGLLQLVYSGATPGQT VTVKVTGAASQPF
LDIQPGEDSSQAIADFIQALDADKADWLEIRSGSVEVHAKVEKVRGSI DKDYGGDVQRFIRELNEVFIDDAYTL
LAGFAIPNQAKTPAIQQECAARGWDCDSETLHKLPGTQHINVDQYACGGGCSGNPYDQ TWGLNPRGWGESAA
LGHNLQVNRLKVGGRSGEISNQIFPLHKDWRVLRREFGQNLDDTRVNYRNAYNLIVAGRAEADPLAGVYKRLW
EDPGTYALNGERMAFYTQWVHYWADLKNLPLQGWDIWTL LLYLHQRQVDKSDWDANKAALGYGTYAQRPGNSGD
ASSTDGNDNLLGLSWLTQRDQRPTFALWGIRTSAQAQVAAYGFAEQPAFFYANNRTNEYSTVKLLDMSQGS
SPAWPFPFGSGHHHHHH
- 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)
- MDKWEKEFRIRSYEPYSNIAEWADKLMTKKYSDDLNP TGISVKAGDDIIVLVGDTYQGNISMQCIWETGTEYK
QTASSGDVYMLNPGVKNLTKMGEGQLFVMYNTELTSTAKPIKIHIPLGSGTVNGFFDLKEHKTDEKYAELLK
KSTHKYFCIRGEKIMFYFHRNKLL EYVPNNILSAIHLWDNIVGWQQELMGIDVDRPSQVNNHLFAISPEGSYMW
WASDYQIGFVYTYLGNILLEDNVMAAEDNAWGPA~~AA~~IGHVHQAAINWASSTESSNNLFSNFI IYKLGKYKSRGN
NGLGSVATARYANGQAWYNMGDATHQNEDETETHMRMNWQLWIYYHRCEYKTD FWTFLFKLMREVNMTGEDPG
KKQLEFAKMASKAANQNLTDFEMWGFEPVNTTIEQYGTYYVSDAMIREAKEYMAQFPAPKHAFQYIEDR
KKSEFPSNDYRYSAVGDVGYYTQFKENQKITKAITAELAGRKVSIQNGDEAVAFELRENDENGKLLYFSTFTTF
FEIPSSILMVNAKLYAVQADGKRILLGSGHHHHHH
- 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)
 MVLELEMRGDSISEAKKRKVWNFQDWQITGLSARAGDKITVYVDVAEGDPTPTLLYKQSLTQHGGATSFQLKP
 GKNEITPEINYESNGIPKDV IQGGDLFF TNYKSDSQKRAPKVRIEGASKYPVFILGKSDENEVMKELEAYVE
 KIKAEPKTTPNIFAVSSNKSLEFVQATYALDWYKKNNKTPKYTAEQWDQYIADAMGFWGFDNSKDVNSDFNFR
 IMPMVKNLSGGAFMNAGNGV GIRPGNQDAILAANKGWGVAAALGHNFDTGGRTIVEVTNNMMPLFFESKYKT
 KTRITDQNIWENNNTYPKVGLDDYSNNELYNKADSTHLAQLAPLWQLYLYDNTFYGKFERQFRERDFGNKNRED
 IYKSWVVAASDAMELDLTEFFARHGIRVDDKVKEDLAKYKPKDKKIYYLNDLAMNYKGDGFTENAKVSVSTSG
 SNGNIKLSFSVDDENKDNILGYEIRRDGKYVGFTSND SFVDTKSNLDEDEGVYVVTPTYDRKLTNLNPIEVNGSG
 HHHHHH

Motivo

de metaloproteasa alterado subrayado

5

SEQ ID NO: 41HELGH (motivo de metaloproteasa)

SEQ ID NO: 42HEIGH (motivo de metaloproteasa)

10

SEQ ID NO: 43GVAHELGHNF (motivo de metaloproteasa)

SEQ ID NO: 44HALGH (motivo de metaloproteasa alterado)

SEQ ID NO: 45AELGH (motivo de metaloproteasa alterado)

15

SEQ ID NO: 46AALGH (motivo de metaloproteasa alterado)

SEQ ID NO: 47GKPRPYSPRPTSHPRPIRV (péptido de glicodrosocina con sitio O-gly en la T)

20

SEQ ID NO: 48YIYGSFK (Péptido no O-glicosilado)

SEQ ID NO: 49KKLVFFA (Péptido no O-glicosilado)

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:
 - (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,
 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.
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:
 - (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;
 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,
 - (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.
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.
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.
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.

- 5 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.
- 10 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.
- 15 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;
(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.
- 20 14. El método según cualquiera de las reivindicaciones 11 a 13,
(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:
(a) Urea de alta concentración molar;
(b) Detergente de alta concentración; o
25 (c) Un polipéptido con actividad endoproteasa específica para proteínas O-glicosiladas según cualquiera de las reivindicaciones 1 a 5.

DIBUJOS

Fig. 1

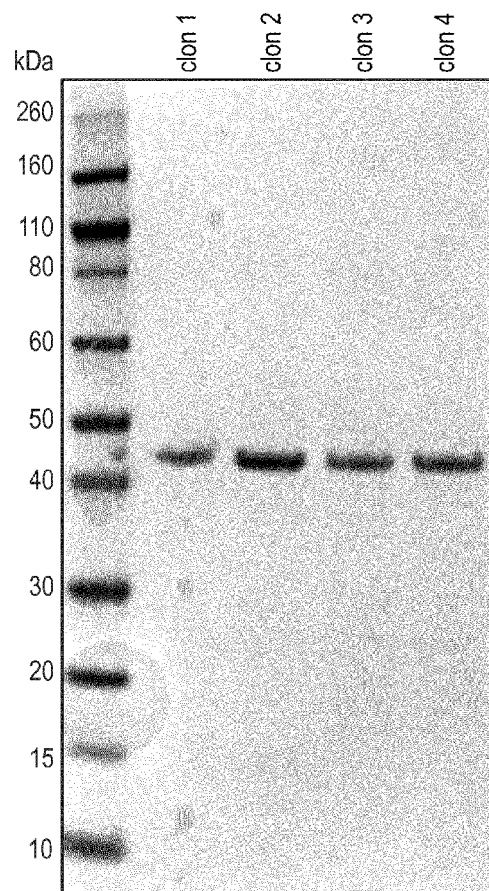


Fig. 2

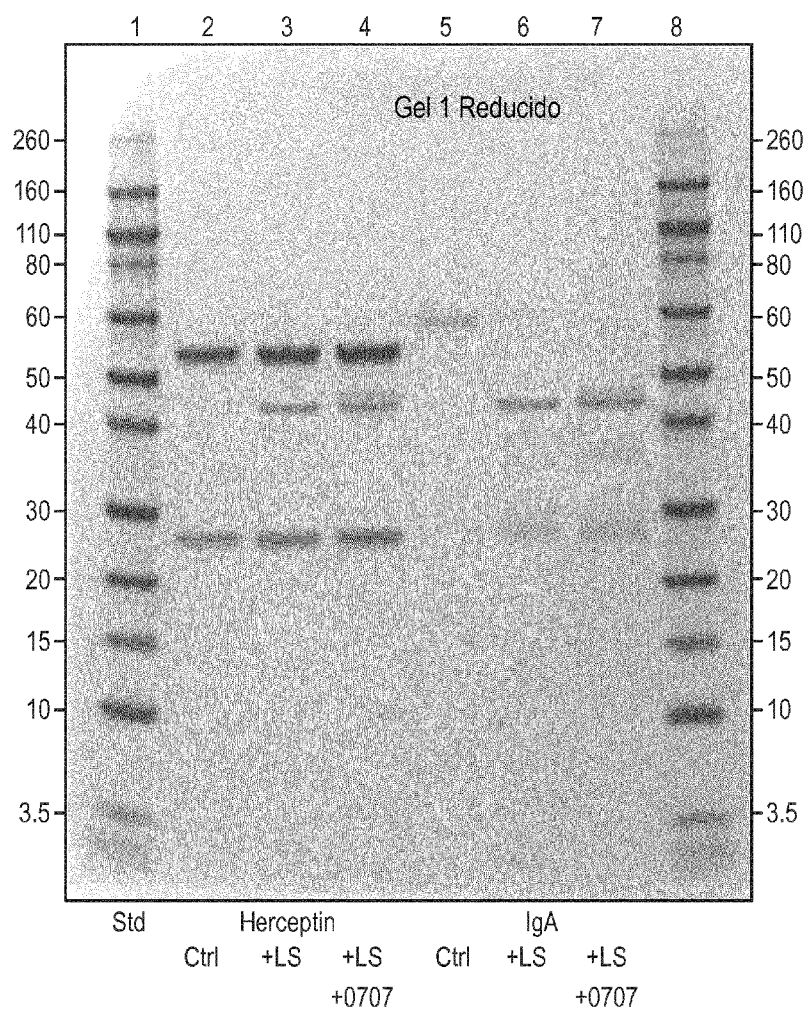


Fig. 3A

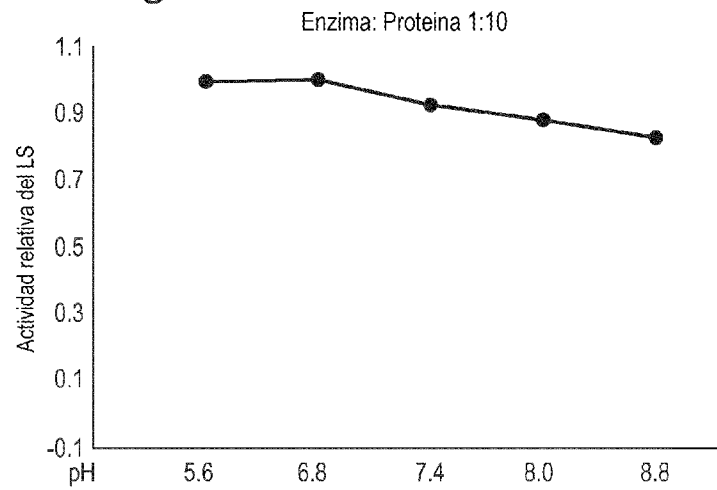


Fig. 3B

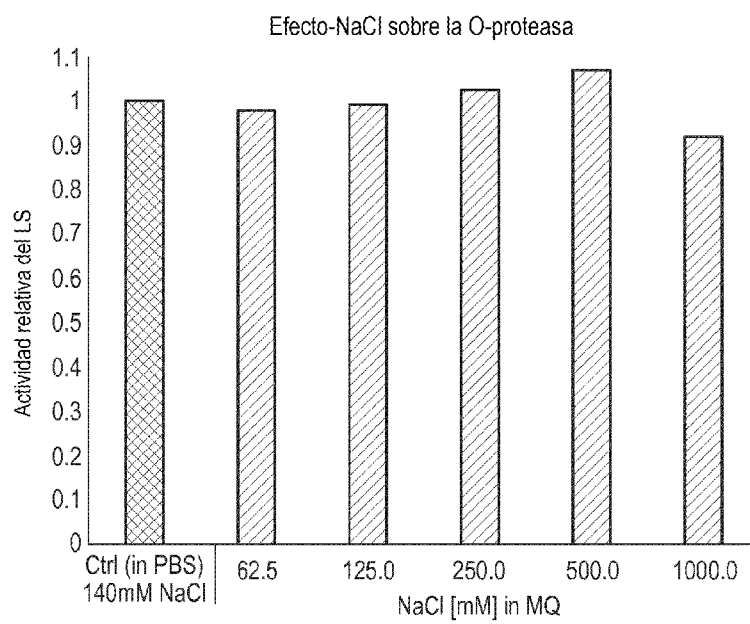


Fig. 3C

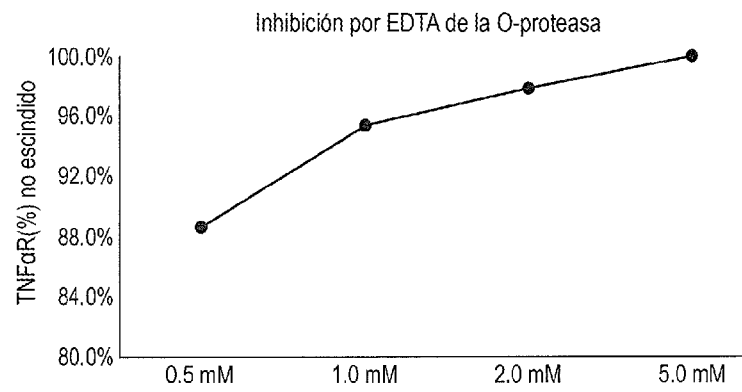


Fig. 3D

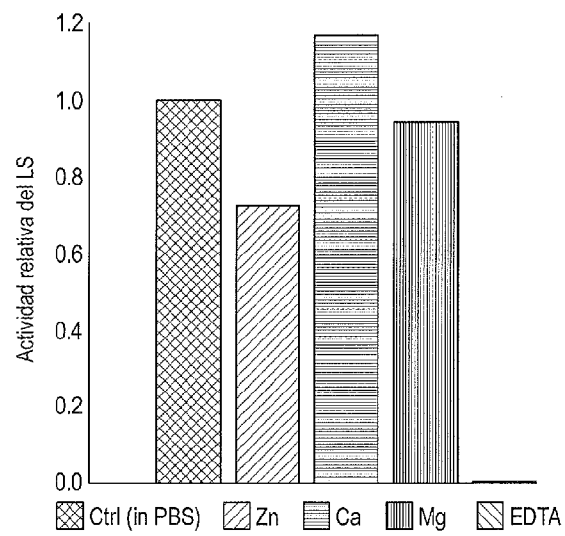


Fig. 4A

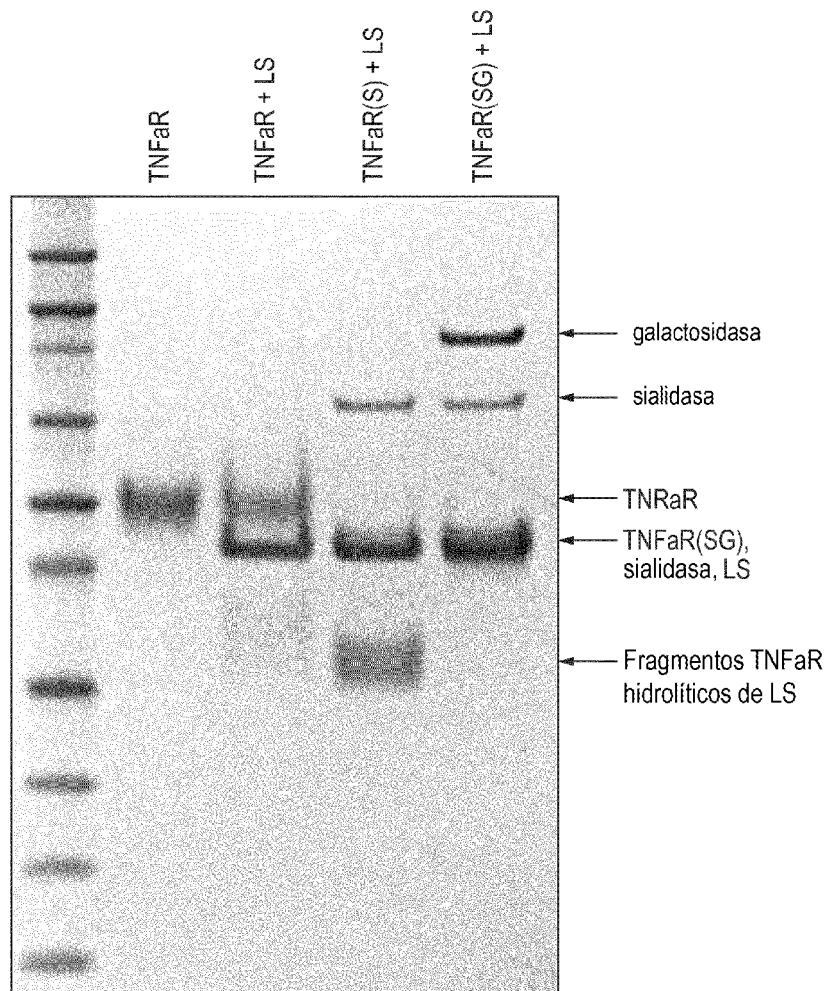


Fig. 4B

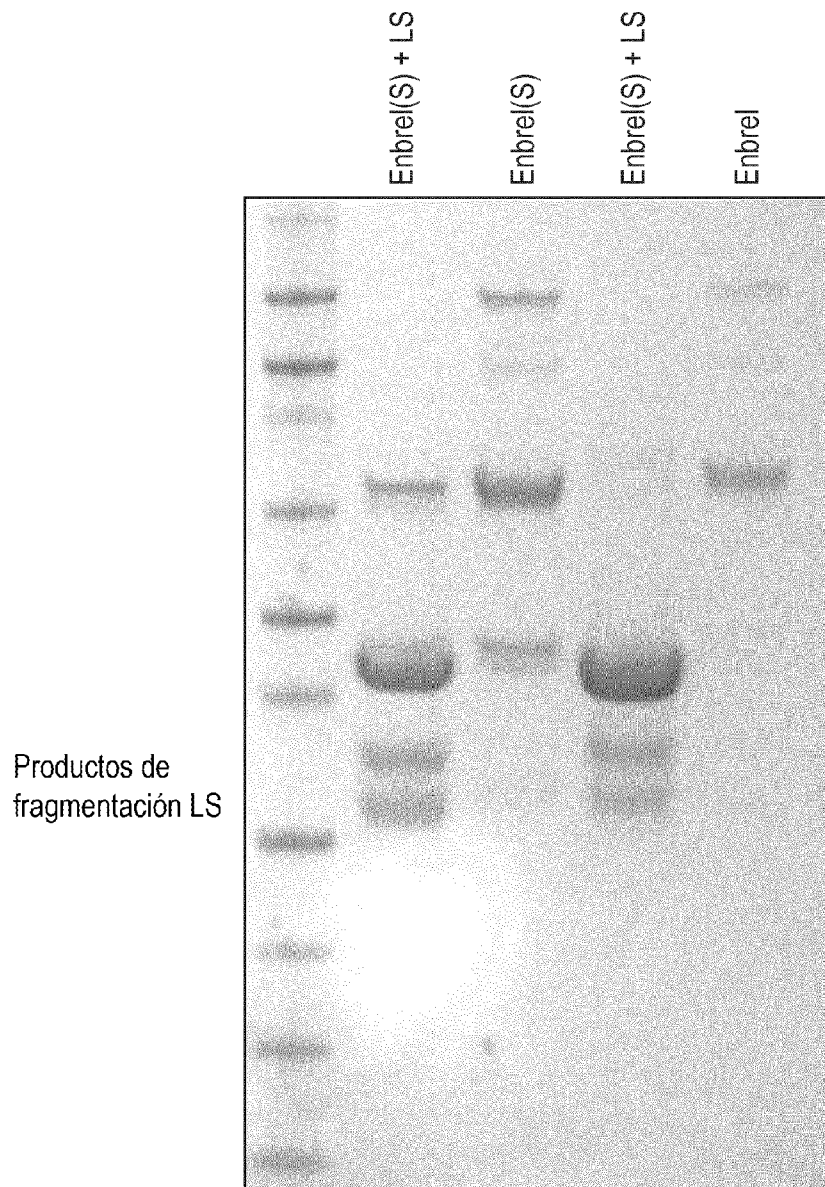


Fig. 4C

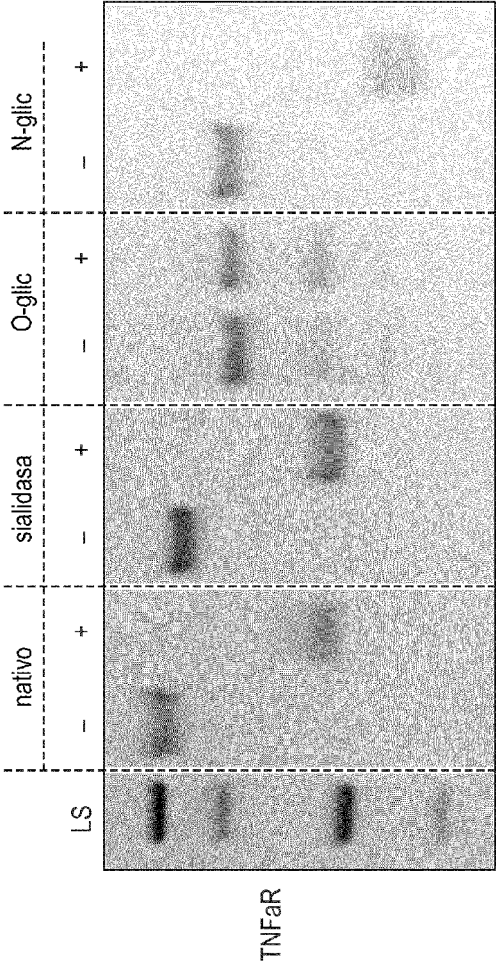
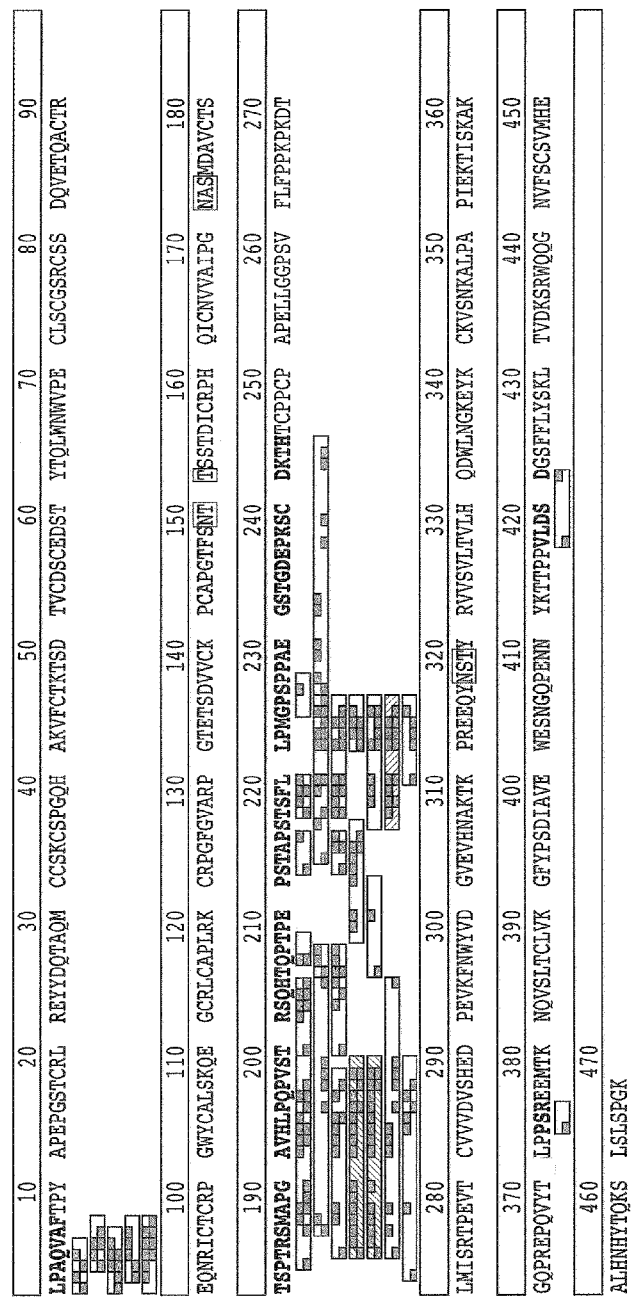


Fig. 5A

10	20	30	40	50	60	70	80	90
LPAQVAFPTY	APEPGSTCRL	REYDQTAQM	CCSKCSPGQH	AKVCTKTS	TVCDSCEDST	YTQLNNWPE	CLSCGSRCS	DQVETQACTR
100	110	120	130	140	150	160	170	180
EQNRICTRP	GWYCALSKQE	GRLCAPLRK	CRPGGVARP	GTETSDVCK	PCAPGTFST	ESSTDICRPH	QICNVVAIPG	NASMDAVCTS
190	200	210	220	230	240	250	260	270
TSPTRSNAPG	AVHLPQPVST	RSQHTQTPTE	PSTAPSTSFL	LPNGSPPPAE	GSTGDEPKSC	DKHTTCPPCP	APELLGGPSV	FLFPPKPKDT
280	290	300	310	320	330	340	350	360
LALSRTPET	CWVDVSHED	PEVKFNWYD	GVEVHNAKTK	PREEQVNSIK	RWVSVLTVLH	QDWLNGKEYK	CKVSNKALPA	PIEKTISKAK
370	380	390	400	410	420	430	440	450
GQPREPQVIT	LPPSREEMTK	NQVSLICLVK	GFYPSDIAVE	WESNGQPENN	YKTTTPPVLD	DGSFFLYSKL	TVDKSRWQOG	NVTFSCSVHHE
460	470							
ALHNHYTQKS	LSLSPGK							

Fig. 5B



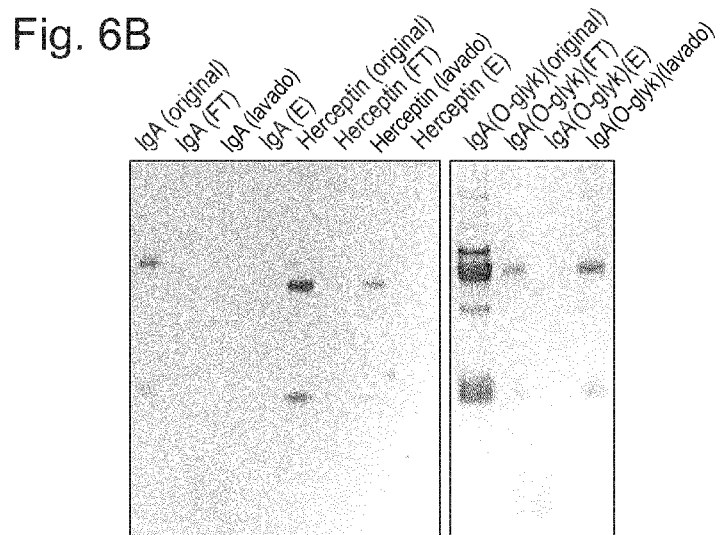
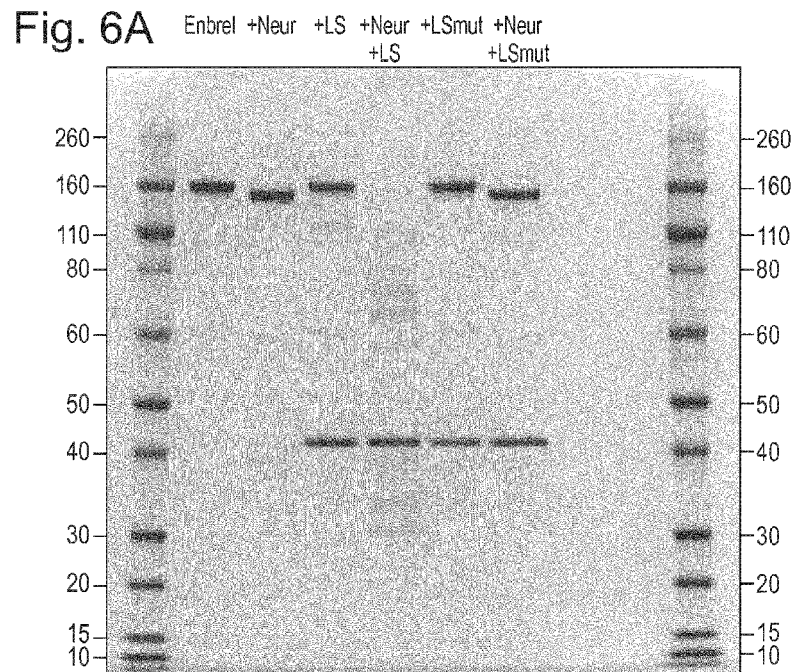


Fig. 7

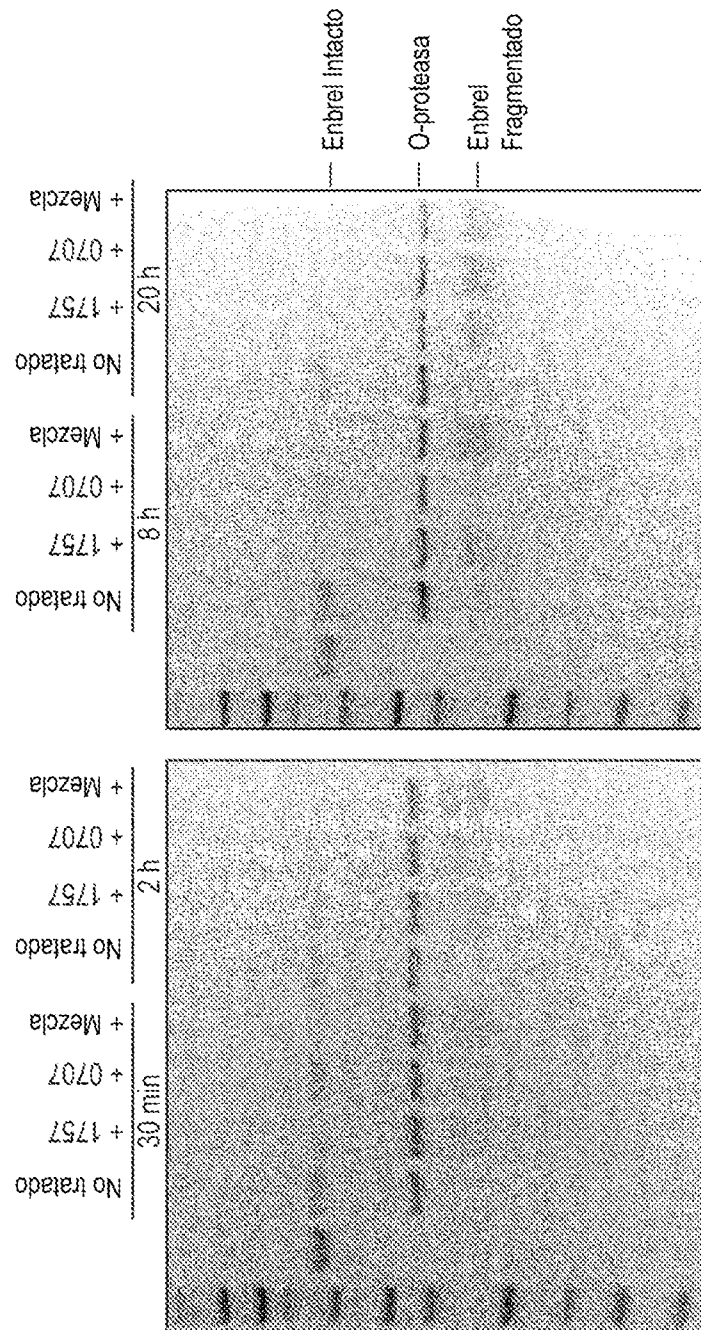


Fig. 8

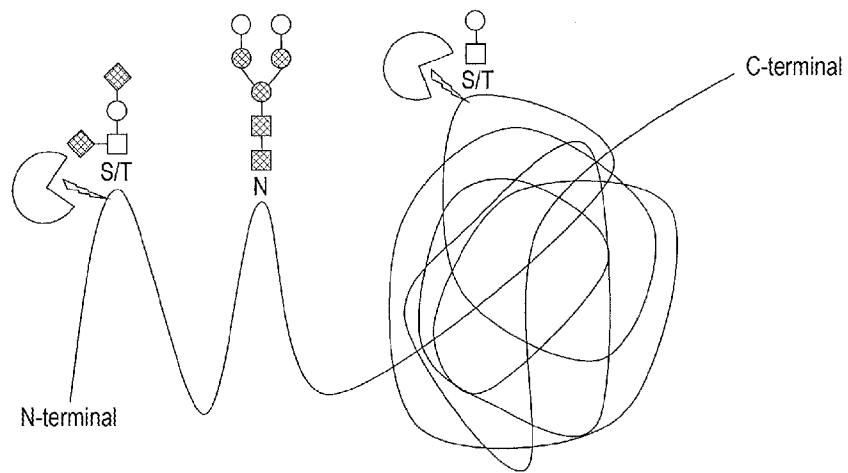


Fig. 9A

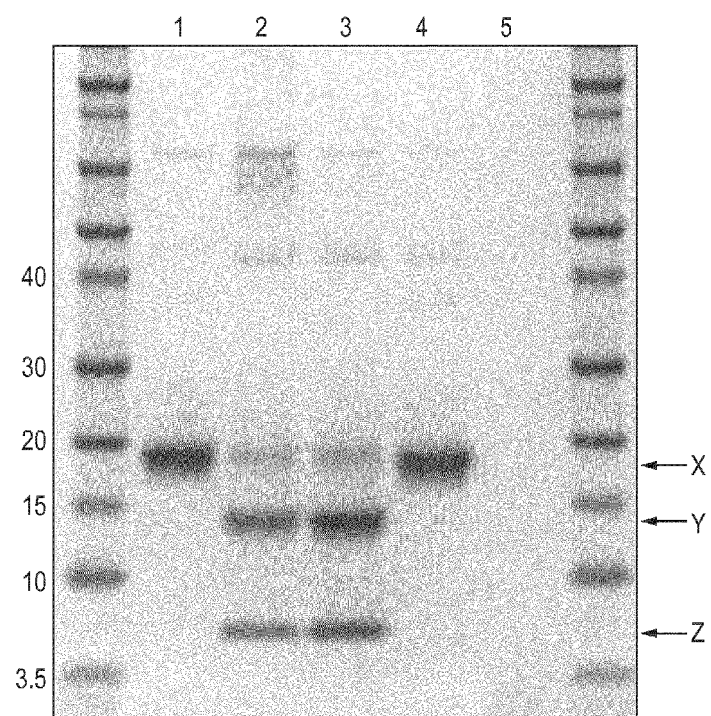


Fig. 9B

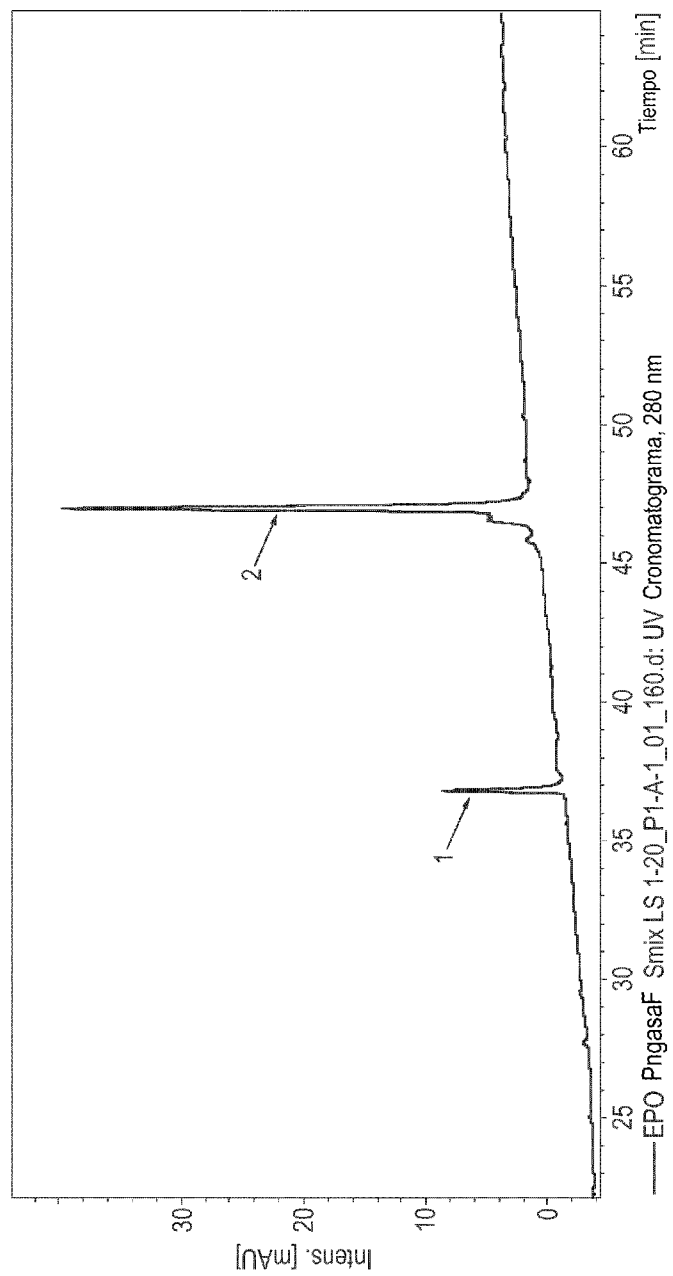


Fig. 9C

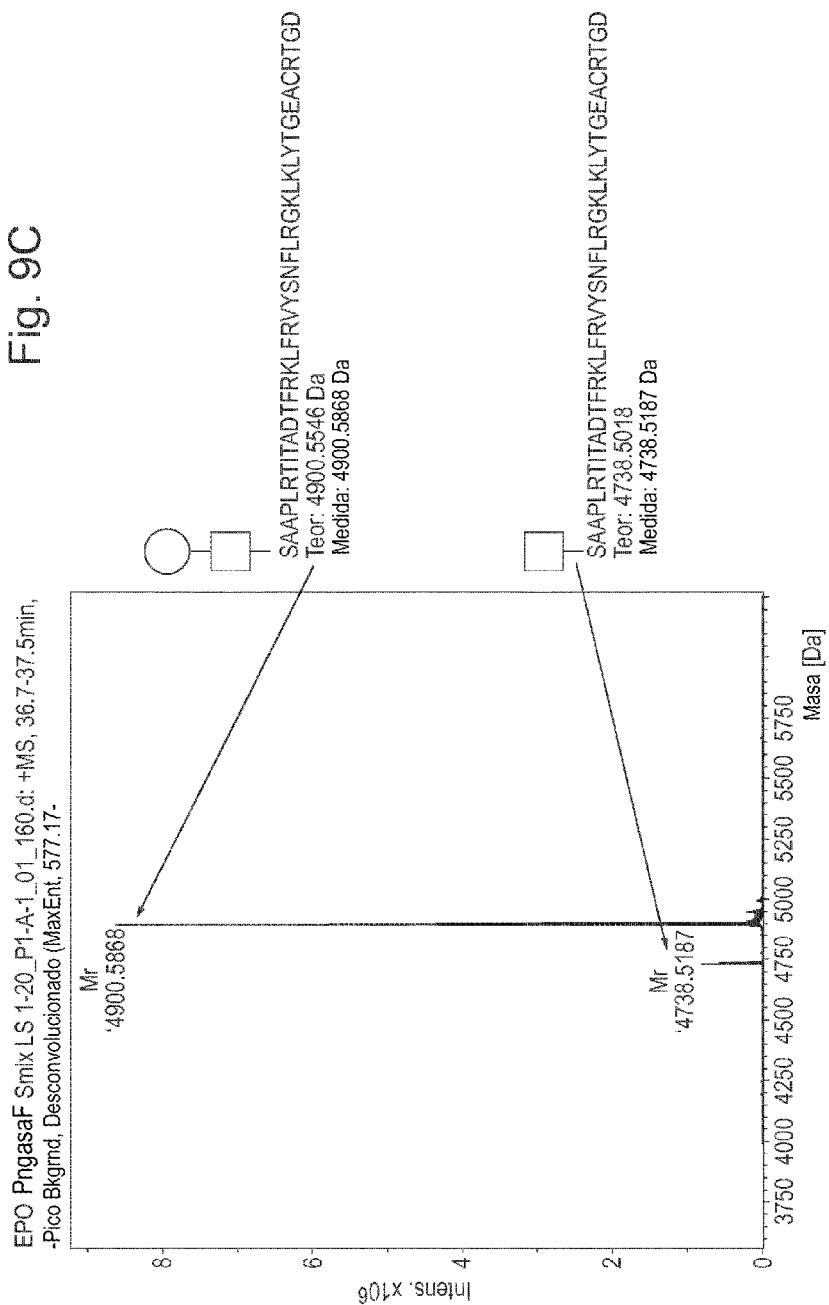


Fig. 9D

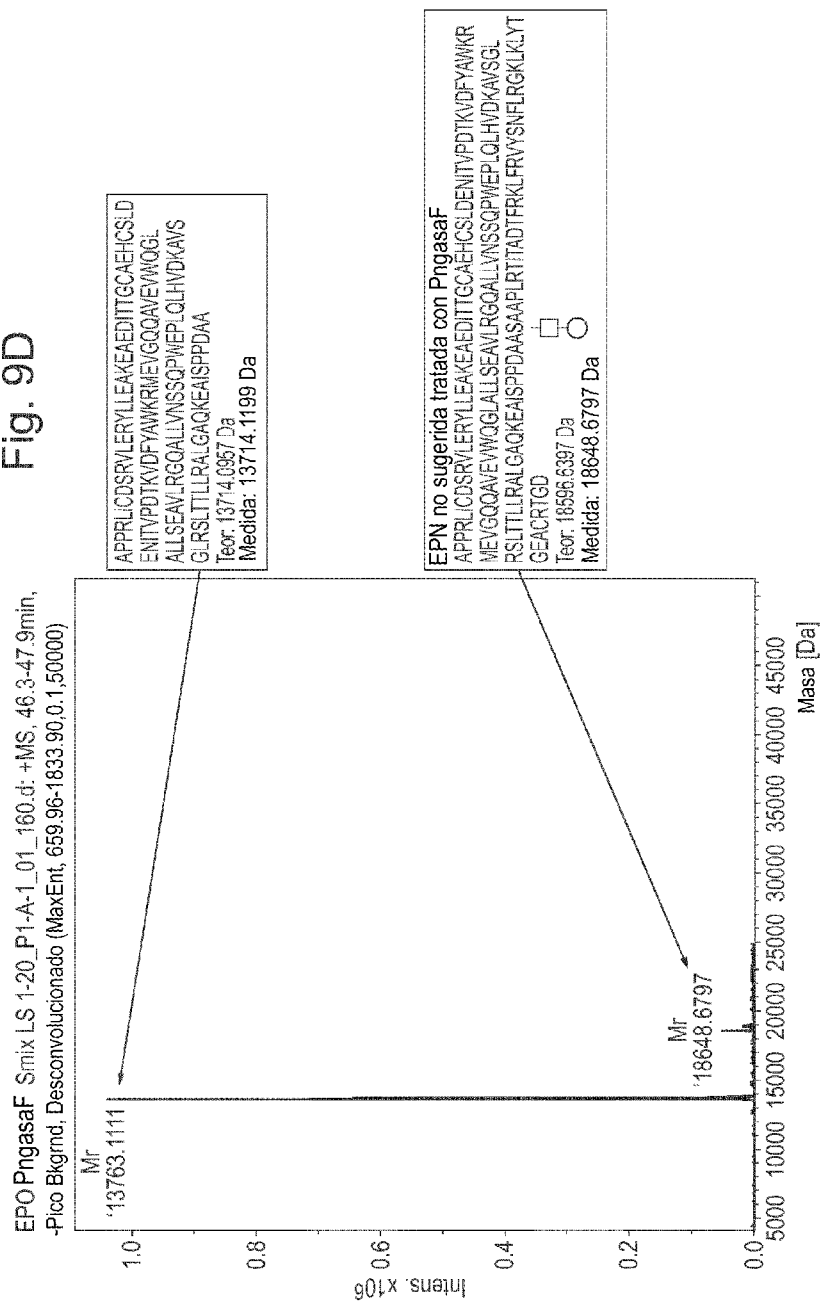


Fig. 10B

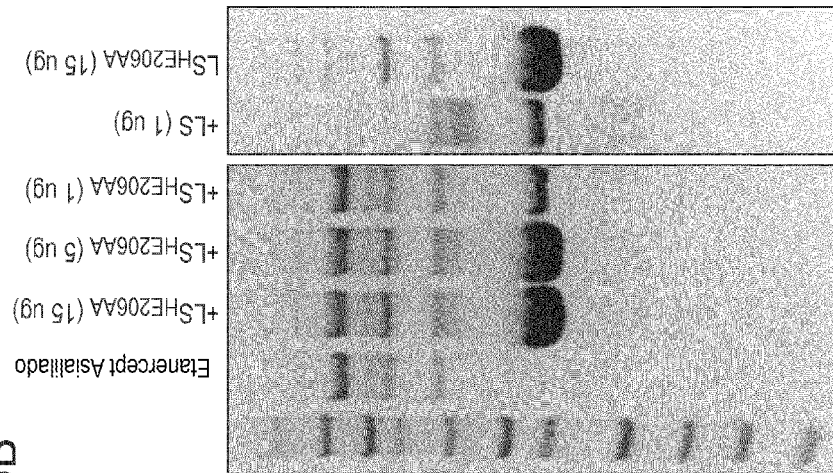


Fig. 10A

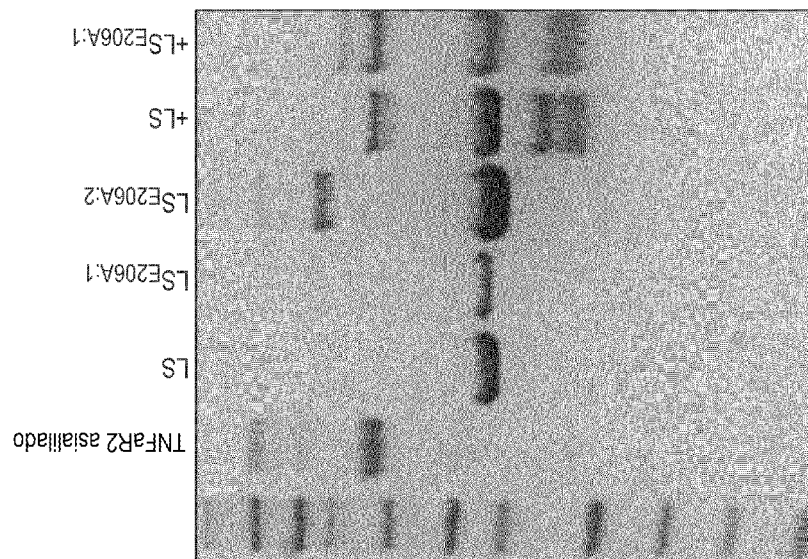
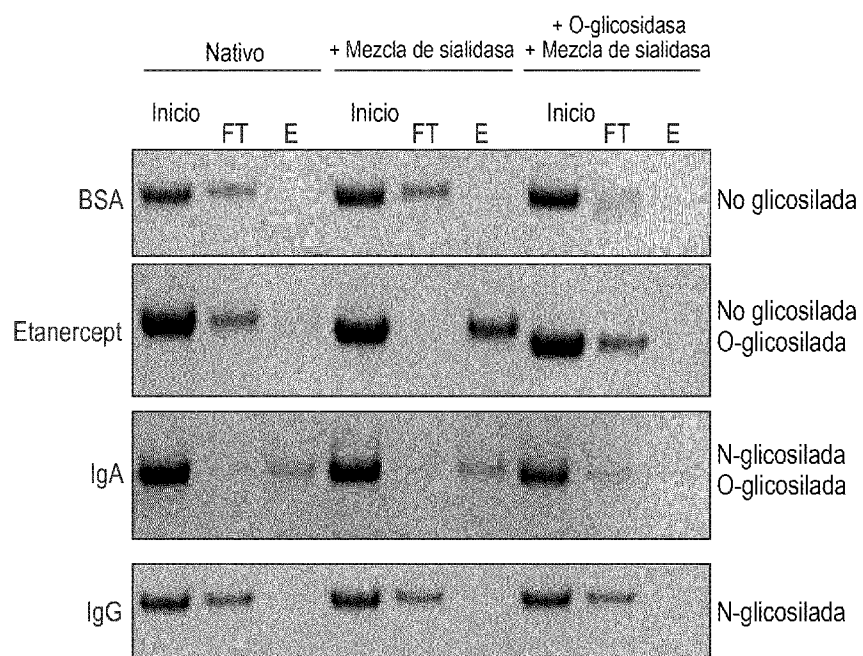


Fig. 11A



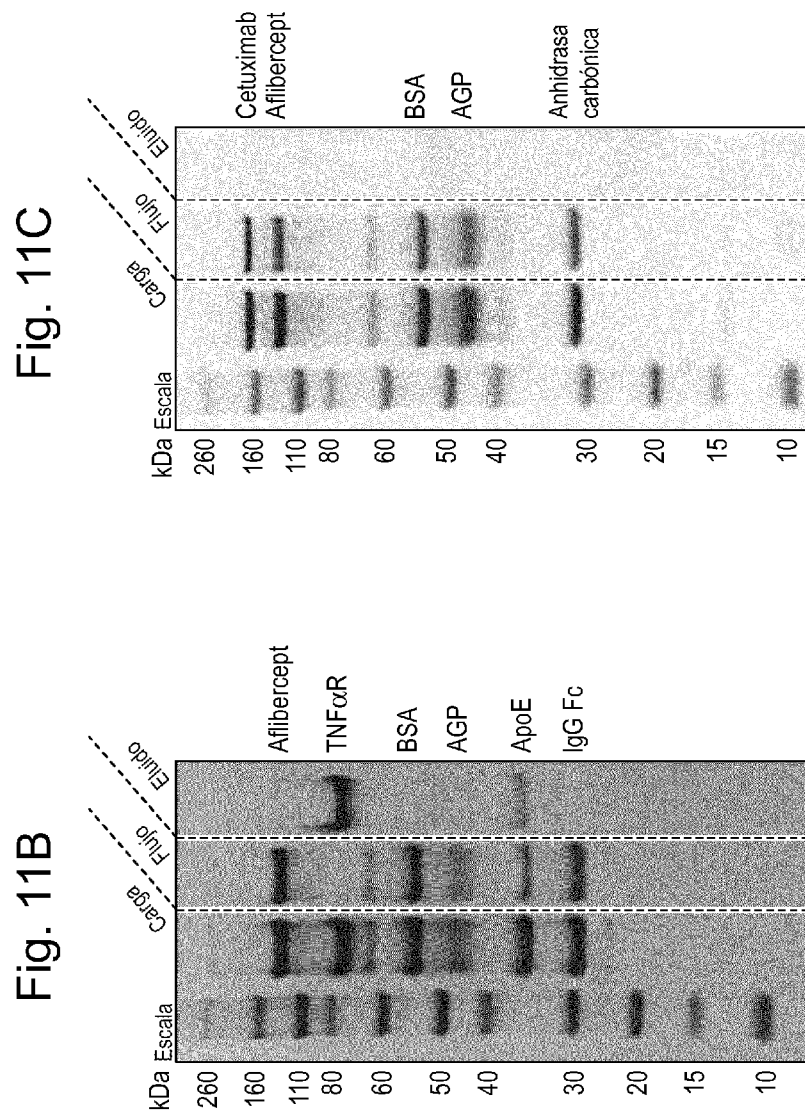


Fig. 12A

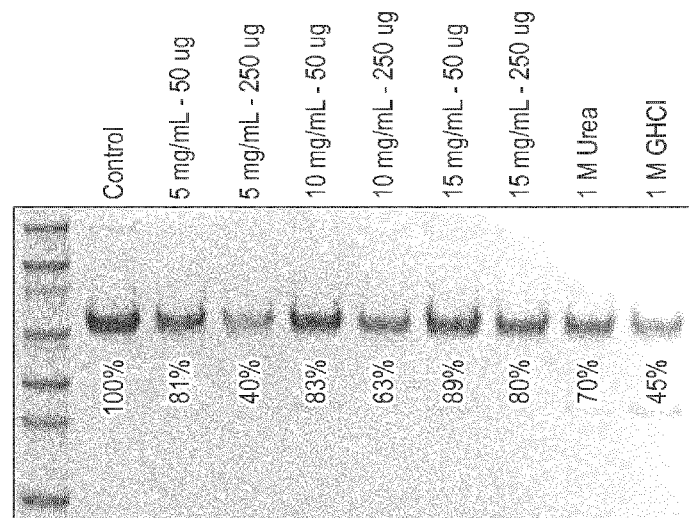


Fig. 12B

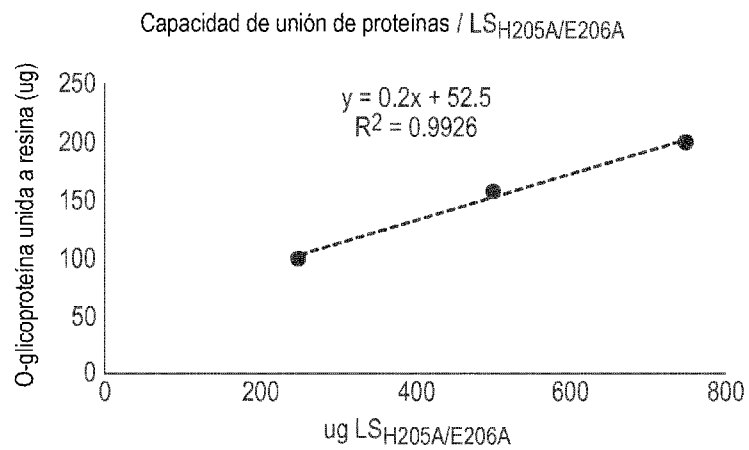


Fig. 13

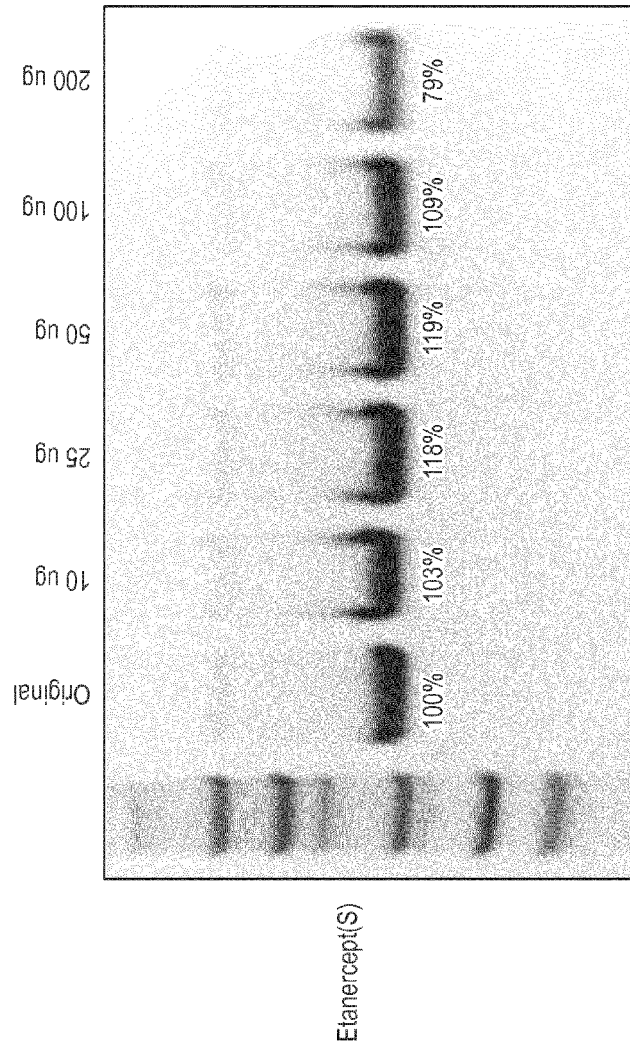


Fig. 14B

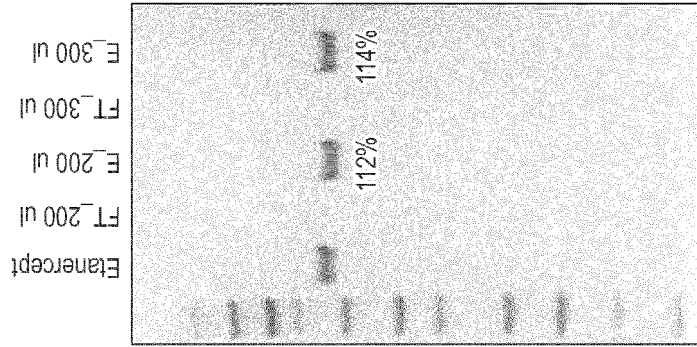


Fig. 14A

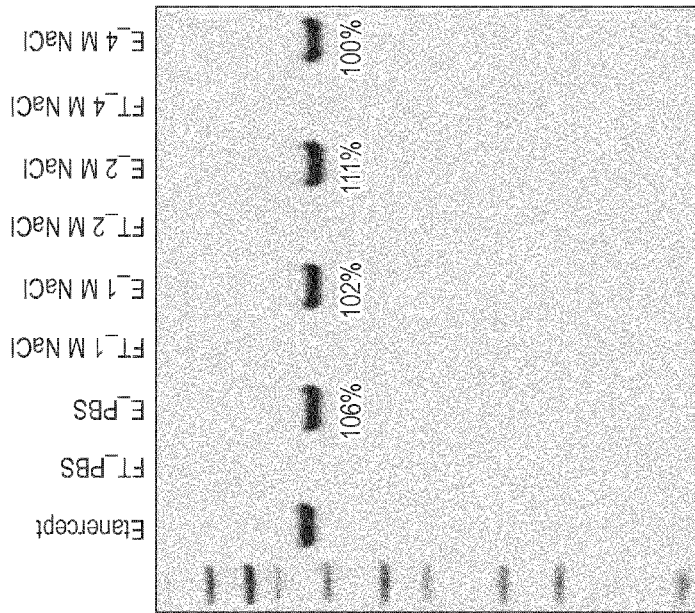


Fig. 14C

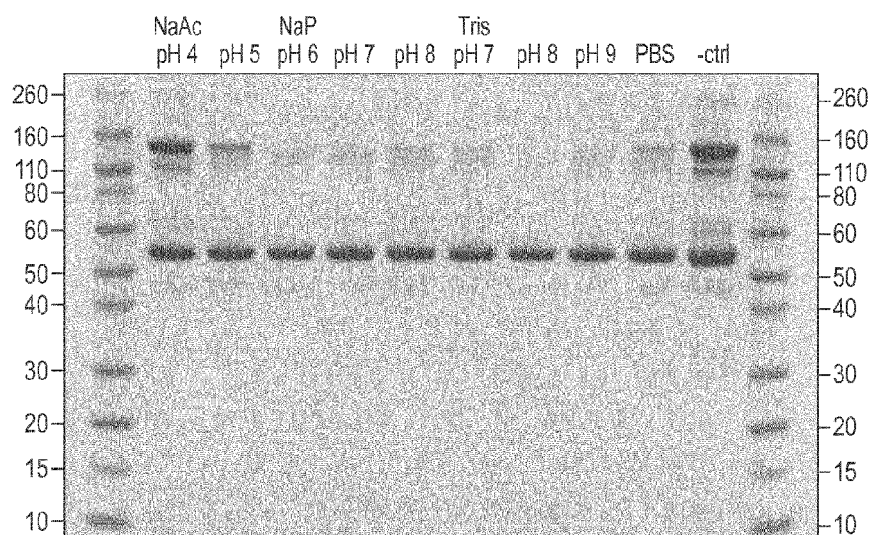


Fig. 14D

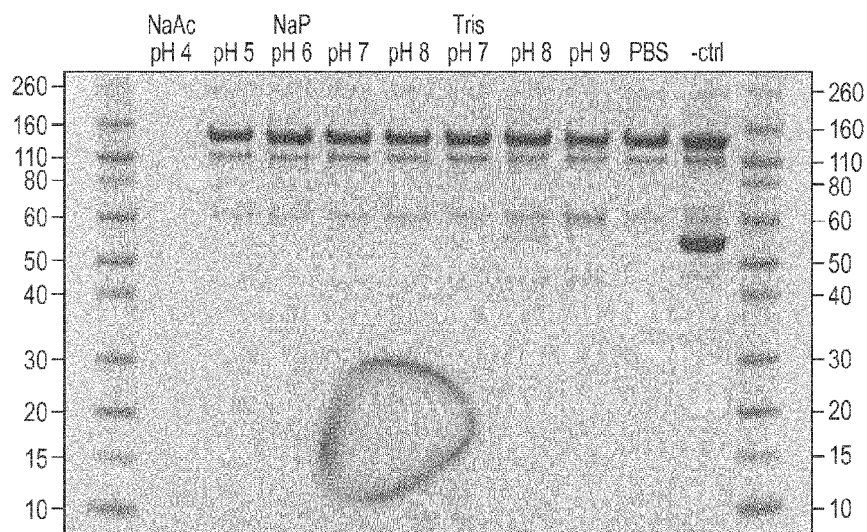


Fig. 15A

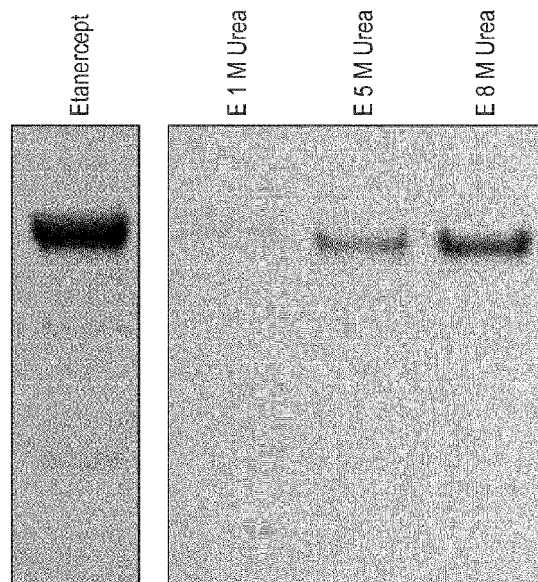


Fig. 15B

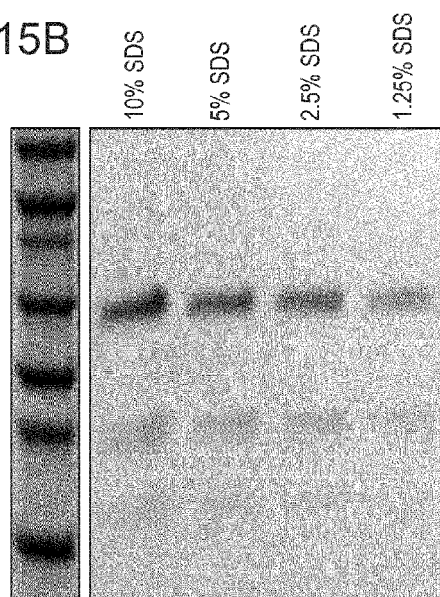


Fig. 16A

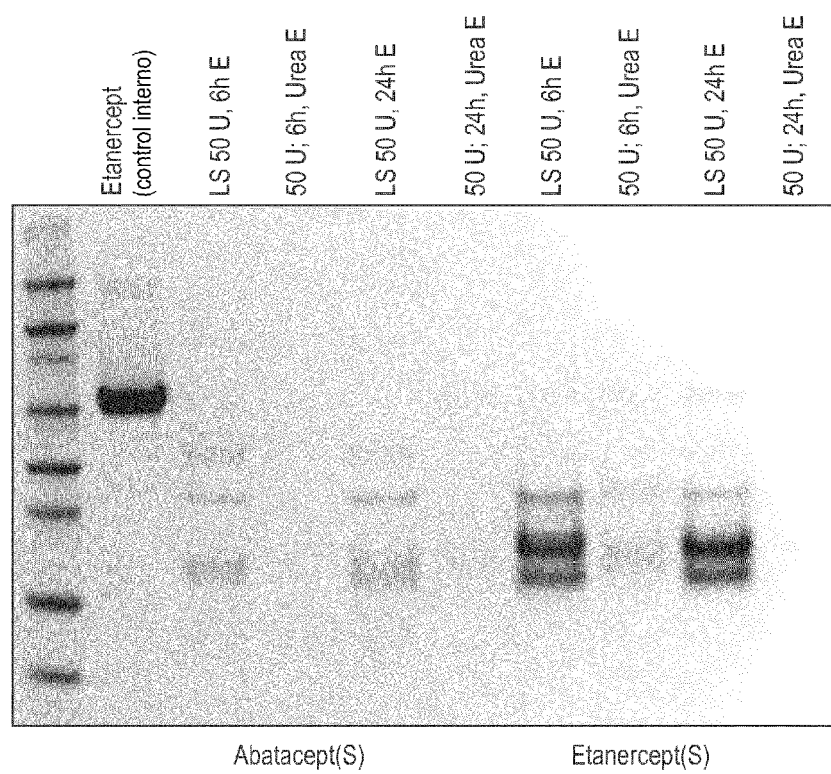
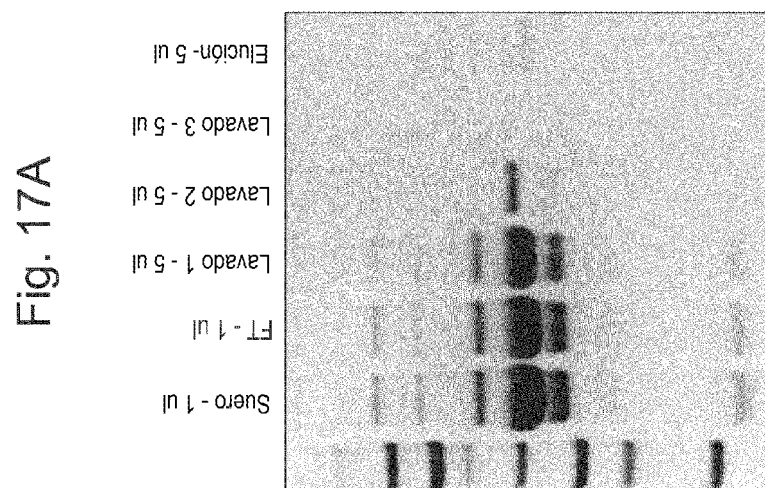
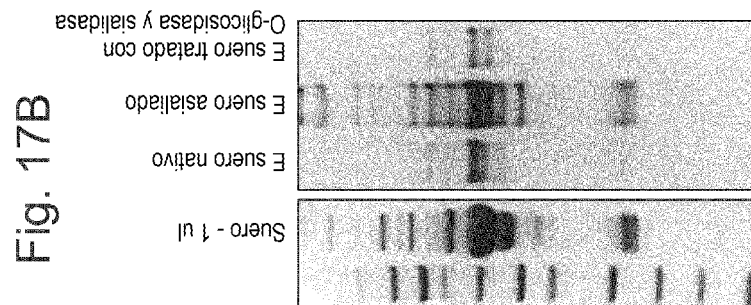
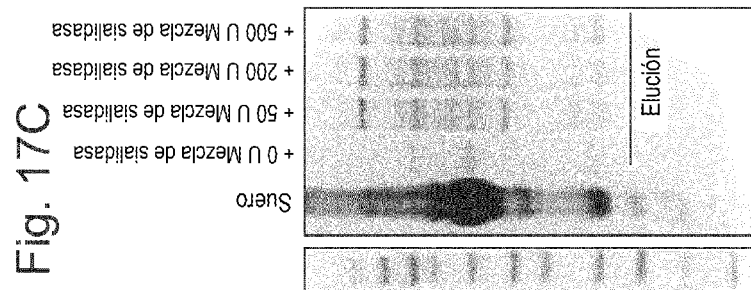


Fig. 16B.

[illegible]

Fig. 16B.2

10	20	30	40	50	60	70	80	90	100	110	120	130
LPAQVPTPY	APRPGSTRL	REYDTDTQM	COSKCSFQH	ANVCKTSD	TVCSCEDST	YQJNWNWPE	CLSGSGRSS	DOVETQACTR	ENRNICICRP	GWYCAISQJE	GGRICAPLAK	CPFGFWAR
140	150	160	170	180	190	200	210	220	230	240	250	260
GTEPSDVCK	PCLGCTFQV	ESNDICRPH	QICNVAITG	WASNDVACTS	TSPTSRMARG	AVHLPQVST	BSQHTQPTPE	PSYAPSTISTL	LMGSPSPAE	GSYDQFASC	DNHTWCPCE	ARELLGGSV
270	280	290	300	310	320	330	340	350	360	370	380	390
ELFFPRKPT	LMISRTPEVT	CYVWVSEED	PEVAFNIVD	GVEVHMATK	PRRQWNSIT	RVSVLVFLH	QMLNCKETK	CVSNKALPA	PIENTISKAK	GQPREQVYT	LPPSEEMTK	NOVSLCTLK
400	410	420	430	440	450	460	470					
GTPSDLAWE	WBSNGQPENN	YATTPPVLD	DGSTFNSKL	TOKSKWQGG	NFTSCVWHE	ALHHHTQKS	ISLSPEK					



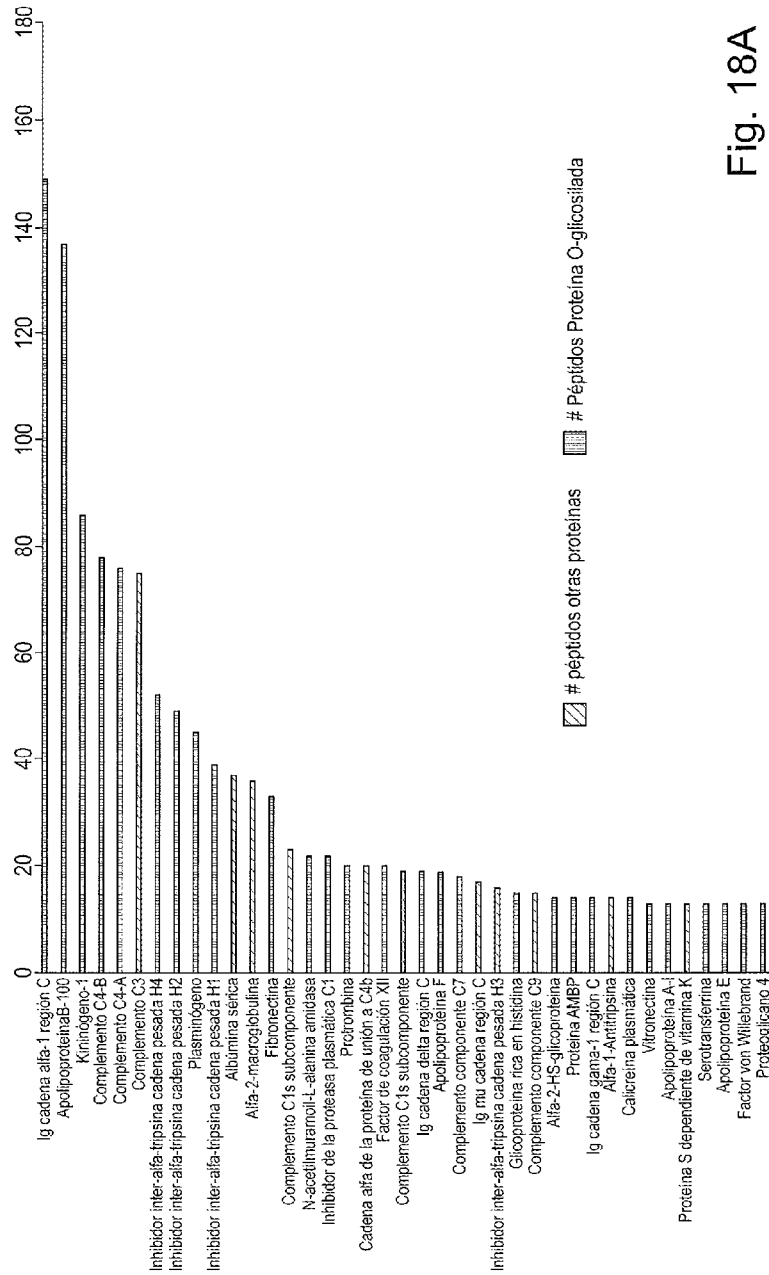


Fig. 18A

Fig. 188

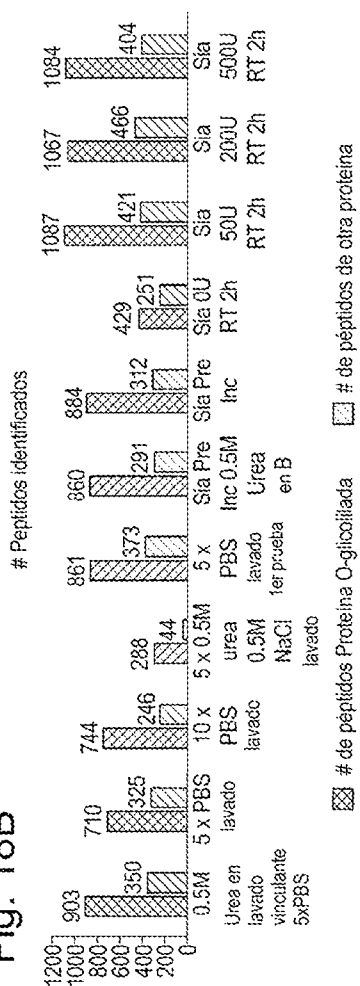


Fig. 18C

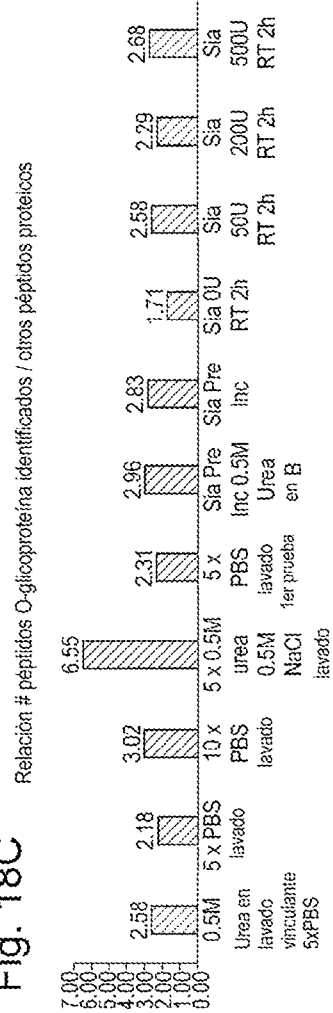


Fig. 19A

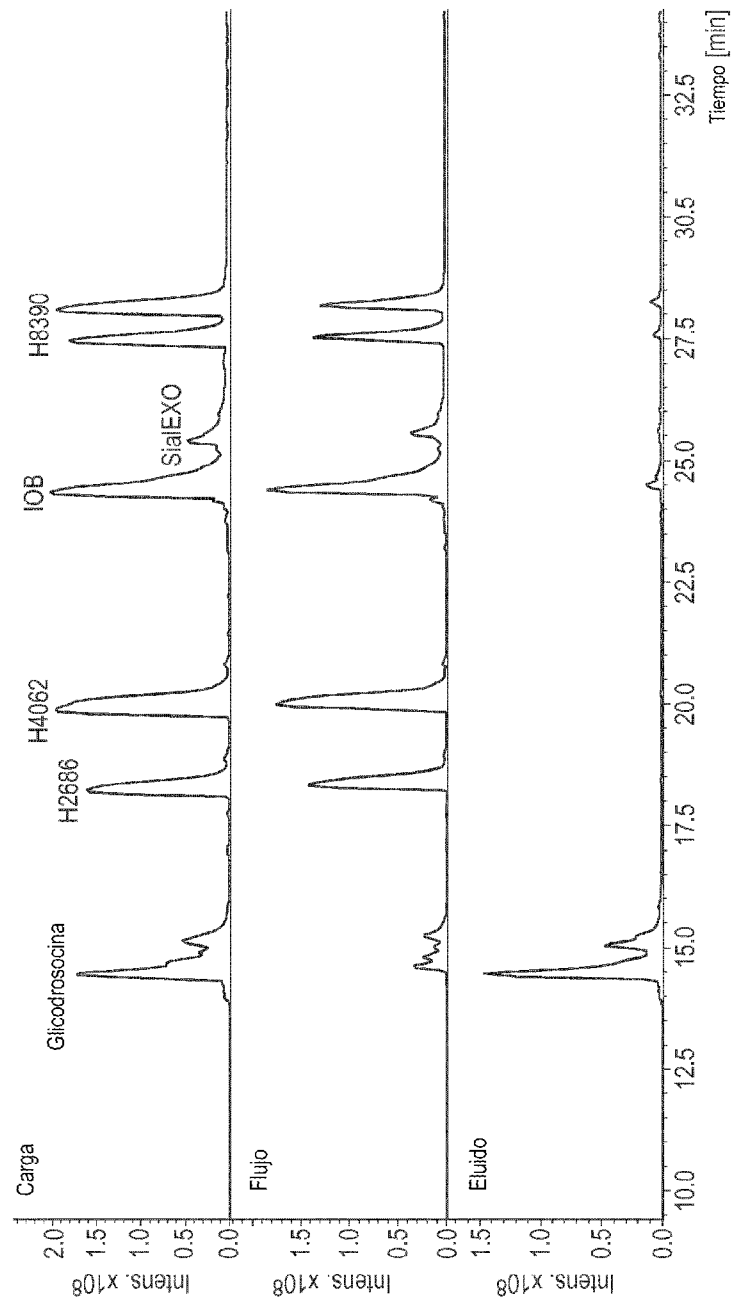
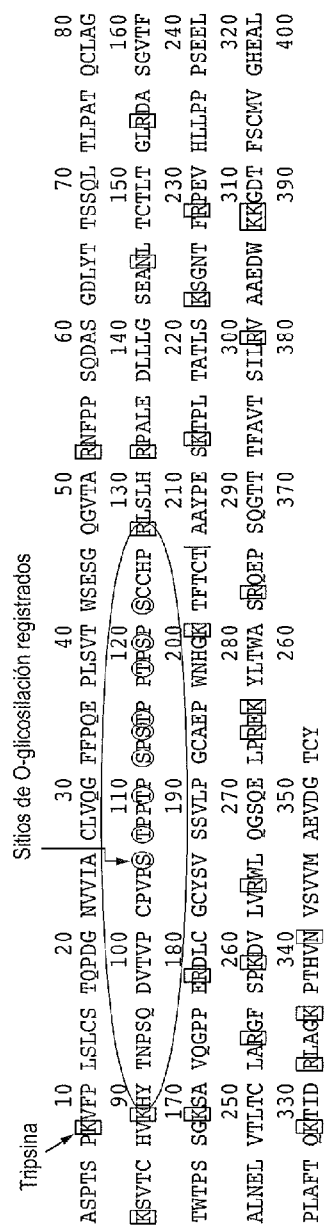


Fig. 19B



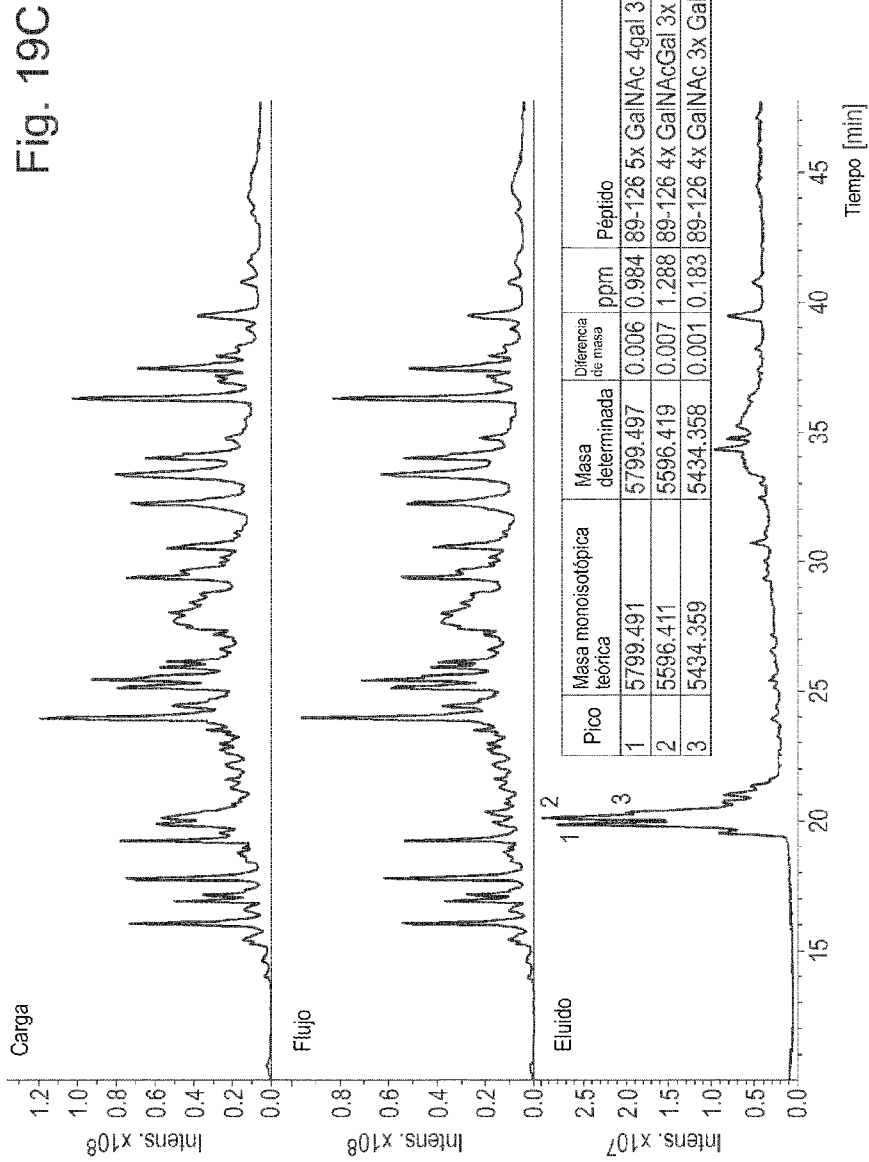


Fig. 20A

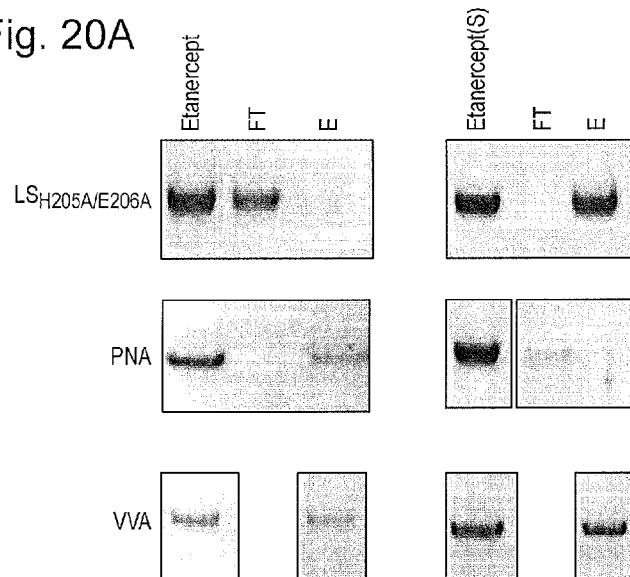


Fig. 20B

