(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 415 955**

(51) Int. CI.:

C12P 21/00 (2006.01)
C12N 9/10 (2006.01)
C12N 9/24 (2006.01)
C12N 1/14 (2006.01)
C12N 5/10 (2006.01)
C12R 1/645 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **20.02.2004 E 10013052 (5)**

(97) Fecha y número de publicación de la concesión europea: **10.04.2013 EP 2316963**

(54) Título: **Biblioteca de ADN combinatoria para producir N-glucanos modificados en eucariotas inferiores**

(30) Prioridad:

20.02.2003 US 371877

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
29.07.2013

(73) Titular/es:

GLYCOFI, INC. (100.0%)
21 Lafayette Street, Suite 200
Lebanon, NH 03766, US

(72) Inventor/es:

GERNGROSS, TILLMAN U.;
WILDT, STEFAN;
CHOI, BYUNG-KWON;
NETT, JUERGEN HERMANN;
BOBROWICZ, PIOTR;
HAMILTON, STEPHEN R. y
DAVIDSON, ROBERT C.

(74) Agente/Representante:

CARPINTERO LÓPEZ, Mario

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

Biblioteca de ADN combinatoria para producir N-glucanos modificados en eucariotas inferiores

Declaración con respecto a la investigación patrocinada federalmente

La presente invención se financió, al menos en parte, con una subvención estatal de los Institutitos Nacionales de la Salud (Subvención de NIH de Fase I n.º: 1R43GM66690-1) y una subvención del departamento de comercio, Número de Acuerdo de Cooperación de NIST-ATP 70NANB2H3046. Por lo tanto, el gobierno de los Estados Unidos puede tener ciertos derechos en la presente invención.

Campo de la invención

La presente invención se refiere a procedimientos y composiciones por los que pueden modificarse genéticamente células huésped eucariotas no humanas, tales como hongos u otras células eucariotas, para producir proteínas glucosiladas (glucoproteínas) que tienen patrones de glucosilación similares a los de glucoproteínas producidas por células animales, especialmente células humanas, que son útiles como agentes terapéuticos humanos o animales.

Antecedentes de la invención

Rutas de glucosilación en humanos y eucariotas inferiores

Después de que el ADN se transcriba y traduzca en una proteína, el procesamiento postraduccional adicional implica la unión de restos de azúcar, un proceso conocido como glucosilación. Diferentes organismos producen diferentes enzimas de glucosilación (glucosiltransferasas y glucosidasas) y tienen diferentes sustratos (azúcares de nucleótido) disponibles, de modo que los patrones de glucosilación así como la composición de los oligosacáridos individuales, incluso de la misma proteína, serán diferentes dependiendo del sistema huésped en el que se exprese la proteína particular. Las bacterias normalmente no glucosilan proteínas y si lo hacen es solamente de una manera muy inespecífica (Moens y Vanderleyden, 1997 Arch Microbiol. 168 (3):169-175). Los eucariotas inferiores tales como hongos filamentosos y levaduras añaden principalmente manosa y azúcares de manosilfosfato. El glucano resultante se conoce como un glucano del tipo de “alto contenido de manosa” o un manano. Las células vegetales y células de insecto (tales como células Sf9) glucosilan proteínas de otra manera más. Por el contrario, en eucariotas superiores tales como seres humanos, la cadena lateral de oligosacárido naciente puede reducirse para retirar varios restos de manosa y alargarse con restos de azúcar adicionales que normalmente no se encuentran en los N-glucanos de eucariotas inferiores. Véase, por ejemplo, R.K. Bretthauer y cols. Biotechnology and Applied Biochemistry, 1999, 30, 193-200; W. Martinet y cols. Biotechnology Letters, 1998, 20, 1171-1177; S. Weikert y cols. Nature Biotechnology, 1999, 17, 1116-1121; M. Malissard y cols. Biochemical and Biophysical Research Communications, 2000, 267, 169-173; Jarvis y cols., Current Opinion in Biotechnology, 1998, 9:528-533; y M. Takeuchi, 1 Trends in Glycoscience and Glycotechnology, 1997, 9, S29-S35.

La síntesis de una estructura de oligosacárido de tipo mamífero comienza con un conjunto de reacciones secuenciales en el transcurso de las que se añaden y retiran restos de azúcar mientras que la proteína se mueve a lo largo de la ruta secretora en el organismo huésped. Las enzimas que residen a lo largo de la ruta de glucosilación del organismo o célula huésped determinan los patrones de glucosilación resultantes de proteínas secretadas. Por lo tanto, el patrón de glucosilación resultante de proteínas expresadas en células huésped eucariotas inferiores difiere sustancialmente del patrón de glucosilación de proteínas expresadas en eucariotas superiores tales como seres humanos y otros mamíferos (Bretthauer, 1999). La estructura de un *N*-glucano fúngico típico se muestra en la **Figura 1A.**

Las primeras etapas de la glucosilación humana pueden dividirse en al menos dos fases diferentes: (i) los oligosacáridos $Glc_3Man_9GlcNAc_2$ unidos a lípidos se ensamblan por medio de un conjunto secuencial de reacciones en la membrana del retículo endoplásmico (RE) y (ii) la transferencia de este oligosacárido desde el doliquil pirofosfato del anclaje lipídico a la proteína sintetizada *de novo*. El sitio de la transferencia específica se define mediante un resto de asparagina (Asn) en la secuencia Asn-Xaa-Ser/Thr en la que Xaa puede ser cualquier aminoácido salvo prolina (Gavel, 1990). Se produce un procesamiento adicional mediante glucosidasas y manosidasas en el RE antes de que la glucoproteína naciente se transfiera al aparato de Golgi temprano, en el que se retiran restos de manosa adicionales mediante alfa (α)-1,2-manosidasas específicas de Golgi. El procesamiento continúa a medida que la proteína avanza a través del Golgi. En el Golgi medio, varias enzimas modificadoras, incluyendo N-acetilglucosaminil Transferasas (GnTI, GnTII, CmTIII, GnTIV y GnTV), manosidasa II y fucosiltransferasas, añaden y retiran restos de azúcar específicos. Finalmente, en el trans-Golgi, las galactosiltransferasas (GalT) y sialiltransferasas (ST) producen una estructura glucoproteica que se libera del Golgi. Es esta estructura, caracterizada por estructuras bi-, tri- y tetra-antenarias, que contienen galactosa, fucosa, N-acetilglucosamina y un alto grado de ácido siálico terminal, la que proporciona a la glucoproteína sus características humanas. La estructura de un *N*-glucano humano típico se muestra en la **Figura 1B.**

En casi todos los eucariotas, las glucoproteínas derivan de un precursor oligosacárido unido a lípido común $Glc_3Man_9GlcNAc_2$-dolicol-pirofosfato. Dentro del retículo endoplásmico, la síntesis y el procesamiento de oligosacáridos unidos a dolicol pirofosfato son idénticos entre todos los eucariotas conocidos. Sin embargo, el

procesamiento adicional del oligosacárido central por parte de células fúngicas, por ejemplo, levadura, una vez que se ha transferido a un péptido que deja el RE y entra en el Golgi, difiere significativamente de los seres humanos a medida que se mueve a lo largo de la ruta secretora e implica la adición de varios azúcares de manosa.

En levaduras, estas etapas las catalizan manosiltransferasas que residen en el Golgi, como Och1p, Mnt1p y Mnn1p, que añaden secuencialmente azúcares de manosa al oligosacárido central. La estructura resultante no es deseable para la producción de proteínas de tipo humano y por lo tanto, es deseable reducir o eliminar la actividad manosiltransferasa. Se ha mostrado que los mutantes de *S. cerevisiae,* deficientes en actividad manosiltransferasa (por ejemplo, mutantes *och1* o *mnn9*) no son letales y presentan un contenido de manosa reducido en el oligosacárido de glucoproteínas de levadura. Otras enzimas de procesamiento de oligosacáridos, tales como manosilfosfato transferasa, también pueden tener que eliminarse dependiendo del patrón de glucosilación endógeno particular del huésped.

Precursores de nucleótidos de azúcares

Los *N*-glucanos de glucoproteínas animales normalmente incluyen galactosa, fucosa y ácido siálico terminal. Estos azúcares no se encuentran en glucoproteínas producidas en levadura y hongos filamentosos. En seres humanos, se sintetiza la serie completa de precursores de azúcares de nucleótido (por ejemplo UDP-*N*-acetilglucosamina; UDP-*N*-acetilgalactosamina, CMP-*N*-acetilneuramínico, UDP-galactosa, GDP-fucosa, etc.) en el citosol y se transportan al Golgi, en el que se unen al oligosacárido central mediante glucosiltransferasas. (Sommers y Hirschberg, 1981 J. Cell Biol. 91 (2): A406-A406; Sommers y Hirschberg 1982 J. Biol. Chem. 257 (18): 811-817; Perez y Hirschberg 1987 Methods in Enzymology 138: 709-715).

Las reacciones de transferencia de glucosilo normalmente producen un producto secundario que es un nucleósido difosfato o un monofosfato. Aunque los monofosfatos pueden exportarse directamente en intercambio con azúcares de nucleósido trifosfatos por medio de un mecanismo antiportador, los difosfonucleósidos (por ejemplo, GDP) tienen que escindirlos las fosfatasas (por ejemplo, GDPasa) para producir nucleósido monofosfatos y fosfato inorgánico antes de exportarse. Esta reacción es importante para una glucosilación eficaz; por ejemplo, se ha descubierto que la GDPasa de *Saccharomyces cerevisiae (S. cerevisiae)* es necesaria para manosilación. Sin embargo, esa GDPasa tiene actividad reducida en un 90 % para UDP (Berninsone y cols., 1994 J. Biol. Chem. 269 (1): 207-211).

Los eucariotas inferiores normalmente carecen de actividad difosfatasa específica de UDP en el Golgi puesto que no usan precursores de UDP-azúcar para la síntesis de glucoproteínas basada en el Golgi. Se ha descubierto que *Schizosaccharomyces pombe*, una levadura que se ha descubierto que añade restos de galactosa a polisacáridos de la pared celular (procedentes de UDP-galactosa) tiene actividad UDPasa específica, lo que indica el requisito potencial de dicha enzima (Beminsone y cols., 1994). Se sabe que el UDP es un inhibidor potente de glucosiltransferasas y la retirada de este producto secundario de glucosilación puede ser importante para evitar la inhibición de la glucosiltransferasa en el lumen del Golgi (Khatara y cols., 1974). Véase Berninsone, P. y cols. 1995. J. Biol. Chem. 270 (24): 14564-14567; Beaudet, L. y cols. 1998 Abc Transporters: Biochemical, Cellular, and Molecular Aspects. 292: 397-413.

Procesamiento secuencial de N-glucanos por actividades enzimáticas compartimentalizadas

Las transferasas de azúcares y glucosidasas (por ejemplo, manosidasas) revisten la superficie interior (luminal) del RE y aparato de Golgi y de este modo proporcionan una superficie “catalítica” que posibilita el procesamiento secuencial de glucoproteínas a medida que avanzan a través de la red del RE y Golgi. Los múltiples compartimentos de las regiones cis, medial y trans del Golgi y la Red del trans-Golgi (TGN), proporcionan las diferentes localidades en las que puede tener lugar la secuencia ordenada de reacciones de glucosilación. A medida que una glucoproteína avanza desde la síntesis en el RE a la maduración completa en el Golgi tardío o TGN, se expone secuencialmente a diferentes glucosidasas, manosidasas y glucosiltransferasas de modo que pueda sintetizarse una estructura de carbohidratos específica. Se ha dedicado mucho trabajo a revelar el mecanismo exacto por el que estas enzimas se conservan y anclan a sus orgánulos respectivos. La imagen evolutiva es compleja pero las pruebas sugieren que la región del tallo, la región transmembrana y la cola citoplásmica, individualmente o de forma concertada, dirigen las enzimas a la membrana de orgánulos individuales y de este modo localizan el dominio catalítico asociado con ese locus (véase, por ejemplo, Gleeson, P.A. (1998) Histochem. Cell Biol. 109, 517-532).

En algunos casos, se descubrió que estas interacciones específicas actuaban entre especies. Por ejemplo, también se mostró que el dominio que abarcaba la membrana de α2,6-ST de ratas, una enzima que se sabe que se localiza en el trans-Golgi del animal, también sitúa un gen indicador (invertasa) en el Golgi de levadura (Schwientek, 1995). Sin embargo, el mismo dominio que abarca la membrana como parte de una α2,6-ST de longitud completa se conservaba en el RE y no se transportaba adicionalmente al Golgi de levadura (Krezdorn, 1994). Una GalT de longitud completa de seres humanos ni siquiera se sintetizaba en levadura, a pesar de la existencia de niveles de transcripción manifiestamente altos. Por el contrario, la región transmembrana de la misma GalT humana fusionada con un indicador de invertasa fue capaz de dirigir la localización en el Golgi de levadura, aunque a niveles de producción bajos. Schwientek y colaboradores han mostrado que la fusión de 28 aminoácidos de una manosiltransferasa de levadura (*MNT1*), una región que contiene una cola citoplásmica, una región transmembrana y ocho aminoácidos de la región de tallo, con el dominio catalítico de GalT humana es suficiente para la localización

en el Golgi de una GalT activa. Otras galactosiltransferasas parecen basarse en interacciones con enzimas residentes en orgánulos particulares debido a que, después de la retirada de su región transmembrana, aún son capaces de situarse apropiadamente.

La localización inapropiada de una enzima de glucosilación puede evitar el funcionamiento apropiado de la enzima en la ruta. Por ejemplo, *Aspergillus nidulans,* que tiene numerosas α-1,2-manosidasas (Eades y Hintz, 2000 Gene 255 (1): 25-34), no añade GlcNAc a $Man_5GlcNAc_2$ cuando se transforma con el gen de conejo GnTI, a pesar del alto nivel global de actividad GnTI (Kalsner y cols., 1995). GnTI, aunque se expresa de forma activa, puede situarse incorrectamente de modo que la enzima no esté en contacto con sus dos sustratos: UDP-GlcNAc y un sustrato $Man_5GlcNAc_2$ productivo (no todas las estructuras de $Man_5GlcNAc_2$ son productivas; véase posteriormente). Como alternativa, el organismo huésped puede no proporcionar un nivel adecuado de UDP-GlcNAc en el Golgi o la enzima puede situarse de forma apropiada pero ser inactiva en su nuevo ambiente. Además, las estructuras de $Man_5GlcNAc_2$ presentes en la célula huésped pueden diferir en estructura de $Man_5GlcNAc_2$ hallado en mamíferos. Maras y colaboradores descubrieron que aproximadamente un tercio de los N-glucanos de celobiohidrolasa I (CBHI) obtenidos de *T. reesei* pueden reducirse a $Man_5GlcNAc_2$ mediante la 1,2 manosidasa de *A. saitoi in vitro*. Menos del 1 % de esos N-glucanos, sin embargo, podrían actuar como un sustrato productivo para GnTI. La mera presencia de $Man_5GlcNAc_2$, por lo tanto, no asegura que pueda conseguirse un procesamiento adicional para dar GlcNAc $Man_5GlcNAc_2$. Lo que se requiere es la formación de una estructura de $Man_5GlcNAc_2$ sensible a GnTI productiva. Aunque podría producirse $Man_5GlcNAc_2$ en la célula (aproximadamente un 27 % en moles), solamente una fracción pequeña podría convertirse en GlcNAc $Man_5GlcNAc_2$ (menos de aproximadamente un 5 %, véase Chiba, WO 01/14522).

Hasta la fecha, no hay una forma fiable de predecir si una glucosiltransferasa o manosidasa expresada de forma heteróloga particular en un eucariota inferior (1), se traducirá de forma suficiente, (2), será catalíticamente activa o (3) se localizará en el orgánulo apropiado dentro de la ruta secretora. Debido a que estos tres puntos son necesarios para afectar a los patrones de glucosilación en los eucariotas inferiores, sería deseable un esquema sistemático para conseguir la función catalítica deseada y retención apropiada de enzimas en ausencia de herramientas predictivas, que en la actualidad no están disponibles.

Producción de glucoproteínas terapéuticas

Un número significativo de proteínas aisladas a partir de seres humanos o animales se modifican postraduccionalmente, siendo la glucosilación una de las modificaciones más significativas. Una estimación del 70 % de todas las proteínas terapéuticas están glucosiladas y por lo tanto, actualmente se cuenta con un sistema de producción (es decir, célula huésped) que sea capaz de glucosilar de una manera similar a los seres humanos. Varios estudios han mostrado que la glucosilación desempeña un papel importante en la determinación de (1) la inmunogenicidad, (2) las propiedades farmacocinéticas, (3) el tráfico y (4) la eficacia de las proteínas terapéuticas. Por lo tanto, no es sorprendente que se hayan dirigido esfuerzos importantes por parte de la industria farmacéutica al desarrollo de procedimientos para obtener glucoproteínas que sean tan “humanoides” o “de tipo humano” como sea posible. Hasta la fecha, la mayoría de las glucoproteínas se fabrican en un sistema huésped de mamífero. Esto puede implicar la modificación por ingeniería genética de tales células de mamífero para potenciar el grado de sialilación (es decir, la adición terminal de ácido siálico) de proteínas expresadas por las células, que se sabe que mejora las propiedades farmacocinéticas de dichas proteínas. Como alternativa, se puede mejorar el grado de sialilación mediante adición *in vitro* de dichos azúcares usando glucosiltransferasas conocidas y sus azúcares de nucleótidos respectivos (por ejemplo, 2,3-sialiltransferasa y ácido CMP-siálico).

Aunque la mayoría de los eucariotas superiores llevan a cabo reacciones de glucosilación que son similares a las halladas en seres humanos, las proteínas humanas recombinantes expresadas en los sistemas de huésped mencionados anteriormente invariablemente difieren de su homólogo humano “natural” (Raju, 2000). Por lo tanto, se ha dirigido un trabajo de desarrollo considerable a encontrar maneras de mejorar el “carácter humano” de las proteínas preparadas en estos sistemas de expresión. Esto incluye la optimización de las condiciones de fermentación y la modificación genética de huéspedes de expresión de proteínas introduciendo genes que codifican enzimas implicadas en la formación de glucoformas de tipo humano (Werner, 1998; Weikert, 1999; Andersen, 1994; Yang, 2000). No se han resuelto problemas inherentes asociados con todos los sistemas de expresión de mamíferos.

Los procedimientos de fermentación basados en cultivo de células de mamífero (por ejemplo, células CHO, murinas o humanas), por ejemplo, tienden a ser muy lentos (los tiempos de fermentación de más de una semana son frecuentes), con frecuencia producen títulos de producto bajos, requieren nutrientes y cofactores caros (por ejemplo, suero fetal bovino), están limitados por la muerte celular programa (apoptosis) y con frecuencia no permiten la expresión de proteínas terapéuticamente valiosas particulares. De forma más importante, las células de mamífero son susceptibles a virus que tienen el potencial de ser patógenos humanos y se requieren controles de calidad rigurosos para asegurar la seguridad del producto. Esto es particularmente preocupante ya que muchos de tales procedimientos requieren la adición de componentes de medios sensibles a la temperatura y complejos que derivan de animales (por ejemplo, suero bovino de ternero), que pueden portar agentes patógenos para seres humanos tales como priones de la encefalopatía espongiforme bovina (BSE) o virus. Además, la producción de compuestos terapéuticos se lleva a cabo preferentemente en un ambiente estéril bien controlado. Una granja de animales,

independientemente de lo limpia que se mantenga, no constituye dicho ambiente, lo que crea un problema adicional en el uso de animales transgénicos para fabricar grandes volúmenes de proteínas terapéuticas.

Por lo tanto, la mayoría, sino todas, de las glucoproteínas terapéuticas producidas en la actualidad se expresan en células de mamífero y se ha dirigido mucho esfuerzo para mejorar (es decir, "humanizar") el patrón de glucosilación de estas proteínas recombinantes. Se han empleado con éxito cambios en la composición del medio así como la coexpresión de genes que codifican enzimas implicadas en la glucosilación humana (véase, por ejemplo, Weikert, 1999).

Producción de glucoproteínas usando microorganismos eucariotas

La falta de un sistema de expresión de mamíferos adecuado es un obstáculo significativo para la producción de bajo coste y segura de glucoproteínas humanas recombinantes para aplicaciones terapéuticas. Sería deseable producir proteínas recombinantes similares a sus homólogos de mamífero, por ejemplo, humanos, en eucariotas inferiores (hongos y levaduras). La producción de glucoproteínas mediante la fermentación de microorganismos ofrecería numerosas ventajas frente a los sistemas existentes. Por ejemplo, los procedimientos basados en fermentación pueden ofrecer (a) la producción rápida de altas concentraciones de proteína; (b) la capacidad de usar condiciones de producción bien controladas estériles; (c) la capacidad de usar medios de cultivo químicamente definidos, sencillos (y sin proteínas); (d) la facilidad de manipulación genética; (e) la ausencia de patógenos humanos o animales contaminantes tales como virus; (f) la capacidad de expresar una amplia diversidad de proteínas, incluyendo las escasamente expresadas en cultivo celular debido a la toxicidad, etc.; y (g) la facilidad de recuperación de proteínas (mediante secreción en el medio). Además, las instalaciones de fermentación para levaduras y hongos son generalmente bastante menos costosas de construir que las instalaciones de cultivo celular. Aunque la estructura de oligosacáridos central transferida a una proteína en el retículo endoplásmico es básicamente idéntica en mamíferos y eucariotas inferiores, se han descubierto diferencias sustanciales en las reacciones de procesamiento posteriores que tienen lugar en el aparato de Golgi de hongos y mamíferos. De hecho, incluso entre diferentes eucariotas inferiores existe una gran diversidad de estructuras de glucosilación. Esto ha evitado históricamente el uso de eucariotas inferiores como huéspedes para la producción de glucoproteínas humanas recombinantes a pesar, por otra parte, de ventajas notables de otro modo frente a los sistemas de expresión de mamíferos.

Las glucoproteínas terapéuticas producidas en un microorganismo huésped tal como levadura que usan la ruta de glucosilación del huésped endógena difieren estructuralmente de las producidas en células de mamífero y normalmente muestran una eficacia terapéutica reducida en gran medida. Tales glucoproteínas son normalmente inmunogénicas en seres humanos y muestran una semivida (y por lo tanto bioactividad) reducida *in vivo* después de la administración (Takeuchi, 1997). Los receptores específicos en seres humanos y animales (es decir, receptores de manosa de macrófagos) pueden reconocer restos de manosa terminales y promover la rápida eliminación de la glucoproteína ajena del torrente sanguíneo. Los efectos adversos adicionales pueden incluir cambios en el plegamiento, solubilidad, susceptibilidad a proteasas, tráfico, transporte, compartimentación, secreción, reconocimiento por otras proteínas o factores, antigenicidad o alergenicidad de las proteínas.

Se han usado con éxito tanto levaduras como hongos filamentosos para la producción de proteínas recombinantes, tanto intracelulares como secretadas (Cereghino, J. L. y J. M. Cregg 2000 *FEMS Microbiology Reviews* 24 (1): 45-66; Harkki, A. y cols. 1989 Bio-Technology 7 (6): 596; Berka, R. M. y cols. 1992 Abstr.Papers Amer. Chem. Soc. 203: 121-BIOT; Svetina, M. y cols. 2000 J. Biotechnol. 76 (2-3): 245-251). Diversas levaduras, tales como *K. lactis, Pichia pastoris, Pichia methanolica* y *Hansenula polymorpha,* han desempeñado papeles particularmente importantes como sistemas de expresión eucariotas debido a que son capaces de proliferar a altas densidades celulares y secretar grandes cantidades de proteína recombinante. De forma similar, se han usado hongos filamentosos, tales como *Aspergillus niger, Fusarium sp, Neurospora crassa* y otros, para producir eficazmente glucoproteínas a escala industrial. Sin embargo, como se ha indicado anteriormente, las glucoproteínas expresadas en cualquiera de estos microorganismos eucariotas difieren sustancialmente en la estructura de *N*-glucanos de las de animales. Esto ha evitado el uso de levaduras u hongos filamentosos como huéspedes para la producción de muchas glucoproteínas terapéuticas.

Aunque la glucosilación en levaduras y hongos es muy diferente a la de humanos, se comparten algunos elementos comunes. La primera etapa, la transferencia de la estructura de oligosacárido central a la proteína naciente, está altamente conservada en todos los eucariotas incluyendo levaduras, hongos, plantas y seres humanos (compárense las **Figuras 1A** y **1B**). El procesamiento posterior del oligosacárido central, sin embargo, difiere significativamente en levaduras e implica la adición de varios azúcares de manosa. Esta etapa se cataliza mediante manosiltransferasas que residen en el Golgi (por ejemplo *OCH1*, *MNT1, MNN1,* etc.), que añaden secuencialmente azúcares de manosa al oligosacárido central. La estructura resultante no es deseable para la producción de proteínas humanoides y por lo tanto, es deseable reducir o eliminar la actividad manosiltransferasa. Se ha mostrado que los mutantes de *S. cerevisiae* deficientes en actividad manosiltransferasa (por ejemplo, mutantes *och1* o *mnn9*) son no letales y presentan un contenido de manosa reducido en el oligosacárido de glucoproteínas de levaduras. Otras enzimas que procesan oligosacáridos, tales como la manosilfosfato transferasa, también pueden tener que eliminarse dependiendo del patrón de glucosilación endógeno particular del huésped. Después de reducir las reacciones de glucosilación endógenas no deseadas, debe introducirse por ingeniería genética la formación de N-

glucanos complejos en el sistema del huésped. Esto requiere la expresión estable de varias enzimas y transportadores de azúcar-nucleótido. Además, estas enzimas se deben situar de modo que se asegure un procesamiento secuencial de la estructura de glucosilación en maduración.

Se han realizado varios intentos de modificar las rutas de glucosilación de microorganismos eucariotas para proporcionar glucoproteínas más adecuadas para su uso como agentes terapéuticos de mamíferos. Por ejemplo, varias glucosiltransferasas se han clonado por separado y se han expresado en *S. cerevisiae* (GalT, GnTI), *Aspergillus nidulans* (GnTI) y otros hongos (Yoshida y cols., 1999, Kalsner y cols., 1995 Glycoconj. J. 12 (3):360-370, Schwientek y cols., 1995). Sin embargo, no se obtuvieron *N*-glucanos que se asemejaran a los obtenidos en células humanas.

Las levaduras producen una diversidad de manosiltransferasas (por ejemplo, 1,3-manosiltransferasas tales como *MNN1* en *S. cerevisiae;* Graham y Emr, 1991 J. Cell. Biol. 114 (2):207-218), 1,2-manosiltransferasas (por ejemplo, familia *KTR*/*KRE* de *S. cerevisiae*), 1,6-manosiltransferasas (por ejemplo, *OCH1* de *S. cerevisiae*), manosilfosfato transferasas y sus reguladores (por ejemplo, *MNN4* y *MNN6* de *S. cerevisiae*) y enzimas adicionales que están implicadas en las reacciones de glucosilación endógenas. Muchos se estos genes se han suprimido individualmente dando lugar a organismos viables que tienen perfiles de glucosilación alterados. Se muestran ejemplos en la **Tabla 1**.

Tabla 1. Ejemplos de cepas de levadura que tienen manosilación alterada

Cepa	N-glucano (tipo silvestre)	Mutación	N-glucano (mutante)	Referencia
S. pombe	$Man_{>9}GlcNAc_2$	OCH1	$Man_8GlcNAc_2$	Yoko-o y cols., 2001 FEBS Lett. 489 (1): 75-80
S. cerevisiae	$Man_{>9}GlcNAc_2$	OCH1/MNN1	$Man_8GlcNAc_2$	Nakanishi-Shindo y cols.,. 1993 J. Biol. Chem. 268 (35):26338-26345
S. cerevisiae	$Man_{>9}GlcNAc_2$	OCH1/MNN1/MNN4	$Man_8GlcNAc_2$	Chiba y cols., 1998 J. Biol. Chem. 273,26298-26304
P. pastoris	Hiperglucosilado	OCH1 (deleción completa)	No hiperglucosilado	Welfide, Publicación de Solicitud Japonesa N.º: 8-336387
P. pastoris	$Man_{>8}GlcNAc_2$	OCH1 (interrupción)	$Man_{>8}GlcNAc_2$	Contreras y cols. WO 02/00856 A2

La Publicación de Solicitud de Patente Japonesa N.º: 8-336387 desvela la deleción de un homólogo de *OCH1* en *Pichia pastoris.* En *S. cerevisiae, OCHI* codifica una 1,6-manosiltransferasa, que añade una manosa a la estructura de glucano $Man_8GlcNAc_2$ para producir $Man_9GlcNAc_2$. La estructura de $Man_9GlcNAc_2$ que contiene tres restos de 1,6 manosa, es después un sustrato para 1,2-, 1,6- y 1,3-manosiltransferasas adicionales *in vivo*, conduciendo a las glucoproteínas hipermanosiladas que son características de *S. cerevisiae* y que normalmente pueden tener 30-40 restos de manosa por N-glucano. Debido a que la Ochlp inicia la transferencia de 1,6 manosa al núcleo de $Man_8GlcNAc_2$, esta se denomina con frecuencia "1,6 manosiltransferasa de inicio" para distinguirla de otras 1,6 manosiltransferasas que actúan posteriormente en el Golgi. En una cepa mutante *och1 mnn1 mnn4* de *S. cerevisiae*, las proteínas glucosiladas con $Man_8GlcNAc_2$ se acumulan y no se produce hipermanosilación. Sin embargo, $Man_8GlcNAc_2$ no es un sustrato para glucosiltransferasas de mamífero, tales como UDP-GlcNAc transferasa I humana y en consecuencia, el uso de esa cepa mutante, en sí mismo, no es útil para producir proteínas de tipo mamífero, es decir, con patrones de glucosilación complejos o híbridos.

Aunque la Publicación de Solicitud de Patente Japonesa N.º: 8-336387 desvela procedimientos para obtener un mutante *och1* de *P. pastoris* que presenta un fenotipo de manosilación reducido, no proporciona datos sobre si la actividad 1,6 manosiltransferasa de inicio que se supone que está codificada por *OCH1* se reduce o elimina. Está bien establecido en el campo de la genética fúngica que los homólogos de genes con frecuencia no desempeñan el mismo papel en sus organismos huésped respectivos. Por ejemplo, el gen *rca-1* de *Neurospora* complementa un mutante de esporulación *flbD* de *Aspergillus* pero no tiene un papel identificable en la esporulación de Neurospora. Shen, W.C. y cols., Genetics 1998; 148 (3): 1031-41. Más recientemente, Contreras (documento WO 02/00856 A2) muestra que, en un mutante *och1* de of *P. pastoris,* al menos 50 % de los glucanos de la pared celular no pueden reducirse a $Man_5GlcNAc_2$ con una α-1,2-manosidasa de *Trichoderma reesei* (véase la Figura 11 del documento WO 02/00856 A2). Puesto que el tipo silvestre presenta un patrón de glucosilación muy similar (Figura 10, Panel 2 del

documento WO02/00856 A2), parece ser que el gen *OCH1* de *P. pastoris no* puede codificar la actividad 1,6-manosiltransferasa de inicio y por lo tanto es diferente de su homólogo genético en *S. cerevisiae.* Por lo tanto, hasta la fecha, no hay pruebas de que la actividad de la α-1,6-manosiltransferasa de inicio se elimine en mutantes *och1* de *P. pastoris,* lo que apoya adicionalmente la noción de que las rutas de glucosilación de *S. cerevisiae* y *P. pastoris* son significativamente diferentes.

Martinet y cols. (Biotechnol. Lett. 1998, 20(12), 1171-1177) informaron sobre la expresión de la α-1,2-manosidasa de *T. reesei* en *P. pastoris.* Se observó algo de reducción de manosa de los *N*-glucanos de una proteína modelo. Sin embargo, la proteína modelo no tenía *N*-glucanos con la estructura de $Man_5GlcNAc_2$, que serían necesarios como un intermedio para la generación de *N*-glucanos complejos. En consecuencia, este sistema no es útil para producir proteínas con patrones de glucosilación complejos o híbridos.

De forma similar, Chiba y cols. (1998) expresaron α-1,2-manosidasa de *Aspergillus saitoi* en la levadura *Saccharomyces cerevisiae.* Se introdujo por ingeniería genética una secuencia de péptido señal (His-Asp-Glu-Leu) (**SEC ID N.º: 5**) en la manosidasa exógena para promover su retención en el retículo endoplásmico. Además, el huésped de levadura era un mutante sin actividades enzimáticas asociadas con la hipermanosilación de proteínas: 1,6-manosiltransferasa (*och1*); 1,3-manosiltransferasa (*mnn1*); y un regulador de manosilfosfato transferasa (*mnn4*). Los *N*-glucanos del huésped triple mutante consistían principalmente en la estructura de $Man_8GlcNAc_2$, en lugar de las formas de alto contenido de manosa halladas en *S. cerevisiae* de tipo silvestre. En presencia de la manosidasa modificada por ingeniería genética, los *N*-glucanos de una proteína modelo (carboxipeptidasa Y) se redujeron para dar una mezcla consistente en un 27 % en moles de $Man_5GlcNAc_2$, un 22 % en moles de $Man_6GlcNAc_2$, un 22 % en moles de $Man_7GlcNAc_2$ y un 29 % en moles de $Man_8GlcNAc_2$. La reducción de glucoproteínas de la pared celular fue menos eficaz, teniendo solamente un 10 % en moles de los *N*-glucanos la estructura de $Man_5GlcNAc_2$ deseada.

Incluso si todos los glucanos $Man_5GlcNAc_2$ tuvieran la forma $Man_5GlcNAc_2$ correcta que pudiera convertirse en $GlcNAcMan_5GlcNAc_2$ mediante GnTI, el sistema anterior no sería eficaz para la producción de proteínas que tuvieran patrones de glucosilación de tipo humano. Si están presentes varios sitios de glucosilación en una proteína deseada, la probabilidad (P) de obtener dicha proteína en una forma correcta sigue la relación $P=(F)^n$, en la que n es igual al número de sitios de glucosilación y F es igual a la fracción de glucoformas deseadas. Una glucoproteína con tres sitios de glucosilación tendría una probabilidad del 0,1 % de proporcionar los precursores apropiados para el procesamiento de N-glucanos complejos e híbridos en todos sus sitios de glucosilación. Por lo tanto, usando el sistema de Chiba para preparar una glucoproteína con un único sitio de *N*-glucosilación, al menos el 73 % en moles tendría una estructura incorrecta. Para una glucoproteína que tuviera dos o tres sitios de N-glucosilación, al menos un 93 o 98 % en moles tendría una estructura incorrecta, respectivamente. Estas eficacias de conversión tan bajas son insatisfactorias para la producción de agentes terapéuticos, particularmente porque la separación de proteínas que tienen diferentes glucoformas es normalmente costosa y difícil.

Chiba y cols. (WO 01/14522) han mostrado altos niveles de estructuras de $Man_5GlcNAc_2$ en el factor de crecimiento de fibroblastos (FGF) recombinante, una glucoproteína soluble secretada producida en *S. cerevisiae.* No está claro, sin embargo, que la $Man_5GlcNAc_2$ detectada se produjera dentro de la célula huésped (es decir *in vivo*) porque la α-1,2 manosidasa se dirigió por fusión con un marcador de localización HDEL (**SEC ID N.º: 5**), un mecanismo que se sabe que tiene fugas (Pelham H.R. (1998) EMBO J. 7, 913-918). Es más probable que FGF se secretara al medio, donde se procesó después mediante la α-1,2 manosidasa que había escapado del mecanismo de recuperación HDEL (**SEC ID N.º: 5**) y se habían filtrado al medio. Como se ha mencionado anteriormente, una proteína intracelular (CPY), expresada en la misma cepa, contenía principalmente glucanos (más del 73 %) que eran $Man_6GlcNAc_2$ y mayores. La mayoría de las estructuras de $Man_5GlcNAc_2$ en FGF, por lo tanto, probablemente se han producido *ex vivo.* Adicionalmente no estás claro si las estructuras de $Man_5GlcNAc_2$ que se produjeron eran sustratos productivos para GnTI.

Como demuestra el trabajo anterior, se pueden reducir estructuras de $Man_8GlcNAc_2$ a un isómero de $Man_5GlcNAc_2$ en *S. cerevisiae,* aunque aún tiene que determinarse una reducción de alta eficacia mayor del 50 % *in vivo*, por introducción por ingeniería genética de una manosidasa fúngica de *A. saitoi* en el retículo endoplásmico (RE). Los defectos de este enfoque son dobles: (1) no está claro si las estructuras de $Man_5GlcNAc_2$ formadas se forman efectivamente *in vivo* (en lugar de haberse secretado y modificado adicionalmente mediante manosidasas fuera de la célula); y (2) no está claro si alguna estructura de $Man_5GlcNAc_2$ formada, si efectivamente se hubiera formado *in vivo*, es de la isoforma correcta para ser un sustrato productivo para la modificación con *N*-glucano posterior mediante GlcNAc transferasa I (Maras y cols., 1997, Eur. J. Biochem. 249, 701-707).

Con el objetivo de proporcionar una glucoproteína de tipo humano derivada de un huésped fúngico, la Patente de Estados Unidos N.º: 5.834.251 desvela un procedimiento para producir una glucoproteína híbrida derivada de *Trichoderma reseei.* Un *N*-glucano híbrido tiene solamente restos de manosa en la rama Manα1-6 de la estructura de manosa central y una o dos antenas complejas en la rama Manα1-3. Aunque esta estructura tiene utilidad, el procedimiento tiene la desventaja de que deben realizarse numerosas etapas enzimáticas *in vitro*, lo que es costoso y consume tiempo. Las enzimas aisladas son caras de preparar y necesitan sustratos costosos (por ejemplo, UDP-GlcNAc). El procedimiento tampoco permite la producción de glucanos complejos en una proteína deseada.

Actividad manosidasa Intracelular Implicada en el recorte de N-glucanos

Se requiere actividad alfa-1,2-manosidasa para el recorte de $Man_8GlcNAc_2$ para formar $Man_5GlcNAc_2$, que es un intermedio principal de la formación de N-glucanos complejos en mamíferos. El trabajo anterior ha mostrado que la α-1,2-manosidasa murina, fúngica y humana truncada pueden expresarse en la levadura metilotrópica *P. pastoris* y presenta actividad de recorte de $Man_8GlcNAc_2$ a $Man_5GlcNAc_2$ (Lal y cols., Glycobiology octubre de 1998; 8 (10): 981-95; Tremblay y cols., Glycobiology junio de 1998; 8 (6): 585-95, Callewaert y cols., 2001). Sin embargo, hasta la fecha no existen informes que muestren el alto nivel de recorte *in vivo* de $Man_8GlcNAc_2$ a $Man_5GlcNAc_2$ en una glucoproteína secretada de *P. pastoris.*

Aunque es útil modificar por ingeniería genética cepas que sean capaces de producir $Man_5GlcNAc_2$ como la estructura de *N*-glucano primaria, cualquier intento de modificar adicionalmente estas estructuras precursoras de alto contenido en manosa para que se asemejen más a glucanos humanos requiere etapas *in vivo* o *in vitro* adicionales. En la Patente de Estados Unidos N.º: 5.834.251 (mencionada anteriormente) se describen procedimientos para humanizar adicionalmente glucanos de fuentes fúngicas de levadura *in vitro*. Como se ha analizado anteriormente, sin embargo, si $Man_5GlcNAc_2$ debe humanizarse adicionalmente *in vivo*, hay que asegurarse de que las estructuras de $Man_5GlcNAc_2$ generadas efectivamente se generen de forma intracelular y no sean el producto de actividad manosidasa en el medio. La formación de *N*-glucanos complejos en levadura u hongos requerirá que se generen altos niveles de $Man_5GlcNAc_2$ dentro de la célula porque solo los glucanos $Man_5GlcNAc_2$ intracelulares pueden procesarse adicionalmente para dar N-glucanos híbridos y complejos *in vivo*. Además, debe demostrarse que la mayoría de las estructuras de $Man_5GlcNAc_2$ generadas efectivamente son un sustrato para GnTI y por lo tanto, permiten la formación de N-glucanos híbridos y complejos.

Además, la mera presencia de una α-1,2-manosidasa en la célula no asegura, por sí misma, la reducción intracelular apropiada de $Man_8GlcNAc_2$ a $Man_5GlcNAc_2$. (Véase, por ejemplo, Contreras y cols., documento WO 02/00856 A2, en el que una manosidasa marcada con HDEL (**SEC ID N.º: 5**) de *T. reesei* se localiza principalmente en el RE y se coexpresa con una proteína indicadora de hemaglutinina de gripe (HA) en la que prácticamente no pudo detectarse $Man_5GlcNAc_2$. Véase también Chiba y cols., 1998 (mencionado anteriormente), en el que una fusión de dominio transmembrana de Och1p/α-1,2-manosidasa quimérica localizada en el RE, Golgi temprano y citosol de *S. cerevisiae*, no tuvo actividad de reducción de manosidasa). En consecuencia, la mera localización de una manosidasa en el RE o Golgi es insuficiente para asegurar la actividad de la enzima respectiva en ese orgánulo diana. (Véase también, Martinet y cols. (1998), mencionado anteriormente, que muestra que la α-1,2-manosidasa de *T. reesei*, aunque se localice intracelularmente, aumentaba en lugar de reducir el grado de manosilación). Hasta la fecha, no hay informe que demuestre la localización intracelular de una α-1,2-manosidasa heteróloga activa en levadura u hongos usando una secuencia de localización transmembrana.

En consecuencia, se necesitan procedimientos para producir glucoproteínas caracterizadas por un alto contenido de $Man_5GlcNAc_2$ intracelular que puedan procesarse adicionalmente para dar estructuras glucoproteicas de tipo humano en células huésped eucariotas no humanas y particularmente en levaduras y hongos filamentosos.

Sumario de la invención

Se han desarrollado líneas celulares y células huésped que tienen rutas de glucosilación modificadas genéticamente que les permiten llevan a cabo una secuencia de reacciones enzimáticas que imitan el procesamiento de glucoproteínas en mamíferos, especialmente en seres humanos. Las proteínas recombinantes expresadas en estos huéspedes modificados por ingeniería genética producen glucoproteínas más similares, si no sustancialmente idénticas, a sus homólogos de mamíferos, por ejemplo, humanos. Las células huésped descritas en el presente documento, por ejemplo, microorganismos eucariotas inferiores y otras células huésped eucariotas no humanas desarrolladas en cultivo, se modifican para producir N-glucanos tales como $Man_5GlcNAc_2$ u otras estructuras producidas a lo largo de las rutas de glucosilación humanas. Esto se consigue usando una combinación de ingeniería genética y/o selección de cepas que: no expresan ciertas enzimas que crean las estructuras no deseables características de las glucoproteínas fúngicas; que expresan enzimas heterólogas seleccionadas para tener actividad óptima en las condiciones presentes en la célula huésped en la que debe conseguirse actividad; o combinaciones de las mismas; en las que el eucariota modificado por ingeniería genética expresa al menos una actividad enzimática heteróloga requerida para producir una glucoproteína “de tipo humano”. Las células huésped descritas en el presente documento pueden modificarse adicionalmente por expresión heteróloga de una o más actividades tales como glucosiltransferasas, transportadores de azúcares y manosidasas, para convertirse en cepas para la producción de glucoproteínas terapéuticas de mamífero, por ejemplo, humanas.

También se describe en el presente documento un procedimiento de producción de glucoproteínas que usa (1) un huésped eucariota inferior tal como un hongo unicelular o filamentoso para modificar la composición de glucosilación y estructuras de las proteínas fabricadas en un organismo huésped (“célula huésped”) de modo que se asemejen más a las estructuras de carbohidrato halladas en proteínas de mamífero, por ejemplo, humanas. El procedimiento permite obtener una célula huésped modificada por ingeniería genética que puede usarse para expresar y dirigir cualquier gen o genes deseables, por ejemplo, uno implicado en la glucosilación, por procedimientos que están bien establecidos en la bibliografía científica y se conocen generalmente por el experto en el campo de la expresión de proteínas. Se crean o seleccionan células huésped con oligosacáridos modificados. Los N-glucanos fabricados en

las células huésped modificadas por ingeniería genética tienen una estructura central de $Man_5GlcNAc_2$ que puede después modificarse adicionalmente por expresión heteróloga de una o más enzimas, por ejemplo, glucosiltransferasas, glucosidasas, transportadores de azúcares y manosidasas, para producir glucoproteínas de tipo humano. Para la producción de proteínas terapéuticas, este procedimiento puede adaptarse para modificar por ingeniería genética líneas celulares en las que puede obtenerse cualquier estructura de glucosilación deseada.

En consecuencia, en una realización, la invención proporciona un procedimiento para producir una glucoproteína de tipo humano en una célula huésped de hongo filamentoso o unicelular. La célula huésped de la invención se selecciona o modifica por ingeniería genética para anular sus actividades 1,6-manosiltransferasa que, de otro modo, añadirían restos de manosa al N-glucano en una glucoproteína. Se introducen una o más enzimas (actividades enzimáticas) en la célula huésped que permiten la producción de una estructura de carbohidrato $Man_5GlcNAc_2$ a un alto rendimiento, por ejemplo, al menos 30 % en moles. En una realización más preferida, al menos un 10 % de la $Man_5GlcNAc_2$ producida dentro de la célula huésped es un sustrato productivo para GnTI y por lo tanto, para reacciones de glucosilación adicionales *in vivo* y/o *in vitro* que producen un N-glucano terminado que es similar o idéntico al formado en mamíferos, especialmente seres humanos.

En otra realización, una molécula de ácido nucleico que codifica una o más enzimas para la producción de una estructura de carbohidrato $Man_5GlcNAc_2$ se introduce en una célula huésped seleccionada o modificada por ingeniería genética para anular sus actividades 1,6-manosiltransferasas. En una realización preferida, al menos una enzima introducida en la célula huésped se selecciona para tener actividad óptima al pH de la localización subcelular en la que se produce la estructura de carbohidrato. En otra realización preferida, al menos una enzima se dirige a un orgánulo subcelular huésped en el que la enzima tendrá actividad óptima, por ejemplo, por medio de una proteína quimérica que comprende un péptido señal de dirección celular no asociado normalmente con la enzima.

También se describen en el presente documento moléculas de ácido nucleico aisladas y vectores que comprenden tales moléculas que codifican una actividad α1,6-manosiltransferasa de inicio aislada de *P. pastoris* o de *K. lactis.* Estas moléculas de ácido nucleico comprenden secuencias que son homólogas del gen *OCH1* en *S. cerevisiae.* Estas y secuencias homólogas son útiles para construir células huésped que no hipermanosilen el N-glucano de una glucoproteína.

En otra realización, la célula huésped se modifica por ingeniería genética para expresar una glucosidasa heteróloga, por ejemplo, introduciendo en el huésped una o más moléculas de ácido nucleico que codifican la glucosidasa. Preferentemente, una molécula de ácido nucleico codifica una o más actividades manosidasa implicadas en la producción de $Man_5GlcNAc_2$ a partir de $Man_8GlcNAc_2$ o $Man_9GlcNAc_2$. En una realización preferida, al menos una de las actividades manosidasa codificadas tiene un pH óptimo a una distancia de 1,4 unidades de pH del pH óptimo medio de otras enzimas representativas en el orgánulo en el que se localiza la actividad manosidasa, o tiene actividad óptima un pH de entre aproximadamente 5,1 y aproximadamente 8,0, preferentemente entre aproximadamente 5,5 y aproximadamente 7,5. Preferentemente, la enzima heteróloga se dirige al retículo endoplásmico, el aparato de Golgi o las vesículas de transporte entre el RE y Golgi del organismo huésped, en el que recorta *N*-glucanos tales como $Man_8GlcNAc_2$ para producir altos niveles de $Man_5GlcNAc_2$. En una realización, la enzima se dirige formando una proteína de fusión entre un dominio catalítico de la enzima y un péptido señal de dirección celular por ejemplo, por la ligación en fase de un fragmento de ADN que codifica un péptido señal de dirección celular con un fragmento de ADN que codifica una enzima de glucosilación o fragmento catalíticamente activo de la misma.

En otra realización más, la ruta de glucosilación de un huésped se modifica para expresar un transportador de azúcares de nucleótido. En una realización preferida, también se expresa una enzima nucleótido difosfatasa. El transportador y la difosfatasa mejoran la eficacia de las etapas de glucosilación modificadas por ingeniería genética, proporcionando los sustratos apropiados para las enzimas de glucosilación en los compartimentos apropiados, reduciendo la inhibición competitiva de producto y promoviendo la retirada de nucleósido difosfatos.

También se describe en el presente documento una biblioteca de ácidos nucleicos combinatoria útil para preparar construcciones de fusión que pueden dirigir una proteína o fragmento polipeptídico deseado, por ejemplo, una enzima implicada en la glucosilación o un dominio catalítico de la misma, a una región subcelular seleccionada de una célula huésped. Como se describe en el presente documento, la biblioteca de ácidos nucleicos combinatoria comprende (a) secuencias de ácido nucleico que codifican diferentes péptidos señal de dirección celular y (b) secuencias de ácido nucleico que codifican diferentes polipéptidos diana. Las secuencias de ácido nucleico de o derivadas de (a) y (b) se ligan entre sí para producir construcciones de fusión, al menos una de las cuales codifica un dominio proteico funcional (por ejemplo, un dominio catalítico de una enzima) ligado en fase con un péptido señal de dirección celular heterólogo, es decir, uno con el que normalmente no se asocia.

La memoria descriptiva también enseña un procedimiento para modificar la ruta de glucosilación de una célula huésped de hongo unicelular o filamentoso usando enzimas implicadas en la modificación de N-glucanos incluyendo glucosidasas y glucosiltransferasas; mediante la transformación de la célula huésped con una biblioteca de ácidos nucleicos (por ejemplo, una combinatoria) de la invención para producir una población de células genéticamente mixta que expresa al menos una y preferentemente dos o más enzimas de glucosilación quiméricas distintas que tienen un dominio catalítico ligado en fase con un péptido señal de dirección celular con el que normalmente no se

asocia. Puede seleccionarse opcionalmente de la población una célula huésped que tenga un fenotipo de glucosilación deseado. Las células huésped modificadas usando la biblioteca y procedimientos asociados de la invención son útiles, por ejemplo, para producir glucoproteínas que tengan un patrón de glucosilación similar o idéntico a las producidas en mamíferos, especialmente seres humanos.

Como se describe en el presente documento, la biblioteca combinatoria especificada en la presente descripción permite la producción de una o una combinación de enzimas de glucosilación catalíticamente activas, que se localizan satisfactoriamente en compartimentos intracelulares en los que actúan eficazmente en la ruta de glucosilación/secretora. Las enzimas preferidas convierten $Man_5(\alpha\text{-}1,2\text{-}Man)_{3\text{-}9}GlcNAc_2$ en $Man_5GlcNAc_2$ con alta eficacia *in vivo*. Además, la invención proporciona cepas huésped eucariotas y en particular, huéspedes de levadura, fúngicos, de insecto, de plantas, de células vegetales, de algas y de células de insecto, capaces de producir intermedios de glucoproteína o productos con $Man_5GlcNAc2$ y/o $GlcNAcMan_5GlcNAc_2$ como el *N*-glucano predominante.

También se describen en el presente documento moléculas recombinantes derivadas de una biblioteca de ácidos nucleicos combinatoria; vectores, incluyendo vectores de expresión, que comprenden tales moléculas recombinantes; proteínas codificadas por las moléculas recombinantes y vectores; células huésped transformadas con las moléculas o vectores recombinantes; y glucoproteínas producidas a partir de tales huéspedes transformados. La presente invención proporciona procedimientos para producir, en hongos unicelulares o filamentosos, intermedios de glucoproteínas o productos con *N*-glucanos predominantemente $Man_5GlcNAc_2$ o $GlcNAcMan_5GlcNAc_2$ unidos covalentemente a sitios de glucosilación apropiados usando la biblioteca combinatoria.

También se describen en el presente documento procedimientos, composiciones y kits para usos terapéuticos y de diagnóstico en los que puede detectarse la presencia o ausencia de $Man_5GlcNAc_2$ y/o $GlcNAcMan_5GlcNAc_2$ en una glucoproteína.

La invención se especifica adicionalmente en las reivindicaciones.

Breve descripción de los dibujos

La **Figura 1A** es un diagrama esquemático de una ruta de *N*-glucosilación fúngica típica.

La **Figura 1B** es un diagrama esquemático de una ruta de *N*-glucosilación humana típica.

La **Figura 2** representa la construcción de una biblioteca de ADN combinatoria de construcciones de fusión. La **Figura 2A** esquematiza la inserción de un fragmento peptídico de dirección en pCR2.1-TOPO (Invitrogen, Carlsbad, CA). La **Figura 2B** muestra la sub-biblioteca de péptidos de dirección generados que tienen sitios de restricción *NotI-AscI*. La **Figura 2C** esquematiza la inserción de una región de dominio catalítico en pJN347, un vector pUC19 modificado. La **Figura 2D** muestra la sub-biblioteca de dominio catalítico generada que tiene sitios de restricción *NotI, AscI* y *PacI*. La **Figura 2E** representa una construcción de fusión particular generada a partir de la sub-biblioteca de péptidos de dirección y la sub-biblioteca de dominios catalíticos.

La **Figura 3 (SEC ID N.º: 45-46** respectivamente, en orden de aparición) ilustra la fase abierta de lectura de α-1,2-manosidasa IA de *M. musculus*. Las secuencias de los cebadores de PCR usados para generar truncamientos N-terminales están subrayadas.

Las **Figuras 4A-4F** ilustran la modificación por ingeniería genética de vectores con múltiples marcadores auxotróficos y la integración genética de proteínas diana en el locus *OCH1* de *P. pastoris*.

Las **Figuras 5A-5E** muestran un análisis de MALDI-TOF que demuestra producción del dominio kringle 3 de glucoproteínas de plasminógeno humano (K3) que tiene $Man_5GlcNAc_2$ como estructura de *N*-glucano predominante en *P. pastoris*. La **Figura 5A** representa el glucano convencional $Man_5GlcNAc_2$ [**a**] (Glyko, Novato, CA) y $Man_5GlcNAc_2 + Na^+$ [**b**]. La **Figura 5B** muestra glucanos liberados por PNGasa de K3 de tipo silvestre. Los *N*-glucanos mostrados son los siguientes: $Man_9GlcNAc_2$ [**d**]; $Man_{10}GlcNAc_2$ [**e**]; $Man_{11}GlcNAc_2$ [**f**]; $Man_{12}GlcNAc_2$ [**g**]. La **Figura 5C** representa la deleción de *ochI* que da como resultado la producción de $Man_8GlcNAc_2$ [**c**] como el *N*-glucano predominante. Las **Figuras 5D** y **5E** muestran la producción de $Man_5GlcNAc_2$ [**b**] después de reducción *in vivo* de $Man_8GlcNAc_2$ con una α-1,2-manosidasa quimérica. El *N*-glucano predominante se indica por un pico con una masa (m/z) de 1253 coherente con su identificación como $Man_5GlcNAc_2$ [**b**].

Las **Figuras 6A-6F** muestran el análisis de MALDI-TOF que demuestra la producción de glucoproteínas IFN-β que tienen $Man_5GlcNAc_2$ como la estructura de *N*-glucano predominante en *P. pastoris*. La **Figura 6A** muestra la $Man_5GlcNAc_2$ convencional [**a**] y $Man_5GlcNAc_2 + Na^+$ [**b**] como el patrón (Glyko, Novato, CA). La **Figura 6B** muestra glucanos liberados por PNGasa a partir de IFN-β de tipo silvestre. La **Figura 6C** representa el bloqueo de *ochI* que produce $Man_8GlcNAc_2$ [**c**]; $Man_9GlcNAc_2$ [**d**]; $Man_{10}GlcNAc_2$ [**e**]; $Man_{11}GlcNAc_2$ [**f**]; $Man_{12}GlcNAc_2$ [**g**]; y sin producción de $Man_5GlcNAc_2$ [**b**]. La **Figura 6D** muestra una cantidad relativamente pequeña de $Man_5GlcNAc_2$ [**b**] entre otros *N*-glucanos intermedios $Man_8GlcNAc_2$ [**c**] con respecto a $Man_{12}GlcNAc_2$ [**g**]. La **Figura 6E** muestra una cantidad significativa de $Man_5GlcNAc_2$ [**b**] en relación con los otros glucanos $Man_8GlcNAc_2$ [**c**] y $Man_9GlcNAc_2$ [**d**] producidos por pGC5 (*MNS1*(m) de *Saccharomyces*/manosidasa IB Δ99 de ratón). La **Figura 6F** muestra la

producción predominante de $Man_5GlcNAc_2$ [**b**] en la glucoproteína secretada IFN-β por pFB8 (*SEC12* (m) de *Saccharomyces/manosidasa* IA Δ187 de ratón). El *N*-glucano se indica por un pico con una masa (m/z) de 1254 coherente con su identificación como $Man_5GlcNAc_2$ [**b**].

La **Figura 7** muestra un cromatograma de cromatografía líquida de alta resolución para: (**A**) patrón de $Man_9GlcNAc_2$ marcado con 2-AB (control negativo); (**B**) sobrenadante de medio de *P. pastoris,* Δ*ochI* transformada con manosidasa de pFB8, que demuestra una falta de actividad manosidasa extracelular en el sobrenadante; y (**C**) patrón de $Man_9GlcNAc_2$ marcado con 2-AB después de la exposición a manosidasa de *T. reesei* (control positivo).

La **Figura 8** muestra un cromatograma de cromatografía líquida de alta resolución para: (**A**) patrón de $Man_9GlcNAc_2$ marcado con 2-AB (control negativo); (**B**) sobrenadante de medio de *P. pastoris,* Δ*ochI* transformada con manosidasa de pGC5, que demuestra una falta de actividad manosidasa extracelular en el sobrenadante; y (C) patrón de $Man_9GlcNAc_2$ marcado con 2-AB después de la exposición a manosidasa de *T. reesei* (control positivo).

La **Figura 9** muestra un cromatograma de cromatografía líquida de alta resolución para: (**A**) patrón de $Man_9GlcNAc_2$ marcado con 2-AB (control negativo); (**B**) sobrenadante de medio de *P. pastoris,* Δ*ochI* transformada con manosidasa de pBC18-5, que demuestra una falta de actividad manosidasa extracelular en el sobrenadante; y (**C**) sobrenadante de medio de *P. pastoris,* Δ*ochI* transformada con pDD28-3, que demuestra actividad en el sobrenadante (control positivo).

Las **Figuras 10A-10B** demuestran la actividad de un transportador de UDP-GlcNAc en la producción de $GlcNAcMan_5GlcNAc_2$ en *P. pastoris.* La **Figura 10A** representa una cepa de *P. pastoris* (YSH-3) con una GnTI humana pero sin el transportador de UDP-GlcNAc, que da como resultado algo de producción de $GlcNAcMan_5GlcNAc_2$ [**b**] pero una producción predominante de $Man_5GlcNAc_2$ [**a**]. La **Figura 10B** representa la adición de transportador UDP-GlcNAc de *K. lactis* en una cepa (PBP-3) con la GnTI humana, que dio como resultado la producción predominante de $GlcNAcMan_5GlcNAc_2$ [**b**]. El pico prominente sencillo de masa (m/z) a 1457 es coherente con su identificación como $GlcNAcMan_5GlcNAc_2$ [**b**] como se muestra en la **Figura 10B.**

La **Figura 11** muestra un pH óptimo de una enzima de manosidasa heteróloga codificada por pBB27-2 (*MNN10* (s) de *Saccharomyces/*manosidasa IB Δ31 de *C. elegans*) expresada en *P. pastoris.*

Las **Figuras 12A-12C** muestran el análisis de MALDI-TOF de *N*-glucanos liberados de un extracto sin células de *K. lactis.* La **Figura 12A** muestra los *N*-glucanos liberados de células de tipo silvestre, que incluyen *N*-glucanos del tipo de alto contenido en manosa. La **Figura 12B** muestra los *N*-glucanos liberados de células con *ochI mnnI* delecionado, que revelan un pico definido de masa (m/z) a 1908 coherente con su identificación como $Man_9GlcNAc_2$ [**d**]. La **Figura 12C** muestra los *N*-glucanos liberados de células con *och1 mnn1* delecionado después de la digestión de α-1,2-manosidasa *in vitro* correspondiente a un pico coherente con $Man_5GlcNAc_2$.

La **Figura 13** representa casetes de ADN-T con uno o más dominios catalíticos de enzimas de glucosilación fusionados en fase con diferentes secuencias líder. Los extremos del ADN-T están marcados por los bordes derecho (rb) e izquierdo (lb). También pueden usarse diversos promotores y terminadores. El marcador seleccionable de plantas también puede variarse. Los bordes derecho e izquierdo se requieren solamente para transformación mediada por *Agrobacterium* y no para bombardeo de partículas o electroporación.

Descripción detallada de la invención

A no ser que se definan de otro modo en el presente documento, los términos científicos y técnicos usados en relación con la presente invención tendrán los significados que se entienden habitualmente por los expertos en la materia. Además, a no ser que se requiera de otro modo por el contexto, los términos singulares incluirán pluralidades y los términos plurales incluirán el singular. Los procedimientos y técnicas de la presente invención se realizan generalmente de acuerdo con procedimientos convencionales bien conocidos en la técnica. Generalmente, las nomenclaturas usadas en relación con y las técnicas de bioquímica, enzimología, biología molecular y celular, microbiología, genética y química proteica y de ácidos nucleicos e hibridación descritas en el presente documento son las que son bien conocidas y se usan habitualmente en la técnica.

Los procedimientos y técnicas de la presente invención se realizan generalmente de acuerdo con procedimientos convencionales bien conocidos en la técnica y como se describe en diversas referencias generales y más específicas que se citan y analizan a lo largo de la presente memoria descriptiva a no ser que se indique otra cosa. Véase, por ejemplo, Sambrook y cols. Molecular Cloning: A Laboratory Manual, 2ª ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel y cols., Current Protocols in Molecular Biology, Greene Publishing Associates (1992 y suplementos hasta 2002); Harlow y Lane Antibodies: A Laboratory Manual Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Introduction to Glycobiology, Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003); Worthington Enzyme Manual, Worthington Biochemical Corp. Freehold, NJ; Handbook of Biochemistry: Section A Proteins vol. I 1976 CRC Press; Handbook of Biochemistry: Section A Proteins vol. II 1976 CRC Press; Essentials of Glycobiology, Cold Spring Harbor Laboratory Press (1999). Las nomenclaturas usadas en relación con y los procedimientos y técnicas de laboratorio de, biología molecular y celular, bioquímica de proteínas, enzimología y química medicinal y farmacéutica descritas en el presente documento son las que son bien conocidas y se usan habitualmente en la técnica.

Se entenderá que los siguientes términos, a no ser que se indique otra cosa, tienen los siguientes significados:

Como se usa en el presente documento, el término “N-glucano” se refiere a un oligosacárido ligado a N, por ejemplo, uno que se une por un enlace de asparagina-N-acetilglucosamina con un resto de asparagina de un polipéptido. Los N-glucanos tienen un núcleo de pentasacárido común de $Man_3GlcNAc_2$ ("Man" se refiere a manosa; "Glc" se refiere a glucosa; y "NAc" se refiere a N-acetilo; GlcNAc se refiere a N-acetilglucosamina). La expresión “núcleo de trimanosa” usada con respecto al N-glucano también se refiere a la estructura de $Man_3GlcNAc_2$ ("Man_3"). Los N-glucanos difieren con respecto al número de ramas (antenas) que comprenden azúcares periféricos (por ejemplo, fucosa y ácido siálico) que se añaden a la estructura central de Man_3. Los N-glucanos se clasifican de acuerdo con sus constituyentes ramificados (por ejemplo, alto contenido de manosa, complejo o híbrido).

Un N-glucano del tipo “de alto contenido de manosa” tiene cinco o más restos de manosa. Un N-glucano de tipo “complejo” normalmente tiene al menos un GlcNAc unido con la rama de 1,3 manosa y al menos una GlcNAc unida con la rama de 1,6 manosa del núcleo de trimanosa. Los N-glucanos complejos también pueden tener restos de galactosa ("Gal") que se han modificado opcionalmente con ácido siálico o derivados ("NeuAc", en el que "Neu" se refiere a ácido neuramínico y "Ac" se refiere a acetilo). Un N-glucano complejo normalmente tiene al menos una rama que termina en un oligosacárido tal como, por ejemplo: NeuNAc-; NeuAca2-6GalNAca1-; NeuAca2-3Galb1-3GalNAca1-; NeuAca2-3/6Galb1-4GlcNAcb1-; GlcNAca1-4Galb1-(solamente mucinas); Fuca1-2Galb1-(grupo sanguíneo H). Pueden aparecer ésteres de sulfato en restos de galactosa, GalNAc y GlcNAc y pueden aparecer ésteres de fosfato en restos de manosa. El NeuAc (Neu: ácido neuramínico; Ac: acetilo) puede estar O-acetilado o reemplazarse por NeuG1 (ácido N-glicolilneuramínico). Los N-glucanos complejos también pueden tener sustituciones intracatenarias que comprenden GlcNAc “de bisección” y fucosa central ("Fuc"). Un N-glucano “hibrido” tiene al menos una GlcNAc en el extremo de la rama de 1,3 manosa del núcleo de trimanosa y cero o más manosas en la rama de 1,6 manosa del núcleo de trimanosa.

El término “predominante” o “predominantemente”, usado con respecto a la producción de N-glucanos, se refiere a una estructura que representa el pico principal detectado por análisis de espectrometría de masas de dispersión de desorción de ionización por láser asistida por matriz (MALDI-TOF).

Las abreviaturas usadas en el presente documento son de uso común en la técnica, véanse, por ejemplo, las abreviaturas de azúcares, indicadas anteriormente. Otras abreviaturas comunes incluyen "PNGasa", que se refiere a una péptido N-glucosidasa F (EC 3.2.2.18); "GlcNAc Tr" o "GnT," que se refiere a enzimas *N*-acetilglucosaminil Transferasa; "NANA" se refiere a ácido *N*-acetilneuramínico.

Como se usa en el presente documento, una “glucoproteína humanizada” o una “glucoproteína de tipo humano” se refiere alternativamente a una proteína que tiene unida a la misma N-glucanos que tienen menos de cuatro restos de manosa, e intermedios de glucoproteína sintéticos (que también son útiles y pueden manipularse adicionalmente *in vitro* o *in vivo*) que tienen al menos cinco restos de manosa. Preferentemente, las glucoproteínas producidas de acuerdo con la invención contienen al menos un 30 % en moles, preferentemente al menos 40 % en moles y más preferentemente, un 50-100 % en moles, del intermedio $Man_5GlcNAc_2$, al menos transitoriamente. Esto puede conseguirse, por ejemplo, modificando por ingeniería genética una célula huésped de la invención para expresar una enzima de glucosilación “mejor”, es decir, más eficaz. Por ejemplo, una manosidasa se selecciona de modo que tenga actividad óptima en las condiciones presentes en el sitio de la célula huésped en el que las proteínas se glucosilan y se introduce en la célula huésped preferentemente por dirección de la enzima a un orgánulo de la célula huésped en el que se desea actividad.

El término “enzima”, cuando se usa en el presente documento en relación con la alteración de la glucosilación de la célula huésped, se refiere a una molécula que tiene al menos una actividad enzimática, e incluye enzimas de longitud completa, fragmentos catalíticamente activos, quiméricos, complejos y similares. Un “fragmento catalíticamente activo” de una enzima se refiere a un polipéptido que tiene un nivel detectable de actividad funcional (enzimática).

Una célula huésped eucariota inferior, cuando se usa en el presente documento en relación con perfiles de glucosilación, se refiere a cualquier célula eucariota que produce de forma habitual N-glucanos que tienen un alto contenido de manosa y por lo tanto, se pretende que incluya algunas células animales o vegetales y más normalmente, células eucariotas inferiores, incluyendo células fúngicas y de algas uni- y multicelulares.

Como se usa en el presente documento, la expresión “ruta de secreción” se refiere a la línea de ensamblaje de diversas enzimas de glucosilación a las que se exponen secuencialmente un precursor de oligosacárido ligado a lípidos y un sustrato de N-glucano, siguiendo el flujo molecular de una cadena polipeptídica naciente desde el citoplasma al retículo endoplásmico (RE) y los compartimentos del aparato de Golgi. Se dice que las enzimas se localizan a lo largo de esta ruta. Se dice que una enzima X que actúa en un glucano ligado a lípidos o en un N-glucano antes que la enzima Y está o actúa “corriente arriba” de la enzima Y; de forma similar, la enzima Y está o actúa “corriente abajo” de la enzima X.

La expresión “péptido de dirección”, como se usa en el presente documento, se refiere a secuencias de nucleótidos o aminoácidos que codifican un péptido señal de dirección celular que media en la localización (o retención) de una

secuencia asociada en localizaciones subcelulares, por ejemplo, orgánulos.

El término “polinucleótido” o “molécula de ácido nucleico” se refiere a una forma polimérica de nucleótidos de al menos 10 bases de longitud. El término incluye moléculas de ADN (por ejemplo, ADNc o ADN genómico o sintético) y moléculas de ARN (por ejemplo, ARNm o ARN sintético), así como análogos de ADN o ARN que contienen análogos nucleotídicos no naturales, enlaces internucleosídicos no nativos o ambos. El ácido nucleico puede estar en cualquier conformación topológica. Por ejemplo, el ácido nucleico puede ser monocatenario, bicatenario, tricatenario, cuádruple, parcialmente bicatenario, ramificado, en horquilla, circular o en una conformación de candado. El término incluye formas mono- y bicatenarias de ADN. Una molécula de ácido nucleico de la presente invención puede incluir hebras tanto sentido como antisentido de ARN, ADNc, ADN genómico y formas sintéticas y polímeros mixtos de los anteriores. Pueden modificarse de forma química o bioquímica o pueden contener bases nucleotídicas no naturales o derivatizadas, como se apreciará fácilmente por los expertos en la materia. Tales modificaciones incluyen, por ejemplo, marcadores, metilación, sustitución de uno o más de los nucleótidos de origen natural con un análogo, modificaciones internucleotídicas tales como enlaces no cargados (por ejemplo, metilfosfonatos, fosfotriésteres, fosforamidatos, carbamatos, etc.), enlaces cargados, (por ejemplo, fosforotioatos, fosforoditioatos, etc.), restos colgantes (por ejemplo, polipéptidos), intercaladores (por ejemplo, acridina, psoraleno, etc.), quelantes, alquilantes y enlaces modificados (por ejemplo, ácidos nucleicos anoméricos alfa, etc.). También se incluyen moléculas sintéticas que imitan a los polinucleótidos en su capacidad para unirse a una secuencia designada mediante enlaces de hidrógeno y otras interacciones químicas. Tales moléculas se conocen en la técnica e incluyen, por ejemplo, aquellas en las que los enlaces peptídicos sustituyen enlaces fosfato en la cadena principal de la molécula.

A no ser que se indique otra cosa, un “ácido nucleico que comprende la SEC ID N.º: X" se refiere a un ácido nucleico, al menos una parte del cual tiene (i) la secuencia de la SEC ID N.º: X, o (ii) una secuencia complementaria a la SEC ID N.º: X. La elección entre las dos se dicta por el contexto. Por ejemplo, si el ácido nucleico se usa como una sonda, la elección entre las dos se dicta por el requisito de que la sonda sea complementaria con la diana deseada.

Un ácido nucleico o polinucleótido “aislado” o “sustancialmente puro” (por ejemplo un ARN, ADN o un polímero mixto) es uno que está sustancialmente separado de otros componentes celulares que acompañan de forma natural al polinucleótido nativo en su célula huésped natural, por ejemplo, ribosomas, polimerasas y secuencias genómicas con las que se asocia de forma natural. El término abarca un ácido nucleico o polinucleótido que (1) se ha retirado de su ambiente de origen natural, (2) no está asociado con todo o una parte de un polinucleótido en el que el “polinucleótido aislado” se encuentra en la naturaleza, (3) está unido operativamente a un polinucleótido con el que no está unido en la naturaleza, o (4) no aparece en la naturaleza. La expresión “aislado” o “sustancialmente puro” también puede usarse en referencia a aislados de ADN recombinantes o clonados, análogos polinucleotídicos sintetizados químicamente o análogos polinucleotídicos que se sintetizan biológicamente por sistemas heterólogos.

Sin embargo, “aislado” no requiere necesariamente que el ácido nucleico o polinucleótido descrito así se haya retirado físicamente en sí mismo de su ambiente nativo. Por ejemplo, una secuencia de ácido nucleico endógena en el genoma de un organismo se considera “aislada” en el presente documento si una secuencia heteróloga (es decir, una secuencia que no está de forma natural adyacente a esta secuencia de ácido nucleico endógena) se coloca adyacente a la secuencia de ácido nucleico endógena, de modo que se altera la expresión de la secuencia de ácido nucleico endógena. A modo de ejemplo, una secuencia promotora no nativa puede sustituir (por ejemplo, por recombinación homóloga) la promotora nativa de un gen en el genoma de una célula humana, de modo que este gen tenga un patrón de expresión alterado. Este gen estaría ahora “aislado” debido a que está separado de al menos algunas de las secuencias que lo flanquean de forma natural.

Un ácido nucleico también se considera “aislado” si contiene cualquier modificación que no se produce de forma natural en el ácido nucleico correspondiente en un genoma. Por ejemplo, una secuencia codificante endógena se considera “aislada” si contiene una inserción, deleción o una mutación puntual introducida de forma artificial, por ejemplo, por intervención humana. Un “ácido nucleico aislado” también incluye un ácido nucleico integrado en un cromosoma de la célula huésped en un sitio heterólogo, una construcción de ácido nucleico presente como un episoma. Además, un “ácido nucleico aislado” puede carecer sustancialmente de otro material celular o carecer sustancialmente de medio de cultivo cuando se produce por técnicas recombinantes, o carecer sustancialmente de precursores químicos u otros compuestos químicos cuando se sintetiza de forma química.

Como se usa en el presente documento, la frase “variante degenerada” de una secuencia de ácido nucleico de referencia abarca secuencias de ácido nucleico que pueden traducirse, de acuerdo con el código genético convencional, para proporcionar una secuencia de aminoácidos idéntica a la traducida a partir de la secuencia de ácido nucleico de referencia.

La expresión “porcentaje de identidad de secuencia” o “idéntico”, en el contexto de secuencias de ácido nucleico, se refiere a los restos en las dos secuencias que son iguales cuando se alinean para una correspondencia máxima. La longitud de comparación de identidad de secuencia puede ser sobre un tramo de al menos aproximadamente nueve nucleótidos, habitualmente al menos aproximadamente 20 nucleótidos, más habitualmente al menos aproximadamente 24 nucleótidos, normalmente al menos aproximadamente 28 nucleótidos, más normalmente al

menos aproximadamente 32 nucleótidos y preferentemente al menos aproximadamente 36 o más nucleótidos. Hay varios algoritmos diferentes conocidos en la técnica que pueden usarse para medir la identidad de secuencia de nucleótidos. Por ejemplo, las secuencias polinucleotídicas pueden compararse usando FASTA, Gap o Bestfit, que son programas del Wisconsin Package Versión 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA proporciona alineamientos y porcentaje de intensidad de secuencia de las regiones de mejor solapamiento entre las secuencias de consulta y de búsqueda (Pearson, 1990). Por ejemplo, el porcentaje de identidad de secuencia entre las secuencias de ácido nucleico puede determinarse usando FASTA con sus parámetros por defecto (un tamaño de palabra de 6 y el factor NOPAM para la matriz de puntuación) o usando Gap con sus parámetros por defecto proporcionado en GCG Versión 6.1.

La expresión “homología sustancial” o “similitud sustancial”, cuando se refiere a un ácido nucleico o fragmento del mismo, indica que, cuando se alinea de forma óptima con inserciones o deleciones de nucleótidos apropiadas con otro ácido nucleico (o su hebra complementaria), hay identidad de secuencia de nucleótidos en al menos aproximadamente un 50 %, más preferentemente un 60 % de las bases de nucleótidos, habitualmente al menos aproximadamente un 70 %, más habitualmente al menos aproximadamente un 80 %, preferentemente al menos aproximadamente un 90 % y más preferentemente al menos aproximadamente un 95 %, 96 %, 97 %, 98 % o 99 % de las bases de nucleótidos, como se mide por cualquier algoritmo bien conocido de identidad de secuencia, tal como FASTA, BLAST o Gap, como se ha analizado anteriormente.

Como alternativa, existe homología o similitud sustancial cuando un ácido nucleico o fragmento del mismo hibrida con otro ácido nucleico, con una hebra de otro ácido nucleico o con la hebra complementaria del mismo, en condiciones de hibridación rigurosas. Las “condiciones de hibridación rigurosas” y “condiciones de lavado rigurosas”, en el contexto de experimentos de hibridación de ácido nucleico, dependen de varios parámetros físicos diferentes. La hibridación de ácidos nucleicos se verá afectada por condiciones tales como la concentración salina, temperatura, disolventes, la composición de bases de las especies que hibridan, la longitud de las regiones complementarias y el número de emparejamientos erróneos de bases nucleotídicas entre los ácidos nucleicos que hibridan, como se apreciará fácilmente por los expertos en la materia. Un experto habitual en la materia sabrá cómo variar estos parámetros para conseguir una rigurosidad particular de hibridación.

En general, la “hibridación rigurosa”, se realiza a aproximadamente 23 ºC por debajo del punto de fusión térmico (T_m) para el hibrido de ADN específico en un conjunto de condiciones particular. El “lavado riguroso” se realiza a temperaturas aproximadamente 5 ºC inferiores a la T_m para el híbrido de ADN específico en un conjunto particular de condiciones. La T_m es la temperatura a la que el 50 % de la secuencia diana hibrida con una sonda perfectamente coincidente. Véase Sambrook y cols., mencionado anteriormente, página 9.51. Para los fines del presente documento, las “condiciones de alta rigurosidad” se definen para la hibridación en fase de solución, como hibridación acuosa (es decir, sin formamida), en SSC 6X (donde SSC 20X contiene NaCl 3,0 M y citrato sódico 0,3 M), SDS 1 % a 65 ºC durante 8-12 horas, seguido de dos lavados en SSC 0,2X, SDS 0,1 % a 65 ºC durante 20 minutos. Se apreciará por los expertos en la materia que la hibridación a 65 ºC se producirá a diferentes velocidades dependiendo de varios factores que incluyen la longitud y el porcentaje de identidad de las secuencias que están hibridando.

El término “mutado” cuando se aplica a secuencias de ácido nucleico significa que los nucleótidos en una secuencia de ácido nucleico pueden insertarse, delecionarse o cambiarse en comparación con una secuencia de ácido nucleico de referencia. Puede realizarse una sola alteración en un locus (una mutación puntual) o pueden insertarse, delecionarse o cambiarse múltiples nucleótidos en un solo locus. Además, pueden realizarse una o más alteraciones en cualquier número de locus dentro de una secuencia de ácido nucleico. Una secuencia de ácido nucleico puede mutarse por cualquier procedimiento conocido en la técnica incluyendo pero sin limitación técnicas de mutagénesis tales como “PCR propensa a errores” (un procedimiento para realizar PCR en condiciones en las que la fidelidad de copia de la ADN polimerasa es baja, de modo que se obtiene una alta tasa de mutaciones puntuales a lo largo de la longitud completa del producto de PCR. Véase, por ejemplo, Leung, D. W. y cols., Technique, 1, pp. 11-15 (1989) y Caldwell, R. C. & Joyce G. F., PCR Methods Applic., 2, pp. 28-33 (1992)); y “mutagénesis dirigida por oligonucleótidos” (un procedimiento que permite la generación de mutaciones específicas de sitio en cualquier segmento de ADN clonado de interés. Véase, por ejemplo, Reidhaar-Olson, J. F. & Sauer, R T. y cols., Science, 241, pp. 53-57 (1988)).

El término “vector”, como se usa en el presente documento, pretende hacer referencia a una molécula de ácido nucleico capaz de transportar otro ácido nucleico con el que se ha unido. Un tipo de vector es un “plásmido”, que se refiere a un bucle de ADN bicatenario circular en el que pueden ligarse segmentos de ADN adicionales. Otros vectores incluyen cósmidos, cromosomas artificiales de bacterias (BAC) y cromosomas artificiales de levadura (YAC). Otro tipo de vector es un vector viral, en el que pueden ligarse segmentos de ADN adicionales en el genoma viral (analizado en más detalle posteriormente). Ciertos vectores tienen capacidad de replicación autónoma en una célula huésped en la que se han introducido (por ejemplo, vectores que tienen un origen de replicación que actúa en la célula huésped). Otros vectores pueden integrarse en el genoma de una célula huésped tras introducción en la célula huésped y de este modo se replican junto con el genoma del huésped. Además, ciertos vectores preferidos son capaces de dirigir la expresión de genes con los que están unidos operativamente. Tales vectores se denominan en el presente documento “vectores de expresión recombinantes” (o simplemente, “vectores de expresión”).

Las secuencias de control de la expresión “unidas operativamente” se refieren a un enlace en el que la secuencia de control de la expresión es contiguo al gen de interés para controlar el gen de interés, así como secuencias de control de la expresión que actúan en *trans* o a una distancia para controlar el gen de interés.

La expresión “secuencia de control de la expresión”, como se usa en el presente documento, se refiere a secuencias polinucleotídicas que son necesarias para afectar a la expresión de secuencias codificantes con las que están unidas operativamente. Las secuencias de control de la expresión son secuencias que controlan la transcripción, acontecimientos postranscripcionales y traducción de secuencias de ácido nucleico. Las secuencias de control de la expresión incluyen secuencias de inicio de la transcripción, terminación, promotoras y potenciadoras; señales de procesamiento de ARN eficaces tales como señales de corte y empalme y poliadenilación; secuencias que estabilizan el ARNm citoplásmico; secuencias que potencian la eficacia de traducción (por ejemplo, sitios de unión de ribosomas); secuencias que potencian la estabilidad de proteínas; y cuando se desee, secuencias que potencian la secreción de proteínas. La naturaleza de tales secuencias de control difiere dependiendo del organismo huésped; en procariotas, tales secuencias de control generalmente incluyen promotor, sitios de unión a ribosomas y secuencia de terminación de la transcripción. La expresión “secuencias de control” pretende incluir, como mínimo, todos los componentes cuya presencia es esencial para la expresión y también puede incluir componentes adicionales cuya presencia es ventajosa, por ejemplo, secuencias líder y secuencias de compañeros de fusión.

La expresión “célula huésped recombinante” (o simplemente “célula huésped”), como se usa en el presente documento, pretende referirse a una célula en la que se ha introducido un ácido nucleico tal como un vector recombinante. Debería entenderse que tales términos pretenden referirse no solamente a la célula objeto particular sino también a la descendencia de dicha célula. Debido a que pueden producirse ciertas modificaciones en generaciones sucesivas debido a mutación o influencias ambientales, dicha descendencia puede, de hecho, no ser idéntica a la célula parental, pero se sigue incluyendo dentro del alcance de la expresión “célula huésped” como se usa en el presente documento. Una célula huésped recombinante puede ser una célula o línea celular aislada desarrollada en cultivo o puede ser una célula que reside en un tejido u organismo vivo.

El término “péptido”, como se usa en el presente documento, se refiere a un polipéptido corto, por ejemplo, uno que tiene normalmente menos de aproximadamente 50 aminoácidos de longitudinal y más normalmente, menos de aproximadamente 30 aminoácidos de longitud. El término como se usa en el presente documento abarca análogos y miméticos que imitan la función estructural y por lo tanto, biológica.

El término “polipéptido”, como se usa en el presente documento, abarca proteínas tanto de origen natural como de origen no natural y fragmentos, mutantes, derivados y análogos de las mismas. Un polipéptido puede ser monomérico o polimérico. Además, un polipéptido puede comprender varios dominios diferentes de los que cada uno tiene una o más actividades distintas.

La expresión “proteína aislada” o “polipéptido aislado” es una proteína o polipéptido que, dependiendo de su origen o fuente de derivación, (1) no está asociado con componentes asociados de forma natural que lo acompañan en su estado nativo, (2) existe en una pureza no encontrada en la naturaleza, pudiendo estimarse la pureza con respecto a la presencia de otro material celular (por ejemplo, está libre de otras proteínas de la misma especie) (3) se expresa por una célula de una especie diferente, o (4) no aparece en la naturaleza (por ejemplo, es un fragmento de un polipéptido hallado en la naturaleza o incluye análogos o derivados de aminoácidos no hallados en la naturaleza o enlaces distintos de enlaces peptídicos convencionales). Por lo tanto, un polipéptido que se sintetiza químicamente o se sintetiza en un sistema celular diferente de la célula de la que procede de forma natural estará “aislado” de sus componentes asociados de forma natural. Un polipéptido o proteína también puede liberarse sustancialmente de componentes asociados de forma natural por aislamiento, usando técnicas de purificación de proteínas bien conocidas en este campo. Según se define de este modo, “aislado” no requiere necesariamente que la proteína, polipéptido, péptido u oligopéptido descrito de este modo se haya retirado físicamente de su ambiente nativo.

La expresión “fragmento polipeptídico”, como se usa en el presente documento, se refiere a un polipéptido que tiene una deleción amino-terminal y/o carboxi-terminal en comparación con un polipéptido de longitud completa. En una realización preferida, el fragmento polipeptídico es una secuencia contigua en la que la secuencia de aminoácidos del fragmento es idéntica a las posiciones correspondientes en la secuencia de origen natural. Los fragmentos normalmente son de al menos 5, 6, 7, 8, 9 o 10 aminoácidos de longitud, preferentemente de al menos 12, 14, 16 o 18 aminoácidos de longitud, más preferentemente de al menos 20 aminoácidos de longitud, más preferentemente de al menos 25, 30, 35, 40 o 45 aminoácidos, incluso más preferentemente de al menos 50 o 60 aminoácidos de longitud y aún más preferentemente de al menos 70 aminoácidos de longitud.

Un “derivado modificado” se refiere a polipéptidos o fragmentos de los mismos que son sustancialmente homólogos en secuencia estructural primaria pero que incluyen, por ejemplo, modificaciones químicas y bioquímicas *in vivo* o *in vitro* o que incorporan aminoácidos que no se encuentran en el polipéptido nativo. Tales modificaciones incluyen, por ejemplo, acetilación, carboxilación, fosforilación, glucosilación, ubiquitinación, marcaje, por ejemplo, con radionúclidos y diversas modificaciones enzimáticas, como se apreciará fácilmente por los expertos en la materia. Se conocen bien en la técnica una diversidad de procedimientos para marcar polipéptidos y de sustituyentes o marcadores útiles para tales fines, e incluyen isótopos radiactivos tales como ^{125}I, ^{32}P, ^{35}S y ^{3}H, ligandos que se unen a antiligandos marcados (por ejemplo, anticuerpos), fluoróforos, agentes quimioluminiscentes, enzimas y

antiligandos que pueden actuar como miembros de pares de unión específicos para un ligando marcado. La elección de marcador depende de la sensibilidad requerida, facilidad de conjugación con el cebador, requisitos de estabilidad e instrumentación disponible. Se conocen bien en la técnica procedimientos para marcar polipéptidos. Véase Ausubel y cols., 1992.

Un "mutante polipeptídico" o "muteína" se refiere a un polipéptido cuya secuencia contiene una inserción, duplicación, deleción, reordenamiento o sustitución de uno o más aminoácidos en comparación con la secuencia de aminoácidos de una proteína nativa o de tipo silvestre. Una muteína puede tener una o más sustituciones puntuales de aminoácidos en las que un aminoácido sencillo en una posición se ha cambiado por otro aminoácido, una o más inserciones y/o deleciones, en las que uno o más aminoácidos se insertan o delecionan, respectivamente, en la secuencia de la proteína de origen natural y/o truncamientos de la secuencia de aminoácidos en uno o los dos extremos amino o carboxilo terminales. Una muteína puede tener la misma, pero preferentemente tiene una actividad biológica diferente en comparación con la proteína de origen natural. Por ejemplo, una muteína puede tener una actividad de unión con NgR o neurona aumentada o reducida. En una realización preferida de la presente invención, un derivado de MAG que es una muteína (por ejemplo, en el dominio de tipo Ig de MAG 5) tiene una actividad inhibidora del crecimiento neuronal reducida en comparación con MAG de tipo silvestre endógeno o soluble.

Una muteína tiene al menos un 70 % de homología de secuencia global con su homólogo de tipo silvestre. Se prefieren aún más muteínas que tengan un 80 %, 85 % o 90 % de homología de secuencia global con la proteína de tipo silvestre. En una realización incluso más preferida, una muteína muestra un 95 % de identidad de secuencia, aún más preferentemente un 97 %, aún más preferentemente un 98 % y aún más preferentemente un 99 % de identidad global de secuencia. La homología de secuencia puede medirse por cualquier algoritmo de análisis de secuencia habitual, tal como Gap o Bestfit.

Son sustituciones de aminoácidos preferidas las que: (1) reducen la susceptibilidad a la proteolisis, (2) reducen la susceptibilidad a oxidación, (3) alteran la afinidad de unión para formar complejos de proteínas, (4) alteran la afinidad de unión o actividad enzimática y (5) confieren o modifican otras propiedades fisicoquímicas o funcionales de tales análogos.

Como se usa en el presente documento, los veinte aminoácidos convencionales y sus abreviaturas siguen el uso convencional. Véase Immunology -A Synthesis (2ª edición, E.S. Golub y D.R. Gren, Eds., Sinauer Associates, Sunderland, Mass. (1991)). Los estereoisómeros (por ejemplo, D-aminoácidos) de los veinte aminoácidos convencionales, aminoácidos no naturales tales como aminoácidos α-,α-disustituidos, N-alquilaminoácidos y otros aminoácidos poco convencionales también pueden ser componentes adecuados para los polipéptidos de la presente invención. Los ejemplos de aminoácidos no convencionales incluyen: 4-hidroxiprolina, γ-carboxiglutamato, ε-N,N,N-trimetil-lisina, ε-N-acetil-lisina, O-fosfoserina, N-acetilserina, N-formilmetionina, 3-metilhistidina, 5-hidroxilisina, s-N-metilarginina y otros aminoácidos e iminoácidos similares (por ejemplo, 4-hidroxiprolina). En la notación de polipéptidos usada en el presente documento, la dirección hacia la izquierda es la dirección amino-terminal y la dirección hacia la derecha es la dirección carboxi-terminal, de acuerdo con el uso convencional y la convención.

Una proteína tiene "homología" o es "homóloga" a una segunda proteína si la secuencia de ácido nucleico que codifica la proteína tiene una secuencia similar a la secuencia de ácido nucleico que codifica la segunda proteína. Como alternativa, una proteína tiene una homología con una segunda proteína si las dos proteínas tienen secuencias de aminoácidos "similares". (Por lo tanto, se define que la expresión "proteínas homólogas" significa que las dos proteínas tienen secuencias de aminoácidos similares). En una realización preferida, una proteína homóloga es una que muestra el 60 % de homología de secuencia con la proteína de tipo silvestre, se prefiere más un 70 % de homología de secuencia. Se prefieren aún más proteínas homólogas que muestran un 80 %, 85 % o 90 % de homología de secuencia con la proteína de tipo silvestre. En una realización aún más preferida, una proteína homóloga muestra un 95 %, 97 %, 98 % o 99 % de identidad de secuencia. Como se usa en el presente documento, se interpreta que la homología entre dos regiones de la secuencia de aminoácidos (especialmente con respecto a similitudes estructurales previstas) implica similitud en la función.

Cuando se usa "homólogo" en referencia a proteínas o péptidos, se reconoce que las posiciones de restos que no son idénticos con frecuencia difieren por sustituciones de aminoácidos conservativas. Una "sustitución de aminoácido conservativa" es una en la que un resto de aminoácido está sustituido por otro resto de aminoácido que tiene una cadena lateral (grupo R) con propiedades químicas similares (por ejemplo, carga o hidrofobia). En general, una sustitución de aminoácido conservativa no cambiará sustancialmente las propiedades funcionales de una proteína. En casos en los que dos o más secuencias de aminoácidos difieren entre sí por sustituciones conservativas, el porcentaje de identidad de secuencia o grado de homología puede ajustarse hacia arriba para corregir con respecto a la naturaleza conservativa de la sustitución. Los medios para realizar este ajuste se conocen bien por los expertos en la materia (véase, por ejemplo, Pearson y cols., 1994).

Los siguientes seis grupos contienen, cada uno, aminoácidos que son sustituciones conservativas entre sí: 1) Serina (S), Treonina (T); 2) Ácido Aspártico (D), Ácido Glutámico (E); 3) Asparagina (N), Glutamina (Q); 4) Arginina (R), Lisina (K); 5) Isoleucina (I), Leucina (L), Metionina (M), Alanina (A), Valina (V) y 6) Fenilalanina (F), Tirosina (Y), Triptófano (W).

La homología de secuencia para polipéptidos, que también se denomina porcentaje de identidad de secuencia, se mide normalmente usando un software de análisis de secuencia. Véase, por ejemplo, el Paquete de Software de Análisis de Secuencia del Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. El software de análisis de proteínas empareja secuencias similares usando una medida de homología asignada a diversas sustituciones, deleciones y otras modificaciones, incluyendo sustituciones de aminoácidos conservativas. Por ejemplo, GCG contiene programas tales como "Gap" y "Bestfit" que pueden usarse con parámetros por defecto para determinar la homología de secuencia o identidad de secuencia entre polipéptidos muy relacionados, tales como polipéptidos homólogos de diferentes especies de organismos o entre una proteína de tipo silvestre y una muteína de la misma. Véase, por ejemplo, GCG Versión 6.1.

Un algoritmo preferido cuando se compara una secuencia de molécula inhibidora con una base de datos que contiene un gran número de secuencias de diferentes organismos es el programa informático BLAST (Altschul, S.F. y cols. (1990) J. Mol. Biol. 215: 403-410; Gish y States (1993) Nature Genet. 3: 266-272; Madden, T.L. y cols. (1996) Meth. Enzymol. 266: 131-141; Altschul, S.F. y cols. (1997) Nucleic Acids Res.25: 3389-3402; Zhang, J. y Madden, T.L. (1997) *Genome Res.* 7: 649-656), especialmente blastp o tblastn (Altschul y cols., 1997). Son parámetros preferidos para BLASTp: valor de expectativa: 10 (por defecto); Filtro: s (por defecto); Coste para abrir un hueco: 11 (por defecto); Coste para extender un hueco: 1 (por defecto); Alineamientos máximos: 100 (por defecto); Tamaño de palabra: 11 (por defecto); N.º de descripciones: 100 (por defecto); Matriz de penalización: BLOWSUM62.

La longitud de las secuencias polipeptídicas comparadas con respecto a la homología generalmente será de al menos aproximadamente 16 restos de aminoácido, habitualmente al menos aproximadamente 20 restos, más habitualmente al menos aproximadamente 24 restos, normalmente al menos aproximadamente 28 restos y preferentemente más de aproximadamente 35 restos. Cuando se busca en una base de datos que contiene secuencias de un gran número de organismos diferentes, es preferible comparar secuencias de aminoácidos. La búsqueda en bases de datos usando secuencias de aminoácidos puede medirse por algoritmos distintos de blastp conocidos en la técnica. Por ejemplo, pueden compararse secuencias polipeptídicas usando FASTA, un programa en GCG Versión 6.1. FASTA proporciona alineamientos y porcentaje de identidad de secuencia de las regiones de mejor solapamiento entre las secuencias de consulta y de búsqueda (Pearson, 1990). Por ejemplo, el porcentaje de identidad de secuencia entre secuencias de aminoácidos puede determinarse usando FASTA con sus parámetros por defecto (un tamaño de palabra de 2 y la matriz de puntuación PAM250), como se proporciona en GCG Versión 6.1.

La expresión “proteína de fusión” se refiere a un polipéptido que comprende un polipéptido o fragmento acoplado a secuencias de aminoácidos heterólogas. Las proteínas de fusión son útiles porque pueden construirse para contener dos o más elementos funcionales deseados de dos o más proteínas diferentes. Una proteína de fusión comprende al menos 10 aminoácidos contiguos de un polipéptido de interés, más preferentemente al menos 20 o 30 aminoácidos, aún más preferentemente al menos 40, 50 o 60 aminoácidos, aún más preferentemente al menos 75, 100 o 125 aminoácidos. Las proteínas de fusión pueden producirse de forma recombinante construyendo una secuencia de ácido nucleico que codifique el polipéptido o un fragmento del mismo en fase con una secuencia de ácido nucleico que codifique una proteína o péptido diferente y expresando después la proteína de fusión. Como alternativa, una proteína de fusión puede producirse químicamente reticulando el polipéptido o un fragmento del mismo con otra proteína.

El término “región”, como se usa en el presente documento, se refiere a una parte físicamente contigua de la estructura primaria de una biomolécula. En el caso de proteínas, una región se define por una parte contigua de la secuencia de aminoácidos de esa proteína.

El término “dominio”, como se usa en el presente documento, se refiere a una estructura de una biomolécula que contribuye a una función conocida o sospechada de la biomolécula. Los dominios pueden ser coextensivos con regiones o partes de los mismos; los dominios también pueden incluir regiones distintas no contiguas de una biomolécula. Los ejemplos de dominios de proteínas incluyen, pero sin limitación, un dominio Ig, un domino extracelular, un dominio transmembrana y un dominio citoplásmico.

Como se usa en el presente documento, el término “molécula” significa cualquier compuesto, incluyendo, pero sin limitación, una molécula pequeña, péptido, proteína, azúcar, nucleótido, ácido nucleico, lípido, etc. y dicho compuesto puede ser natural o sintético.

A no ser que se defina de otro modo, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que se entiende habitualmente por un experto habitual en la materia a la que pertenece la presente invención. Más adelante se describen procedimientos y materiales ejemplares, aunque también pueden usarse procedimientos y materiales similares o equivalentes a los descritos en el presente documento en la práctica de la presente invención y resultarán evidentes para los expertos en la materia. En caso de conflicto prevalecerá la presente memoria descriptiva, incluyendo las definiciones. Los materiales, procedimientos y ejemplos son solamente ilustrativos y no se pretende que sean limitantes.

A lo largo de la presente memoria descriptiva y de las reivindicaciones, se entenderá que la palabra “comprender” o variaciones tales como “comprende” o “que comprende”, implica la inclusión de un número entero indicado o grupo

de números enteros pero no la exclusión de ningún otro número entero o grupo de números enteros.

Procedimientos para producir células huésped que tienen oligosacáridos modificados $Man_5GlcNAc_2$ para la generación de N-Glucanos de tipo humano

Se describe en el presente documento un procedimiento para producir una glucoproteína que tiene glucosilación de tipo humano en una célula huésped eucariota no humana. Como se describe en más detalle posteriormente, como una célula huésped de partida se selecciona una célula huésped de hongos unicelulares o filamentosos que no expresa de forma natural, o que se modifica por ingeniería genética para no expresar, una o más enzimas implicadas en la producción de estructuras de alto contenido de manosa. Una célula huésped seleccionada tal se modifica por ingeniería genética para expresar una o más enzimas u otros factores requeridos para producir glucoproteínas de tipo humano. En una cepa huésped deseada puede modificarse por ingeniería genética una enzima o más de una enzima cada vez. Además, puede usarse una molécula de ácido nucleico que codifique una o más enzimas o actividades para modificar por ingeniería genética una cepa huésped de los procedimientos de la invención. Como se describe, se crea una biblioteca de moléculas de ácido nucleico que codifican enzimas potencialmente útiles (por ejemplo, enzimas quiméricas que comprenden un fragmento de enzima catalíticamente activo ligadas en fase con una secuencia de dirección subcelular heteróloga) (por ejemplo, por ligación de sub-bibliotecas que comprenden fragmentos enzimáticos y secuencias de dirección subcelular) y puede seleccionarse una cepa que tenga una o más enzimas con actividades óptimas o que produzca las glucoproteínas más “de tipo humano” transformando células huésped diana con uno o más miembros de la biblioteca.

En consecuencia, los procedimientos descritos en el presente documento permiten obtener, *in vivo,* estructuras de $Man_5GlcNAc_2$ en alto rendimiento, al menos de forma transitoria, con el fin de modificarlas adicionalmente para producir N-glucanos complejos. Un esquema exitoso para obtener estructuras de $Man_5GlcNAc_2$ adecuadas en rendimientos apropiados en una célula huésped de hongos unicelulares o filamentosos generalmente implica dos enfoques paralelos: (1) reducir de estructuras de alto contenido de manosa obtenidas por actividades manosiltransferasa endógenas, si las hubiera y (2) retirar la 1,2-α-manosa mediante manosidasas para producir niveles altos de estructuras $Man_5GlcNAc_2$ adecuadas que pueden reaccionar adicionalmente dentro de la célula huésped para formar glucoformas de tipo humano complejas.

En consecuencia, una primera etapa implica la selección o creación de una célula huésped de hongos unicelulares o filamentosos capaz de producir una estructura precursora específica de $Man_5GlcNAc_2$ que es capaz de aceptar GlcNAc *in vivo* mediante la acción de una GlcNAc transferasa I (“GnTT”). En una realización, el procedimiento implica fabricar o usar una célula huésped de hongos unicelulares o filamentosos con la actividad 1,6 manosiltransferasa eliminada con respecto al N-glucano en una glucoproteína. Preferentemente, la célula huésped tiene la actividad 1,6 manosiltransferasa de inicio eliminada (véase posteriormente). Dicha célula huésped carecerá de una o más enzimas implicadas en la producción de estructuras de alto contenido de manosa que no son deseables para producir glucoproteínas de tipo humano.

Después se introducen una o más actividades enzimáticas en dicha célula huésped para producir N-glucanos dentro de la célula huésped caracterizados por tener al menos un 30 % en moles de las estructura de carbohidrato $Man_5GlcNAc_2$ (“Man_5”). Son necesarias estructuras $Man_5GlcNAc_2$ para la formación de N-glucanos complejos: debe formarse $Man_5GlcNAc_2$ *in vivo* con un alto rendimiento (por ejemplo, en un exceso del 30 %), al menos transitoriamente, puesto que las reacciones posteriores de glucosilación de tipo mamífero y humano requieren la estructura de $Man_5GlcNAc_2$ o un derivado de la misma.

Esta etapa también requiere la formación de una estructura isomérica particular de $Man_5GlcNAc_2$ dentro de la célula a un alto rendimiento. Aunque las estructuras $Man_5GlcNAc_2$ son necesarias para formación de N-glucanos complejos, su presencia no es en absoluto suficiente. Esto se debe a que $Man_5GlcNAc_2$ puede aparecer en diferentes formas isoméricas que pueden o no actuar como un sustrato para GlcNAc transferasa I. Como la mayoría de las reacciones de glucosilación no son completas, una proteína glucosilada particular generalmente contiene una serie de estructuras de carbohidrato diferentes (es decir glucoformas) en su superficie. Por lo tanto, la mera presencia de cantidades muy pequeñas (es decir, menos del 5 %) de una estructura particular como $Man_5GlcNAc_2$ es de poca relevancia práctica para producir glucoproteínas de tipo mamífero o humano. Lo que se requiere es la formación de un intermedio de $Man_5GlcNAc_2$ aceptor de GlcNAc transferasa I (**Figura 1B**) en alto rendimiento (es decir, por encima del 30 %). La formación de este intermedio es necesaria para permitir la síntesis *in vivo* posterior de N-glucanos complejos en proteínas glucosiladas de interés (proteínas diana).

En consecuencia, parte o toda la $Man_5GlcNAc_2$ producida por la célula huésped seleccionada debe ser un sustrato productivo para actividades enzimáticas a lo largo de una ruta de glucosilación de mamífero, por ejemplo, puede actuar como un sustrato para una actividad GlcNAc transferasa I *in vivo,* formando de este modo el intermedio de N-glucano de tipo humano $GlcNAcMan_5GlcNAc_2$ en la célula huésped. En una realización preferida, al menos un 10 %, más preferentemente al menos un 30 % y más preferentemente un 50 % o más del intermedio de $Man_5GlcNAc_2$ producido en la célula huésped de la invención es un sustrato productivo para GnTI *in vivo.* Debe entenderse que si, por ejemplo, $GlcNAcMan_5GlcNAc_2$ se produce en un 10 % y $Man_5GlcNAc_2$ se produce en un 25 % en una proteína diana, la cantidad total de $Man_5GlcNAc_2$ producida de forma transitoria es del 35 % debido a que $GlcNAcMan_5GlcNAc_2$ es un producto de $Man_5GlcNAc_2$.

Un experto habitual en la materia puede seleccionar células huésped de la naturaleza, por ejemplo, hongos existentes u otros eucariotas inferiores que producen niveles significativos de $Man_5GlcNAc_2$ *in vivo.* Hasta la fecha, sin embargo, no se ha mostrado que ningún eucariota inferior proporcione tales estructuras *in vivo* por encima del 1,8 % de los N-glucanos totales (véase, por ejemplo, Maras y cols., 1997). Como alternativa, dichas células huésped pueden modificarse por ingeniería genética para producir la estructura de $Man_5GlcNAc_2$ *in vivo.* Pueden usarse procedimientos tales como los descritos en la Patente de Estados Unidos N.º: 5.595.900 para identificar la ausencia o presencia de glucosiltransferasas particulares, manosidasas y transportadores de azúcares de nucleótido en una célula huésped diana u organismo de interés.

Inactivación de enzimas de glucosilación de células huésped no deseadas

Los procedimientos de la invención se dirigen a preparar células huésped que producen glucoproteínas que tienen estructuras de N-glucanos y preferentemente de tipo humano, alteradas. En una realización preferida, los procedimientos se dirigen a preparar células huésped en las que se enriquecen precursores de oligosacáridos en $Man_5GlcNAc_2$. Preferentemente, se usa una célula huésped de hongos unicelulares o filamentosos que no expresa una o más enzimas implicadas en la producción de estructuras de alto contenido de manosa. Dicha célula huésped puede encontrarse en la naturaleza o puede modificarse por ingeniería genética, por ejemplo, comenzando con o derivada de uno de los muchos mutantes ya descritos en levaduras. Por lo tanto, dependiendo de la célula huésped seleccionada, tendrá que delecionarse uno o varios genes que codifican enzimas que se sabe que son características de reacciones de glucosilación no humanas. Dichos genes y sus proteínas correspondientes se han caracterizado de forma exhaustiva en varios eucariotas inferiores (por ejemplo, *S. cerevisiae, T. reesei, A. nidulans* etc.), proporcionado de este modo una lista de glucosiltransferasas conocidas en eucariotas inferiores, sus actividades y su secuencia genética respectiva. Estos genes probablemente se seleccionarán del grupo de manosiltransferasas, por ejemplo, 1,3 manosiltransferasas (por ejemplo, *MNN1* en *S. cerevisiae*) (Graham, 1991), 1,2 manosiltransferasas (por ejemplo, la familia *KTRlKRE* de *S. cerevisiae*), 1,6 manosiltransferasas (*OCH1* de *S. cerevisiae*), manosilfosfato transferasas y sus reguladores (*MNN4* y *MNN6* de *S. cerevisiae*) y enzimas adicionales que están implicadas en reacciones de glucosilación aberrantes, es decir no humanas. Muchos de estos genes de hecho se han delecionado individualmente dando lugar a fenotipos viables con perfiles de glucosilación alterados. Se muestran ejemplos en la **Tabla 1**.

Son células huésped eucariotas inferiores preferidas de la invención, como se describe en el presente documento para ejemplificar las etapas de manipulación requeridas, mutantes de hipermanosilación-menos *(och1*) de *Pichia pastoris* o *K. lactis.* Al igual que otros eucariotas inferiores, *P. pastoris* procesa estructuras de $Man_9GlcNAc_2$ en el RE con una α-1,2-manosidasa para producir $Man_8GlcNAc_2$ (**Figura 1A**). A través de la acción de varias manosiltransferasas, esta estructura se convierte después en estructuras hipermanosiladas ($Man_{>9}GlcNAc_2$), también conocidas como mananos. Además, se ha descubierto que *P. pastoris* es capaz de añadir grupos fosfato no terminales, a través de la acción de manosilfosfato transferasas, a la estructura de carbohidrato. Esto difiere de las reacciones realizadas en células de mamífero, que implican la retirada en lugar de la adición de azúcares de manosa. Es de particular importancia para eliminar la capacidad de la célula huésped eucariota, por ejemplo, del hongo, hipermanosilar una estructura de $Man_8GlcNAc_2$ existente. Esto puede conseguirse seleccionando una célula huésped que no hipermanosile o modificando por ingeniería genética dicha célula.

Se han identificado genes que están implicados en el procedimiento de hipermanosilación, por ejemplo, en *Pichia pastoris y* creando mutaciones en estos genes, se puede reducir la producción de glucoformas “no deseables”. Dichos genes pueden identificarse por homología con manosiltransferasas existentes o sus reguladores (por ejemplo, *OCH1, MNN4, MNN6, MNN1*) hallados en otros eucariotas inferiores tales como C. *albicans, Pichia angusta* o *S. cerevisiae* o mutando la cepa huésped y seleccionando un fenotipo de glucosilación con manosilación reducida. Basándose en homologías entre manosiltransferasas y manosilfosfatotransferasas conocidas, se pueden diseñar cebadores de PCR **(SEC ID N.º: 7, 8, 47** y **4** de izquierda a derecha, respectivamente, por orden de aparición) (se muestran ejemplos de estos en la **Tabla 2**), o usar genes o fragmentos génicos que codifican tales enzimas como sondas para identificar homólogos en bibliotecas de ADN del organismo diana o un organismo relacionado. Como alternativa, se puede identificar un homólogo funcional que tiene actividad manosiltransferasa por su capacidad para complementar fenotipos de glucosilación particulares en organismos relacionados.

Tabla 2. Cebadores de PCR

Cebador de PCR A	Cebador de PCR B	Gen o Genes Diana en *P. pastoris*	Homólogos
ATGGCGAAGGCAGA TGGCAGT	**TTAGTCCTTCCAAC TTCCTTC**	1,6-manosiltransferasa	OCH1 *S. cerevisiae*, *Pichia albicans*
TAYTGGMGNGTNGA RCYNGAYATHAA	**GCRTCNCCCCANCK YTCRTA**	1,2 manosiltransferasas	familia KTR/KRE, *S. cerevisiae*
Leyenda: M = A o C, R = A o G, W = A o T, S = C o G, Y = C o T, K = G o T, V = A o C o G, H = A o C o T, D = A o G o T, B = C o G o T, N = G o A o T o C.			

Para obtener el gen o genes que codifican actividad 1,6-manosiltransferasa en *P. pastoris,* por ejemplo, se llevarían a cabo las siguientes etapas: los mutantes *OCH1* de *S. cerevisiae* son sensibles a la temperatura y proliferan lentamente a temperaturas elevadas. Por lo tanto, se pueden identificar homólogos funcionales *OCH1* en *P. pastoris* complementando un mutante *OCH1* de S. *cerevisiae* con una biblioteca de ADNc o ADN de *P. pastoris*. Están disponibles mutantes de S. *cerevisiae,* por ejemplo, de la Universidad de Stanford y están disponibles en el mercado en ResGen, una Invitrogen Corp. (Carlsbad, CA). Los mutantes que presentan un fenotipo de proliferación normal a temperatura elevada después de haberse transformado con una biblioteca de ADN de *P. pastoris*, probablemente portarán un homólogo de *OCH1* de *P. pastoris.* Dicha biblioteca puede crearse digiriendo parcialmente ADN cromosómico de *P. pastoris* con una enzima de restricción adecuada y después de inactivar la enzima de restricción, ligando el ADN digerido en un vector adecuado, que se ha digerido con una enzima de restricción compatible.

Los vectores adecuados incluyen, por ejemplo, pRS314, un plásmido de bajo número de copias (CEN6/ARS4) basado en pBluescript que contiene el marcador Trp1 (Sikorski, R. S. y Hieter, P., 1989, Genetics 122, pg 19-27) y pFL44S, un plásmido de alto número de copias (2μ) basado en un pUC19 modificado que contiene el marcador URA3 (Bonneaud, N. y cols., 1991, Yeast 7, pg. 609-615). Tales vectores se usan habitualmente por investigadores académicos y están disponibles vectores similares de varios distribuidores diferentes (por ejemplo, Invitrogen (Carlsbad, CA); Pharmacia (Piscataway, NJ); New England Biolabs (Beverly, MA)). Otros ejemplos adicionales incluyen pYES/GS, el plásmido de expresión de levadura basado en origen de replicación 2μ de Invitrogen, o el vehículo de clonación Yep24 de New England Biolabs.

Después de la ligación del ADN cromosómico y el vector, se puede introducir la biblioteca de ADN por transformación en una cepa de *S. cerevisiae* con una mutación específica y seleccionar la corrección del fenotipo correspondiente. Después de subclonar y secuenciar el fragmento de ADN que es capaz de restaurar el fenotipo de tipo silvestre, se puede usar este fragmento para eliminar la actividad del producto génico codificado por *OCH1* en *P. pastoris* usando mutagénesis *in vivo* y/o técnicas de recombinación buen conocidas por los expertos en la materia.

Como alternativa, si se conoce la secuencia genómica completa de una célula huésped particular, por ejemplo, hongo, de interés, se pueden identificar tales genes simplemente buscando en las bases de datos de ADN públicamente disponibles, que están disponibles en varias fuentes, tales como NCBI, Swissprot. Por ejemplo, al buscar en una secuencia genómica o base de datos dada secuencias de un gen de 1,6 manosiltransferasa conocido (por ejemplo, *OCH1* de *S. cerevisiae*), se pueden identificar genes de alta homología en dicho genoma de la célula huésped que pueden (pero no necesariamente) codificar proteínas que tienen actividad 1,6-manosiltransferasa. La homología de secuencias de ácido nucleico sola no es suficiente para demostrar, sin embargo, que se ha identificado y aislado un homólogo que codifica una enzima que tiene la misma actividad. Hasta la fecha, por ejemplo, no existen datos que muestren que una deleción en *OCH1* en *P. pastoris* elimine la actividad 1,6-manosiltransferasa de inicio crucial. (Martinet y cols. *Biotech. Letters* 20 (12) (diciembre de 1998): 1171-1177; Contreras y cols. documento WO 02/00856 A2). Por lo tanto, ningún dato demuestra que el homólogo del gen *OCH1* de *P. pastoris* codifique realmente esa función. Esa demostración se proporciona por primera vez en el presente documento.

Se han identificado homólogos de varias manosiltransferasas de *S. cerevisiae* en *P. pastoris* usando estos enfoques. Los genes homólogos con frecuencia tienen funciones similares a genes implicados en la manosilación de proteínas en *S. cerevisiae* y por lo tanto, su deleción puede usarse para manipular el patrón de glucosilación en *P. pastoris* o, por analogía, en cualquier otra célula huésped, por ejemplo, células de hongo, planta, insecto o animal, con rutas de glucosilación similares.

La creación de bloqueo de genes, una vez que se ha determinado una secuencia génica diana dada, es una técnica bien establecida en este campo y puede llevarse a cabo por un experto habitual en la materia (véase, por ejemplo, R. Rothstein, (1991) Methods in Enzymology, vol. 194, p. 281). La elección de un organismo huésped puede verse

influida por la disponibilidad de buenas técnicas de transformación y de interrupción génica.

Si deben suprimirse por bloqueo varias manosiltransferasas, el procedimiento desarrollado por Alani y Kleckner, (Genetics 116:541-545 (1987)), por ejemplo, permite el uso repetido de un marcador seleccionable, por ejemplo, el marcador *URA3* en levadura, para eliminar secuencialmente toda la actividad manosiltransferasa endógena no deseable. Esta técnica se ha refinado por otros, pero básicamente implica el uso de dos secuencias de ADN repetidas, que flanquean un marcador contraseleccionable. Por ejemplo, *URA3* puede usarse como marcador para asegurar la selección de un transformante que tiene una construcción integrada. Flanqueando el marcador *URA3* con repeticiones directas se pueden seleccionar primero transformantes que tienen la construcción integrada y han interrumpido de este modo el gen diana. Después del aislamiento de los transformantes y su caracterización, se puede contraseleccionar en un segundo ciclo los que son resistentes a ácido 5-fluoroorótico (5-FOA). Las colonias que son capaces de sobrevivir en placas que contienen 5-FOA han perdido el marcador *URA3* de nuevo a través de un acontecimiento de cruce en el que están implicadas las repeticiones mencionadas anteriormente. Este enfoque permite, por lo tanto, el uso repetido del mismo marcador y facilita la interrupción de múltiples genes sin requerir marcadores adicionales. También pueden usarse técnicas similares para la eliminación secuencial de genes adaptados para su uso en otras células huésped eucariotas con otros marcadores seleccionables y contraseleccionables.

La eliminación de manosiltransferasas específicas, tales como 1,6 manosiltransferasa *(OCH1)* o manosilfosfatotransferasas (*MMN6*, o genes que complementan mutantes *lbd*) o reguladores *(MNN4)* en *P. pastoris* permite crear cepas modificadas por ingeniería genética de este organismo que sintetizan principalmente $Man_8GlcNAc_2$ y que pueden usarse para modificar adicionalmente el patrón de glucosilación para asemejarse a estructuras de glucoforma más complejas, por ejemplo, las producidas en células de mamífero, por ejemplo humanas. Una realización preferida de este procedimiento usa secuencias de DhIA que codifican actividades de glucosilación bioquímicas para eliminar funciones bioquímicas similares o idénticas en *P. pastoris* para modificar la estructura de glucosilación de glucoproteínas producidas en la cepa de *P. pastoris* alterada genéticamente.

Pueden usarse procedimientos usados para modificar por ingeniería genética la ruta de glucosilación en levaduras como se ejemplifica en el presente documento en hongos filamentosos para producir un sustrato preferido para modificación posterior. Pueden desarrollarse estrategias para modificar rutas de glucosilación en *A. niger* y otros hongos filamentosos, por ejemplo, usando protocolos análogos a los descritos en el presente documento para modificar por ingeniería genética cepas para producir glucoproteínas de tipo humano en levadura. Se modifican o eliminan actividades génicas no deseadas implicadas en la actividad 1,2 manosiltransferasa, por ejemplo, homólogos de *KTR*/*KRE*. Un hongo filamentoso, tal como *Aspergillus,* es un huésped preferido porque carece de la actividad 1,6 manosiltransferasa y como tal, no se esperaría una actividad génica de hipermanosilación, por ejemplo, *OCH1,* en este huésped. Por el contrario, en el huésped se introducen otras actividades deseadas (por ejemplo, α-1,2-manosidasa, transportador de UDP-GlcNAc, glucosiltransferasa (GnT), galactosiltransferasa (GalT) y sialiltransferasa (ST)) implicadas en la glucosilación usando los procedimientos de dirección de la invención.

Modificación por ingeniería genética o selección de huéspedes que tienen la actividad α-1,6 manosiltransferasa de Inicio Reducida

En una realización preferida, el procedimiento de la invención implica fabricar o usar una célula huésped que tiene reducida o eliminada la actividad de una α-1,6-manosiltransferasa de inicio, es decir, una enzima específica de inicio que inicia la manosilación de cadena exterior en la rama α-1,3 de la estructura central de $Man_3GlcNAc_2$. En *S. cerevisiae,* esta enzima se codifica por el gen *OCH1*. La interrupción del gen *OCH1* en *S. cerevisiae* da como resultado un fenotipo en el que los azúcares ligados a N carecen completamente de la cadena exterior de polimanosa. Los enfoques anteriores para obtener glucosilación de tipo mamífero en cepas fúngicas tienen la inactivación requerida de *OCH1* (véase, por ejemplo, Chiba, 1998). La interrupción de la actividad α-1,6-manosiltransferasa de inicio en una célula huésped de la invención puede ser, sin embargo, opcional (dependiendo de la célula huésped seleccionada), puesto que la enzima Och1p requiere una $Man_8GlcNAc_2$ intacta para un inicio de cadena exterior de manosa eficaz. Por lo tanto, las células huésped seleccionadas o producidas de acuerdo con la presente invención que acumulan oligosacáridos que tienen siete o menos restos de manosa pueden producir N-glucanos hiperglucosilados que probablemente sean malos sustratos para Ochlp (véase, por ejemplo, Nakayama, 1997).

El gen *OCH1* se clonó a partir de *P. pastoris* **(Ejemplo 1)** y *K. lactis* **(Ejemplo 16)**, como se ha descrito. Las secuencias de aminoácidos y ácido nucleico del gen *OCH1* de *K. lactis* se exponen en las SEC ID N.º: 41 y 42. Usando cebadores específicos de gen, se realizó una construcción de cada clon para delecionar el gen *OCH1* del genoma de *P. pastoris* y *K. lactis* **(Ejemplos 1 y 16,** respectivamente). Se obtuvieron de este modo células huésped con la actividad α-1,6-manosiltransferasa de inicio eliminada y modificadas por ingeniería genética para producir N-glucanos que tienen una estructura de carbohidrato $Man_5GlcNAc_2$ (véanse, por ejemplo, las **Figuras 5** y **6**; **Ejemplos 11 y 16**).

Por lo tanto, también se describe en el presente documento una molécula de ácido nucleico aislada que tiene una secuencia de ácido nucleico que comprende o consiste en al menos 45, o al menos 50, o al menos 60 y/o 75 o más restos de nucleótidos del gen *OCH1* de *K. lactis* (SEC ID N.º: 41) y homólogos, variantes y derivados de la misma.

También se describen en el presente documento moléculas de ácido nucleico que hibridan en condiciones rigurosas con las moléculas de ácido nucleico anteriormente descritas. De forma similar, se desvelan polipéptidos aislados (incluyendo muteínas, variantes alélicas, fragmentos, derivados y análogos) codificados por las moléculas de ácido nucleico de la invención. También se describen vectores, incluyendo vectores de expresión, que comprenden las moléculas de ácido nucleico anteriores como se describe adicionalmente en el presente documento. De forma similar, se proporcionan células huésped transformadas con las moléculas de ácido nucleico o vectores según se describen en las reivindicaciones.

Células huésped de la invención

Una célula huésped preferida de la invención es una levadura o un hongo unicelular y multicelular o filamentoso, por ejemplo levadura. Sin embargo, una amplia, diversidad de células huésped de mamíferos, no humanos, se pueden alterar para expresar una glucoproteína similar a la humana (Ejemplos 17 y 18). Asimismo, una diversidad de células huésped de mamífero, no humanas se puede alterar para expresar glucoproteínas más similares a las humanas o alteradas según se describe en el presente documento. Como alguien experto en la técnica apreciará, cualquier célula huésped eucariota (incluyendo una célula humana) se puede usar en conjunción con una biblioteca descrita en el presente documento para expresar una o más proteínas quiméricas que están dirigidas a una localización celular, por ejemplo, un orgánulo, en la célula huésped donde la actividad de la proteína está modificada y preferentemente está potenciada. Una proteína tal es preferentemente –pero no necesita ser necesariamente- una enzima implicada en una glucosilación de proteínas, como se ejemplifican en el presente documento. Se prevé que cualquier secuencia codificante de proteína pueda marcarse y seleccionarse para actividad modificada en una célula huésped eucariota usando los procedimientos descritos en el presente documento.

Son particularmente útiles eucariotas inferiores que sean capaces de producir glucoproteínas que tengan el *N*-glucano $Man_5GlcNAc_2$ unido debido a que (a) al carecer de un alto grado de manosilación (por ejemplo, más de 8 manosas por *N*-glucano, o especialmente 30-40 manosas), muestran inmunogenicidad reducida en seres humanos; y (b) el N-glucano es un sustrato para reacciones de glucosilación adicionales para formar una glucoforma aún más de tipo humano, por ejemplo, mediante la acción de GlcNAc transferasa **I** (**Figura 1B**; β**1**,2 GnTI) para formar $GlcNAcMansGlcNAc_2$. Se obtiene un rendimiento de más de un 30 % en moles, más preferentemente un rendimiento del 50-100 % en moles, teniendo las glucoproteínas con N-glucanos una estructura de $Man_5GlcNAc_2$. En una realización preferida, se muestra que más del 50 % de la estructura de $Man_5GlcNAc_2$ es un sustrato para una actividad GnTI y puede actuar como dicho sustrato *in vivo.*

Los eucariotas inferiores preferidos de la invención incluyen, pero sin limitación: *Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia guercuum, Pichia pijperi, Pichia stiptis, Pichia methanolica, Pichia sp., Saccharomyces cerevisiae, Saccharomyces sp., Hansenula polymorpha, Kluyveromyces sp., Kluyveromyces lactis, Candida albicans, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Trichoderma reseei, Chrysosporfum lucknowense, Fusarium sp., Fusarium gramineum, Fusarium venenatum* y *Neurospora crassa.*

En cada realización anterior, el procedimiento se dirige a preparar una célula huésped en la que los precursores de oligosacáridos se enriquecen en $Man_5GlcNAc_2$. Estas estructuras son deseables debido a que pueden después procesarse por tratamiento *in vitro*, por ejemplo, usando el procedimiento de Maras y Contreras, Patente de Estados Unidos N.º: 5.834.251. En una realización preferida, sin embargo, se procesan precursores enriquecidos en $Man_5GlcNAc_2$ por al menos una reacción de glucosilación adicional *in vivo*, con glucosidasas (por ejemplo, α-manosidasas) y glucosiltransferasas (por ejemplo, GnTI) -para producir N-glucanos de tipo humano. Por ejemplo se procesan preferentemente precursores de oligosacáridos enriquecidos en $Man_5GlcNAc_2$, para dar los que tienen estructuras centrales $GlcNAcMan_xGlcNAc_2$, en las que X es 3, 4 o 5 y es preferentemente 3. Los N-glucanos que tienen una estructura central $GlcNAcMan_xGlcNAC_2$, en la que X es mayor de 3, pueden convertirse en $GlcNAcMan_3GlcNAc_2$, por ejemplo, por tratamiento con una actividad α-1,3 y/o α-1,6 manosidasa, donde sea aplicable. El procesamiento adicional de $GlcNAcMan_3GlcNAc_2$ por tratamiento con glucosiltransferasas (por ejemplo, GnTII) produce estructuras centrales de $GlcNAc2Man_3GlcNAc_2$ que pueden después modificarse, según se desee, por ejemplo, por tratamiento *ex vivo* o por expresión heteróloga en la célula huésped de enzimas de glucosilación adicionales, incluyendo glucosiltransferasas, transportadores de azúcares y manosidasas (véase posteriormente), para convertirse en N-glucanos de tipo humano.

Las glucoproteínas de tipo humano preferidas que pueden producirse de acuerdo con la invención incluyen las que comprenden *N*-glucanos que tienen siete o menos, o tres o menos, restos de manosa; y que comprenden uno o más azúcares seleccionados del grupo que consiste en galactosa, GlcNAc, ácido siálico y fucosa. Preferentemente, las células huésped de la invención son capaces de producir N-glucanos según se describe en las reivindicaciones.

Formación de N-glucanos complejos

La formación de síntesis de N-glucanos complejos es un procedimiento secuencial por el que se retiran restos de azúcares específicos y se unen a la estructura del oligosacárido central. En eucariotas superiores, esto se consigue exponiendo el sustrato secuencialmente a diversas enzimas de procesamiento. Estas enzimas llevan a cabo reacciones específicas dependiendo de su localización particular dentro de la cascada de procesamiento completa.

Esta “línea de ensamblaje” consiste en RE, Golgi temprano, medial y tardío y la red del trans Golgi, todos con su ambiente de procesamiento específico. Para recrear el procesamiento de glucoproteínas humanas en el Golgi y RE de eucariotas inferiores, tienen que expresarse numerosas enzimas (por ejemplo, glucosiltransferasas, glucosidasas, fosfatasas y transportadores) y dirigirse específicamente a estos orgánulos y preferentemente, en una localización tal que actúen de la forma más eficaz en relación con su ambiente así como con otras enzimas de la ruta.

Debido a que un objetivo de los procedimientos descritos en el presente documento es conseguir una cepa de producción de proteínas robusta que sea capaz de comportarse bien en un procedimiento de fermentación industrial, la integración de múltiples genes en el cromosoma de la célula huésped implica una planificación cuidadosa. Como se ha descrito anteriormente, preferentemente se delecionan uno o más genes que codifican enzimas que se sabe que son características de reacciones de glucosilación no humanas. La cepa de células modificada por ingeniería genética se transforma con una serie de genes diferentes que codifican actividades deseadas y estos genes se introducen por transformación de una manera estable, asegurando de este modo que la actividad deseada se mantiene durante todo el procedimiento de fermentación.

Cualquier combinación de las siguientes actividades enzimáticas puede modificarse por ingeniería genética de forma sencilla o múltiple en el huésped usando procedimientos de la invención: sialiltransferasas, manosidasas, fucosiltransferasas, galactosiltransferasas, GlcNAc transferasas, transportadores específicos de RE y Golgi (por ejemplo, transportadores de simporte y antiporte para UDP-galactosa y otros precursores), otras enzimas implicadas en el procesamiento de oligosacáridos y enzimas implicadas en la síntesis de precursores de oligosacáridos activados tales como UDP-galactosa y ácido CMP-N-acetilneuramínico. Preferentemente, se introducen actividades enzimáticas en una o más moléculas de ácido nucleico (véase también posteriormente). Pueden introducirse moléculas de ácido nucleico de forma sencilla o múltiple, por ejemplo, en el contexto de una biblioteca de ácido nucleico tal como una biblioteca combinatoria de la invención. Debe entenderse, sin embargo, que las actividades enzimáticas sencillas o múltiples pueden introducirse en una célula huésped de cualquier manera, incluyendo pero sin limitación procedimientos de suministro de proteínas y/o mediante el uso de una o más moléculas de ácido nucleico sin usar necesariamente una biblioteca de ácido nucleico o biblioteca combinatoria de la invención.

Expresión de glucosiltransferasas para producir N-glucanos complejos:

Con información de la secuencia de ADN, el experto en la materia puede clonar moléculas de ADN que codifican actividades GnT (por ejemplo, **Ejemplos 3 y 4**). Usando técnicas convencionales bien conocidas por los expertos en la materia, pueden insertarse moléculas de ácido nucleico que codifican GnTI, II, III, IV o V (o que codifican fragmentos catalíticamente activos de las mismas) en vectores de expresión apropiados sometidos al control transcripcional de promotores y otras secuencias de control de la expresión capaces de dirigir la transcripción en una célula huésped seleccionada de la invención, por ejemplo, un huésped fúngico tal como *Pichia sp., Kluyveromyces sp.* y *Aspergillus sp.,* como se describe en el presente documento, de modo que una o más de estas enzimas GnT de mamífero pueda expresarse de forma activa en una célula huésped de selección para la producción de una glucoproteína compleja de tipo humano (por ejemplo, **Ejemplos 15, 17 y 18**).

Se han clonado varias glucosiltransferasas individuales y se han expresado en *S. cerevisiae* (GalT, GnTI), *Aspergillus nidulans* (GnTI) y otros hongos, sin demostrar, sin embargo, el resultado deseado de la “humanización” en el patrón de glucosilación de los organismos (Yoshida, 1995; Schwientek, 1995; Kalsner, 1995). Se ha especulado que la estructura de carbohidratos requerida para aceptar azúcares por la acción de tales glucosiltransferasas no estaba presente en cantidades suficientes, lo que con mucha probabilidad, contribuyó a la falta de formación de N-glucanos complejos.

Un procedimiento preferido de la invención proporciona la expresión funcional de una GnT, tal como GnTI, en el aparato de Golgi temprano o medial, además de asegurar un aporte suficiente de UDP-GlcNAc (por ejemplo, mediante expresión de un transportador de UDP-GlcNAc; véase posteriormente).

Procedimientos para proporcionar precursores de azúcar de nucleótido al aparato de Golgi:

Para que una glucosiltransferasa actúe de forma satisfactoria en el Golgi, la enzima requiere una concentración suficiente de un azúcar de nucleótido apropiado, que es el donador de alta energía del resto de azúcar añadido a una glucoproteína naciente. En seres humanos, la serie completa de precursores de azúcares de nucleótido (por ejemplo UDP-N-acetilglucosamina, UDP-N-acetilgalactosamina, ácido CMP-N-acetiluramínico, UDP-galactosa, etc.) se sintetiza generalmente en el citosol y se transporta al Golgi, en el que se unen al oligosacárido central por glucosiltransferasas.

Para replicar este proceso en células huésped no humanas tales como eucariotas inferiores, tienen que expresarse transportadores específicos de azúcares de nucleósido en el Golgi para asegurar niveles adecuados de precursores de azúcares de nucleótido (Sommers, 1981; Summers, 1982; Perez, 1987). Pueden proporcionarse azúcares de nucleótido a los compartimentos apropiados, por ejemplo, expresando en el microorganismo huésped un gen exógeno que codifique un transportador de azúcar de nucleótido. La elección de la enzima transportadora está influenciada por la naturaleza de la glucosiltransferasa exógena que se use. Por ejemplo, una GlcNAc transferasa puede requerir un transportador de UDP-GlcNAc, una fucosiltransferasa puede requerir un transportador de GDP-

fucosa, una galactosiltransferasa puede requerir un transportador de UDP-galactosa y una sialiltransferasa puede requerir un transportador de ácido CMP-siálico.

La proteína transportadora añadida transporta un azúcar de nucleótido desde el citosol al aparato de Golgi, en el que el azúcar de nucleótido puede reaccionar con la glucosiltransferasa, por ejemplo, para elongar un *N*-glucano. La reacción libera un nucleósido difosfato o monofosfato, por ejemplo, UDP, GDP o CMP. Los nucleósidos monofosfatos pueden exportarse directamente desde el Golgi en intercambio con azúcares nucleósido trifosfatos por un mecanismo de antiporte. La acumulación de un nucleósido difosfato, sin embargo, inhibe la actividad adicional de una glucosiltransferasa. Puesto que esta reacción parece ser importante para una glucosilación eficaz, frecuentemente es deseable proporcionar una copia expresada de un gen que codifique una nucleótido difosfatasa. La difosfatasa (específica para UDP o GDP según sea apropiada) hidroliza el difosfonucleósido para producir un nucleósido monofosfato y fosfato inorgánico.

Más adelante se describen enzimas transportadoras adecuadas, que son normalmente de origen de mamífero. Tales enzimas pueden introducirse por ingeniería genética en una célula huésped seleccionada los procedimientos de la invención (véanse también los **Ejemplos 7-10**).

En otro ejemplo, la α 2,3- o α 2,6-sialiltransferasa cubre restos de galactosa con ácido siálico en el trans-Golgi y TGN de seres humanos llevando a una forma madura de la glucoproteína (**Figura 1B**). Para rediseñar esta etapa de procesamiento en una levadura u hongo modificado metabólicamente se requerirá (1) actividad α 2,3- o α 2,6-sialiltransferasa y (2) un aporte suficiente de ácido CMP-N-acetilneuramínico, en el Golgi tardío de levadura (**Ejemplo 6**). Para obtener suficiente actividad α 2,3-sialiltransferasa en el Golgi tardío, por ejemplo, el dominio catalítico de una sialiltransferasa conocida (por ejemplo de seres humanos) tiene que dirigirse al Golgi tardío en hongos (véase anteriormente). De forma similar, los transportadores tienen que modificarse por ingeniería genética para permitir el transporte de ácido CMP-N-acetilneuramínico al Golgi tardío. No existe en la actualidad ningún indicio de que los hongos sinteticen o puedan incluso transportar suficientes cantidades de ácido CMP-N-acetilneuramínico al Golgi. En consecuencia, para asegurar el aporte adecuado de sustrato para las glucosiltransferasas correspondientes, hay que modificar metabólicamente la producción de ácido CN1P-siálico en el hongo.

UDP-N-acetilglucosamina

Se ha clonado el ADNc del transportador de UDP-N-acetilglucosamina humano que se reconoció a través de una búsqueda de homología en la base de datos de marcadores de secuencia expresados (dbEST) (Ishida, 1999 J. Biochem.126 (1): 68-77). El transportador de membrana de Golgi de mamíferos UDP-N-acetilglucosamina se clonó por corrección fenotípica con ADNc de células de riñón canino (MDCK) de un mutante de *Kluyveromyces lactis* recientemente caracterizado deficiente en transporte del Golgi del azúcar de nucleótido anterior (Guillen, 1998). Los resultados demuestran que el gen del transportador de UDP-GloNAc del Golgi de mamíferos tiene toda la información necesaria para que la proteína se exprese y se dirija funcionalmente al aparato de Golgi de levadura y para que dos proteínas con secuencias de aminoácidos muy diferentes puedan transportar el mismo soluto dentro de la misma membrana de Golgi (Guillen, 1998).

En consecuencia, se puede incorporar la expresión de un transportador de UDP-GlcNAc en una célula huésped por medio de una construcción de ácido nucleico que puede contener, por ejemplo: (1) una región por la que se mantiene la construcción transformada en la célula (por ejemplo, origen de replicación o una región que medie en la integración cromosómica), (2) un gen marcador que permite la selección de células que se han transformado, incluyendo marcadores contraseleccionables y reciclables tales como *ura3* o *T-uf13* (Soderholm, 2001) u otros marcadores de selección bien caracterizados (por ejemplo, *his4, bla, Sh ble* etc.), (3) un gen o fragmento del mismo que codifique un transportador de UDP-GlcNAc funcional (por ejemplo, de *K. lactis,* (Abeijon, (1996) Proc. Natl. Acad. Sci. EE.UU. 93:5963-5968), o de *H. sapiens* (Ishida, 1996) y (4) un promotor que active la expresión de la biblioteca de construcciones de fusión de dominio catalítico/localización mencionada anteriormente.

GDP-Fucosa

El transportador de GDP-fucosa de la membrana del Golgi de hígado de rata se ha identificado y purificado por Puglielli, L. y C. B. lerschberg (Puglielli, 1999 J. Biol. Chem. 274 (50): 35596-35600). El gen correspondiente no se ha identificado, sin embargo, puede usarse secuenciación N-terminal para el diseño de sondas oligonucleotídicas específicas para el gen correspondiente. Estos oligonucleótidos pueden usarse como sondas para clonar el gen que codifica el transportador de GDP-fucosa.

UDP-Galactosa

Dos genes heterólogos, *gmal2*(+) que codifica la alfa 1,2-galactosiltransferasa (alfa 1,2 GalT) de *Schizosaccharomyces pombe* y *(hUGT2)* que codifica el transportador de UDP-galactosa humana (UDP-Gal), se han expresado funcionalmente en *S. cerevisiae* para examinar las condiciones intracelulares requeridas para la galactosilación. La correlación entre la galactosilación de proteínas y la actividad de transporte de UDP-galactosa indicó que un aporte exógeno del transportador de UDP-Gal, en lugar de alfa 1,2 GalT, desempeñaba un papel clave

para la galactosilación eficaz en *S. cerevisiae* (Kainuma, 1999 Glycobiology 9 (2): 133-141). De forma similar, se clonó un transportador de UDP-galactosa de *S. pombe* (Aoki, 1999 J. Biochem. 126 (5): 940-950; Segawa, 1999 Febs Letters 451 (3): 295-298).

Ácido CMP-N-acetilneuramínico (ácido CMP-Siálico).

Se ha clonado el transportador de ácido CMP-siálico humano (hCST) y se ha expresado en células Lec 8 CHO (Aoki, 1999; Eckhardt, 1997). La expresión funcional del transportador de ácido CMP-siálico murino se consiguió en *Saccharomyces cerevisiae* (Berninsone, 1997). Se ha descubierto ácido siálico en algunos hongos, aunque no está claro si el sistema del huésped seleccionado será capaz de proporcionar niveles suficientes de ácido CMP-Siálico. El ácido siálico puede proporcionarse en el medio o, como alternativa, también pueden integrarse en el genoma del huésped rutas fúngicas implicadas en la síntesis de ácido siálico.

Expresión de difosfatasas:

Cuando se transfieren azúcares a una glucoproteína; se libera un nucleósido difosfato o monofosfato de los precursores de azúcar de nucleótido. Aunque pueden exportarse monofosfatos directamente en intercambio con azúcares nucleósido trifosfato por un mecanismo de antiporte, los difosfonucleósidos (por ejemplo GDP) tienen que escindirse mediante fosfatasas, (por ejemplo, GDPasa) para producir nucleósido monofosfatos y fosfato inorgánico antes de exportarse. Esta reacción parece ser importante para una glucosilación eficaz, puesto que se ha descubierto que la GDPasa de *S. cerevisiae* es necesaria para la manosilación. Sin embargo, la enzima solamente tiene un 10 % de la actividad hacia UDP (Berninsone, 1994). Los eucariotas inferiores con frecuencia no tienen actividad difosfatasa específica de UDP en el Golgi, puesto que no usan precursores de UDP-azúcar para la síntesis de glucoproteínas en el Golgi. Se descubrió que *Schizosaccharomyces pombe,* una levadura que añade restos de galactosa a polisacáridos de la pared celular (a partir de UDP-galactosa) tenía actividad UDPasa específica, lo que sugiere adicionalmente la necesidad de dicha enzima (Berninsone, 1994). Se sabe que el UDP es un inhibidor potente de glucosiltransferasas y la retirada de este producto secundario de glucosilación es importante para evitar la inhibición de glucosiltransferasa en el lumen del Golgi (Khatara y cols. 1974).

Procedimientos para alterar N-glucanos en un huésped expresando una actividad enzimática diana a partir de una molécula de ácido nucleico

La presente invención proporciona adicionalmente un procedimiento para producir una glucoproteína de tipo humano en una célula huésped no humana que comprende la etapa de introducir en la célula una o más moléculas de ácido nucleico que codifican una enzima o enzimas para la producción de la estructura de carbohidrato $Man_5GlcNAc_2$. En una realización preferida, una molécula de ácido nucleico que codifica una o más actividades manosidasa implicadas en la producción de $Man_5GlcNAc_2$ a partir de $Man_8GlcNAc_2$ o $Man_9GlcNAc_2$ se introduce en el huésped. La invención se refiere adicionalmente a procedimientos para preparar glucoproteínas alteradas en una célula huésped que comprenden la etapa de introducir en la célula huésped una molécula de ácido nucleico que codifique una o más enzimas o actividades de glucosilación. Se seleccionan actividades enzimáticas preferidas del grupo que consiste en UDP-GlcNAc transferasa, UDP-galactosiltransferasa, GDP-fucosiltransferasa, CMP-sialiltransferasa, transportador de UDP-GlcNAc, transportador de UDP-galactosa, transportador de GDP-fucosa, transportador de CMP-ácido siálico y nucleótido difosfatasas. En una realización particularmente preferida, el huésped se selecciona o modifica por ingeniería genética para expresar dos o más actividades enzimáticas en las que el producto de una actividad aumenta niveles de sustrato de otra actividad, por ejemplo, una glucosiltransferasa y un transportador de azúcar correspondiente, por ejemplo, actividades transportadoras de GnTI y UDP-GlcNAc. En otra realización preferida, el huésped se selecciona o se modifica por ingeniería genética para expresar una actividad para retirar productos que pueden inhibir reacciones de glucosilación posteriores, por ejemplo una actividad difosfatasa específica de UDP o GDP.

Los procedimientos preferidos de la invención implican expresar una o más actividades enzimáticas de una molécula de ácido nucleico en una célula huésped y comprenden la etapa de dirigir al menos una actividad enzimática a una localización subcelular deseada (por ejemplo, un orgánulo) formando una proteína de fusión que comprende un dominio catalítico de la enzima y un péptido señal de dirección celular, por ejemplo, un péptido señal heterólogo que no se liga normalmente a o se asocia con el dominio catalítico. La proteína de fusión se codifica por al menos una construcción genética ("construcción de fusión") que comprende un fragmento de ácido nucleico que codifica un péptido señal de dirección celular ligado en la misma fase de lectura de traducción ("en fase") con un fragmento de ácido nucleico que codifica una enzima (por ejemplo, enzima de glucosilación), o fragmento catalíticamente activo de la misma.

El componente de péptido señal de dirección de la construcción o proteína de fusión deriva preferentemente de un miembro del grupo que consiste en: proteínas unidas a la membrana del RE o Golgi, señales de recuperación, proteína de membrana de Tipo II, proteínas de membrana de Tipo I, transportadores de azúcares de nucleótido que abarcan la membrana, manosidasas, sialiltransferasas, glucosidasas, manosiltransferasas y fosfomanosiltransferasas.

El componente de dominio catalítico de la construcción o proteína de fusión deriva preferentemente de una

glucosidasa, manosidasas o una actividad glucosiltransferasa derivada de un miembro del grupo que consiste en GnTI, GnTII, GnTIII, GnTIV, GnTV, GNTV, GalT, Fucosiltransferasa y Sialiltransferasa. El dominio catalítico preferentemente tiene un pH óptimo dentro de 1,4 unidades de pH del pH óptimo medio de otras enzimas representativas en el orgánulo en el que se localiza la enzima, o tiene actividad óptima a un pH entre 5,1 y 8,0. En una realización preferida, el dominio catalítico codifica una manosidasa seleccionada del grupo que consiste en manosidasa IA de *C. elegans*, manosidasa IB de *C. elegans*, manosidasa IA de *D. melanogaster*, manosidasa IB de *H. sapiens*, manosidasa I de *P. citrinum*, manosidasa IA de ratón, manosidasa IB de ratón, manosidasa IA de *A. nidulans*, manosidasa IB de *A. nidulans*, manosidasa IC de *A. nidulans*, manosidasa II de ratón, manosidasa II de C. *elegans*, manosidasa II de *H. sapiens* y manosidasa III.

Selección de una enzima de glucosilación: pH óptimo y localización subcelular

En una realización de la invención, se prepara una glucoproteína de tipo humano eficazmente en una célula huésped eucariota no humana introduciendo en un compartimento subcelular de la célula una enzima de glucosilación seleccionada para tener un pH óptimo similar a los pH óptimos de otras enzimas en el compartimento subcelular diana. Por ejemplo, la mayoría de las enzimas que son activas en el RE y aparato de Golgi de *S. cerevisiae* tienen pH óptimos que están entre aproximadamente 6,5 y 7,5 (véase la **Tabla 3**). Debido a que la glucosilación de proteínas es un procedimiento altamente evolucionado y eficaz, el pH interno del RE y el Golgi probablemente también esté en el intervalo de aproximadamente 6-8. Todos los enfoques anteriores para reducir la manosilación por la acción de manosidasas recombinantes en huéspedes fúngicos, sin embargo, han introducido enzimas que tienen un pH óptimo de aproximadamente pH 5,0 (Martinet y cols., 1998 y Chiba y cols., 1998). A pH 7,0, la actividad determinada *in vitro* de esas manosidasas se reduce a menos del 10 %, que probablemente es una actividad insuficiente en su punto de uso, concretamente, el RE y Golgi temprano, para la producción eficaz *in vivo* de $Man_5GlcNAc_2$ en N-glucanos.

En consecuencia, una realización preferida de la presente invención dirige una enzima de glucosilación seleccionada (o dominio catalítico de la misma), por ejemplo, una α-manosidasa, a una localización subcelular en la célula huésped (por ejemplo, un orgánulo) en la que el pH óptimo de la enzima o dominio está a una distancia de 1,4 unidades de pH del pH óptimo medio de otras enzimas marcadoras representativas localizadas en el mismo orgánulo u orgánulos. El pH óptimo de la enzima a dirigir a un orgánulo específico debería coincidir con el pH óptimo de otras enzimas halladas en el mismo orgánulo para maximizar la actividad por enzima unitaria obtenida. La **Tabla 3** resume la actividad de manosidasas de diversas fuentes y sus pH óptimos respectivos. La **Tabla 4** resume sus localizaciones subcelulares típicas.

(Tabla 3 a continuación)

Tabla 3. Manosidasas y su pH óptimo

Fuente	Enzima	pH óptimo	Referencia
Aspergillus saitoi	α-1,2-manosidasa	5,0	Ichishima y cols., 1999 Biochem. J. 339(Pt 3): 589-597
Trichoderma reesei	α-1,2-manosidasa	5,0	Madras y cols., 2000 J. Biotechnol. 77(2-3): 255-263
Pereicillium citrinum	α-D-1,2-manosidasa	5,0	Yoshida y cols., 1993 Biochem. J. 290(Pt 2):349-354
C. elegans	α-1,2-manosidasa	5,5	véase la **Figura 11**
Aspergillus nidulans	α-1,2-manosidasa	6,0	Eades y Hintz, 2000
Homo sapiens IA(Golgi)	α-1,2-manosidasa	6,0	
Homo sapiens IB (Golgi)	α-1,2-manosidasa	6,0	
Células de insecto Lepidóptero	α-1,2-Man_6-manosidasa Tipo I	6,0	Ren y cols., 1995 Biochem. 34(8): 2489-2495
Homo sapiens	α-D-manosidasa	6,0	Chandrasekaran y cols., 1984 Cancer Res. 44(9):4059-68
Xanthomonas manihotis	α-1,2,3-manosidasa	6,0	Patente de Estados Unidos N.º: 6.300.113
Ratón IB (Golgi)	α-1,2-manosidasa	6,5	Schneikert y Herscovics, 1994 Glycobiology. 4(4):445-50
Bacillus sp. (secretado)	α-D-1,2-manosidasa	7,0	Mazuyama y cols., 1994 Carbohydrate Res. 251:89-98

En una realización preferida, una enzima particular o dominio catalítico se dirige a una localización subcelular en la célula huésped por medio de una construcción de fusión quimérica que codifica una proteína que comprende un péptido señal de dirección celular normalmente no asociado con el dominio enzimático. Preferentemente, una enzima o dominio se dirige al RE, el Golgi temprano, medial o tardío o el aparato trans-Golgi de la célula huésped.

En una realización más preferida, la enzima de glucosilación diana es una manosidasa, glucosiltransferasa o una glucosidasa. En una realización especialmente preferida, la actividad manosidasa se dirige al RE o cis-Golgi, en los que se producen las reacciones tempranas de glucosilación. Aunque este procedimiento es útil para producir una glucoproteína de tipo humano en una célula huésped no humana, se apreciará que el procedimiento también es útil de forma más general para modificar perfiles de carbohidratos de una glucoproteína en cualquier célula huésped eucariota, incluyendo células huésped humanas.

Se conocen bien secuencias de dirección que median la retención de proteínas en ciertos orgánulos de la ruta secretora de la célula huésped y se describen en la bibliografía científica y en bases de datos públicas, como se analiza en más detalle posteriormente con respecto a bibliotecas para la selección de secuencias de dirección y enzimas diana. Tales secuencias de dirección subcelulares pueden usarse solas o en combinación para dirigir una enzima de glucosilación seleccionada (o dominio catalítico de la misma) a una localización subcelular particular en una célula huésped, es decir, especialmente a una en la que la enzima tendrá una actividad mejorada u óptima basándose en el pH óptimo o la presencia de otros factores estimuladores.

Cuando se intenta reducir estructuras de alto contenido de manosa para producir $Man_5GlcNAc_2$ en el RE o el aparato de Golgi de una célula huésped tal como *S. cerevisiae,* por ejemplo, se puede seleccionar cualquier enzima o combinación de enzimas que (1) tenga un pH óptimo suficientemente cercano (es decir entre pH 5,2 y pH 7,8) y (2) se sabe que genera, sola o en conjunto, la estructura de $Man_5GlcNAc_2$ isomérica específica requerida para aceptar la adición posterior de GlcNAc por GnTI. Cualquier enzima o combinación de enzimas que se muestre que genere una estructura que pueda convertirse en $GlcNAcMan_5GlcNAc_2$ mediante GnTI *in vitro* constituiría una elección apropiada. Este conocimiento puede obtenerse a partir de la bibliografía científica o de forma experimental.

Por ejemplo, se puede determinar si una manosidasa potencial puede convertir $Man_8GlcNAc_2$-2AB (2-

aminobenzamida) en $Man_5GlcNAc_2$-AB y después verificar que la estructura de $Man_5GlcNAc_2$-2AB obtenida puede actuar como sustrato para GnTI y UDP-GlcNAc para proporcionar $GlcNAcMan_5GlcNAc_2$ *in vitro*. La manosidasa IA de una fuente humana o murina, por ejemplo, sería una elección apropiada (véase, por ejemplo, el **Ejemplo 11**). Los ejemplos descritos en el presente documento usan oligomanosa ligada a N marcada con 2-aminobenzamida seguido de análisis de HPLC para realizar esta determinación.

Tabla 4. Localización celular y pH óptimos de diversas enzimas relacionadas con glucosilación de S. cerevisiae.

Gen	Actividad	Localización	pH óptimo	Referencia o referencias
KTR1	α-1,2 manosiltransferasa	Golgi	7,0	Romero y cols., 1997 Biochem. J. 321 (Pt2): 289-295
MNS1	α- 1,2- manosidasa	RE	6,5	
CWH41	glucosidasa I	RE	6,8	
-	manosiltransferasa	Golgi	7-8	Lehele y Tanner, 1974 Biochim. Biophys. Acta 350 (1): 225-235
KRE	α-1,2 manosiltransferasa	Golgi	6,5-9,0	Romero y cols., 1997

En consecuencia, una enzima de glucosilación tal como una enzima α-1,2-manosidasa usada de acuerdo con la invención tiene una actividad óptima a un pH de entre 5,1 y 8,0. En una realización preferida, la enzima tiene una actividad óptima a un pH de entre 5,5 y 7,5. La enzima manosidasa de *C. elegans*, por ejemplo, funciona bien en los procedimientos de la invención y tiene un pH óptimo aparente de aproximadamente 5,5. Las manosidasas preferidas incluyen las enumeradas en la **Tabla 3** que tienen pH óptimos apropiados, por ejemplo, *Aspergillus nidulans*, *Homo sapiens* IA (Golgi), *Homo sapiens* IB (Golgi), células de insecto *Lepidóptero* (IPLB-SF21AE), *Homo sapiens,* IB de ratón (Golgi), *Xanthomonas manihotis*, *Drosophila melanogaster* y *C. elegans.*

El experimento que ilustra el pH óptimo para una enzima α-1,2-manosidasa se describe en el **Ejemplo 14.** Una proteína de fusión quimérica BB27-2 (*MNN10* (s) de *Saccharomyces*/manosidasa IB Δ31 de *C. elegans*), que se filtra al medio se sometió a diversos intervalos de pH para determinar la actividad óptima de la enzima. Los resultados del experimento muestran que la α-1,2-manosidasa tiene un pH óptimo de aproximadamente 5,5 para su función (**Figura 11**).

En una realización preferida, se expresa un gen de manosidasa clonado individual en el organismo huésped. Sin embargo, en algunos casos puede ser deseable expresar varios genes de manosidasa diferentes, o varias copias de un gen particular, para conseguir la producción adecuada de $Man_5GlcNAc_2$. En casos en los que se usen múltiples genes, todas las manosidasas codificadas preferentemente tienen pH óptimos dentro del intervalo preferido de aproximadamente 5,1 a aproximadamente 8,0, o especialmente entre aproximadamente 5,5 y aproximadamente 7,5. Las actividades manosidasa preferidas incluyen α-1,2-manosidasas derivadas de ratón, ser humano, *Lepidoptera, Aspergillus nidulans,* o *Bacillus* sp., *C. elegans, D. melanogaster, P. citrinum, X. laevis* o *A. nidulans*.

Alteración *in vivo* de la glucosilación de células huésped usando una biblioteca de ADN combinatoria

Ciertos procedimientos de la invención se llevan a cabo preferentemente (pero no necesariamente) usando una o más bibliotecas de ácidos nucleicos. Una característica ejemplar de una biblioteca de ácido nucleico combinatoria de la invención es que comprende secuencias que codifican péptidos señal de dirección celular y secuencias que codifican proteínas que se fijarán como diana (por ejemplo, enzimas o dominios catalíticos de las mismas, incluyendo pero sin limitación las que median en la glucosilación).

Como se describe en el presente documento, una biblioteca de ácidos nucleicos combinatoria comprende: (a) al menos dos secuencias de ácido nucleico que codifican diferentes péptidos señal de dirección celular; y (b) al menos una secuencia de ácido nucleico que codifica un polipéptido que se fijará como diana. También se describe en el presente documento que una biblioteca de ácidos nucleicos combinatoria comprende: (a) al menos una secuencia de ácido nucleico que codifica un péptido señal de dirección celular; y (b) al menos dos secuencias de ácido nucleico que codifican un polipéptido para dirigir a una célula huésped. Como se describe adicionalmente posteriormente, una secuencia de ácido nucleico derivada de (a) y una secuencia de ácido nucleico derivada de (b) se ligan para producir una o más construcciones de fusión que codifican un péptido señal de dirección celular unido funcionalmente a un dominio polipeptídico de interés. Un ejemplo de un enlace funcional es cuando el péptido señal de dirección celular está ligado con el dominio polipeptídico de interés en la misma fase de lectura de traducción ("en fase").

También se describe en el presente documento una biblioteca de ADN combinatoria que expresa una o más proteínas de fusión que comprenden péptidos señal de dirección celular ligados en fase con dominios enzimáticos catalíticos. La proteína de fusión codificada posiblemente comprende un dominio catalítico de una enzima implicada en la modificación de tipo mamífero o humano de N-glucanos. Como se describe en el presente documento, el dominio catalítico deriva de una enzima seleccionada del grupo que consiste en manosidasas, glucosiltransferasas y otras glucosidasas que se ligan en fase con uno o más péptidos señal de dirección. El dominio enzimático puede ser exógeno y/o endógeno para la célula huésped. Un péptido señal posiblemente es uno normalmente asociado con una proteína que experimenta transporte del RE al Golgi.

La biblioteca de ADN combinatoria descrita en el presente documento puede usarse para producir y localizar enzimas *in vivo* implicadas en la modificación de N-glucanos de tipo mamífero o humano. Las construcciones de fusión de la biblioteca de ADN combinatoria se modifican por ingeniería genética de modo que las enzimas codificadas se localicen en el RE, Golgi o la red del trans-Golgi de la célula huésped en la que están implicadas en la producción de *N*-glucanos particulares en una glucoproteína de interés. La localización de N-glucanos que modifican enzimas de la presente invención se consigue a través de un mecanismo de anclaje o a través de interacción proteína-proteína en la que el péptido de localización construido a partir de la biblioteca de ADN combinatoria se localiza en un orgánulo deseado de la ruta secretora tal como el RE, Golgi o la red del trans-Golgi.

Un ejemplo de un *N*-glucano útil, que se produce eficazmente y en suficientes cantidades para la modificación adicional por reacciones de glucosilación de tipo humano (complejas) es $Man_5GlcNAc_2$. Se necesita una cantidad suficiente de $Man_5GlcNAc_2$ en una glucoproteína de interés para el procesamiento de tipo humano adicional *in vivo* (por ejemplo, más de un 30 % en moles). El intermedio $Man_5GlcNAc_2$ puede usarse como un sustrato para la modificación de N-glucano adicional para producir $GlcNAcMan_5GlcNAc_2$ (Figura 1B; véase anteriormente). En consecuencia, la biblioteca de ADN combinatoria descrita en el presente documento puede usarse para producir enzimas que posteriormente producen $GlcNAcMan_5GlcNAc_2$, u otros N-glucanos complejos deseados, en una cantidad útil.

Un aspecto adicional de las construcciones de fusión producidas usando la biblioteca de ADN combinatoria descrita en el presente documento es que permiten una actividad de reducción de *N*-glucanos intracelulares suficiente y con frecuencia casi completa en la célula huésped modificada por ingeniería genética. Las construcciones de fusión preferidas producidas por la biblioteca de ADN combinatoria descritas en el presente documento codifican una enzima de glucosilación, por ejemplo, una manosidasa, que se localiza eficazmente en un compartimento de la célula huésped intracelular y de este modo muestra muy poca y preferentemente ninguna actividad extracelular. Se muestra que las construcciones de fusión preferidas de la presente invención que codifican una enzima manosidasa se localizan donde se modifican los N-glucanos, es decir, el RE y el Golgi. Las enzimas de fusión de la presente invención se dirigen a dichos orgánulos particulares en la ruta secretora en los que se sitúan y actúan sobre *N*-glucanos tales como $Man_8GlcNAc_2$ para producir $Man_5GlcNAc_2$ en una glucoproteína de interés.

Las enzimas producidas por la biblioteca de ADN combinatoria descrita en el presente documento pueden modificar *N*-glucanos en una glucoproteína de interés como se muestra para proteínas K3 o IFN-β expresadas en *P. pastoris,* como se muestra en la **Figura 5** y la **Figura 6,** respectivamente (véanse también los **Ejemplos 2 y 11**). Se apreciará, sin embargo, que pueden glucosilarse de esta manera otros tipos de glucoproteínas, sin limitación incluyendo eritropoyetina, citocinas tales como interferón-α, interferón-β, interferón-γ, interferón-ϖ y CSF-granulocitos, factores de coagulación tales como el factor VIII, el factor IX y la proteína humana C, cadena α del receptor de IgE soluble, IgG, fragmentos de IgG, IgM, interleucinas, urocinasa, quimasa e inhibidor de tripsina de urea, proteína de unión a IgG, factor de crecimiento epidérmico, factor de liberación de hormonas del crecimiento, proteína de fusión a anexina V, angioestatina, factor de crecimiento endotelial vascular 2, factor inhibidor del progenitor mieloide 1, osteoprotegerina, α-1 antitripsina, DNasa II, α-fetoproteínas, AAT, rhTBP-1 (onercept, también conocido como proteína de unión a TNF 1), TACI-Ig (activador transmembrana y modulador de calcio y elemento de interacción con el ligando de ciclofilina), FSH (hormona foliculoestimulante), GM-CSF, GLP-1 con y sin FC (proteína de tipo glucagón 1), agonista del receptor de IL-1, sTNFr (enbrel, también conocido como fusión de Fc del receptor de TNF soluble) ATIII, rhTrombina, glucocerebrosidasa y CTLA4-Ig (antígeno asociado con linfocitos T citotóxicos 4-Ig).

Construcción de una biblioteca de ADN combinatoria de construcciones de fusión:

Una biblioteca de ADN combinatoria de construcciones de fusión presenta uno o más péptidos señal de dirección celular ("péptidos de dirección") generalmente derivados de dominios N-terminales de proteínas nativas (por ejemplo, realizando deleciones C-terminales). Algunos péptidos de dirección, sin embargo, derivan del extremo C-terminal de proteínas nativas (por ejemplo, *SEC12).* Se usan preferentemente proteínas unidas a la membrana del RE o el Golgi como fuente de secuencias peptídicas de dirección. Estas proteínas tienen secuencias que codifican una cola citosólica (ct), un dominio transmembrana (tmd) y una región de tallo (sr) que varía de longitud. Estas regiones se pueden reconocer por alineamientos de secuencias proteicas y comparaciones con homólogos conocidos y/u otras proteínas localizadas (por ejemplo, comparando gráficos de hidrofobia).

Los péptidos de dirección se indican en el presente documento como corto (s), medio (m) y largo (l) en relación con las partes de una membrana de tipo II. La secuencia del péptido de dirección indicada como corta (s) se

corresponde con el dominio transmembrana (tmd) de la proteína unida a la membrana. La secuencia de péptido de dirección indicada como larga (l) corresponde a la longitud del dominio transmembrana (tmd) y la región de tallo (sr). La secuencia peptídica de dirección indicada como media (m) corresponde al dominio transmembrana (tmd) y aproximadamente a la mitad de la longitud de la región del tallo (sr). Las regiones de dominio catalítico se indican en el presente documento por el número de deleción de nucleótido con respecto a su enzima de glucosilación de tipo silvestre.

Sub-bibliotecas

En algunos casos, una biblioteca de ácidos nucleicos combinatoria como se describe en el presente documento puede ensamblarse directamente a partir de genes existentes o de tipo silvestre. Como se describe en el presente documento, la biblioteca de ADN se ensambla a partir de la fusión de dos o más sub-bibliotecas. Por la ligación en fase de las sub-bibliotecas, es posible crear un gran número de construcciones genéticas nuevas que codifican dominios de proteína diana útiles tales como los que tienen actividades de glucosilación.

Sub-bibliotecas de dominios catalíticos que codifican actividades de glucosilación

Una sub-biblioteca útil incluye secuencias de ADN que codifican enzimas tales como glucosidasas (por ejemplo, manosidasas), glucosiltransferasas (por ejemplo, fucosiltransferasas, galactosiltransferasas, glucosiltransferasas), GlcNAc transferasas y sialiltransferasas. Pueden seleccionarse dominios catalíticos a partir del huésped para modificar por ingeniería genética, así como a partir de otros organismos relacionados o no relacionados. Las enzimas de mamífero, de planta, de insecto, de reptiles, de algas o fúngicas son todas útiles y deberían seleccionarse para representar un amplio espectro de propiedades bioquímicas con respecto a temperatura y pH óptimos. En una realización preferida, los genes se truncan para proporcionar fragmentos de los que algunos codifican los dominios catalíticos de las enzimas. Retirando secuencias de dirección endógenas, las enzimas pueden después redireccionarse y expresarse en otros lugares celulares.

La elección de tales dominios catalíticos puede guiarse por el conocimiento del ambiente particular en el que el dominio catalítico va a estar activo posteriormente. Por ejemplo, si una enzima de glucosilación particular va a estar activa en el Golgi tardío y todas las enzimas conocidas del organismo huésped en el Golgi tardío tienen un cierto pH óptimo, o se sabe que el Golgi tardío tiene un pH particular, entonces se selecciona un dominio catalítico que muestra actividad adecuada, preferentemente máxima a ese pH, como se ha analizado anteriormente.

Dirección de sub-bibliotecas de secuencias peptídicas

Otra sub-biblioteca útil incluye secuencias de ácidos nucleicos que codifican péptidos señal de dirección que dan como resultado la localización de una proteína en una localización particular dentro del RE, Golgi o red de trans-Golgi. Estos péptidos de dirección pueden seleccionarse a partir del organismo huésped para modificarse por ingeniería genética así como de otros organismos relacionados o no relacionados. Generalmente tales secuencias se dividen en tres categorías: (1) secuencias N-terminales que codifican una cola citosólica (ct), un dominio transmembrana (tmd) y parte de o toda una región tallo (sr), que juntos o individualmente anclan proteínas a la membrana interna (del lumen) del Golgi; (2) señales de recuperación que generalmente se encuentran en el extremo C-terminal tales como el tetrapéptido HDEL (**SEC ID N.º: 5**) o KDEL (**SEC ID N.º: 6**); y (3) regiones que abarcan la membrana de diversas proteínas, por ejemplo, transportadores de azúcares de nucleótido, que se sabe que se localizan en el Golgi.

En el primer caso, cuando el péptido de dirección consiste en diversos elementos (ct, tmd y sr), la biblioteca se diseña de modo que se representen la ct, el tmd y diversas partes de la región troncal. En consecuencia, la sub-biblioteca de secuencias peptídicas de dirección debería incluir secuencias ct, tmd y/o sr de las proteínas unidas a la membrana del RE o Golgi. En algunos casos puede ser deseable proporcionar la sub-biblioteca con diversas longitudes de secuencia sr. Esto puede conseguirse mediante PCR usando cebadores que se unen al extremo 5' del ADN que codifica la región citosólica y empleando una serie de cebadores opuestos que se unen a diversas partes de la región troncal.

Otras fuentes útiles más de secuencias peptídicas de dirección incluyen péptidos de señal de recuperación, por ejemplo los tetrapéptidos HDEL (**SEC ID N.º: 5**) o KDEL (**SEC ID N.º: 6**), que se encuentran normalmente en el extremo C-terminal de proteínas que se transportan de forma retrógrada al RE o Golgi. Otras fuentes más de secuencias peptídicas de dirección incluyen (a) proteínas de membrana de tipo II, (b) las enzimas enumeradas en la **Tabla 3**, (c) transportadores de azúcares de nucleótido que abarcan la membrana que se localizan en el Golgi y (d) secuencias mencionadas en la **Tabla 5**. (La señal de HDEL en la columna 1, celda 8 se muestra en la **SEC ID N.º: 5**).

Tabla 5. Fuentes de secuencias de dirección compartimental útiles

Gen o Secuencia	Organismo	Función	Localización del Producto Génico
MNSI	*A. nidulans*	α-1,2-manosidasa	RE
MNSI	*A. niger*	α-1,2-manosidasa	RE
MNSI	*S. cerevisiae*	α-1,2-manosidasa	RE
GLSI	*S. cerevisiae*	glucosidasa	RE
GLSI	*A. niger*	glucosidasa	RE
GLSI	*A. nidulans*	glucosidasa	RE
HDEL en el extremo C-terminal	*Universal en hongos*	señal de recuperación	RE
SEC12	*S. cerevisiae*	proteína de vesícula de COPII	RE/Golgi
SEC12	*A. niger*	proteína de vesícula de COPII	RE/Golgi
OCH1	*S. cerevisiae*	1,6-manosiltransferasa	Golgi (cis)
OCH1	*P. pastoris*	1,6-manosiltransferasa	Golgi (cis)
MNN9	*S. cerevisiae*	complejo de 1,6-manosiltransferasa	Golgi
MNN9	*A. niger*	sin determinar	Golgi
VAN1	*S. cerevisiae*	sin determinar	Golgi
VAN1	*A. niger*	sin determinar	Golgi
ANP1	*S. cerevisiae*	sin determinar	Golgi
HOCI	*S. cerevisiae*	sin determinar	Golgi
MNN10	*S. cerevisiae*	sin determinar	Golgi
MNN10	*A. niger*	sin determinar	Golgi
MNN11	*S. cerevisiae*	sin determinar	Golgi (cis)
MNN11	*A. niger*	sin determinar	Golgi (cis)
MNT1	*S. cerevisiae*	1,2-manosiltransferasa	Golgi (cis, medial)
KTR1	*P. pastoris*	sin determinar	Golgi (medial)
KRE2	*P. pastoris*	sin determinar	Golgi (medial)
KTR3	*P. pastoris*	sin determinar	Golgi (medial)
MNN2	*S. cerevisiae*	1,2-manoilltransferasa	Golgi (medial)
KTR1	*S. cerevisiae*	sin determinar	Golgi (medial)
KTR2	*S. cerevisiae*	sin determinar	Golgi (medial)
MNN1	*S. cerevisiae*	1,3-manosiltransferasa	Golgi (trans)
MNN6	*S. cerevisiae*	Fosfomanosiltransferasa	Golgi (trans)
2,6 ST	*H. sapiens*	2,6-sialiltransferasa	red trans-Golgi
UDP-Gal T	*S. pombe*	transportador de UDP-Gal	Golgi

En cualquier caso, se prefiere en gran medida que se seleccionen secuencias peptídicas de dirección que sean apropiadas para la actividad o las actividades enzimáticas particulares para actuar de forma óptima dentro de la secuencia de reacciones de glucosilación deseadas. Por ejemplo, en el desarrollo de un microorganismo modificado capaz de sialilación terminal de *N*-glucanos nacientes, un proceso que sucede en el Golgi tardío en seres humanos, es deseable usar una sub-biblioteca de secuencias peptídicas de dirección derivadas de proteínas del Golgi tardías. De forma similar, la reducción de $Man_8GlcNAc_2$ mediante una α-1,2-manosidasa para proporcionar $Man_5GlcNAc_2$ es una etapa temprana en la formación de N-glucanos complejos en seres humanos (**Figura 1B**). Por lo tanto, es deseable hacer que esta reacción se produzca en el RE o Golgi temprano de un microorganismo huésped modificado por ingeniería genética. Se usa una sub-biblioteca que codifica señales de retención en RE y Golgi temprano.

Después se construye una serie de construcciones de proteínas de fusión (es decir, una biblioteca de ADN combinatoria) uniendo funcionalmente una o una serie de secuencias peptídicas de dirección con una o una serie de secuencias que codifican dominios catalíticos. En una realización preferida, esto se consigue mediante ligación en fase de una sub-biblioteca que comprende ADN que codifica secuencias peptídicas de dirección (anteriormente) con una sub-biblioteca que comprende ADN que codifica enzimas de glucosilación o fragmentos catalíticamente activos de las mismas (véase posteriormente).

La biblioteca resultante comprende genes sintéticos que codifican proteínas de fusión que contienen secuencias peptídicas de dirección. En algunos casos es deseable proporcionar una secuencia peptídica de dirección en el extremo N-terminal de una proteína de fusión, o en otros casos en el extremo C-terminal. En algunos casos, las secuencias peptídicas de dirección pueden insertarse dentro de la fase abierta de lectura de una enzima, siempre que no se interrumpa la estructura proteica de dominios plegados individuales. Se construye cada tipo de proteína de fusión (de una manera dirigida por etapas o semialeatoria) y pueden seleccionarse construcciones óptimas tras la transformación de células huésped y caracterización de patrones de glucosilación en células transformadas usando procedimientos de la invención.

Generación de diversidad de secuencia adicional

El procedimiento de la presente realización es más eficaz cuando un ácido nucleico, por ejemplo, una biblioteca de ADN transformada en el huésped contiene una gran diversidad de secuencias, aumentando de este modo la probabilidad de que al menos un transformante muestre el fenotipo deseado. Las mutaciones de un solo aminoácido, por ejemplo, pueden alterar drásticamente la actividad de enzimas de procesamiento de glucoproteínas (Romero y cols., 2000). En consecuencia, antes de la transformación, una biblioteca de ADN o una sub-biblioteca constituyente puede someterse a una o más técnicas para generar diversidad de secuencia adicional. Por ejemplo, pueden realizarse uno o más ciclos de combinación de genes, PCR propensa a errores, mutagénesis *in vitro* u otros procedimientos para generar diversidad de secuencia, para obtener una mayor diversidad de secuencias dentro de la reserva de construcciones de fusión.

Secuencias de control de la expresión

Además de las secuencias de fase abierta de lectura descritas anteriormente, es generalmente preferible proporcionar cada construcción de biblioteca con secuencias de control de la expresión, tales como promotores, terminadores de transcripción, potenciadores, sitios de unión a ribosomas y otras secuencias funcionales que pueden ser necesarias para asegurar la transcripción y traducción eficaz de las proteínas de fusión tras transformación de construcciones de fusión en el organismo huésped.

Se seleccionan componentes de vectores adecuados, por ejemplo, marcadores seleccionables, secuencias de control de la expresión (por ejemplo, promotor, potenciadores, terminadores y similares) y opcionalmente, secuencias requeridas para la replicación autónoma en una célula huésped, en función de los que se selecciona la célula huésped particular. Los criterios de selección para componentes de vectores adecuados para usar en una célula huésped de mamífero o eucariota inferior particular son rutinarios. Las células huésped eucariotas inferiores preferidas de la invención incluyen *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp., *Saccharomyces cerevisiae, Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis, Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger, Aspergillus oryzae, Trichoderma reesei, Chrysosporium lucknowense, Fusarium sp. Fusarium gramineum, Fusarium venenatum* y *Neurospora crassa.* Cuando el huésped es *Pichia pastoris*, los promotores adecuados incluyen, por ejemplo, los promotores *AOX1*, *AOX2*, GAPDH y *P40.*

Marcadores seleccionables

También es preferible proporcionar cada construcción con al menos un marcador seleccionable, tal como un gen para transmitir resistencia a fármacos o para complementar una lesión metabólica del huésped. La presencia del marcador es útil en la selección posterior de transformantes; por ejemplo, en levadura pueden usarse los genes *URA3, HIS4, SUC2, G418, BLA* o *SH BLE*. Se conoce una multitud de marcadores seleccionables y están disponibles para su uso en levadura, hongos, plantas, insectos, mamíferos y otras células huésped eucariotas.

Transformación

La biblioteca de ácidos nucleicos se transforma después en los organismos huésped. En levadura, puede usarse cualquier procedimiento conveniente de transferencia de ADN, tal como electroporación, el procedimiento de cloruro de litio o el procedimiento de esferoplastos. En hongos filamentosos y células vegetales, los procedimientos convencionales incluyen bombardeo de partículas, electroporación y transformación mediada por agrobacterium. Para producir una cepa estable adecuada para cultivo de alta densidad (por ejemplo, fermentación en levadura), es deseable integrar las construcciones de biblioteca de ADN en el cromosoma huésped. En una realización preferida, la integración se produce mediante recombinación homóloga, usando técnicas bien conocidas en la técnica. Por ejemplo, se proporcionan elementos de biblioteca de ADN con secuencias flanqueantes homólogas a secuencias del organismo huésped. De esta manera, la integración se produce en un sitio definido en el genoma del huésped, sin alteración de genes deseables o esenciales.

Es especialmente deseable que el ADN de la biblioteca se integre en el sitio de un gen no deseado en un cromosoma huésped, efectuando la alteración o deleción del gen. Por ejemplo, la integración en los sitios de los genes *OCH1*, *MNN1* o *MNN4* permite la expresión del ADN de biblioteca deseado evitando a la vez la expresión de enzimas implicadas en la hipermanosilación de glucoproteínas en levaduras. En otros casos, puede introducirse ADN de biblioteca en el huésped mediante una molécula de ácido nucleico, plásmido, vector (por ejemplo, vector viral o retroviral), cromosoma y puede introducirse como una molécula de ácido nucleico autónoma o por integración homóloga o aleatoria en el genoma del huésped. En cualquier caso, es generalmente deseable incluir con cada construcción de ADN de biblioteca al menos un gen marcador seleccionable para permitir la selección rápida de organismos huésped que se hayan transformado de forma estable. Son especialmente adecuados genes marcadores reciclables tales como *ura3*, que pueden seleccionarse de forma positiva o negativa.

Procedimiento de exploración y selección

Después de transformación de la cepa huésped con la biblioteca de ADN, se seleccionan transformantes que presenten un fenotipo de glucosilación deseado. La selección puede realizarse en una sola etapa o mediante una serie de etapas de enriquecimiento fenotípico y/o depleción fenotípica usando cualquiera de una diversidad de ensayos o procedimientos de detección. La caracterización fenotípica puede llevarse a cabo de forma manual o usando equipo de exploración de alto rendimiento automáticos. Habitualmente, un microorganismo huésped presenta *N*-glucanos proteicos en la superficie celular, en la que se localizan diversas glucoproteínas.

Alguien puede rastrear las células que tienen la concentración más alta de GlcNAc terminal en la superficie celular, por ejemplo, o las células que secreten la proteína con el mayor contenido de GlcNAc terminal. Un rastreo tal puede basarse en un procedimiento visual, como un procedimiento de tinción, la capacidad de unirse a anticuerpos de unión a GlcNAc terminales específicos o lectinas conjugadas con un marcador (tales lectinas están disponibles de E.Y. Laboratories Inc., San Mateo, CA), la capacidad reducida de lectinas específicas para unirse a restos de manosa terminales, la capacidad para incorporar un azúcar marcado de forma radiactiva *in vivo*, unión alterada a colorantes o superficies cargadas, o puede conseguirse usando un dispositivo de Separación de Células Activadas por Fluorescencia (FACS) junto con una lectina o anticuerpo marcado con fluoróforo (Guillen, 1998).

En consecuencia, las células intactas pueden rastrearse con respecto a un fenotipo de glucosilación deseado exponiendo las células a una lectina o anticuerpo que se une específicamente al *N*-glucano deseado. En el mercado está disponible una amplia diversidad de lectinas específicas de oligosacáridos (por ejemplo, de EY Laboratories, San Mateo, CA). Como alternativa, están disponibles en el mercado anticuerpos para *N*-glucanos humanos o animales específicos o pueden producirse usando técnicas convencionales. Una lectina o anticuerpo apropiado puede conjugarse con una molécula indicadora, tal como un cromóforo, fluoróforo, radioisótopo o una enzima que tenga un sustrato cromogénico (Guillen y cols., 1998. Proc. Natl. Acad. Sci. EE.UU. 95 (14): 7888-7892)).

Después puede realizarse la exploración usando procedimientos analíticos tales como espectrofotometría, fluorometría, separación de células activadas por fluorescencia, o conteo de centelleo. En otros casos, puede ser necesario analizar glucoproteínas aisladas o *N*-glucanos a partir de las células transformadas. El aislamiento de proteínas puede llevarse a cabo por técnicas conocidas en este campo. En una realización preferida, una proteína indicadora se secreta al medio y se purifica por cromatografía de afinidad (por ejemplo, cromatografía de afinidad Ni o afinidad de glutatión S-transferasa). En casos en los que se prefiere un *N*-glucano aislado, puede usarse una enzima tal como endo-β-*N*-acetilglucosaminidasa (Genzyme Co., Boston, MA; New England Biolabs, Beverly, MA) para escindir los *N*-glucanos de glucoproteínas. Después pueden analizarse proteínas aisladas o *N*-glucanos por cromatografía líquida (por ejemplo, HPLC), espectroscopía de masas, u otros medios adecuados. La Patente de Estados Unidos N.º: 5.595.900 enseña varios procedimientos por los que pueden identificarse células con estructuras de carbohidratos extracelulares deseadas. En una realización preferida, se usa espectrometría de masas MALDI-TOF para analizar los N-glucanos escindidos.

Antes de la selección de un transformante deseado, puede ser deseable reducir la población transformada de células que tienen fenotipos no deseados. Por ejemplo, cuando el procedimiento se usa para modificar por ingeniería genética una actividad manosidasa funcional en las células, los transformantes deseados tendrán niveles más bajos de manosa en una glucoproteína celular. La exposición de la población transformada a un radioisótopo

letal de manosa en el medio reduce la población de transformantes que tienen el fenotipo no deseado, es decir niveles altos de manosa incorporada (Huffaker TC y Robbins PW., Proc Natl Acad Sci EE.UU. 1983 diciembre; 80 (24): 7466-70). Como alternativa, puede usarse una lectina citotóxica o anticuerpo, dirigido contra un *N*-glucano no deseable, para reducir una población transformada de fenotipos no deseados (por ejemplo, Stanley P y Siminovitch L. Somatic Cell Genet julio 1977; 3 (4):391-405). La Patente de Estados Unidos N.º: 5.595.900 enseña varios procedimientos por los que pueden identificarse células con una estructuras de carbohidratos extracelulares deseadas. Llevar a cabo repetidas veces esta estrategia permite la modificación por ingeniería genética secuencial de glucanos cada vez más complejos en eucariotas inferiores.

Para detectar células huésped que tengan en su superficie un alto grado del intermedio de N-glucano de tipo humano $GlcNAcMan_3GlcNAc_2$, por ejemplo, alguien puede seleccionar transformantes que permitan la transferencia más eficaz de GlcNAc mediante GlcNAc Transferasa de UDP-GlcNAc en un ensayo celular *in vitro*. Esta exploración puede llevarse a cabo cultivando células que albergan la biblioteca transformada sometidas a presión selectiva en una placa de agar y transfiriendo colonias individuales a una placa de microtitulación de 96 pocillos. Después de cultivar las células, las células se centrifugan, las células se resuspenden en tampón y después de la adición de UDP-GlcNAc y GnT II, la liberación de UDP se determina mediante HPLC o un ensayo ligado a enzima para UDP. Como alternativa, se puede usar UDP-GlcNAc marcada de forma radiactiva y GnT II, lavar las células y después buscar la liberación de GlcNAc radiactiva por N-acetilglucosaminidasa. Todo esto puede llevarse a cabo manualmente o de forma automática a través del uso de un equipo de exploración de alto rendimiento. Se espera que los transformantes que liberan más UDP en el primer ensayo, o más GlcNAc marcado radiactivamente en el segundo ensayo, tengan un mayor grado de $GlcNAcMan_3GlcNAc_2$ en su superficie y de este modo constituyan el fenotipo deseado. Pueden adaptarse ensayos similares para contemplar también N-glucanos en proteínas secretadas.

Como alternativa, se puede usar cualquier otro rastreo adecuado tal como un ensayo de unión de lectina que sea capaz de revelar patrones de glucosilación alterados en la superficie de células transformadas. En este caso la unión reducida de lectinas específica con manosas terminales puede ser una herramienta de selección adecuada. La lectina de *Galantus nivalis* se une específicamente a α-1,3 manosa terminal, que se espera que se reduzca si está presente suficiente actividad manosidasa II en el Golgi. También se puede enriquecer con respecto a los transformantes deseados llevando a cabo una etapa de separación cromatográfica que permita la retirada de células que contengan un alto contenido de manosa terminal. Esta etapa de separación se llevaría a cabo con una columna de lectina que se une específicamente a células con un contenido de manosa terminal alto (por ejemplo, lectina de *Galantus nivalis* unida a agarosa, Sigma, St. Louis, MO) sobre las que tienen un contenido de manosa terminal bajo.

Además, alguien puede crear directamente tales construcciones de proteínas de fusión, a medida que se pone a disposición en la bibliografía científica información adicional sobre la localización de enzimas que modifican carbohidratos activos en diferentes huéspedes eucariotas inferiores. Por ejemplo, se sabe que β1,4-GalTr puede fusionarse al dominio de membrana de *MNT*, una manosiltransferasa de *S. cerevisiae* y localizarse en el aparato de Golgi conservando a la vez su actividad catalítica (Schwientek y cols., 1995). Si *S. cerevisiae* o un organismo relacionado es el huésped para modificar por ingeniería genética alguien puede incorporar directamente tales hallazgos en la estrategia global para obtener N-glucanos complejos a partir de tal huésped. Se han identificado varios de estos fragmentos génicos en *P. pastoris* que están relacionados con glucosiltransferasas en *S. cerevisiae* y por lo tanto pueden usarse para ese fin.

Alteración de la glucosilación de células huésped usando construcciones de fusión de bibliotecas combinatorias

La construcción de una biblioteca de ADN combinatoria como se describe en el presente documento se ilustra de forma esquemática en la **Figura 2** y se describe en el **Ejemplo 11.** La construcción de fusión puede unirse operativamente con una multitud de vectores, tales como vectores de expresión bien conocidos en la técnica. Se ensambló una amplia diversidad de tales construcciones de fusión usando actividades representativas como se muestra en la **Tabla 6.** Pueden ensamblarse combinaciones de dominios peptídicos/catalíticos de dirección para su uso en la dirección de actividades manosidasa, glucosiltransferasa y glucosidasa en el RE, el Golgi y la red del trans-Golgi de acuerdo con la invención. Sorprendentemente, el mismo dominio catalítico puede desde no tener ningún efecto hasta tener un efecto muy profundo en los patrones de N-glucosilación, dependiendo del tipo de péptido de dirección usado (véase, por ejemplo, la **Tabla 7, Ejemplo 11**).

Construcciones de fusión de manosidasa

Un ejemplo representativo de una construcción de fusión de manosidasa derivada de una biblioteca de ADN combinatoria de la invención es pFB8, con un péptido de dirección *SEC12*(m) de *Saccharomyces* (nucleótidos 988-1296 de *SEC12* de SwissProt P11655) ligado en fase con una deleción de aminoácido N-terminal 187 de una α-manosidasa IA de ratón (Genbank AN 6678787). La nomenclatura usada en el presente documento, por lo tanto, se refiere a la región de dominio peptídico/catalítico de dirección de una enzima de glucosilación como *SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón. La proteína de fusión codificada se localiza en el RE por medio de la secuencia peptídica de dirección de *SEC12* conservando a la vez su actividad de dominio catalítico de manosidasa y es capaz de producir N-glucanos *in vivo* que tienen una estructura de $Man_5GlcNAc_2$ (**Ejemplo 11; Figura 6F, Figura**

7B).

La construcción de fusión **pGC5,** *MNS1*(m) de *Saccharomyces*/manosidasa IB Δ99 de ratón, es otro ejemplo de una construcción de fusión que tiene actividad de reducción de manosidasa intracelular (**Ejemplo 11; Figura 5D, Figura 8B**). La construcción de fusión **pBC18-5** (*VAN1*(s) de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*) es otro ejemplo más de una construcción de fusión eficaz capaz de producir N-glucanos que tienen una estructura de $Man_5GlcNAc_2$ *in vivo*. Creando una biblioteca de ADN combinatoria de estas y otras construcciones de fusión de manosidasa de acuerdo con la invención, un experto en la materia puede distinguir y seleccionar las construcciones que tienen una actividad de reducción intracelular óptima entre las que tienen actividad relativamente baja o no tienen ninguna actividad. Los procedimientos que usan bibliotecas de ADN combinatorias de la invención son ventajosos debido a que solo unas pocas construcciones de fusión de manosidasas selectas pueden producir un *N*-glucano particularmente deseado *in vivo.*

Además, la actividad de reducción de manosidasa puede ser específica de una proteína particular de interés. Así, debe entenderse adicionalmente que no todas las construcciones de fusión de dominio catalítico de manosidasa/péptido de dirección pueden actuar igualmente bien para producir la glucosilación apropiada en una glucoproteína de interés. En consecuencia, una proteína de interés puede introducirse en una célula huésped transfectada con una biblioteca de ADN combinatoria para identificar una o más construcciones de fusión que expresan una actividad manosidasa óptima para la proteína de interés. Un experto en la materia será capaz de producir y seleccionar una construcción o construcciones de fusión óptimas usando el enfoque de biblioteca de ADN combinatoria descrito en el presente documento.

Resulta evidente, además, que pueden prepararse otras construcciones de fusión que muestran dominios catalíticos de manosidasa activos localizados (o de forma más general, dominios de cualquier enzima) usando técnicas tales como las ejemplificadas en el **Ejemplo 11** y descritas en el presente documento. Será un asunto de experimentación rutinaria para el experto en la materia preparar y usar la biblioteca de ADN combinatoria de la presente invención para optimizar, por ejemplo, la producción de $Man_5GlcNAc_2$ a partir de una biblioteca de construcciones de fusión en un vector de expresión particular introducido en una célula huésped particular.

Construcciones de fusión de glucosiltransferasa

De forma similar, se preparó una biblioteca de ADN combinatoria de glucosiltransferasa usando los procedimientos descritos en el presente documento. Una biblioteca de ADN combinatoria de secuencias derivadas de actividades de glucosiltransferasa I (GnTI) se ensambló con péptidos de dirección y se exploró con respecto a la producción eficaz en una célula huésped eucariota inferior de una estructura de N-glucano $GlcNAcMan_5GlcNAc_2$ en una glucoproteína marcadora. Se identificó una construcción de fusión que mostró producir $GlcNAcMan_5GlcNAc_2$ (**pPB104**), *MNN9*(s) de *Saccharomyces*/GnTI Δ38 humana (**Ejemplo 15**). Se ensambló una amplia diversidad de dichas construcciones de fusión de GnTI (**Ejemplo 15, Tabla 10**). Otras combinaciones de dominios catalíticos de GnTI/péptido de dirección pueden ensamblarse fácilmente realizando una biblioteca de ADN combinatoria. También resulta evidente para un experto en la técnica que pueden fabricarse otras construcciones de fusión que muestren actividad glucosiltransferasa como se demuestra en el **Ejemplo 15.** Será un asunto de experimentación rutinaria para un experto en la técnica usar el procedimiento de biblioteca de ADN combinatoria descrito en el presente documento para optimizar la producción de $GlcNAcMan_5GlcNAc_2$ usando una construcción de fusión seleccionada en un vector de expresión y línea de células huésped particular.

Como se ha indicado anteriormente para construcciones de fusión de manosidasa, no todas las construcciones de fusión de dominio catalítico de péptido de dirección/GnTI actuarán igualmente bien para producir la glucosilación apropiada en una glucoproteína de interés como se describe en el presente documento. Sin embargo, un experto en la materia será capaz de producir y seleccionar construcción/construcciones de fusión óptimas usando un enfoque de biblioteca de ADN como se describe en el presente documento. El **Ejemplo 15** ilustra una realización preferida de una biblioteca de ADN combinatoria que comprende péptidos de dirección y construcciones de fusión de dominio catalítico de GnTI implicadas en la producción de glucoproteínas con estructura predominantemente de $GlcNAcMan_5GlcNAc_2$.

Uso de múltiples construcciones de fusión para alterar la glucosilación de células huésped

En otro ejemplo de uso de los procedimientos de la invención y las bibliotecas descritas en el presente documento para alterar la glucosilación de células huésped, se transformó una cepa de *P. pastoris* con una deleción de *OCH1* que expresa una proteína indicadora (K3) con múltiples construcciones de fusión aisladas de bibliotecas combinatorias de la invención para convertir N-glucanos de alto contenido de manosa en N-glucanos de tipo humano (**Ejemplo 15**). En primer lugar, la construcción de fusión de manosidasa **pFB8** (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) se usó para transformar una cepa de *P. pastoris* carente de actividad manosiltransferasa de inicio 1,6 (es decir deleción *och1*; **Ejemplo 1**). En segundo lugar, se construyó **pPB103** que comprendía un gen *MNN2-2* de *K. lactis* (Genbank AN AF106080) que codificaba un transportador de UDP-GlcNAc para aumentar adicionalmente la producción de $GlcNAcMan_5GlcNAc_2$. La adición del transportador de UDP-GlcNAc aumentó la producción de $GlcNAcMan_5GlcNAc_2$ significativamente en la cepa de *P. pastoris* como se ilustra en la **Figura 10B.** En tercer lugar, se introdujo **pPB104** que comprendía *MNN9*(s) de *Saccharomyces*/GnTI Δ38 humano

en la cepa. Esta cepa de *P. pastoris* se denomina "**PBP-3.**"

Se entiende por un experto en la materia que las células huésped tales como las cepas de levadura anteriormente descritas pueden transformarse secuencialmente y/o cotransformarse con uno o más vectores de expresión. También se entiende que el orden de transformación no es particularmente relevante en producir la glucoproteína de interés. El experto en la materia reconoce que las modificaciones rutinarias de los procedimientos desvelados en el presente documento pueden proporcionar resultados mejorados en la producción de la glucoproteína de interés.

La importancia de usar una secuencia peptídica de dirección particular con una secuencia de dominio catalítico particular resulta fácilmente evidente a partir de los experimentos descritos en el presente documento. La biblioteca de ADN combinatoria proporciona una herramienta para construir fusiones de enzima que están implicadas en la modificación de N-glucanos en una glucoproteína de interés, que es especialmente útil en la producción de glucoproteínas de tipo humano. (Cualquier fusión enzimática, sin embargo, puede seleccionarse usando bibliotecas y procedimientos de la invención). Los transformantes deseados que expresan α-1,2-manosidasa activa dirigida de forma apropiada producen K3 con *N*-glucanos de la estructura de $Man_5GlcNAc_2$ como se muestra en las **Figuras 5D y 5E.** Esto confiere una masa molecular reducida al glucano escindido en comparación con la K3 de la cepa de deleción de *OCHI* parental, como se detectó por espectrometría de masas MALDI-TOF en la **Figura 5C.**

De forma similar, se usó el mismo enfoque para producir otra glucoproteína secretada: IFN-β que comprende predominantemente $Man_5GlcNAc_2$. La $Man_5GlcNAc_2$ se retiró por digestión con PNGasa (Papac y cols. 1998 Glycobiology 8, 445-454) y se sometió a MALDI-TOF como se muestra en la **Figura 6A - 6F.** Un pico prominente individual en 1254 (m/z) confirma producción de Man_5GlcNA_2 en IFN-β en la **Figura 6E (pGC5)** (*MNS1(*m) de *Saccharomyces*/manosidasa IB Δ99 de ratón) y **6F (pFB8)** (*SEC12* (m)de *Saccharomyces*/manosidasa IA Δ187 de ratón). Además, en la cepa de *P. pastoris* PBP-3 que comprende **pFB8** (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón), **pPB104** (*MNN9* (s) de *Saccharomyces/*GnTI Δ38 humana) y **pPB103** (*gen MNN2-2* de *K. lactis*), se detectó el N-glucano híbrido $GlcNAcMan_5GlcNAc_2$ [**b**] por MALDI-TOF **(Figura 10).**

Después de identificar transformantes con un alto grado de reducción de manosa, se realizaron experimentos adicionales para confirmar que se producía actividad manosidasa (recorte) *in vivo* y no era predominantemente el resultado de actividad extracelular en el medio de cultivo **(Ejemplo 13**; **Figuras 7-9).**

Células huésped

Aunque la presente invención se ejemplifica usando un organismo huésped *P. pastoris*, los expertos en la materia entenderán que otras células huésped eucariotas, incluyendo otras especies de huéspedes de hongos y levadura, pueden alterarse como se describe en el presente documento para producir glucoproteínas de tipo humano. Se entiende que las técnicas descritas en el presente documento para identificación y alteración de genes de glucosilación de células huésped no deseables, por ejemplo *OCH1,* son aplicables para estos y/u otros genes homólogos o funcionalmente relacionados en otras células huésped eucariotas tales como otras cepas de levadura y hongos. Como se describe en el **Ejemplo 16**, se delecionaron genes *ochl mnn1* de *K. lactis* para modificar por ingeniería genética una célula huésped, lo que condujo a N-glucanos que se convertían completamente en $Man_5GlcNAc_2$ mediante la 1,2-manosidasa **(Figura 12C)**.

El gen *MNN1* se clonó a partir de *K. lactis* como se describe en el **Ejemplo 16.** Las secuencias de ácido nucleico y aminoácidos deducidas del gen *MNN1* de *K. lactis* se muestran en las SEC ID N.º: 43 y 44, respectivamente. Usando cebadores específicos de genes, se realizó una construcción para delecionar el gen *MNN1* del genoma de *K. lactis* **(Ejemplo 16).** Las células huésped con actividades *och1* y *mnn1* suprimidas producen N-glucanos que tienen una estructura de carbohidrato $Man_9GlcNAc_2$ (véase, por ejemplo, la **Figura 10).** Tales células huésped pueden modificarse por ingeniería genética usando adicionalmente, por ejemplo, procedimientos y bibliotecas de la invención, para producir glucoproteínas de tipo mamífero o humano.

Así, la revelación proporciona adicionalmente una molécula de ácido nucleico aislada que tiene una secuencia de ácido nucleico que comprende o consiste en al menos cuarenta y cinco, preferentemente al menos 50, más preferentemente al menos 60 y más preferentemente 75 o más restos de nucleótidos del gen *MNN1* de *K. lactis* (SEC ID N.º: 43) y homólogos, variantes y derivados de los mismos. Se proporcionan también moléculas de ácido nucleico que hibridan en condiciones rigurosas con las moléculas de ácido nucleico descritas anteriormente. De forma similar, se proporcionan polipéptidos aislados (incluyendo muteínas, variantes alélicas, fragmentos, derivados y análogos) codificados por las moléculas de ácido nucleico de la invención. Además, también se describen vectores, incluyendo vectores de expresión, que comprenden una molécula de ácido nucleico de la invención, como se describe adicionalmente en el presente documento. Se proporcionan de forma similar, en una realización de la invención, células huésped transformadas con las moléculas o vectores de ácidos nucleicos descritos en el presente documento.

Otro aspecto de la presente invención se refiere así a una cepa huésped de hongos unicelulares o filamentosos que expresan glucoproteínas que comprenden N-glucanos modificados que se asemejan a los realizados por células humanas. Se contempla también llevar a cabo los procedimientos descritos en el presente documento en especies distintas de levaduras y células fúngicas. Se contempla que una biblioteca de ácidos nucleicos combinatoria descrita

en el presente documento se puede usar para seleccionar construcciones que modifican la ruta de glicosilación en cualquier sistema de células huésped eucariotas. Por ejemplo, las bibliotecas combinatorias descritas en el presente documento se pueden usar también en plantas, algas e insectos y en otras células huésped eucariotas, incluyendo células de mamíferos y células humanas, para localizar proteínas, incluyendo enzimas de glucosilación o dominios catalíticos de las mismas, en una localización deseada a lo largo de una vía secretora de células huésped. Como se describe en el presente documento las enzimas de glicosilación o los dominios catalíticos y similares están dirigidos a una localización subcelular a lo largo de la ruta secretora de células huésped donde son capaces de funcionar y preferentemente, donde se han diseñado o seleccionado para funcionar de forma más eficaz.

Como se describe en los Ejemplos 17 y 18, pueden modificarse por ingeniería genética células de plantas y de insectos para alterar la glucosilación de proteínas expresadas usando la biblioteca combinatoria y los procedimientos descritos en el presente documento. Además, también puede modificarse la glucosilación en células de mamífero, incluyendo células humanas usando la biblioteca combinatoria y los procedimientos descritos en el presente documento. Puede ser posible, por ejemplo, optimizar una actividad enzimática particular o modificar de otro modo las proporciones relativas de diversos N-glucanos realizados en una célula huésped de mamífero usando la biblioteca combinatoria y los procedimientos descritos en el presente documento.

Son ejemplos de modificaciones de glucosilación que pueden verse afectadas usando un procedimiento descrito en el presente documento: (1) modificación por ingeniería genética de una célula huésped eucariota para recortar residuos de manosa de $Man_8GlcNAc_2$ para producir un *N*-glucano $Man_5GlcNAc_2$; (2) modificación por ingeniería genética de una célula huésped eucariota para añadir un residuo de *N*-acetilglucosamina (GlcNAc) a $Man_5GlcNAc_2$ mediante la acción de la GlcNAc transferasa I; (3) modificación por ingeniería genética de una célula huésped eucariota para expresar funcionalmente una enzima tal como una *N*-acetilglucosaminiltransferasa (GnTI, GnTII, GnTIII, GnTIV, GnTV, GnTVI), manosidasa II, fucosiltransferasa (FT), galactosiltransferasa (GalT) o una sialiltransferasa (ST).

Repitiendo el procedimiento, pueden modificarse por ingeniería genética rutas de glucosilación cada vez más complejas en un huésped diana, tal como un hongo unicelular o filamentoso. En una realización preferida, el organismo huésped según se reivindica se transforma dos o más veces con bibliotecas de ADN que incluyen secuencias que codifican actividades de glucosilación. Debe realizarse la selección de fenotipos deseados después de cada ronda de transformación o, como alternativa, después que hayan tenido lugar varias transformaciones. De esta manera, las rutas de glucosilación complejas pueden modificarse por ingeniería genética rápidamente.

Reacciones de glucosilación secuenciales

También se describe en el presente documento que tales bibliotecas de dominios catalíticos/péptidos de dirección se diseñan para incorporar información existente sobre la naturaleza secuencial de reacciones de glucosilación en eucariotas superiores. Las reacciones que se sabe que tienen lugar temprano en el curso del procesamiento de glucoproteínas requieren la dirección de enzimas que catalizan tales reacciones a una parte temprana del Golgi o el RE. Por ejemplo, la reducción de $Man_5GlcNAc_2$ a $Man_5GlcNAc_2$ mediante manosidasas es una etapa temprana en la formación de N-glucanos complejos. Debido a que el procesamiento de proteínas se inicia en el RE y después continúa a través del Golgi temprano, medial y tardío, es deseable que esta reacción haya tenido lugar en el RE o Golgi temprano. Cuando se diseña una biblioteca para localización de manosidasa I, por ejemplo, alguien intenta así combinar señales de dirección de RE y Golgi temprano con el dominio catalítico de manosidasa I.

Sitios de integración

Como un objetivo último de estos esfuerzos de ingeniería genética está una cepa de producción de proteína robusta que sea capaz de actuar bien en un procedimiento de fermentación industrial, la integración de múltiples genes en el cromosoma huésped (por ejemplo, fúngico) preferentemente implica planificación cuidadosa. La cepa modificada por ingeniería genética puede tener que transformarse probablemente con una serie de genes diferentes y estos genes tendrán que transformarse de una manera estable para asegurar que se mantenga la actividad deseada durante el procedimiento de fermentación. Como se describe en el presente documento, puede manipularse cualquier combinación de diversas actividades enzimáticas deseadas por ingeniería genética en el huésped de expresión de proteína fúngico, por ejemplo, sialiltransferasas, manosidasas, fucosiltransferasas, galactosiltransferasas, glucosiltransferasas, GlcNAc transferasas, transportadores específicos de RE y Golgi (por ejemplo transportadores simporte y antiporte de UDP-galactosa y otros precursores), otras enzimas implicadas en el procesamiento de oligosacáridos y enzimas implicadas en la síntesis de precursores de oligosacáridos activados tales como UDP-galactosa y CMP-ácido N-acetilneuramínico. Se muestran ejemplos de procedimientos preferidos para modificar glucosilación en una célula huésped eucariota inferior, tal como *Pichia pastoris,* en la **Tabla 6**. (Los péptidos señal HDEL y KDEL en la segunda fila de la tercera columna se muestran en las **SEC ID N.º: 5 y 6**, respectivamente).

Tabla 6. Algunas realizaciones preferidas para modificar la glucosilación en un microorganismo eucariota inferior

Estructura deseada	*Actividades Catalíticas Adecuadas*	*Fuentes Adecuadas de Secuencias de Localización*	*Deleciones Génicas Adecuadas*	*Transportadores Adecuados y/o Fosfatasas*
$Man_5GlcNAc_2$	α-1,2-manosidasa (murina, humana, *Bacillus* sp., *A. nidulans)*	Mns1 (N-terminal, *S. cerevisiae)* Och1 (N-terminal, *S. cerevisiae, P. pastoris*) Ktr1 Mnn9 Mnt1 *(S. cerevisiae)* KDEL, HDEL (C-terminal)	*OCH1 MNN4* *MNN6*	ninguno
$GlcNAcMan_5GlcNAc_2$	GlcNAc Transferasa I, (humana, murina, de rata, etc.)	Och1 (N-terminal, *S cerevisiae, P. pastoris*) KTR1 (N-terminal) Mnn1 (N-terminal, *S. cerevisiae)* Mnt1 (N-terminal, *S. cerevisiae)* GDPasa (N-terminal, *S. cerevisiae)*	*OCH1 MNN4 MNN6*	Transportador de *OCH1 UDP-GlcNAc* (humano, murino, *K. lactis)* UDPasa (humana)
$GlcNAcMan_3GlcNAc_2$	manosidasa II	Ktr1 Mnn1 (N-terminal, *S. cerevisiae)* Mnt1(N-terminal, *S. cerevisiae)* Kre2/Mnt1 *(S. cerevisiae)* Kre2 *(P. pastoris)* Ktr1 *(S. cerevisiae)* Ktr1 *(P. pastoris)* Mnn1 *(S. cerevisiae)*	*OCH1 MNN4 MNN6*	Transportador de UDP-GlcNAc (humano, murino, *K. lactis)* UDPasa (humana)
$GlcNAc_{(2-4)}Man_3GlcNAc_2$	GlcNAc Transferasa II, III, IV, V (humana, murina)	Mnn1 (N-terminal, *S. cerevisiae)* Mnt1 (N-terminal, *S. cerevisiae)* Kre2/Mnt1 *(S. cerevisiae)* Kre2 *(P. pastoris)* Ktr1 *(S. cerevisiae)* Ktr1 *(P. pastoris)* Mnn1 *(S. cerevisiae)*	*OCH1 MNN4 MNN6*	Transportador de UDP-GlcNAc (humano, murino, *K. lactis)* UDPasa (humana)
$Gal_{(1-4)}GlcNAc_{(2-4)}$-$Man_3GlcNAc_2$	β-1,4-Galactosil transferasa (humana)	Mnn1 (N-terminal, *S. cerevisiae)* Mnt1(N-terminal, *S. cerevisiae)* Kre2/Mnt1 *(S. cerevisiae)* Kre2 *(P. pastoris)* Ktr1 *(S. cerevisiae)* Ktr1 *(P. pastoris)* Mnn1 *(S. cerevisiae)*	*OCH1 MNN4 MNN6*	Transportador de UDP-Galactosa (humano, *S. pombe)*
$NANA_{(1-4)}$-$Gal_{(1-4)}GlcNAc_{(2-4)}$-$Man_3GlcNAc_2$	α-2,6-Sialiltransferasa (humana) α-2,3-Sialiltransferasa	KTR1 MNN1 (N-terminal, *S. cerevisiae)* MNT1 (N-terminal, *S. cerevisiae)* Kre2/Mnt1 *(S. cerevisiae)* Kre2 *(P. pastoris)* Ktr1 *(S. cerevisiae)* Ktr1 *(P. pastoris)* MNN1 *(S. cerevisiae)*	*OCH1 MNN4 MNN6*	Transportador de CMP-ácido siálico (humano)

Como cualquier estrategia para modificar por ingeniería genética la formación de N-glucanos complejos en una célula huésped tal como un eucariota inferior implica tanto la eliminación como la adición de actividades glucosiltransferasa particulares, un esquema exhaustivo intentará coordinar ambos requisitos. Los genes que codifican enzimas que no son deseables actúan como sitios de integración potenciales para genes que son deseables. Por ejemplo, la actividad 1,6 manosiltransferasa es una evidencia de glucosilación en muchos eucariotas inferiores conocidos. El gen que codifica la alfa-1,6 manosiltransferasa (*OCH1*) se ha clonado a partir de *S. cerevisiae* y las mutaciones en el gen dan lugar a un fenotipo viable con manosilación reducida. El locus del gen que codifica la actividad alfa-1,6 manosiltransferasa es por lo tanto una diana perfecta para la integración de genes que codifican la actividad glucosiltransferasa. En una manera similar, alguien puede elegir una serie de otros sitios de integración cromosómica que, basándose en un acontecimiento de interrupción de un gen en ese locus, es de esperar que: (1) mejoren la capacidad de las células para glucosilar de una manera más humana, (2) mejoren la capacidad de las células para secretar proteínas, (3) reduzcan la proteolisis de proteínas extrañas y (4) mejoren otras características del procedimiento que faciliten la purificación o el propio proceso de fermentación.

Glucoproteínas diana

Los procedimientos descritos en el presente documento son útiles para producir glucoproteínas, especialmente glucoproteínas usadas terapéuticamente en seres humanos. Las glucoproteínas que tienen glucoformas específicas pueden ser especialmente útiles, por ejemplo, en la dirección de proteínas terapéuticas. Por ejemplo, se ha mostrado que la manosa-6-fosfato dirige proteínas al lisosoma, que pueden ser esenciales para la función apropiada de varias enzimas relacionadas con trastornos de almacenamiento lisosomal tales como enfermedad de Gaucher, Hunter, Hurler, Scheie, Fabry y Tay-Sachs, por mencionar solamente algunas. De forma similar, la adición de uno o más restos de ácido siálico a una cadena lateral de glucano puede aumentar el tiempo de vida de una glucoproteína terapéutica *in vivo* después de la administración. En consecuencia, las células huésped (por ejemplo, eucariotas inferiores o de mamífero) pueden modificarse por ingeniería genética para aumentar el grado de ácido siálico terminal en glucoproteínas expresado en las células. Como alternativa, el ácido siálico puede conjugarse con la proteína de interés *in vitro* antes de la administración usando una ácido siálico transferasa y un sustrato apropiado. Pueden emplearse cambios en el medio de cultivo además de la expresión de actividades enzimáticas implicadas en la glucosilación de tipo humano para producir glucoproteínas que se asemejen más a las formas humanas (S. Weikert y cols., Nature Biotechnology, 1999, 17, 1116-1121; Werner, Noe y cols, 1998 Arzneimittelforschung 48 (8): 870-880; Weikert, Papacetal., 1999; Andersen y Goochee 1994 *Cur. Opin. Biotechnol.* 5: 546-549; Yang y Butler 2000 Biotechnol. Bioengin. 68 (4): 370-380). Se ha mostrado que modificaciones de glucanos específicas de anticuerpos monoclonales (por ejemplo, la adición de una GlcNAc de bisección) mejoran la citotoxicidad celular dependiente de anticuerpos (Umana P. y cols. 1999), lo que puede ser deseable para la producción de anticuerpos u otras proteínas terapéuticas.

Las proteínas terapéuticas se administran normalmente por inyección, por vía oral, pulmonar u otros medios. Los ejemplos de glucoproteínas diana adecuadas que pueden producirse de acuerdo con la invención incluyen, sin limitación: eritropoyetina, citocinas tales como interferón-α, interferón-β, interferón-γ, interferón-ω y CSF de granulocitos, factores de coagulación tales como factor VIII, factor IX y proteína humana C, cadena α del receptor de IgE soluble, IgG, fragmentos de IgG, IgM, interleucinas, urocinasa, quimasa e inhibidor de tripsina de urea, proteína de unión a IGF, factor de crecimiento epidérmico, factor de liberación de hormonas del crecimiento, proteína de fusión de anexina V, angiostatina, factor de crecimiento endotelial vascular 2, factor inhibidor de progenitor mieloide 1, osteoprotegerina, antitripsina α-1, DNasa II, α-fetoproteínas, AAT, rhTBP-1 (onercept, también conocido como proteína de unión a TNF 1), TACI-Ig (activador transmembrana y modulador de calcio y elemento de interacción con ligando de ciclofilina), FSH (hormona foliculoestimulante), GM-CSF, GLP-1 con y sin FC (proteína de tipo glucagón 1), agonista del receptor de IL-1, sTNFr (enbrel, también conocido como fusión de Fc del receptor de TNF soluble) ATIII, rhTrombina, glucocerebrosidasa y CTLA4-Ig (Antígeno 4 asociado a Linfocitos T Citotóxicos-Ig).

Los siguientes son ejemplos que ilustran las composiciones y procedimientos de la presente invención. Estos ejemplos no deben considerarse limitantes: los ejemplos se incluyen solo con fines ilustrativos.

Ejemplo 1

Clonación e interrupción del gen *OCH1* en *P. pastoris*

Generación de un mutante de *OCH1* de *P. pastoris:*

Se amplificó una ORF de 1215 pb del gen de *OCH1* de *P. pastoris* que codificaba una α-1,6 manosiltransferasa potencial a partir de ADN genómico de *P. pastoris* (cepa X-33, Invitrogen, Carlsbad, CA) usando los oligonucleótidos 5'-ATGGCGAAGGCA-GATGGCAGT-3' (SEC ID N.º: 7) y 5'-TTAGTCCTTCCAACTTCCTTC-3' (SEC ID N.º: 8) que se diseñaron basándose en la secuencia de *OCH1* de *P. pastoris* (Publicación de Solicitud de Patente Japonesa N.º: 8-336387). Posteriormente, se amplificaron 2685 pb cadena arriba y 1175 pb cadena abajo de la ORF del gen *OCH1* a partir de una biblioteca de ADN genómico de *P. pastoris* (Boehm, T. y cols. Yeast mayo de 1999; 15 (7): 563-72) usando los oligonucleótidos internos 5'-ACTGCCATCTGCCTTCGCCAT-3' (SEC ID N.º: 9) en el gen *OCH1* y los oligonucleótidos 5'-GTAATACGACTCACTATAGGGC-3'T7 (SEC ID N.º: 10) y 5'-AATTAAC-CCTCACTAAAGGG-3' T3 (SEC ID N.º: 11) en la cadena principal del plásmido que porta la biblioteca lambda ZAP II (Stratagene, La Jolla, CA). El fragmento de 5075 pb resultante se clonó en el vector pCR2.1-TOPO (Invitrogen, Carlsbad, CA) y se designó pBK9.

Después de ensamblar una construcción de bloqueo génico que sustituía la fase de lectura de *OCH1* con un gen de resistencia *HIS4*, *P. pastoris* se transformó y se exploraron las colonias con respecto a la sensibilidad a la temperatura a 37 ºC. Los mutantes *OCH1* de *S. cerevisiae* son sensibles a la temperatura y proliferan lentamente a temperaturas elevadas. Por lo tanto, se pueden identificar homólogos funcionales de *OCH1* en *P. pastoris* complementando un mutante de *OCH1* de *S. cerevisiae* con una biblioteca de ADNc o ADN de *P. pastoris*. Se sometieron aproximadamente 20 cepas sensibles a la temperatura adicionalmente a una exploración por PCR de colonias identificando colonias con un gen *och1* suprimido. Se obtuvieron varias deleciones de *och1*.

El pBK9.1 linealizado, que tiene la secuencia de 2,1 kb cadena arriba y la secuencia de 1,5 kb cadena abajo del casete del gen *OCH1* que porta gen *HIS4* de *Pichia*, se introdujo por transformación en *P. pastoris* BK1 [GS115 (his4 Invitrogen Corp., San Diego, CA) que porta el gen IFN-β humano en el locus de *AOX1*] suprimiendo el gen *OCH1* de tipo silvestre. La exploración inicial de los transformantes se realizó usando medio deficiente en histidina seguido de siembras en placas repetidas seleccionando las colonias sensibles a la temperatura. Veinte de doscientas colonias

positivas para histidina mostraron un fenotipo sensible a la temperatura a 37 ºC. Excluyendo la integración aleatoria de pBK9.1 en el genoma de *Pichia*, los 20 aislados sensibles a la temperatura se sometieron a PCR de colonias usando cebadores específicos para la secuencia cadena arriba del sitio de integración y para la *ORF* de *HIS4*. Dos de veinte colonias eran defectuosas para *och1* y se analizaron adicionalmente usando una transferencia de Southern y una transferencia de Western indicando la interrupción funcional de *och1* por la construcción knock-out de *och1*. El ADN genómico se digirió usando dos enzimas de restricción separadas *BglII* y *ClaI* confirmando el bloqueo de *och1* y confirmando la integración en la fase abierta de lectura. La transferencia de Western mostró mutantes de *och1* que carecían de una banda discreta producida en el tipo silvestre de GS 115 a 46,2 kDa.

Ejemplo 2

Modificación por ingeniería genética de *P. pastoris* con α-1,2-manosidasa para producir precursores de IFN-β que contienen $Man_5GlcNAc_2$

Se requiere una α-1,2-manosidasa para el recorte de $Man_8GlcNAc_2$ produciendo $Man_5GlcNAc_2$, un intermedio esencial para formación de *N*-glucanos complejos. Aunque la producción de un precursor de $Man_5GlcNAc_2$ es esencial, no es necesariamente suficiente para la producción de glucanos híbridos y complejos debido a que el isómero específico de $Man_5GlcNAc_2$ puede o no ser un sustrato para GnTI. Un mutante de *och1* de *P. pastoris* se modifica por ingeniería genética expresando interferón-β humano secretado sometido al control de un promotor de *aox*. Se construye una biblioteca de ADN mediante la ligación en fase del dominio catalítico de manosidasa IB humana (una α-1,2-manosidasa) con una sub-biblioteca que incluye secuencias que codifican péptidos de localización en Golgi temprano y RE. La biblioteca de ADN se introduce después por transformación en el organismo huésped, dando como resultado una población genéticamente mixta en la que los transformantes individuales expresan cada uno interferón así como un gen de manosidasa sintética de la biblioteca. Se cultivan colonias transformantes individuales y la producción del interferón se induce mediante la adición de metanol. En estas condiciones, más del 90 % de la proteína secretada es interferón β glucosilado.

Los sobrenadantes se purifican retirando sales y contaminantes de bajo peso molecular por cromatografía de fase inversa de sílice C_{18}. Los transformantes deseados que expresan α-1,2-manosidasa activa apropiadamente dirigida producen interferón que incluye *N*-glucanos de la estructura de $Man_5GlcNAc_2$, que tiene una masa molecular reducida en comparación con el interferón-β de la cepa parental. El interferón-β purificado se analiza por espectrometría de masas MALDI-TOF y se identifican colonias que expresan la forma deseada de interferón-β.

Ejemplo 3

Generación de una cepa mutante de *och1* que expresa una α-1,2-manosidasa, GnTI y GnTII para producción de una glucoproteína de tipo humano.

La fase abierta de lectura de 1215 pb del gen de *OCH1* de *P. pastoris* así como 2685 pb cadena arriba y 1175 pb cadena abajo se amplificaron por PCR (véase también el documento WO 02/00879), se clonaron en el vector pCR2.1-TOPO (Invitrogen) y se designaron pBK9. Creando una cepa bloqueada en *och1* que contuviera múltiples marcadores autotróficos, se digirieron 100 μg de pJN329, un plásmido que contenía un alelo mutante *och1::URA3* flanqueado con sitios de restricción *Sfi*I con *Sfi*I y se usaron transformando la cepa JC308 de *P. pastoris* (Cereghino y cols. Gene 263 (2001) 159-169) por electroporación. Después de la incubación en medio definido carente de uracilo durante 10 días a temperatura ambiente, se seleccionaron 1.000 colonias y se volvieron a sembrar en estrías. Se sometieron clones de URA^+ que eran incapaces de proliferar a 37 ºC, pero que proliferaban a temperatura ambiente, a PCR de colonias ensayando la integración correcta del alelo mutante de *och1::URA3*. Un clon que mostraba el patrón de PCR esperado se denominó YJN153. El dominio Kringle 3 de plasminógeno humano (K3) se usó como una proteína modelo. Un plásmido marcado con Neo^R que contenía el gen K3 se transformó en cepa YJN153 y una cepa resultante, que expresaba K3, se denominó BK64-1.

El plásmido pPB103, que contiene el gen *MNN2-2* de *Kluyveromyces lactis* que codifica un transportador de UDP-N-acetilglucosamina de Golgi se construyó clonando un fragmento de *Bgl*II-*Hind*III romo a partir del vector pDL02 (Abeijon y cols. (1996) Proc. Natl. Acad. Sci. EE.UU. 93: 5963-5968) en pBLADE-SX de extremos romos y digerido con *Bgl*II y *Bam*HI que contenía el gen *ADE1* de *P. pastoris* (Cereghino y cols. (2001) Gene 263: 159-169). Este plásmido se linealizó con *Eco*NI y se usó transformando la cepa BK64-1 por electroporación y una cepa que se confirmó que contenía el *MNN2-2* por análisis de PCR se denominó PBP 1.

Se generó una biblioteca de construcciones de manosidasa, que comprendía fusiones en fase de los dominios líder de varias proteínas de membrana de tipo I o tipo II de *S. cerevisiae* y *P. pastoris* fusionadas con los dominios catalíticos de varios genes de α-1,2-manosidasa de fuentes humanas, de ratón, de mosca, de gusano y de levadura (véase, por ejemplo, el documento WO02/00879, incorporado en el presente documento por referencia). Esta biblioteca se creó en un vector de integración de *HIS4* de *P. pastoris* y se exploró linealizando con *Sal*I, transformando por electroporación la cepa PBP1 y analizando los glucanos liberados de la proteína indicadora K3. Una construcción activa seleccionada fue una quimera de los 988-1296 nucleótidos (extremo C-terminal) del gen *SEC12* de levadura fusionado con una deleción N-terminal del gen α-1,2-manosidasa IA de ratón (**Figura 3**), del que faltaban los 187 nucleótidos. Una cepa de *P. pastoris* que expresaba esta construcción se denominó PBP2.

Se generó una biblioteca de construcciones de GnTI, que comprendía fusiones en fase de la misma biblioteca líder con los dominios catalíticos de los genes GnTI de fuentes humanas, de gusano, de rana y de mosca (documento WO 02/00879). Esta biblioteca se creó en un vector de integración de *ARG4* de *P. pastoris* y se rastreó linealizando con *Aat*II, transformando por electroporación la cepa PBP2 y analizando los glucanos liberados de K3. Una construcción activa seleccionada fue una quimera de los primeros 120 pb del gen de *MNN9* de *S. cerevisiae* fusionado con una deleción del gen GnTI humano, del que faltaban los primeros 154 pb. Una cepa de *P. pastoris* que expresaba esta construcción se denominó PBP3.

Se generó una biblioteca de construcciones de GnTII, que comprendía fusiones en fase de la biblioteca líder con los dominios catalíticos de genes GnTII a partir de fuentes humanas y de rata (documento WO 02/00879). Esta biblioteca se creó en un vector de integración de *P. pastoris* que contenía el gen NST^R que confería resistencia al fármaco nurseotricina. Los plásmidos de la biblioteca se linealizaron con *Eco*RI, se usaron transformando la cepa RDP27 por electroporación y las cepas resultantes se exploraron por análisis de los glucanos liberados de K3 purificado.

Materiales para las siguientes reacciones

MOPS, cacodilato sódico, cloruro de manganeso, UDP-galactosa y CMP-ácido N-acetilneuramínico se obtuvieron de Sigma. El ácido trifluoroacético (TFA) era de Sigma/Aldrich, Saint Louis, MO. La α2,6-sialiltransferasa de rata recombinante de *Spodoptera frugiperda* y la β1,4-galactosiltransferasa de leche bovina eran de Calbiochem (San Diego, CA). La proteína N-glucosidasa F, las manosidasas y los oligosacáridos eran de Glyko (San Rafael, CA). La resina DEAE ToyoPearl se obtuvo en TosoHaas. La resina "HisBind" quelante metálica se obtuvo en Novagen (Madison, WI). Las placas de eliminación de lisado de 96 pocillos se obtuvieron en Promega (Madison, WI). Las placas de 96 pocillos de unión a proteína se obtuvieron en Millipore (Bedford, MA). Las sales y agentes tamponantes se obtuvieron en Sigma (St. Louis, MO). Las matrices MALDI se obtuvieron en Aldrich (Milwaukee, WI).

Purificación de proteínas

Se purificó Kringle 3 usando un formato de 96 pocillos en un robot de manipulación de muestras Beckman BioMek 2000 (Beckman/Coulter Ranch Cucamonga, CA). Kringle 3 se purificó a partir de medio de expresión usando un marcador de hexahistidina C-terminal. La purificación robótica es una adaptación del protocolo proporcionado por Novagen para su resina HisBind. En resumen, se vierte un volumen sedimentado de 150 ul (μl) de resina en los pocillos de una placa de unión a lisado de 96 pocillos, se lava con 3 volúmenes de agua y se carga con 5 volúmenes de NiSO4 50 mM y se lava con 3 volúmenes de tampón de unión (imidazol 5 mM, NaCl 0,5 M, Tris-HCl 20 mM pH 7,9). El medio de expresión de proteína se diluye 3:2, medio/PBS (PO4 60 mM, KCl 16 mM, NaCl 822 mM pH 7,4) y se carga en las columnas. Después de escurrir, las columnas se lavan con 10 volúmenes de tampón de unión y 6 volúmenes de tampón de lavado (imidazol 30 mM, NaCl 0,5M, Tris-HCl 20 mM pH 7,9) y la proteína se eluye con 6 volúmenes de tampón de elución (imidazol 1 M, NaCl 0,5 M, Tris-HCl 20 mM pH 7,9). Las glucoproteínas eluidas se evaporan hasta sequedad por liofilización.

Liberación de glucanos ligados a N

Los glucanos se liberan y separan de las glucoproteínas por una modificación de un procedimiento anteriormente presentado (Papac y cols. A. J. S. (1998) Glycobiology 8, 445-454). Los pocillos de una placa MultiScreen IP de 96 pocillos (membrana Immobilon-P) (Millipore) se humectan con 100 µl de metanol, se lavan con 3 X 150 µl de agua y 50 µl de tampón RCM (urea 8 M, Tris 360 mM, EDTA 3,2 mM pH 8,6), drenando con vacío suave después de cada adición. Las muestras de proteína secadas se disuelven en 30 µl de tampón RCM y se transfieren a los pocillos que contienen 10 µl de tampón RCM. Los pocillos se drenan y se lavan dos veces con tampón RCM. Las proteínas se reducen mediante adición de 60 μl de DTT 0,1 M en tampón RCM durante 1 hora a 37 ºC. Los pocillos se lavan tres veces con 300 µl de agua y se carboximetilan mediante la adición de 60 μl de ácido yodoacético 0,1 M durante 30 minutos en la oscuridad a temperatura ambiente. Los pocillos se lavan de nuevo tres veces con agua y las membranas se bloquean mediante la adición de 100 µl de PVP 360 al 1 % en agua durante 1 hora a temperatura ambiente. Los pocillos se drenan y se lavan tres veces con 300 µl de agua y se desglucosilan mediante la adición de 30 μl de NH_4HCO_3 10 mM pH 8,3 que contiene una miliunidad de N-glucanasa (Glyko). Después de 16 horas a 37 ºC, la solución que contiene los glucanos se retiró por centrifugación y se evaporó hasta sequedad.

Espectrometría de masas de dispersión de ionización de desorción por láser asistida por matriz

Se determinaron los pesos moleculares de los glucanos usando un espectrómetro de masas MALDI-TOF lineal Voyager DE PRO (Applied Biosciences) usando extracción retardada. Los glucanos secados de cada pocillo se disolvieron en 15 µl de agua y se aplicaron puntualmente 0,5 µl en placas de muestra de acero inoxidable y se mezclaron con 0,5 µl de matriz S-DHB (ácido dihidrobenzoico 9 mg/ml, ácido 5-metoxisalicílico 1 mg/ml en agua/acetonitrilo 1:1 TFA al 0,1 %) y se dejaron secar.

Se generaron iones por irradiación con un láser de nitrógeno pulsado (337 nm) con un tiempo de pulso de 4 ns. El instrumento se operó en el modo de extracción retardada con un retardo de 125 ns y una tensión de aceleración de 20 kV. La tensión de rejilla fue del 93,00 %, la tensión del cable guía fue del 0,10 %, la presión interna fue menor de

0,67 x 10^{-5} kPa (5 x 10^{-7} torr) y el umbral de masa inferior fue de 875 Da. Se generaron espectros a partir de la suma de 100-200 pulsos de láser y se adquirieron con un digitalizador de 2 GHz. Como un patrón de peso molecular externo se usó el oligosacárido $Man_5GlcNAc_2$. Todos los espectros se generaron con el instrumento en el modo de ion positivo. La precisión de la masa estimada de los espectros fue del 0,5 %.

Ejemplo 4

Modificación por ingeniería genética de una cepa para producir galactosiltransferasa

Reacción de galactosiltransferasa

Se purificaron aproximadamente 2 mg de proteína (r-K3:hPg [PBP6-5]) por cromatografía de afinidad de níquel, se dializaron exhaustivamente frente a TFA al 0,1 % y se liofilizaron hasta sequedad. La proteína se volvió a disolver en 150 μl de MOPS 50 mM, $MnCl_2$ 20 mM, pH 7,4. Después de la adición de 32,5 μg (533 nmol) de UDP-galactosa y 4 mU de β1,4-galactosiltransferasa, la muestra se incubó a 37 ºC durante 18 horas. Las muestras se dializaron después frente a TFA al 0,1 % para el análisis por espectrometría de masas MALDI-TOF.

El espectro de la proteína que reaccionó con galactosiltransferasa mostró un aumento de la masa coherente con la adición de dos restos de galactosa en comparación con el espectro de una muestra proteica similar incubada sin enzima. A continuación, las muestras de proteína se redujeron, se carboximetilaron y desglucosilaron con PNGasa F. Los N-glucanos recuperados se analizaron por espectrometría de masas MALDI-TOF. La masa del glucano predominante de la proteína que había reaccionado con galactosiltransferasa fue mayor que la del glucano de control por una masa coherente con la adición de dos restos de galactosa (325,4 Da).

Ejemplo 5

Modificación por ingeniería genética de una cepa para expresar manosidasa II funcional y activa

Generando una glucoforma de tipo humano, un microorganismo se modifica por ingeniería genética expresando una enzima manosidasa II que retira las dos manosas terminales que quedan de la estructura $GlcNAcMan_5GlcNAc_2$ (véase la **Figura 1B**). Una biblioteca de ADN que incluye secuencias que codifican señales de localización del Golgi cis y medial se fusiona en fase a una biblioteca que codifica dominios catalíticos de manosidasa II. El organismo huésped es una cepa, por ejemplo una levadura, que es deficiente en hipermanosilación (por ejemplo, un mutante de *och1*) y proporciona *N*-glucanos que tienen la estructura $GlcNAcMan_5GlcNAc_2$ en el Golgi y/o RE. Después de la transformación, se seleccionan los organismos que tienen el fenotipo de glucosilación deseado. En un procedimiento se usa un ensayo *in vitro*. La estructura deseada $GlcNAcMan_3GlcNAc_2$ (pero no la no deseada $GlcNAcMan_5GlcNAc_2$) es un sustrato para la enzima GlcNAc Transferasa II (véase la **Figura 1B**). En consecuencia, pueden ensayarse colonias individuales usando esta enzima *in vitro* en presencia del sustrato, UDP-GlcNAc. La liberación de UDP se determina mediante HPLC o un ensayo enzimático para UDP. Como alternativa, puede usarse UDP-GlcNAc marcado radiactivamente o MALDI-TOF.

Los ensayos *in vitro* anteriores se realizan convenientemente en colonias individuales usando un equipo de rastreo de alto rendimiento. Como alternativa se usa un ensayo de unión de lectina. En este caso la unión reducida de lectinas específicas para manosas terminales permite la selección de transformantes que tengan el fenotipo deseado. Por ejemplo, la lectina de *Galantus nivalis* se une específicamente a la α-1,3-manosa terminal, cuya concentración se reduce en presencia de la actividad manosidasa II expresada operativamente. En un procedimiento adecuado, se usa la lectina de *G. nivalis* unida a un soporte de agarosa sólido (disponible de Sigma Chemical, St. Louis, MO) reduciendo la población transformada de células que tienen altos niveles de α-1,3-manosa terminal.

Ejemplo 6

Modificación por ingeniería genética de una cepa para expresar sialiltransferasa

Las enzimas α2,3-sialiltransferasa y α2,6-sialiltransferasa añaden ácido siálico terminal a restos de galactosa en *N*-glucanos humanos nacientes, formando glucoproteínas maduras (véase "α 2,3 ST; α2,6 ST" en la **Figura 1B**). En células humanas, las reacciones se producen en el trans-Golgi o TGN. En consecuencia, se construye una biblioteca de ADN por la fusión en fase de secuencias que codifican dominios catalíticos de sialiltransferasa con secuencias que codifican señales de localización de trans-Golgi o TGN (Malissard y cols. Biochem Biophys Res Commun 7 ene 2000, 7; 267 (1):169-73; Borsig y cols. Biochem Biophys Res Commun, 5 de mayo de 1995; 210 (1):14-20). El organismo huésped es una cepa, por ejemplo una levadura, que es deficiente en hipermanosilación (por ejemplo, un mutante de *och1*), que proporciona *N*-glucanos que tienen restos de galactosa terminales en el Golgi tardío o TGN y proporciona una concentración suficiente de CMP-ácido siálico en el Golgi tardío o TGN. Después de la transformación, se seleccionan los transformantes que tienen el fenotipo deseado, por ejemplo, usando un anticuerpo fluorescente específico para *N*-glucanos que tienen un ácido siálico terminal. Además, las cepas se modifican por ingeniería genética produciendo los precursores de CMP-NANA.

Reacción de sialiltransferasa

Después de resuspender las proteínas (que han reaccionado con galactosiltransferasa) (**Ejemplo 4**) en 10 μl de tampón de cacodilato sódico 50 mM pH 6,0, se añadieron 300 μg (488 nmol) de CMP-N-ácido acetilneuramínico (CMP-NANA) disuelto en 15 μl del mismo tampón y 5 μl (2 mU) de α-2,6 sialiltransferasa recombinante. Después de la incubación a 37 ºC durante 15 horas, se añadieron 200 μg adicionales de CMP-NANA y 1 mU de sialiltransferasa. Las muestras de proteína se incubaron durante 8 horas adicionales y después se dializaron y analizaron mediante MALDI-TOF-MS como anteriormente. El espectro de la glucoproteína que había reaccionado con sialiltransferasa mostró un aumento de la masa cuando se comparó con la del material de partida (la proteína después de la reacción de galactosiltransferasa). Los N-glucanos se liberaron y se analizaron como anteriormente. El aumento de la masa de los dos aductos iónicos del glucano predominante fue coherente con la adición de dos restos de ácido siálico (580 y 583 Da).

Ejemplo 7

Modificación por ingeniería genética de una cepa para expresar transportador de UDP-GlcNAc

El ADNc de transportador de UDP-GlcNAc de Golgi humano se ha clonado por Ishida y colaboradores (Ishida, N. y cols. 1999 J. Biochem. 126 (1): 68-77). Guillen y colaboradores han clonado el transportador de UDP-GlcNAc de Golgi de riñón canino por corrección fenotípica de un mutante de *Kluyveromyces lactis* deficiente en transportador de UDP-GlcNAc del Golgi (Guillen, E. y cols. 1998). Por lo tanto, un gen de transportador de UDP-GlcNAc del Golgi de mamífero tiene toda la información necesaria para que la proteína se exprese y se dirija funcionalmente al aparato de Golgi de levadura. Estos u otros genes transportadores clonados pueden modificarse por ingeniería genética en un organismo huésped proporcionando sustratos de UDP-GlcNAc para reacciones de GnT eficaces en el Golgi y/o RE del huésped. La **Figura 10B** demuestra el efecto de una cepa que expresa un transportador de UDP-GlcNAc de *K. lactis*. En comparación con la **Figura 10A**, que carece de un transportador de UDP-GlcNAc, el efecto de añadir un transportador de UDP-GlcNAc muestra un aumento drástico en la producción de $GlcNAcMan_5GlcNAc_2$.

Ejemplo 8

Modificación por ingeniería genética de una cepa para expresar transportador de GDP-Fucosa

El transportador de GDP-fucosa de membrana del Golgi de hígado de rata lo han identificado y purificado Puglielli, L. y C. B. Hirschberg 1999 J. Biol. Chem. 274 (50): 35596-35600. El gen correspondiente puede identificarse usando técnicas convencionales, tales como secuenciación *N*-terminal y transferencia de Southern usando una sonda de ADN degenerada. El gen intacto se expresa después en un microorganismo huésped que también expresa una fucosiltransferasa.

Ejemplo 9

Modificación por ingeniería genética de una cepa para expresar transportador de UDP-galactosa

Se ha clonado transportador de UDP-galactosa humano (UDP-Gal) y se ha mostrado que es activo en *S. cerevisiae.* (Kainuma, M. y cols. 1999 Glycobiology 9 (2): 133-141). Se ha clonado un segundo transportador de UDP-galactosa humano (HUGTl) y se ha expresado funcionalmente en células de ovario de hámster chino. Aoki, K. y cols. 1999 J. Biochem. 126 (5): 940-950. De forma similar, Segawa y colaboradores han clonado un transportador de UDP-galactosa a partir de *Schizosaccharomyces pombe* (Segawa, H. y cols. 1999 Febs Letters 451(3): 295-298). Estas y otras secuencias que codifican actividades de transportador de UDP-galactosa pueden introducirse en una célula huésped directamente o pueden usarse como un componente de una sub-biblioteca de la invención modificando por ingeniería genética una cepa que tenga la actividad transportadora de UDP-galactosa aumentada.

Ejemplo 10

Modificación por ingeniería genética de una cepa para expresar transportador de CMP-ácido siálico

Aoki y colaboradores (1999) han clonado transportador de CMP-ácido siálico humano (hCST) y lo han expresado en células Lec 8 CHO. También se ha conseguido la clonación molecular del transportador de CMP-ácido siálico de hámster (Eckhardt y Gerardy Schahn 1997 Eur. J. Biochem. 248 (1): 187-192). La expresión funcional del transportador de CMP-ácido siálico murino lo consiguieron en *Saccharomyces cerevisiae* Berninsone, P. y cols. 1997 J. Biol. Chem. 272 (19): 12616-12619. Estas u otras secuencias que codifican actividades de transportador de CMP-ácido siálico pueden introducirse en una célula huésped directamente o pueden usarse como un componente de una sub-biblioteca de la invención modificando por ingeniería genética una cepa que tenga la actividad transportadora de CMP-ácido siálico aumentada.

Ejemplo 11

Modificación por ingeniería genética de *P. pastoris* para producir Man_5GlcNA_2 como la estructura de *N*-glucano predominante usando una biblioteca de ADN combinatoria

Se modificó por ingeniería genética un mutante en *och1* de *P. pastoris* (véanse los **Ejemplos 1 y 3)** de forma que

expresó y secretó proteínas tales como el dominio kringle 3 de plasminógeno humano (K3) sometido al control del promotor de *AOXI* inducible. El dominio kringle 3 de plasminógeno humano (K3) se usó como una proteína modelo. Un fragmento de ADN que codificaba el K3 se amplificó usando turbo polimerasa Pfu (Strategene, La Jolla, CA) y se clonó en sitios *EcoRI* y *XbaI* de pPICZαA (Invitrogen, Carlsbad, CA), dando como un resultado un marcador 6-His C-terminal. Mejorando la eficacia de glucosilación ligada a N de K3 (Hayes y cols. 1975 J. Arch. Biochem. Biophys. 171, 651-655), Pro_{46} se reemplazó con Ser_{46} usando mutagénesis dirigida. El plásmido resultante se designó pBK64. La secuencia correcta de la construcción de PCR se confirmó por secuenciación de ADN.

Se construyó una biblioteca de ADN combinatoria por la ligación en fase de dominios catalíticos de α-1,2-manosidasa IB (Genbank AN 6678787) y IA (Genbank AN 6754619) murina con una sub-biblioteca que incluía secuencias que codificaban péptidos Cop II de localización en vesícula, RE y Golgi temprano de acuerdo con la Tabla 6. La biblioteca de ADN combinada se usó generando construcciones de fusión individuales, que se introdujeron después en el organismo huésped que expresaba K3 por transformación, dando como resultado una población genéticamente mixta en la que cada uno de los transformantes individuales expresa K3, así como un gen de señal de localización/fusión de manosidasa de la biblioteca. Los transformantes individuales se cultivaron y la producción de K3 se indujo por transferencia a un medio que contenía metanol. En estas condiciones, después de 24 horas de inducción, más del 90 % de la proteína en el medio era K3. La proteína indicadora K3 se purificó a partir del sobrenadante retirando sales y contaminantes de bajo peso molecular por cromatografía de afinidad de Ni. Después de la purificación de afinidad, la proteína se desaló por cromatografía de exclusión de tamaño en una resina Sephadex G10 (Sigma, St. Louis, MO) y bien se sometió directamente al análisis de MALDI-TOF descrito posteriormente o bien los *N*-glucanos se retiraron por digestión con PNGasa como se describe posteriormente (Liberación de N-glucanos) y se sometieron a análisis de MALDI-TOF Miele y cols. 1997 Biotechnol. Appl. Biochem. 25,151-157.

Siguiendo este enfoque, se obtuvo un conjunto diverso de transformantes; algunos no mostraron ninguna modificación de los *N*-glucanos en comparación con la cepa de *ochl* bloqueado; y otros mostraron un alto grado de recorte de manosa (Figura 5D, 5E). Los transformantes deseados expresando α-1,2-manosidasa activa apropiadamente dirigida produjeron K3 con *N*-glucanos de la estructura de $Man_5GlcNAc_2$. Esto confiere una masa molecular reducida a la glucoproteína en comparación con el K3 de la cepa de deleción de *ochl* parental, una diferencia que se detectó fácilmente por espectrometría de masas de MALDI-TOF (Figura 5). La **Tabla 7** indica los niveles de producción de $Man_5GlcNAc_2$ relativos.

Tabla 7. Una biblioteca de ADN combinatoria representativa de secuencias de localización/dominios catalíticos/que muestran niveles relativos de producción de $Man_5GlcNAc_2$

		Secuencias peptídicas de dirección				
		MNS1(s)	***MNS1(m)***	***MNS1(l)***	***SEC12(s)***	***SEC12(m)***
Catalítico	**Manosidasa 1A Δ187 de ratón**	**FB4 ++**	**FB5 +**	**FB6 -**	**FB7 ++**	**FB8 ++++**
	Manosidasa 1B Δ58 de ratón	**GB4 ++**	**GB5 +**	**GB6 +**	**GB7 ++**	**GB8 +**
	Manosidasa 1B Δ99 de ratón	**GC4 -**	**GC5 +++**	**GC6 +**	**GC7 +**	**GC8 +**
	Manosidasa 1B Δ170 de ratón	**GD4 -**	**GD5 -**	**CD6 -**	**GD7 +**	**GD8 +**

Tabla 8. Otra biblioteca de ADN combinatoria de secuencias de localización/dominios catalíticos que muestran niveles relativos de producción de $Man_5GlcNAc_2$

		Secuencias peptídicas de dirección					
Dominios Catalíticos		***VAN1(s)***	***VAN1(m)***	***VAN(l)***	***MNN10(s)***	***MNN10(m)***	***MNN10(l)***
	Manosidasa 1B Δ80 de *C. elegans*	**BC18-5 +++++**	**BC19 ++++**	**BC20 +++**	**BC27 +++++**	**BC28 +++++**	**BC29 +++**
	Manosidasa 1B Δ31 de *C. elegans*	**BB18 +++++**	**BB19 +++++**	**BB20 ++++**	**BB18 +++++**	**BB19 +++++**	**BB20 ++++**

Los péptidos de dirección se seleccionaron a partir de *MNSI* (SwissProt P32906) en *S. cerevisiae* (largo, medio y corto) (véase anteriormente Bibliotecas de Ácidos Nucleicos; Biblioteca de ADN Combinatoria de Construcciones de Fusión) y *SEC12* (SwissProt P11655) en *S. cerevisiae* (988-1140 nucleótidos: corto) y (988-1296: medio). Aunque la mayoría de las secuencias peptídicas de dirección eran deleciones N-terminales, algunas secuencias peptídicas de dirección, tales como *SEC12* eran deleciones C-terminales. Los dominios catalíticos usados en este experimento se seleccionaron a partir de manosidasa 1A de ratón con una deleción N-terminal de 187 aminoácidos; y manosidasa 1B de ratón con una deleción de 58, 99 y 170 aminoácidos. El número de (+), como se usa en el presente documento, indica los niveles relativos de producción de Man_5GlcNA_2. La notación (-) indica que no hay producción aparente de Man_5GlcNA_2. La notación (+) indica menos del 10 % de producción de Man_5GlcNA_2. La notación (++) indica aproximadamente un 10-20 % de producción de Man_5GlcNA_2. La notación con (+++) indica aproximadamente un 20-40 % de producción de Man_5GlcNA_2. La notación con (++++) indica aproximadamente un 50 % de producción de Man_5GlcNA_2. La notación con (+++++) indica más del 50 % de producción de Man_5GlcNA_2.

La **Tabla 9** muestra la cantidad relativa de $Man_5GlcNAc_2$ en K3 secretado. Se generaron seiscientos ocho (608) cepas diferentes de *P. pastoris,* Δ*och1* transformándolas con una sola construcción de una biblioteca genética combinatoria que se generó fusionando diecinueve (19) dominios catalíticos de α-1,2 manosidasa con treinta y dos (32) líderes de RE y cis-Golgi fúngicos.

Tabla 9.

Cantidad de $Man_5GlcNAc_2$ en K3 secretada (% de glucanos totales)	Número de construcciones (%)
N.D.*	19 (3,1)
0-10 %	341 (56,1)
10-20 %	50 (8,2)
20-40 %	75 (12,3)
40-60 %	72 (11,8)
Más del 60 %	51 (8,4) †
Total	608 (100)

* Varias construcciones de fusión no se ensayaron debido a que los plásmidos correspondientes no podían propagarse en *E. coli* antes de su transformación en *P. pastoris*.

† Los clones con el mayor grado de recorte de $Man_5GlcNAc_2$ (30/51) se analizaron adicionalmente con respecto a la actividad manosidasa en el sobrenadante del medio. La mayoría (28/30) presentaron una actividad manosidasa detectable en el sobrenadante (por ejemplo, **Figura 4B**). Solo dos construcciones presentaron niveles de $Man_5GlcNAc_2$ altos, careciendo a la vez de actividad manosidasa en el medio (por ejemplo, **Figura 4C**).

La **Tabla 7** muestra dos construcciones **pFB8** y **pGC5**, entre otras, que presentan Man_5GlcNA_2. La **Tabla 8** muestra una construcción más preferida, **pBC18-5**, una secuencia de péptido de dirección *VAN1*(s) de *S. cerevisiae* (de SwissProt 23642) ligada en fase con una deleción N-terminal de 80 aminoácidos de manosidasa IB de *C. elegans* (Genbank AN CAA98114) (Van1(s) de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*). Esta construcción de fusión también produce una estructura de Man_5GlcNA_2 predominante, como se muestra en la **Figura 5E**. Se mostró que esta construcción producía más del 50 % de Man_5GlcNA_2 (+++++).

Generación de una biblioteca de localización/manosidasa combinatoria:

La generación de una biblioteca de ADN combinatoria de dominios catalíticos de α-1,2-manosidasa fusionados con péptidos de dirección requería la amplificación de dominios de manosidasa con diversas longitudes de deleciones *N*-terminales de un número de organismos. Consiguiendo este objetivo, las fases abiertas de lectura de longitud completa (ORF) de α-1,2-manosidasas se amplificaron mediante PCR a partir de ADNc o ADN genómico obtenido a partir de las siguientes fuentes: *Homo sapiens, Mus musculus, Drosophila melanogaster, Caenorhabditis elegans, Aspergillus nidulans* y *Penicillium citrinum.* En cada caso, se incubó ADN en presencia de cebadores oligonucleotídicos específicos para la secuencia de manosidasa deseada además de reactivos requeridos realizando la reacción de PCR. Por ejemplo, amplificando la ORF de la α-1,2-manosidasa IA de *M. musculus*, se incubaron el cebador 5' ATGCCCGTGGGGGGCCTGTTGCCGCTCTTCAGTAGC (SEC ID N.º: 12) y el cebador 3'-TCATTTCTCTTTGCCATCAATTTCCTTCTTCTGTTCACGG (SEC ID N.º: 13) en presencia de ADN polimerasa Pfu (Stratagene, La Jolla, CA) y se amplificaron en las condiciones recomendadas por Stratagene usando los parámetros de realización de ciclos: 94 ºC durante 1 min (1 ciclo); 94 ºC durante 30 s, 68 ºC durante 30 s, 72 ºC durante 3 min (30 ciclos). Después de la amplificación, la secuencia de ADN que codificaba la ORF se incubó a 72 ºC durante 5 min con 1 U de ADN polimerasa Taq (Promega, Madison, WI) antes de ligación en pCR2.1-TOPO (Invitrogen, Carlsbad, CA) y se usó transformando *E. coli,* químicamente competente TOP10, como se recomienda por Invitrogen. El producto de PCR clonado se confirmó por secuenciación de ABI usando cebadores específicos para la ORF de manosidasa.

Generando los truncamientos N-terminales deseados de cada manosidasa, se usó la ORF completa de cada manosidasa como molde en un ciclo posterior de reacciones de PCR en las que la posición de hibridación del cebador 5' fue específica para el extremo 5' terminal del truncamiento deseado y el cebador 3' permaneció específico para el extremo 3' original de la ORF. Facilitando la subclonación del fragmento de manosidasa truncado en el vector de expresión de levadura pJN347 (**Figura 2C'**), se introdujeron por ingeniería genética sitios de restricción *AscI* y *PacI* en cada producto de truncamiento, en los extremos 5' y 3' terminales respectivamente. El número y posición de truncamientos *N*-terminales generados para cada ORF de manosidasa dependía de la posición de la región transmembrana (TM) en relación con el dominio catalítico (CD). Por ejemplo, si la región de tallo localizada entre la TM y CD era menor de 150 pb, entonces se generaba solamente un truncamiento para esta proteína. Si, sin embargo, la región de tallo era mayor de 150 pb, entonces se generaban uno o dos truncamientos más dependiendo de la longitud de la región de tallo.

En el presente documento se describe un ejemplo de cómo se generaron truncamientos para la manosidasa IA de *M. musculus* (Genbank AN 6678787), con un enfoque similar usándose para las otras manosidasas. La **Figura 3** ilustra la ORF de la α-1,2-manosidasa IA de *M. musculus* estando los dominios catalíticos transmembrana predichos destacados en negrita. Basándose en esta estructura, se diseñaron tres cebadores 5' (posiciones de hibridación subrayadas en la **Figura 3**) generando las deleciones N-terminales Δ65, Δ105 y Δ187. Usando la deleción N-terminal Δ65 como un ejemplo, el cebador 5' usado fue 5'-**GGCGCGCCGA**CTCCTCCAAGCTGCTCAGCGGGGTCCTGTTCCAC-3' (SEC ID N.º: 14) (con el sitio de restricción AscI destacado en negrita) junto con el cebador 3' 5'-CCT**TAATTAA**TCATTTCTCTTTGCCATCAATTTCCTTCTTCTGTTCACGG-3' (SEC ID N.º: 15) (con el sitio de restricción PacI destacado en negrita). Estos dos cebadores se usaron amplificando un fragmento de 1561 pb en las condiciones perfiladas anteriormente amplificando la ORF de manosidasa IA de *M. musculus* de longitud completa. Además, como el producto obtenido para la ORF de longitud completa, el producto truncado también se incubó con ADN polimerasa Taq, se ligó en pCR2.1-TOPO (Invitrogen, Carlsbad, CA), se introdujo por transformación en TOP10 y se secuenció con ABI. Después de haberse amplificado y confirmado la secuencia del fragmento de manosidasas truncado, el plásmido resultante, pCR2.1-Δ65mMannIA, se digirió con *AscI* y *PacI* en tampón N.º: 4 de New England Biolabs (Beverly, MA) durante 16 h a 37 ºC. En paralelo, el pJN347 (**Figura 2C**) se digirió con las mismas enzimas y se incubó como se ha descrito anteriormente. Después de la digestión, tanto la cadena principal del pJN347 (**Figura 2C**) como el dominio catalítico truncado se extrajeron con gel y se ligaron usando el Kit Quick Ligation (New England Biolabs, Beverly, MA), como lo recomendaban los fabricantes y se usaron transformando células DH5α químicamente competentes (Invitrogen, Carlsbad, CA). Se usó PCR de colonias confirmando la generación de la construcción de Manosidasa IAΔ65 de ratón-pJN347.

Habiendo generado una biblioteca de dominios catalíticos de α-1,2-manosidasa truncados en el vector de expresión de levadura pJN347 (**Figura 2C**), la etapa que queda en la generación de la biblioteca de péptido de dirección/dominio catalítico fue clonar en fase las secuencias peptídicas de dirección (**Figura 2**). Tanto las construcciones de pJN347-manosidasa (**Figura 2D**) como las construcciones de péptido de dirección-pCR2.1TOPO (**Figura 2B**) se incubaron durante una noche a 37 ºC en tampón N.º: 4 de New England Biolabs en presencia de las

enzimas de restricción *NotI* y *AscI*. Después de la digestión, tanto el armazón de pJN347-manosidasa como las regiones de péptido de dirección se extrajeron en gel y se ligaron usando el Kit Quick Ligation (New England Biolabs, Beverly, MA)), como recomendaban los fabricantes y se introdujeron por transformación en células DH5α químicamente competentes (Invitrogen, Carlsbad, CA). Posteriormente, las construcciones de péptido de dirección-pJN347/manosidasa se secuenciaron por ABI confirmando que las fusiones generadas estaban en fase. El tamaño estimado de la biblioteca de péptido de dirección/alfa-1,2-manosidasa final contiene más de 1300 construcciones generadas por el enfoque descrito anteriormente. La **Figura 2** ilustra la construcción de la biblioteca de ADN combinatoria.

Modificación por ingeniería genética de una cepa de *OCH1* inactivada de *P. pastoris* con múltiples marcadores auxotróficos.

La primera etapa en la construcción de plásmidos implicó crear un conjunto de plásmidos universales que contenían regiones de ADN del gen *KEX1* de *P. pastoris* (Boehm y cols. Yeast mayo 1999; 15 (7):563-72) como ocupantes de las regiones 5' y 3' de los genes a inactivarse. Los plásmidos también contenían el Ura-blaster de *S. cerevisiae* (Alani y cols., Genetics 116, 541-545, 1987) como un ocupante para los marcadores auxotróficos y un casete de expresión con un sitio de clonación múltiple para inserción de un gen extraño. Se amplificó un fragmento de 0,9 kb de la región KEX1-5' de *P. pastoris* por PCR usando los cebadores GGCGAGCTCGGCCTACCCGGCCAAGGCTGAGATCATTTGTCCAGCTTCA GA (SEC ID N.º: 16) y GCCCACGTCGACGGATCCGTTTAAACATCGATTGGAGAGGCTGACACC GCTACTA (SEC ID N.º: 17) y ADN genómico de *P. pastoris* como un molde y se clonó en los sitios *SacI, SalI* de pUC19 (New England Biolabs, Beverly, MA). El plásmido resultante se cortó con *BamHI* y *SalI y* se clonó un fragmento de 0,8 kb de la región *KEX1-3'* que se había amplificado usando los cebadores CGGGATCCACTAGTATTTAAATCATATGTGCGAGTGTACAACTCTTCCC ACATGG (SEC ID N.º: 18) y GGACGCGTCGACGGCCTACCCGGCCGTACGAGGAATTTCTCGG ATGACTCTTTTC (SEC ID N.º: 19) en los sitios abiertos creando pJN262. Este plásmido se cortó con *BamHI* y el fragmento *BamHI, BglII* de 3,8 kb de pNKY51 (Alani *y cols.* 1987) se insertó en las dos orientaciones posibles dando como resultado los plásmidos pJN263 (**Figura 4A**) y pJN284 (**Figura 4B**).

Se creó un casete de expresión con *NotI* y *PacI* como sitios de clonación. El promotor GAPDH de *P. pastoris* se amplificó usando los cebadores CGGGATCCCTCGAGAGATCTTTTTTGTAGAAATGTCTTGGTGCCT (SEC ID N.º: 20) y GGACATGCATGCACTAGTGCGGCCGCCACGTGATAGTTGTTCA ATTGATTGAAATAGGGACAA (SEC ID N.º: 21) y el plásmido pGAPZ-A (Invitrogen) como molde y se clonó en los sitios *BamHI, SphI* de pUC19 (New England Biolabs, Beverly, MA) (**Figura 4B**). El plásmido resultante se cortó con *SpeI* y *SphI* y la región terminadora de la transcripción CYC1 (TT') que se había amplificado usando los cebadores CCTTGCTAGCTTAATTAACCGCGGCACGTCCGACGGCGGCCCA CGGGTCCCA (SEC ID N.º: 22) y GGACATGCATGCGGATCCCTTAAGAGCCGGCAGCTTGCAAATT AAAGCCTTCGAGCGTCCC (SEC ID N.º: 23) y el plásmido pPICZ-A (Invitrogen) como un molde se clonó en los sitios abiertos creando pJN261 (**Figura 4B**).

Se creó un plásmido bloqueo para el gen *OCHI* de *P. pastoris* digiriendo pJN263 con *SalI* y *SpeI* y un fragmento de ADN de 2,9 kb de la región *OCH1-5'*, que se había amplificado usando los cebadores GAACCACGTCGACGGCCATTGCGGCCAAAACCTTTTTTCCTATT CAAACACAAGGCATTGC (SEC ID N.º: 24) y CTCCAATACTAGTCGAAGATTATCTTCTACGGTGCCTGGACTC (SEC ID N.º: 25) y ADN genómico de *P. pastoris* como un molde, se clonó en los sitios abiertos (**Figura 4C**). El plásmido resultante se cortó con *EcoRI* y *PmeI* y se insertó un fragmento de ADN de 1,0 kb de la región de *OCHI-3'* que se había generado usando los cebadores TGGAAGGTTTAAACAAAGCTAGAGTAAAATAGATATAGCGAGATTAGAGAATG (SEC ID N.º: 26) y AAGAATTCGGCTGGAAGGCCTTGTACCTTGATGTAGTTCCCGTTTTCATC (SEC ID N.º: 27) generando pJN298 (**Figura 4C**). Permitiendo la posibilidad de usar simultáneamente el plásmido introduciendo un nuevo gen, el casete de expresión de *BamHI* de pJN261 (**Figura 4B**) se clonó en el sitio *BamHI* único de pJN298 (**Figura 4C**) creando pJN299 (**Figura 4E**).

El casete de Ura3-blaster de *P. pastoris* se construyó usando una estrategia similar a la descrita en Lu. P. y cols. 1998 (Cloning and disruption of the β-isopropylmalate dehydrogenase gene (Leu2) of Pichia stipidis with URA3 and recovery of the double auxotroph. Appl. Microbiol. Biotechnol. 49,141-146). Se insertó un fragmento *PstI, SpeI* de 2,0 kb de *URA3* de *P. pastoris* en los sitios *PstI, XbaI* de pUC19 (New England Biolabs, Beverly, MA) creando pJN306 (**Figura 4D**). Después, se clonó un fragmento de ADN de *SacI, PvuII* de 0,7 kb de la fase abierta de lectura de *lacZ* en los sitios ***SacI,*** *SmaI* produciendo pJN3.08 (**Figura 4D**). Después de la digestión de pJN308 (**Figura 4D**) con *PstI y* tratamiento con ADN polimerasa T4, el fragmento *SacI-PvuII* de *lacZ* que se había dejado con extremos romos con ADN polimerasa T4 se insertó generado pJN315 (**Figura 4D**). El casete *lacZ*/URA3 se liberó por digestión con *SacI* y *SphI,* se dejó con extremos romos con ADN polimerasa T4 y se clonó en la cadena principal de p3N299 que se había digerido con *PmeI* y *AflII* y se dejó con extremos romos con ADN polimerasa T4. El plásmido resultante se denominó pJN329 (Figura 4E).

Se creó un plásmido de expresión marcado con *HIS4* cortando pJN261 (**Figura 4F**) con EcoICRI (**Figura 4F**). Se ligó después en el sitio abierto un fragmento de 2,7 kb del gen *HIS4* de *Pichia pastoris* que se había amplificado usando los cebadores GCCCAAGCCGGCCTTAAGGGATCTCCTGATGACTGACTCACTGATAATA AAAATACGG (SEC ID N.º: 28) y GGGCGCGTATTTAAATACTAGTGGATCTATCGAATCTAAATGTAAGTTA AAATCTCTAA (SEC

ID N.º: 29) cortados con *NgoMIV* y *SwaI* y después dejados con extremos romos usando ADN polimerasa T4. Este plásmido se denominó pJN337 (**Figura 4F**). Construyendo un plásmido con un sitio de clonación múltiple adecuado para la construcción de bibliotecas de fusión, se cortó pJN337 con *NotI* y *PacI* y los dos oligonucleótidos GGCCGCCTGCAGATITAAATGAATTCGGCGCGCCITAAT (SEC ID N.º: 30) y TAAGGCGCGCCGAATTCATTTAAATCTGCAGGGC (SEC ID N.º: 31), que se habían hibridado *in vitro*, se ligaron en los sitios abiertos, creando pJN347 (**Figura 4F**).

Creando una cepa de *och1* de bloqueo que contenía múltiples marcadores auxotróficos, se digirieron 100 μg de pJN329 con *SfiI* y se usaron transformando la cepa de *P. pastoris* JC308 (Cereghino y cols. Gene 263 (2001) 159-169) por electroporación. Después de la transformación, las placas deficientes en URA se incubaron a temperatura ambiente durante 10 días. Se seleccionaron mil (1.000) colonias y se volvieron a sembrar en estrías. Las 1.000 colonias se sembraron después en estrías en 2 conjuntos de placas deficientes en URA. Un conjunto se incubó a temperatura ambiente, mientras que el segundo conjunto se incubó a 37 ºC. Los clones que fueron incapaces de proliferar a 37 ºC, pero proliferaron a temperatura ambiente, se sometieron a PCR de colonias ensayando con respecto al bloqueo de *OGHI* correcto. Un clon que mostró la señal de PCR esperada (aproximadamente 4,5 kb) se designó YJN1531

Ejemplo 12

Caracterización de la biblioteca de ADN combinatoria

Se dejaron proliferar los transformantes positivos rastreados por PCR de colonias que confirmó la integración de la construcción de manosidasa en el genoma de *P. pastoris* posteriormente a temperatura ambiente en 50 moles de medio de complejo de metanol tamponado con BMGY (consistente en extracto de levadura al 1 %, peptona al 2 %, tampón fosfato potásico 100 mM, pH 6,0, base nitrogenada de levadura al 1,34 %, biotina al 4 X 10^{-5} % y glicerol al 1 % como medio de cultivo), hasta DO_{600nm} 2-6, punto en el que se lavaron con 10 ml de BMMY (medio complejo con metanol tamponado consistente en extracto de levadura al 1 %, peptona al 2 %, tampón fosfato potásico 100 mM, pH 6,0, base nitrogenada de levadura al 1,34 %, biotina al 4 X 10^{-5} % y metanol al 1,5 % como medio de cultivo) antes de la inducción de la proteína indicadora durante 24 horas a temperatura ambiente en 5 ml de BMMY. En consecuencia, la proteína indicadora se aisló y analizó por espectrometría de masas y HPLC caracterizando su estructura de glucanos. Usando los péptidos de dirección de la Tabla 6, los dominios catalíticos de manosidasa localizados bien en el RE o bien en el Golgi mostraron un nivel significativo de recorte de un glucano que contenía predominantemente $Man_8GlcNAc_2$ a un glucano que contenía predominantemente $Man_5GlcNAc_2$. Esto resulta evidente cuando la estructura del glucano de la glucoproteína indicadora se compara entre la de bloqueo de *och1* de *P. pastoris* en las Figuras **5C, 6C** y la misma cepa transformada con construcciones de manosidasa de *M. musculus* como se muestra en las Figuras **5D, 5E, 6D-6F.** Las **Figuras 5 y 6** muestran la expresión de construcciones generadas a partir de la biblioteca de ADN combinatoria que muestran actividad manosidasa significativa en *P. pastoris.* La expresión de pGC5 (*MNS1* (m) de *Saccharomyces*/manosidasa IB Δ99 de ratón) (**Figura 5D, 6E**) produjo una proteína que tiene aproximadamente un 30 % de todos los glucanos reducidos a $Man_5GlcNAc_2$, mientras que la expresión de pFB8 *(SE12*(m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) (**Figura 6F**) produjo aproximadamente un 50 % de $Man_5GlcNAc_2$ y la expresión de pBC18-5 *(VAN1(s)* de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*) (**Figura 5E**) produjo un 70 % de $Man_5GlcNAc_2$.

Liberación de *N*-glucanos

Los glucanos se liberaron y separaron de las glucoproteínas mediante una modificación de un procedimiento anteriormente presentado (Papac y cols. 1998 *Glycobiology* 8, 44.5-454). Después de que las proteínas se hubieran reducido y carboximetilado y las membranas se hubieran bloqueado, los pocillos se lavaron tres veces con agua. La proteína se desglucosiló mediante la adición de 30 μl de NH_4HCO_3 10 mM pH 8,3 que contenía una miliunidad de N-glucanasa (Glyko, Novato, CA). Después de 16 horas a 37 ºC, la solución que contenía los glucanos se retiró por centrifugación y se evaporó a sequedad.

Espectrometría de masas de dispersión de ionización de desorción por láser asistida por matriz

Después de que los *N*-glucanos se hubieran liberado por digestión con PNGasa, se analizaron por espectrometría de masas de dispersión de ionización de desorción por láser asistida por matriz. Los pesos moleculares de los glucanos se determinaron usando un espectrómetro de masas MALDI-TOF lineal Voyager DE PRO (Applied Biosciences) usando extracción retardada. Los glucanos secados de cada pocillo se disolvieron en 15 μl de agua, se aplicaron puntualmente 0,5 μl en placas de muestra de acero inoxidable y se mezclaron con 0,5 μl de matriz S-DHB (ácido dihidroxibenzoico 9 mg/ml, ácido 5-metoxisalicílico 1 mg/ml en agua/acetonitrilo 1:1 TFA al 0,1 %) y se dejaron secar. Se generaron iones por irradiación con un láser de nitrógeno de impulsos (337 mm) con un tiempo de pulso de 4 ns. El instrumento se manejó en el modo de extracción retardada con un retardo de 125 ns y una tensión de aceleración de 20 kV. La tensión en rejilla fue del 93,00 %, la tensión en cable de guía fue del 0,1 %, la presión interna fue menor de 0,67 x 10^{-7} kPa y el umbral de masa inferior fue de 875 Da. Se generaron espectros a partir de la suma de 100-200 pulsos de láser y se adquirieron con un digitalizador de 500 MHz. Se usó oligosacárido $Man_5GlcNAc_2$ como un patrón de peso molecular externo. Todos los espectros se generaron con el instrumento en el modo de ion positivo.

Ejemplo 13: recorte *in vivo* por alfa-1,2-manosidasa

Asegurando que las nuevas cepas obtenidas por ingeniería genética del **Ejemplo 11** produjeron efectivamente la estructura de $Man_5GlcNAc_2$ deseada *in vivo*, los sobrenadantes celulares se ensayaron con respecto a la actividad manosidasa (véanse las **Figuras 7-9**). Para cada construcción/cepa huésped descrita más adelante, se realizó HPLC a 30 ºC con una columna de 4,0 mm x 250 mm de resina Econosil-NH_2 de Altech (Avondale, PA, Estados Unidos) (5 μm) a un caudal de 1,0 ml/min durante 40 minutos. En las **Figuras 7 y 8**, se mostró que se producía la degradación de $Man_9GlcNAc_2$ convencional [b], dando como resultado un pico que se correlaciona con $Man_8GlcNAc_2$. En la **Figura 7**, el patrón de $Man_9GlcNAc_2$ [b] eluyó a 24,61 minutos y $Man_5GlcNAc_2$ [a] eluyó a 18,59 minutos. En la **Figura 8**, $Man_9GlcNAc_2$ eluyó a 21,37 minutos y $Man_5GlcNAc_2$ a 15,67 minutos. En la **Figura 9**, se mostró que el $Man_8GlcNAc_2$ [b] convencional eluía a 20,88 minutos.

Se dejaron proliferar células de *P. pastoris* que comprendían el plásmido **pFB8** *(SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) a 30 ºC en BMGY hasta una DO_{600} de aproximadamente 10. Las células se recogieron por centrifugación y se transfirieron a BMMY induciendo la producción de K3 (kringle 3 de plasminógeno humano) sometida al control de un promotor de *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación produciendo un sobrenadante esencialmente transparente. Se retiró una alícuota del sobrenadante para ensayos de manosidasa y el resto se usó para la recuperación de K3 soluble secretado. Una etapa de purificación sencilla usando cromatografía de sepharose-CM y un gradiente de elución de NaAc 25 mM, pH 5,0 a NaAc 25 mM, pH 5, NaCl 0,1 M, dio como resultado un K3 puro al 95 % que eluía entre NaCl 300-500 mM. El análisis de *N*-glucano de los glucanos derivados de K3 se muestra en la **Figura 6F**. La alícuota anteriormente retirada del sobrenadante se ensayó adicionalmente con respecto a la presencia de actividad manosidasa secretada. Se añadió un patrón disponible en el mercado de oligomanosa 9 de tipo ligada a N marcada con 2-aminobenzamida (Man9-2-AB) (Glyko, Novato, CA) a: BMMY (**Figura 7A**), el sobrenadante de la alícuota anterior (**Figura 7B**) y BMMY que contenía 10 ng de α-1,2-manosidasa 75 mU/ml de *Trichoderma reesei* (obtenida de Contreras y cols., WO02/00856 A2) (**Figura 7C**). Después de incubación durante 24 horas a temperatura ambiente, las muestras se analizaron mediante HPLC de amino sílice determinando el grado de reducción de manosidasa.

Se cultivaron y ensayaron de forma similar células de *P. pastoris* que comprendían el plásmido **pGC5** *(MNS1*(m) de *Saccharomyces*/manosidasa IB Δ99 de ratón). Las células se cultivaron a temperatura ambiente en BMGY hasta una DO_{600} de aproximadamente 10. Las células se recogieron por centrifugación y se transfirieron a BMMY induciendo la producción de K3 sometida al control de un promotor de *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación produciendo un sobrenadante esencialmente transparente. Se retiró una alícuota del sobrenadante para ensayos de manosidasa y el resto se usó para la recuperación de K3 soluble secretado. Una sola etapa de purificación usando cromatografía de CM-sefarosa y un gradiente de elución de NaAc 25 mM, pH 5,0 a NaAc 25 mM, pH 5,0, NaCl 1 M, dio como resultado un K3 puro al 95 % que eluía entre NaCl 300-500 mM. El análisis de *N*-glucanos de los glucanos derivados de K3 se muestra en la **Figura 5D.** La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente con respecto a la presencia de la actividad manosidasa secretada como se muestra en la **Figura 8B**. Se añadió un patrón disponible en el mercado de Man9-2-AB (Glyko, Novato, CA) a: BMMY (**Figura 8A**), el sobrenadante de la alícuota anterior (**Figura 8B**) y BMMY que contenía 10 ng de α-1,2-manosidasa 75 mU/ml de *Trichoderma reesei* (obtenida de Contreras y cols., WO 02/00856 A2) (**Figura 8C**). Después de la incubación durante 24 horas a temperatura ambiente, las muestras se analizaron mediante HPLC de amino sílice determinando el grado de reducción de manosidasa.

Se usó Man9-2-AB como sustrato y resulta evidente que después de 24 horas de incubación, la actividad manosidasa estaba prácticamente ausente en el sobrenadante del producto de digestión de la cepa pFB8 (*SEC12* de *Saccharomyces* (m)/manosidasa IA Δ187 de ratón) (**Figura 7B**) y del producto de digestión de la cepa pGC5 (*MNS1*(m) de *Saccharomyces*/manosidasa IB Δ99 de ratón) (**Figura 8B**), mientras que el control positivo (α-1,2-manosidasa purificada de *T. reesei* obtenida de Contreras) conduce a la conversión completa de $Man_9GlcNAc_2$ en $Man_5GlcNAc_2$ en las mismas condiciones, como se muestra en **Figuras 7C y 8C.** Estos son datos concluyentes que muestran reducción de manosidasa *in vivo* en la cepa pGC5 de *P. pastoris*; y en la cepa pFB8, lo que es claramente diferente de lo que se ha indicado hasta la fecha (Contreras y cols., WO 02/00856 A2).

La **Figura 9** confirma adicionalmente la localización y la actividad de la enzima manosidasa. Se dejó proliferar ***P. pastoris* que comprendía pBC18-5** (*VAN1*(s) de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*) a temperatura ambiente en BMGY hasta una DO_{600} de aproximadamente 10. Las células se recogieron por centrifugación y se transfirieron a BMMY induciendo la producción de K3 sometida al control de un promotor de *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación produciendo un sobrenadante esencialmente transparente. Se retiró una alícuota del sobrenadante para ensayos de manosidasa y el resto se usó para la recuperación de K3 soluble secretado. Una sola etapa de purificación usando cromatografía de CM-sefarosa y un gradiente de elución de NaAc 25 mM, pH 5,0 a NaAc 25 mM, pH 5,0, NaCl 1 M, dio como resultado K3 puro al 95 % que eluía entre NaCl 300-500 mM. En la **Figura 5E** se muestra análisis de *N*-glucanos de los glucanos derivados de K3. La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente con respecto a la presencia de la actividad manosidasa secretada como se muestra en la **Figura 9B**. Se añadió un patrón disponible en el mercado de Man8-2-AB (Glyko, Novato, CA) a: BMMY (**Figura 9A**), el sobrenadante de la alícuota anterior pBC18-5 *(VAN1(s)* de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*) (**Figura 9B**) y medio que contenía BMMY de una construcción

de fusión diferente pDD28-3 (*MNN10*(m) de *Saccharomyces* (de SwissProt 50108)/manosidasa IB Δ99 de *H. sapiens*) (**Figura 9C**). Después de la incubación durante 24 horas a temperatura ambiente, las muestras se analizaron mediante HPLC de amino sílice determinando el alcance de reducción de manosidasa. La **Figura 9B** demuestra actividad manosidasa intracelular en comparación con una construcción de fusión pDD28-3 (*MNN10*(m) de *Saccharomyces*/manosidasa IB Δ99 de *H. sapiens*) que muestra un resultado negativo (**Figura 9C**).

Ejemplo 14: Ensayo de pH óptimo de α-1,2-manosidasa modificada por ingeniería genética

Se dejaron proliferar células de *P. pastoris* que comprendían el plásmido pBB27-2 (*MNN10* (s) de *Saccharomyces* (de SwissProt 50108)/manosidasa IB Δ31 de *C. elegans*) a temperatura ambiente en BMGY a una DO_{600} de aproximadamente 17. Aproximadamente 80 μl de estas células se inocularon en 600 μl de BMGY y se dejaron proliferar durante una noche. Posteriormente, las células se recogieron por centrifugación y se transfirieron a BMMY induciendo la producción de K3 (kringle 3 de plasminógeno humano) sometida al control de un promotor de *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación produciendo un sobrenadante esencialmente transparente (pH 6,43). El sobrenadante se retiró para ensayos de pH óptimo de manosidasa. Se añadió $MangGlcNAc_2$ marcado con fluorescencia (0,5 mg) a 20 μl de sobrenadante ajustado a diversos valores de pH (**Figura 11**) y se incubó durante 8 horas a temperatura ambiente. Después de la incubación, la muestra se analizó mediante HPLC usando una columna de sílice unida a amino, perla de 5 micrómetros, Econosil NH2 4,6 x 250 mm (Altech, Avondale, PA). El caudal fue de 1,0 ml/min durante 40 minutos y la columna se mantuvo a 30 ºC. Después de eluir de forma isocrática (A al 68 %:B al 32 %) durante 3 minutos, se empleó un gradiente de disolvente lineal (A al 68 %:B al 32 % a A al 40 %:B al 60 %) durante 27 minutos eluyendo los glucanos (18). Disolvente A (acetonitrilo) y disolvente B (formato de amonio, 50 mM, pH 4,5). La columna se equilibró con disolvente (A al 68 %:B al 32 %) durante 20 minutos entre funcionamientos.

Ejemplo 15: modificación por ingeniería genética de *P. pastoris* para producir *N*-glucanos con la estructura $GlcNAcMan_5GlcNAc_2$

Se requiere actividad GlcNAc Transferasa I para la maduración de *N*-glucanos complejos e híbridos (Patente de Estados Unidos N.º: 5.834.251). $Man_5GlcNAc_2$ solamente puede reducirse por la manosidasa II, una etapa necesaria en la formación de glucoformas humanas, después de la adición de *N*-acetilglucosamina al resto de α-1,3-manosa terminal del tallo de trimanosa mediante la GlcNAc Transferasa I (Schachter, 1991 Glycobiology 1 (5): 453-461). En consecuencia, se preparó una biblioteca de ADN combinatoria que incluía fragmentos de ADN que codificaban dominios catalíticos dirigidos de forma adecuada de genes de GlcNAc Transferasa I de *C. elegans* y *Homo sapiens*; y secuencias de localización de *GLS, MNS, SEC, MNN9,VAN1, ANP1, HOC1, MNN10, MNN11, MNT1, KTR1, KTR2, MNN2, MNN5, YURI, MNN1 y MNN6* de α-1,2-manosiltransferasas potenciales de *S. cerevisiae* y *P. pastoris* basándose en la homología de D2, D9 y J3 de *S. cerevisiae*, que son homólogos de *KTR*. La Tabla 10 incluye, pero sin limitación, secuencias peptídicas de dirección tales como *SEC* y *OCH1*, de GnTI *P. pastoris* y *K. lactis*, (véanse **la Tabla 6 y la Tabla 10**)

Tabla 10 Una biblioteca combinatoria representativa de secuencias peptídicas de dirección/dominio catalítico para UDP-*N*-acetilglucosaminil transferasa I (GnTI)

		Péptido de Dirección				
		OCHI(s)	***OCHI(m)***	**OCHT(I)**	***MNN9(s)***	***MNN9(m)***
Dominio catalítico	**Humano, GnTI, Δ38**	**PB105**	**PB106**	**PB107**	**PB104**	**N/A**
	Humano, GnTI, Δ86	**NB12**	**NB13**	**NB14**	**NB15**	**NB**
	***C. elegans,* GnTI, Δ88**	**OA12**	**OA13**	**OA14**	**OA15**	**OA16**
	***C. elegans.* GnTI, Δ35**	**PA12**	**PA13**	**PA14**	**PA15**	**PA16**
	***C. elegans,* GnTI, Δ63**	**PB12**	**PB13**	**PB14**	**PB15**	**PB16**
	***X. Leavis,* GnTI, Δ33**	**QA12**	**QA13**	**QA14**	**OA15**	**QA16**
	***X. Leavis,* GnTI, Δ103**	**QB12**	**OB 13**	**QB14**	**QB15**	**QB16**

Se seleccionaron secuencias peptídicas de dirección de *OCHI* en *P. pastoris* (larga, media y corta) (véase el **Ejemplo 11**) y *MNN9* (SwissProt P39107) en *S. cerevisiae* corta y media. Se seleccionaron dominios catalíticos a partir de GnTI humana con una deleción *N*-terminal de 38 y 86 aminoácidos, GnTI de *C. elegans* (gly-12) con una deleción de 35 y 63 aminoácidos, así como GnTI de *C. elegans* (gly-14) con una deleción N-terminal de 88 aminoácidos y GnTI de *X. leavis* con una deleción *N*-terminal de 33 y 103 aminoácidos, respectivamente.

Una parte del gen que codifica *N*-acetilglucosaminil Transferasa I humana (MGATI, N.º de acceso NM002406), sin los primeros 154 pb, se amplificó por PCR usando oligonucleótidos 5'-TGGCAGGCGCGCCTCAGTCAGCGCTCTCG-3' (SEC ID N.º: 32) y 5'-AGGTTAATTA AGTGCTAATTCCAGCTAGG-3' (SEC ID N.º: 33) y vector pHG4.5 (ATCC N.º: 79003) como molde. El producto de PCR resultante se clonó en pCR2.1-TOPO y se confirmó la secuencia correcta. Después de la digestión con *AscI* y *PacI*, la GnTI trucada se insertó en el plásmido pJN346 creando pNA. Después de la digestión de pJN271 con *NotI* y *AscI*, el inserto de 120 pb se ligó a pNA generando una fusión en fase del dominio transmembrana de *MNN9* con la GnTI, creando pNA15.

El organismo huésped es una cepa de *P. pastoris* que es deficiente en hipermanosilación (por ejemplo, un mutante *och1*), proporciona el sustrato UDP-GlcNAc en el Golgi y/o RE (es decir contiene un transportador de UDP-GlcNAc funcional) y proporciona *N*-glucanos de la estructura de $Man_5GlcNAc_2$ en el Golgi y/o RE (por ejemplo, pFB8 de *P. pastoris* (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) de anteriormente). En primer lugar, se transformó pFB8 de *P. pastoris* con pPB103 que contenía el gen *MNN2*-2 de *Kluyveromyces lactis* (Genbank AN AF106080) (que codifica el transportador de UDP-GlcNAc) clonado en los sitios *BamHI* y *BglII* del plásmido pBLADE-SX (Cereghino y cols. Gene 263 (2001) 159-169). Después, la biblioteca de ADN combinatoria anteriormente mencionada que codifica una combinación de genes de localización/GnTI exógenos o endógenos se usó para la transformación y se seleccionaron y analizaron las colonias con respecto a la presencia de la construcción de GnTI por PCR de colonias. La eficacia de transformación e integración de los autores de la invención estuvo generalmente por encima del 80 % y puede omitirse la exploración por PCR una vez que se han establecido parámetros de transformación robustos.

Purificación de proteínas

K3 se purificó a partir del medio por cromatografía de afinidad de Ni usando un formato de 96 pocillos en un robot de laboratorio Beckman BioMek 2000. La purificación robótica es una adaptación del protocolo proporcionado por Novagen para su resina HisBind. Puede realizarse otro procedimiento de exploración usando un anticuerpo de unión a GlenNAc terminal específico, o una lectina tal como la lectina GSII de *Griffonia simplificolia*, que se une a GlcNAc terminal (EY Laboratories, San Mateo, CA). Estas exploraciones pueden automatizarse usando lectinas o anticuerpos que se han modificado con marcadores fluorescentes tales como FTTC o analizarse por MALDI-TOF.

K3 secretado puede purificarse por cromatografía de afinidad de Ni, puede cuantificarse y pueden unirse cantidades iguales de proteína a una placa de 96 pocillos de alta unión a proteínas. Después de bloquear con BSA, las placas pueden sondarse con una lectina de GSII-FACS y explorarse con respecto a la respuesta fluorescente máxima. Un procedimiento preferido detectando las proteínas glucosiladas anteriores implica la exploración por espectrometría de masas MALDI-TOF después de la purificación de afinidad de K3 secretado del sobrenadante de transformantes cultivados en 96 pocillos. Se seleccionaron colonias transformadas y se dejaron proliferar a una DO_{600} de 10 en una placa de 96 pocillos de 2 ml en BMGY a 30 ºC. Las células se recogieron por centrifugación, se lavaron en BMMY y se resuspendieron en 250 µl de BMMY. Después de 24 horas de inducción, las células se retiraron por centrifugación, el sobrenadante se recuperó y K3 se purificó del sobrenadante por cromatografía de afinidad de Ni. Los *N*-glucanos se liberaron y analizaron por espectrometría de masas con extracción retardada MALDI-TOF como se describe en el presente documento.

En resumen, los procedimientos de la invención producen cepas de *P. pastoris* que producen $GlcNAcMan_5GlcNAc_2$ en alto rendimiento, como se muestra en la **Figura 10B.** Al menos el 60 % de los N-glucanos son $GlcNAcMan_5GlcNAc_2$. Hasta la fecha, no existe ningún informe que describa la formación de $GlcNAcMan_5GlcNAc_2$ en glicoproteínas solubles secretadas en ninguna levadura. Los resultados presentados en el presente documento muestran que la adición del transportador de UDP-GlcNAc junto con la actividad de GnTI produce una estructura predominante $GlcNAcMan_5GleNAc_2$, que se confirma por el pico a 1457 (m/z) (**Figura 10B**).

Construcción de cepa PBP-3:

La cepa de *P. pastoris* que expresa K3, (Δ*och1*, arg-, ade-, his-), se transformó sucesivamente con los siguientes vectores. En primer lugar, **pFB8** (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) se introdujo por transformación en la cepa de *P. pastoris* por electroporación. En segundo lugar, se introdujo por transformación **pPB103** que contenía el gen de *MNN2*-2 de *Kluyveromyces lactis* (Genbank AN AF106080) (que codifica el transportador de UDP-GlcNAc) clonado en el plásmido pBLADE-SX (Cereghino y cols., Gene 263 (2001) 159-169) digerido con enzimas *Bam*HI y *BglII* en la cepa de *P. pastoris*. En tercer lugar, se introdujo por transformación **pPB104** que contenía el gen que codificaba *MNN9*(s) de *Saccharomyces*/GnTI Δ38 humana clonado como fragmento *Not*I-*Pac*I en pJN336 en la cepa de *P. pastoris*.

Ejemplo 16: modificación por ingeniería genética de células *K. lactis* para producir *N*-glucanos con la estructura de $Man_5GlcNAc_2$

Identificación e interrupción del gen *OCH1* de *K. lactis*

El gen *OCH1* de la levadura *S. cerevisiae* en proceso de gemación codifica una 1,6-manosiltransferasa que es

responsable de la primera adición de manosa localizada en el Golgi a la estructura de *N*-glucano de $Man_5GlcNAc_2$ en proteínas secretadas (Nakayama y cols., J Biol Chem.; 268 (35): 26338-45 (15 diciembre, 1993)). Esta transferencia de manosa se reconoce generalmente como la etapa clave inicial en la polimanosilación específica fúngica de estructuras de *N*-glucanos (Nakanishi-Shindo y cols., 1993; Nakayama y cols., 1992; Morin-Ganet y cols., Traffic 1 (1): 56-68. (enero 2000)). La deleción de este gen en *S. cerevisiae* da como resultado una estructura de N-glucano significativamente más corta que no incluye esta polimanosilación típica o un defecto de la proliferación a temperaturas elevadas (Nakayama y cols., EMBO J.;(7): 2511-9 (julio 1992)).

La secuencia *Ochlp* de *S. cerevisiae* se alineó con homólogos conocidos de *Candida albicans* (N.º de acceso de Genbank AAL49987) y *P. pastoris* (B. K. Choi y cols. en preparación) junto con las proteínas *Hocl* de *S. cerevisiae* (Neiman y cols., Genetics, 145 (3): 637-45 (Mar 1997) y *K. lactis* (base de datos PENDANT EST) que son manosiltransferasas relacionadas pero distintas. Se usaron regiones de alta homología que estaban en común entre homólogos de *Och1p* pero distintas de los homólogos de *Hoc1p* diseñando pares de cebadores degenerados que se dirigieron contra ADN genómico de la cepa MG1/2 de *K. lactis* (Bianchi y cols., Current Genetics 12, 185-192 (1987)). La amplificación por PCR con los cebadores RCD33 (CCAGAA-GAATTCAATTYTGYCARTGG) (SEC ID N.º: 34) y RCD34 (CAGTGAAAATACCTGGNCCNGTCCA) (SEC ID N.º: 35) dio como resultado un producto de 302 pb que se clonó y secuenció y se mostró que la traducción predicha tenía un alto grado de homología con proteínas *Och1* (>55 % con *Och1p* de *S. cerevisiae*).

El producto de PCR de 302 pb se usó sondeando una transferencia de bandas de Southern de ADN genómico de la cepa de *K. lactis* (MG1/2) con alta rigurosidad (Sambrook y cols., 1989). Se observó hibridación en un patrón coherente con un gen individual indicando que este segmento de 302 pb corresponde a una parte del genoma de *K. lactis* y *K. lactis* (*KlOCH1*) contiene una sola copia del gen. Clonando el gen de *KlOCH1* completo, se usó transferencia de bandas de Southern mapeando el locus genómico. En consecuencia, se clonó un fragmento *Bam*HI/*Pst*I de 5,2 kb por digestión del ADN genómico y ligación de los fragmentos en el intervalo de 5,2 kb en pUC19 (New England Biolabs, Beverly, MA) creando una biblioteca subgenómica de *K. lactis*. Esta biblioteca subgenómica se usó transformando *E. coli* y se ensayaron varios cientos de clones por PCR de colonias usando RCD 33/34. El clon de 5,2 kb que contenía el gen de *KlOCH1* predicho se secuenció y se observó una fase abierta de lectura de 1362 pb que codificaba una proteína predicha que es idéntica en un 46,5 % al gen de *OCH1* de *S. cerevisiae*. La secuencia de 5,2 kb se usó fabricando cebadores para la construcción de un alelo de deleción *och1::KAN*^{R} usando un procedimiento de solapamiento de PCR (Davidson y cols., *Microbiology.* 148(Pt 8): 2607-15. Agosto de 2002). Este alelo de deleción se introdujo por transformación en dos cepas de *K. lactis* y se seleccionaron colonias resistentes a G418. Estas colonias se exploraron tanto por PCR como por sensibilidad a temperatura obteniendo una cepa con deleción para la ORF de *OCH1*. Los resultados del experimento muestran cepas que revelan un patrón de PCR mutante, que se caracterizaron por análisis de proliferación a diversas temperaturas y análisis de carbohidratos de *N*-glucanos de proteínas secretadas y de la pared celular después de digestión con PNGasa. La mutación *och1* confirió una sensibilidad a temperatura que permitió que las cepas proliferaran a 30 ºC pero no a 35 ºC. La **Figura 12A** muestra un análisis de MALDI-TOF de una cepa de *K. lactis* de tipo silvestre que producía N-glucanos de $Man_8GlcNAc_2$ [c] y superiores.

Identificación, clonación e interrupción del gen *MNN1* de *K. lactis*

MNN1 de *S. cerevisiae* es el gen estructural para la α-1,3-manosiltransferasa del Golgi. El producto de *MNN1* es una proteína de membrana de 762 aminoácidos de tipo II (Yip y cols., Proc Natl Acad Sci EE.UU. 91 (7): 2723-7 (1994)). Los oligosacáridos tanto ligados a *N* como ligados a *O* aislados de mutantes *mnn1* carecen de enlaces α-1,3-manosa (Raschke y cols., J Biol Chem., 248 (13): 4660-6. (10 de julio, 1973).

La secuencia de *Mnnlp* de *S. cerevisiae* se usó buscando las secuencias genómicas traducidas de *K. lactis* (PEDANT). Se identificó una secuencia de ADN de 405 pb que codificaba un fragmento proteico potencial de similitud significativa con *Mnn1p*. Un segmento interno de esta secuencia se amplificó por PCR posteriormente con los cebadores KMN1 (TGCCATCTTTTAGGTC-CAGGCCCGTTC) (SEC ID N.º: 36) y KMN2 (GATCCCACGACGCATCGTATTTCTTTC), (SEC ID N.º: 37) y se usó sondeando una transferencia de Southern de ADN genómico de la cepa de *K. lactis* (MG1/2). Basándose en los datos de hibridación de Southern se clonó un fragmento *BamHI-PstI* de 4,2 kb generando una biblioteca seleccionada por tamaño como se describe en el presente documento. Se identificó un solo clon que contenía el gen *MMN1* de *K. lactis* por PCR de colonias completas usando cebadores KMN1 (SEC ID N.º: 36) y KMN2 (SEC ID N.º: 37) y se secuenció. Dentro de este clon se identificó una ORF de 2241 pb que codificaba una proteína predicha que era idéntica al 34 % al gen de *MNN1* de *S. cerevisiae*. Se diseñaron cebadores para la construcción de un alelo de deleción de *mnnl::NAT*^{R} usando el procedimiento de solapamiento de PCR (Davidson y cols. 2002).

Este alelo de interrupción se usó transformando una cepa de *K. lactis* por electroporación, se seleccionaron transformantes resistentes a Nurseotricina y se amplificaron por PCR para la inserción homóloga del alelo de interrupción. Las cepas que revelan un patrón de PCR mutante pueden someterse a análisis de carbohidratos de *N*-glucano de un gen indicador conocido.

La **Figura 12B** representa los *N*-glucanos de la cepa de deleción *och1mnn1* de *K. lactis* observada después de la digestión con PNGasa mediante MALDI-TOF como se describe en el presente documento. El pico predominante a

1908 (m/z) indicado como **[d]** es coherente con la masa de $Man_9GlcNAc_2$.

Ejemplo 17: modificación por ingeniería genética de células vegetales para expresar GlcNAc transferasas o galactosiltransferasas

Se requiere GlcNAc transferasa IV para la adición de β1,4 GlcNAc al resto de α-1,6 manosa y los residuos de α-1,3 manosa en N-glucanos complejos en seres humanos. Hasta la fecha no se ha detectado GlcNAc transferasa IV en o aislada de plantas. Por lo tanto, una planta transgénica que sea capaz de añadir N-glucanos de tipo humano debe modificarse por ingeniería genética expresando GlcNAc transferasa IV. Por lo tanto, la célula huésped vegetal o planta transgénica también debe localizar una GlcNAc transferasa IV expresada en el compartimento intracelular correcto en el huésped de modo que la enzima pueda añadir la β1,4 GlcNAc a los restos de manosa apropiados.

Hay algunas pruebas de que las glucosiltransferasas de mamíferos y plantas tienen señales de dirección similares. Por ejemplo, se ha mostrado que una α-2,6-sialiltransferasa de rata de longitud completa se sitúa correctamente en la red del trans-Golgi en arabidopsis transgénica aunque no necesariamente está activa (Wee E y cols. Plant Cell octubre de 1998; 10 (10): 1759-68). También se mostró que una construcción de fusión que tenía cincuenta y dos aminoácidos N-terminales de α-2,6-sialiltransferasa fusionados con una proteína indicadora verde fluorescente (GFP) se localizaba correctamente en el Golgi de la planta (Boevink y cols. Plant J agosto 1998; 15 (3): 441-7). Dos proteínas de mamífero, TGN30 y furina y AtELP, una proteína de membrana integral de arabidopsis (Sanderfoot y cols. Proc Natl Acad Sci EE.UU 18 de agosto de 1998; 95 (17): 9920-5), que se localizan en la red del trans-Golgi, contienen cada una un motivo tetrapeptídico de tirosina que las dirige al Golgi, probablemente mediante un mecanismo de reciclaje a través de la membrana plasmática. Aunque los mamíferos y plantas parecen compartir algunos mecanismos comunes relacionados con la dirección de proteínas, las glucosilasas exógenas pueden no dirigirse correctamente en una célula vegetal, sin embargo, la localización no necesariamente equivale a la actividad enzimática. Es por lo tanto esencial idear un medio dirigiendo correctamente en una célula vegetal estas enzimas y/u otras enzimas que participen en la formación de N-glucanos de tipo humano, complejos.

Las enzimas de glucosilación son proteínas de membrana integrales que residen en el retículo endoplásmico y el aparato de Golgi. Las señales de dirección y localización normalmente están contenidas en los dominios citoplasmáticos y/o transmembrana y en algunos casos están contenidas en algunos aminoácidos del lumen. Por ejemplo, se requieren cincuenta y dos aminoácidos que componen el dominio transmembrana, nueve aminoácidos citoplasmáticos y veintiséis aminoácidos del lumen de la α-2,6-sialiltransferasa dirigiendo GFP a la red del trans-Golgi (Boevink y cols. Plant J agosto de 1998; 15 (3): 441-7).

Por lo tanto, se genera una biblioteca de secuencias que codifican péptidos señal de dirección celular que comprenden solamente los dominios citoplasmático y transmembrana o los dominios citoplasmático, transmembrana y del lumen de proteínas específicas del retículo endoplásmico y Golgi, como se describe en el **Ejemplo 11.** Las secuencias peptídicas de dirección pueden seleccionarse de proteínas vegetales, de levadura o animales residentes en el RE y Golgi. Una proteína relacionada con la glucosilación, por ejemplo, una enzima (o dominio catalítico de la misma) tal como una glucosilasa o enzima integral de membrana puede fusionarse en fase con la biblioteca de secuencias peptídicas de dirección e introducirse en plantas **(Figura 13)**. Las secuencias peptídicas de dirección de plantas pueden ser más eficaces en la localización de las enzimas quiméricas en el RE y Golgi, aunque también pueden ser eficaces las secuencias peptídicas de dirección de hongos y mamíferos. Por ejemplo, se ha mostrado que los 77 aminoácidos N-terminales de la *N*-acetilglucosaminil-transferasa I del tabaco dirigen correctamente una proteína indicadora al Golgi (Essl D. y cols., FEBS Lett 18 junio de 1999; 453 (1-2): 169-73). En una realización, uno o más fragmentos N-terminales que comprenden estos 77 aminoácidos (o subconjuntos de estos aminoácidos) se fusionan con uno o más fragmentos que contienen un dominio catalítico de la GlcNAc transferasa IV. Al menos una proteína de fusión resultante sitúa correctamente una GlcNAc transferasa IV funcional en el aparato de Golgi en una célula vegetal, como se demuestra supervisando el estado de glucosilación de una glucoproteína indicadora residente o introducida en la célula huésped vegetal usando técnicas descritas en el presente documento.

Otra enzima vegetal que se ha mostrado que se localiza en el Golgi es la GlcNAc transferasa II de *Arabidopsis* (Strasser R y cols., Glycoconj J diciembre de 1999; 16 (12): 787-91). Por lo tanto, en otra realización, uno o más fragmentos diferentes del péptido de dirección de GlcNAc transferasa II de arabidopsis se fusionan con un dominio catalítico de GlcNAc transferasa IV y las construcciones de fusión se producen y ensayan como se ha descrito anteriormente. La β1,2-xilosiltransferasa específica de plantas de *Arabidopsis thaliana* es otra proteína que se localiza en el Golgi y su localización y retención en el Golgi depende de sus secuencias citoplasmática y transmembrana (Dimberger y cols., Plant Mol Biol septiembre de 2002; 50 (2): 273-81). Por lo tanto, en otra realización, se fusionan uno o más fragmentos que comprenden las secuencias citoplasmática y transmembrana de β1,2-xilosiltransferasa con uno o más fragmentos que comprenden un dominio catalítico de GlcNAc transferasa IV y las construcciones de fusión resultantes se usan transformando células vegetales y se ensayan con respecto a su capacidad produciendo un N-glucano de tipo humano y modulando de otro modo la glucosilación en la célula huésped vegetal.

Debido a que la GlcNAc transferasa IV o Galactosiltransferasa de un organismo pueden actuar más eficazmente en un huésped vegetal específico que la de otro organismo, se fusionan fragmentos que comprenden GlcNAc transferasa IV (o dominios catalíticos) de diversos organismos eucariotas en fase con la biblioteca de secuencias

peptídicas de dirección a retículo endoplásmico (RE) y Golgi y después se introducen en plantas. El uso de una biblioteca de ácidos nucleicos que codifican dominios enzimáticos aislados o derivados de diferentes especies aumenta las probabilidades de glucosilación eficaz, además de la localización correcta y la glucosilación mediante GlcNAc transferasa IV.

Los procedimientos y bibliotecas de ácidos nucleicos combinatorias de la invención pueden usarse introduciendo o localizar, secuencialmente o en masa, múltiples enzimas requeridas glucosilando proteínas en una célula vegetal con N-glucanos de tipo humano. Puesto que diferentes especies vegetales pueden requerir diferentes condiciones de crecimiento, los protocolos para transformación pueden variar dependiendo de las especies que se transformen (Potrykus, "Gene transfer methods for plants and cell cultures." Ciba Found Symp 1990; 154: 198-208; análisis 208-12). Los procedimientos usados habitualmente generando plantas transgénicas incluyen transformación mediada por Agrobacterium, bombardeo de partículas (Sanford, J.C. y cols., Biolistic plant transformation. Physiol. Plant. 1990, 79: 206-209) y electroporación.

Procedimiento de Agrobacterium

Los dominios catalíticos de la GlcNAc transferasa IV se fusionan en fase con múltiples secuencias peptídicas de dirección diferentes que se sabe que dirigen proteínas al RE y Golgi en plantas. Cada una de estas construcciones de fusión se introduce sometida al control de promotores expresados de forma ubicua tales como el promotor 35S de CaMV, promotores de ubiquitina o actina, promotores específicos de tejido o promotores inducibles. También se usa una región terminadora específica de planta. Este casete (Promotor::péptido de dirección-GlcNAc transferasa IV::terminador) se clona en un vector adecuado para la transformación mediada por Agrobacterium (**Figura 13**). El vector también contiene un marcador seleccionable que permite seleccionar plantas transformadas. Los marcadores seleccionables habituales usados incluyen los que dan como resultado resistencia a kanamicina, higromicina y basta. La construcción se introduce en Agrobacterium mediante procedimientos de transformación bien establecidos, que están disponibles en la técnica. Se genera de este modo una biblioteca de Agrobacterium de GlcNAc transferasa IV dirigida al Golgi.

Puede transformarse tejido embrionario y meristemático y se pueden regenerar plantas transgénicas. Transformando tejidos, los explantes de tejido (estos podrían ser plúmula y radicales de semillas germinadas) se empapan en primer lugar y se revisten con un inóculo de *Agrobacterium*. Después se cultivan en placas que contienen el inóculo formando una masa indiferenciada de células denominada el callo. Las células vegetales transformadas se seleccionan añadiendo al medio la kanamicina, higromicina o basta relevante (dependiendo del marcador seleccionable usado en la construcción). Las células vegetales transformadas bien pueden dejarse proliferar en cultivo y permanecer indiferenciadas o bien se tratan con medio de regeneración de brotes y elongación de brotes. Los explantes que se diferencian se transfieren a medio de enraizamiento generando plantas transgénicas. Algunas plantas tales como *Arabidopsis* pueden transformarse por inmersión de las flores en una solución de *Agrobacterium*. Las semillas de las plantas transformadas germinan en placas que contienen la selección de herbicida o antibióticos relevante. Las plantas transgénicas son las que crecen en los medios de selección. Las plantas transgénicas se rastrean después en busca de las que tienen proteínas glucosiladas de forma apropiada (es decir, las que tienen N-glucanos de tipo humano complejos) aislando glicoproteínas de extractos vegetales y analizando los patrones de glicoproteínas como se describe en otros sitios en el presente documento, por ejemplo, usando un anticuerpo específico o lectina. Aunque el procedimiento de *Agrobacterium* es económico y sencillo, se limita a ciertas especies de plantas. En consecuencia, las plantas que no pueden transformarse usando *Agrobacterium* pueden transformarse por balística o electroporación.

Procedimiento de bombardeo de partículas y electroporación

En comparación con la transformación mediada por Agrobacterium, estos procedimientos tienen una mayor tendencia a insertar múltiples copias del transgén en el genoma. Esto podría dar como resultado silenciamiento génico y cosupresión. Sin embargo, a diferencia de la transformación mediada por *Agrobacterium*, estos procedimientos no están limitados por especies y son, por lo tanto, útiles cuando no puede emplearse un procedimiento de *Agrobacterium* generando plantas transgénicas. En el procedimiento de bombardeo de partículas, las células vegetales cultivadas se bombardean con partículas de oro o wolframio muy pequeñas que se han revestido con ADN (Promotor::péptido de dirección-GlcNAc transferasa IV-terminador::marcador seleccionable) (**Figura 13**) (no se requiere rb e Ib), mientras que en el procedimiento de electroporación, las células vegetales en una solución de ADN (Promotor::péptido de dirección-GlcNAc transferasa IV-terminador::marcador seleccionable) se tratan con un pulso eléctrico que perfora la célula permitiendo que capte ADN. Las células se cultivan después y se dejan recuperar. Los transformantes estables se seleccionan cultivando y regenerando plantas en medio de selección apropiado.

Modificación por ingeniería genética de soja para expresar GlcNAc transferasa IV usando un sistema de transformación mediado por *Agrobacterium* de nódulos de los cotiledones de soja

Se genera una biblioteca de Agrobacterium de GlcNAc transferasa IV dirigida al Golgi como se ha descrito anteriormente. Se transforman explantes de soja con la biblioteca usando un protocolo descrito por Hinchee y cols. (BiolTechnology 1988. 6: 915). Se expresa una proteína indicadora con un marcador His, se purifica y después se

analiza. Las plantas transgénicas se ensayan con respecto a proteínas con los restos de α-1,6 manosa y α-1,3 manosa usando, por ejemplo, espectroscopía de masas.

Modificación por ingeniería genética de guisantes para expresar GlcNAc transferasa IV usando bombardeo de partículas

Se revisten partículas de wolframio u oro con una biblioteca de plásmido de GlcNAc transferasa IV y se usan como microproyectiles bombardeando callos derivados de tejido embrionario de guisantes como se ha descrito (Molnar y cols., Symposium on Recent Advances in Plant Biotechnology, 4-11 de Septiembre, 1999, Stara Lesna, República de Eslovaquia). Se expresa una proteína indicadora con un marcador His, se purifica y después se analiza. Las plantas transgénicas se ensayan con respecto a proteínas con los restos de α-1,6 manosa y α-1,3 manosa usando, por ejemplo, MALDI.

Modificación por ingeniería genética de plantas para expresar GlcNAc transferasa I

La GlcNAc transferasa I está implicada en la adición de GlcNAc al residuo de α-1,3 manosa terminal formando $Man_5GlcNAc_2$, una etapa esencial en la maduración de N-glucanos complejos. Aunque la GlcNAc transferasa I se ha aislado a partir de plantas y parece tener la misma función que su homólogo de mamífero, puede no ser la enzima más eficaz para glucosilación de proteínas de mamífero o exógenas y puede no encontrarse en todas las especies vegetales. Como la adición de GlcNAc al resto de α-1,3 manosa terminal es una etapa de control en la ruta de glucosilación de mamíferos, es ventajoso tener plantas transgénicas que puedan llevar a cabo esta etapa de forma eficaz. Creando plantas transgénicas que expresen GlcNAc transferasa I que pueda actuar eficazmente promoviendo la formación de N-glucanos complejos, se fusiona una biblioteca de GlcNAc transferasa I aislada o derivada de diversos organismos en fase con múltiples secuencias peptídicas de dirección al Golgi de plantas de acuerdo con los procedimientos descritos en el presente documento. La biblioteca combinatoria creada de este modo se introduce en una célula vegetal u organismo como se ha descrito anteriormente para la GlcNAc transferasa IV.

Modificación por ingeniería genética de maíz para expresar GlcNAc transferasa I usando bombardeo de partículas

Puede obtenerse maíz transgénico usando un protocolo similar al usado generando guisantes que expresan GlcNAc transferasa IV. Aquí se revisten partículas de wolframio u oro con biblioteca plasmídica de GlcNAc transferasa I y se usan bombardeando callos derivados de tejido embrionario de maíz, por ejemplo, usando un protocolo específico para la generación de maíz transgénico (Gordon-Kamm WJ y cols., Plant Cell julio de 1990; 2 (7): 603-618)). Las plantas transgénicas se ensayan con respecto a proteínas que tengan GlcNAc en el resto de α-1,3 manosa terminal, por ejemplo, usando anticuerpos específicos o ensayando la unión reducida de los N-glucanos con ciertas lectinas o usando MALDI-TOF.

Otras referencias útiles usando células huésped vegetales de acuerdo con la invención incluyen: Christou P. Plant Mol Biol septiembre de 1997; 35 (1-2): 197-203; Chowrira GM y cols. Mol Biotechnol febrero de 1995; 3 (1): 17-23; Dimberger y cols., Plant Mol Biol septiembre de 2002; 50 (2): 273-81; Frame BR y cols. Plant Physiol mayo de 2002; 129 (1): 13-22; Gomord V y cols., Biochimie junio de 1999; 81 (6): 607-18; Laursen CM y cols. Plant Mol Biol enero de 1994; 24 (1): 51-61; Orci L y cols. J Cell Biol 18 de septiembre de 2000; 150 (6): 1263-70; Newell CA. Mol Biotechnol septiembre de 2000; 16 (1): 53-65; Pawlowski Wp y cols. Mol Biotechnol agosto de 1996; 6 (1): 17-30; Schroeder HE y cols. Plant Physiol marzo de 1993; 101 (3); 751-757; Sorokin, AP y cols. Plant. Sci. 28 de julio 2000; 156 (2): 227-233; Strasser R y cols. Glycoconj J diciembre de 1999; 16 (12): 787-91; y Tomes DT y cols. Plant Mol Biol febrero de 1990; 14 (2): 261-8.

Modificación por ingeniería genética de células vegetales para producir β1,4-galactosiltransferasas

La β1,4-galactosiltransferasa es una glucosiltransferasa humana importante que está ausente en plantas. Lerouge P y cols. *Plant Mol Biol s*eptiembre de 1998; 38 (1-2): 31-48. En mamíferos, la β1,4-galactosiltransferasa se localiza en el Golgi y es responsable de la transferencia de restos de galactosa a la N-acetilglucosamina terminal de la $Man_3GlcNAc_2$ central de N-glucanos complejos. En plantas, el núcleo de $Man_3GlcNAc_2$ contiene residuos de β1,2-xilosa y α1,3-fucosa y carece de la β1,4-galactosa. Las modificaciones de xilosa y fucosa están implicadas en alergias y actúan como epítopos antigénicos y por lo tanto, no son modificaciones deseables de proteínas terapéuticas.

Las modificaciones de galactosa llevadas a cabo por la β1,4-galactosiltransferasa pueden ser importantes para el funcionamiento apropiado de las proteínas terapéuticas. En mamíferos, la β1,4-galactosiltransferasa actúa después de la N-acetilglucosaminiltransferasa I y la N-acetilglucosaminiltransferasa II y se ha mostrado que inicia la ramificación del N-glucano complejo. Lerouge P y cols. Plant Mol Biol septiembre de 1998; 38 (1-2): 31-48. Palacpac N y cols. Proc Natl Acad Sci EE.UU. 13 de abril de 1999; 96 (8): 4692-7. En células de tabaco, se ha mostrado que la expresión de β1,4-galactosiltransferasa humana da como resultado N-glucanos galactosilados con modificaciones de fucosa y xilosa reducidas. Bakker H y cols. Proc Natl Acad Sci EE.UU. 27 de febrero de 2001; 98 (5): 2899-904 Fujiyama K y cols. Biochem Biophys Res Commun 30 de noviembre de 2001; 289 (2): 553-7. Palacpac N y cols.

Proc Natl Acad Sci EE.UU. 13 de abril de 1999; 96 (8): 4692-7. En estos estudios, se clonó un fragmento de 1,2 kb de β1,4-galactosiltransferasa humana cadena abajo del promotor del virus del mosaico de la coliflor (35SCaMV), se introdujo en el vector binario pGA482 y finalmente en células de tabaco. Palacpac N y cols. Proc Natl Acad Sci EE.UU. 13 de abril de 1999; 96 (8): 4692-7.

Se transformaron células de tabaco usando el procedimiento de agrobacterium descrito por Rempel y cols. (Rempel, H. C. y cols. 1995. Transgenic Res. 4 (3): 199-207). También se ha descrito la transformación de células de tabaco (An, G 1985. Plant Physiol. 79: 568-570). La expresión de la β1,4-galactosiltransferasa sometida al 35SCaMV dio como resultado una expresión ubicua del gen en células de tabaco. Las células de tabaco que expresaban β1,4-galactosiltransferasa humana mostraron la presencia de N-glucanos galactosilados. (Palacpac N y cols. Proc Natl Acad Sci EE.UU. 13 de abril de 1999 13; 96 (8): 4692-7). Bakker y cols. mostraron que el cruce de plantas de tabaco que expresaban β1,4-galactosiltransferasa humana con plantas que expresaban la cadena pesada y ligera de un anticuerpo de ratón producía plantas en las que el anticuerpo mostraba un 30 % de galactosilación (Bakker H y cols. Proc Natl Acad Sci EE.UU. 27 de febrero de 2001; 98 (5): 2899-904).

Puede construirse una biblioteca de ADN combinatoria obteniendo la línea de β1,4-galactosiltransferasa para la adición de residuos de galactosa. La biblioteca de ADN combinatoria puede producir eficazmente líneas que son más eficaces en la adición de residuos de galactosa. Una vez que se ha obtenido una línea tal esta puede cruzarse fácilmente con líneas que expresan otras enzimas de glucosilación y con las que expresan proteínas terapéuticas produciendo proteínas terapéuticas con glucosilación de tipo humano. La línea final después puede dejarse proliferar como plantas y recogerse extrayendo proteínas o puede cultivarse como células vegetales en cultivos en suspensión produciendo proteínas en biorreactores. Expresando las proteínas terapéuticas usando la biblioteca de péptidos señal, es posible retener la proteína terapéutica dentro de las células o haberla secretado en el medio. Las células de tabaco que expresan β1,4-galactosiltransferasa secretan N-glucanos galactosilados (Ryo Misaki y cols. Glycobiology 17 de diciembre de 2002; 10: 1093). Aunque la isozima C de peroxidasa de rábano rusticano expresada en plantas de tabaco que expresaban β1,4-galactosiltransferasa contenía modificaciones de xilosa y fucosa, no pudo detectarse modificación de xilosa o fucosa en isozima C de peroxidasa de rábano rusticano expresada en células de tabaco que expresaban β1,4-galactosiltransferasa (células GT6). (Fujiyama Ky cols. Biochem Biophys Res Commun 30 de noviembre de 2001; 289 (2): 553-7). Esto indica que puede ser ventajoso expresar proteínas terapéuticas en líneas celulares en lugar de expresarlas en plantas completas.

Modificación por ingeniería genética de plantas para producir sialiltransferasa

En mamíferos, la sialiltransferasa es una enzima del trans-Golgi que añade residuos de ácido siálico terminales a polipéptidos glucosilados. Hasta la fecha, no se han detectado restos de ácido siálico terminales en plantas (Wee E y cols. Plant Cell octubre de 1998; 10 (10): 1759-68). Wee y cols. expresaron la α-2,6-sialiltransferasa de rata en arabidopsis transgénica y mostraron que la enzima se localizaba apropiadamente en el golgi y era funcional. Wee y cols. demostraron que las membranas aisladas de arabidopsis transgénica, cuando se incubaban con CMP-^{3}H-ácido siálico y aceptor de asialofetuína, daban como resultado la adición de residuos de ácido siálico, mientras que la membrana aislada de arabidopsis de tipo silvestre no lo hacía. Aunque expresar la α-2,6-sialiltransferasa de ratas en arabidopsis dio como resultado una enzima funcional que era capaz de incorporar restos de ácido siálico, fusionar las enzimas de mamífero α-2,3-sialiltransferasa y α-2,6-sialiltransferasa a una diversidad de péptidos en tránsito usando el enfoque de la biblioteca de la presente invención descrito anteriormente puede dar como resultado una sialilación más eficaz en otras especies vegetales. Wee E y cols. tuvieron que aislar membranas e incubarlas con CMP-^{3}H-ácido siálico y aceptor de asialofetuina ya que arabidopsis no tiene CMP-ácido siálico o su transportador. Superando esta etapa adicional y consiguiendo la adición de ácido siálico en la planta, la ruta biosintética de CMP-ácido siálico y el transportador de CMP-ácido siálico pueden coexpresarse en plantas transgénicas que expresan α-2,3-sialiltransferasa y α-2,6-sialiltransferasa. Como alternativa, el transportador de CMP-ácido siálico puede coexpresarse con α-2,3-sialiltransferasa y α-2,6-sialiltransferasa en células vegetales cultivadas en cultivo en suspensión y suministrarse al medio CMP-ácido siálico u otros precursores de CMP-ácido siálico.

Expresión de α-2,3-sialiltransferasa y α-2,6-sialiltransferasa en lemna

Como se describe en la Patente de Estados Unidos N.º: 6.040.498, la lemna (lenteja de agua) puede transformarse usando procedimientos tanto de agrobacterium como balísticos. Usando protocolos descritos en la patente, la lemna se transformará con una biblioteca de α-2,3-sialiltransferasa y/o α-2,6-sialiltransferasa dirigidas al golgi y una biblioteca de transportadores de CMP-ácido siálico de mamífero. Las plantas transgénicas pueden ensayarse con respecto a proteínas con restos de ácido siálico terminales.

Expresión de α-2,3-sialiltransferasa y α-2,6-sialiltransferasa en células de tabaco

También pueden introducirse alfa-2,3-sialiltransferasa y/o α-2,6-sialiltransferasa y/o una biblioteca de transportadores de CMP-ácido siálico de mamífero en células de tabaco cultivadas en cultivo en suspensión como se describe para las β1,4-galactosiltransferasas. Puede añadirse CNT-ácido siálico al medio. Tanto las células como el medio de cultivo (proteínas secretadas) pueden ensayarse con respecto a las proteínas con restos de ácido siálico terminales.

Ejemplo 18: modificación por ingeniería genética de células de insecto para producir glucosiltransferasas

Las células de insecto proporcionan otro mecanismo produciendo glicoproteínas, pero las glicoproteínas resultantes no son glicoproteínas de tipo humano complejas. Marz y cols. 1995 Glycoproteins, 29: 543-563; Jarvis 1997 The Baculoviruses 389-431. Otra característica de la presente invención es proporcionar enzimas en células de insectos, que se dirigen a los orgánulos en la ruta secretora. En una realización preferida, se dirigen enzimas tales como glucosiltransferasas, galactosiltransferasas y sialiltransferasas al RE, Golgi o la red del trans-Golgi en células de insectos lepidópteros (Sf9). Se ha mostrado expresión de β1,4-galactosiltransferasa de mamífero en células Sf9. Hollister y cols. Glycobiology. 1998 8 (5): 473-480. Estas enzimas se dirigen por medio de una proteína quimérica que comprende un péptido señal de dirección celular normalmente no asociado con la enzima. Las proteínas quiméricas se preparan construyendo una biblioteca de ácidos nucleicos que comprende secuencias de dirección como se describen en el presente documento y las enzimas de glucosilación. Habitualmente se usa expresión de baculovirus en células de insecto consiguiendo una transformación estable añadiendo glucosiltransferasas de mamífero en células de insecto. Hollister y cols. Glycobiology. 2001 11 (1): 1-9.

Tabla 11: Recursos de Secuencias de ADN y Proteínas

1. El Instituto de Bioinformática Europea (EBI) es un centro para la investigación y servicios de bioinformática-http: //www.ebi.ac.uk/
2. Base de datos Swissprot: http: //www.expasy.ch/spr
3. Lista de glucosiltransferasas conocidas y su origen.
4. ADNc humano, Kumar y cols. (1990) Proc. Natl. Acad. Sci. EE.UU. 87: 9948-9952
5. Gen humano, Hull y cols. (1991) Biochem. Biophys. Res. Commun. 176: 608-615
6. ADNc de ratón, Kumar y cols. (1992) Glycobiology 2: 383-393
7. Gen de ratón, Pownall y cols. (1992) Genomics 12: 699-704
8. Gen murino (flanqueante 5', no codificante), Yang y cols. (1994) Glycobiology 5: 703-712
9. ADNc de conejo, Sarkar y cols. (1991) Proc. Natl. Acad. Sci. EE.UU. 88: 234-238
10. ADNc de rata, Fukada y cols. (1994) Biosci. Biotechnol. Biochem. 58: 200-201 1,2 (GnTII) EC 2.4.1.143
11. Gen humano, Tan y cols. (1995) Bur. J. Biochem. 231: 317-328
12. ADNc de rata, D'Agostaro y cols. (1995) J. Biol. Chem. 270: 15211-15221
13. β1,4 (GnTIII) EC 2.4.1.144
14. ADNc humano, Ihara y cols. (1993). J. Biochem.113: 692-698
15. Gen murino, Bhaumik y cols. (1995) Gene 164: 295-300
16. ADNc de rata, Nishikawa y cols. (1992) J. Biol. Chem. 267: 18199-18204 β1,4 (GnTIV) EC 2.4.1.145
17. ADNc humano, Yoshida y cols. (1998) Glycoconjugate Journal 15: 1115-1123
18. ADNc bovino, Minowa y cols., Patente Europea EP 0 905 232 β1,6 (GnT V) EC 2.4.1.155
19. ADN humano, Saito y cols. (1994) Biochem. Biophys. Res. Commun. 198: 318-327
20. ADNc de rata , Shoreibah y cols. (1993) J. Biol. Chem. 268: 15381-15385. β1,4 Galactosiltransferasa, EC 2.4.1.90 (LacNAc sintetasa) EC 2.4.1.22 (lactosa sintetasa)
21. ADNc bovino, D'Agostaro y cols. (1989) Eur. J. Biochem. 183: 211-217
22. ADNc bovino (parcial), Narimatsu y cols. (1986) Proc. Natl. Acad. Sci. EE.UU. 83: 4720-4724
23. ADNc bovino (parcial), Masibay y Qasha (1989) Proc. Natl. Acad. Sci. EE.UU. 86: 5733-5377
24. ADNc bovino (extremo 5'), Russo y cols. (1990) J. Biol. Chem. 265: 3324
25. ADNc de pollo (parcial), Ghosh y cols. (1992) Biochem. Biophys. Res. Commun. 1215-1222
26. ADNc humano, Masri y cols. (1988) Biochem. Biophys. Res. Commun. 157: 657-663

27. ADNc humano, (células HeLa) Watzele & Berger (1990) Nucl. Acids Res. 18: 7174

28. ADNc humano, (parcial) Uejima y cols. (1992) Cancer Res. 52: 6158-6163

29. ADNc humano, (carcinoma) Appert y cols. (1986) Biochem. Biophys. Res. Commun. 139: 163-168

30. Gen humano, Mengle-Gaw y cols. (1991) Biochem. Biophys. Res. Commun. 176: 1269-1276.

31. ADNc murino, Nakazawa y cols. (1988) J. Biochem. 104: 165-168

32. ADN murino, Shaper y cols. (1988) J. Biol. Chem. 263: 10420-10428

33. ADNc murino (novel), Uehara y Muramatsu no publicado

34. Gen murino, Hollis y cols. (1989) Biochem. Biophys. Res. Commun. 162: 1069-1075

35. Proteína de rata (parcial), Bendiak y cols. (1993) Eur. J. Biochem. 216: 405-417 2,3-Sialiltransferasa, (ST3Gal II) (ligado a N) (Gal-1,3/4-GlcNAc) EC 2.4.99.6

36. ADNc humano, Kitagawa y Paulson (1993) Biochem. Biophys. Res. Commun. 194: 375-382

37. ADNc de rata, Wen y cols. (1992) J. Biol. Chem. 267: 21011-21019 2,6-Sialiltransferasa, (ST6Gal I) EC 2.4.99.1

38. Pollo, Kurosawa y cols. (1994) Eur. J. Biochem 219: 375-381

39. ADNc humano (parcial), Lance y cols. (1989) Biochem. Biophys. Res. Commun. 164: 225- 232

40. ADNc humano, Grundmann y cols. (1990) Nucl. Acids Res. 18: 667

41. ADNc humano, Zettlmeisl y cols. (1992) Patente EPO475354-A/3

42. ADNc humano, Stamenkovic y cols. (1990) J. Exp. Med. 172: 641-643 (CD75)

43. ADNc humano, Bast y cols. (1992) J. Cell Biol, 116: 423-435

44. Gen humano (parcial), Wang y cols. (1993) J. Biol. Chem. 268: 4355-4361

45. Gen humano (flanqueante 5'), Aasheim y cols. (1993) Eur. J. Biochem. 213: 467-475

46. Gen humano (promotor), Aas-Eng y cols. (1995) Biochim. Biophys. Acta 1261: 166-169

47. ADNc de ratón, Hamamoto y cols. (1993) Bioorg. Med. Chem. 1: 141-145

48. ADNc de rata, Weinstein y cols. (1987) J. Biol. Chem. 262: 17735-17743

49. ADNc de rata (fragmentos de transcrito), Wang y cols. (1991) Glycobiology 1: 25-31, Wang y cols. (1990) J. Biol. Chem. 265: 17849-17853

50. ADNc de rata (extremo 5'), O'Hanlon y cols. (1989) J. Biol. Chem. 264: 17389-17394; Wang y cols. (1991) Glycobiology 1: 25-31

51. Gen de rata (promotor), Svensson y cols. (1990) J. Biol. Chem. 265: 20863-20688

52. ARNm de rata (fragmentos), Wen y cols. (1992) J. Biol. Chem. 267: 2512-2518

En la bibliografía se describen procedimientos adicionales y reactivos que pueden usarse en los procedimientos modificando la glucosilación, tal como en la Patente de Estados Unidos N.º: 5.955.422, Patente de Estados Unidos N.º: 4.775.622, Patente de Estados Unidos N.º: 6.017.743, Patente de Estados Unidos N.º: 4.925.796, Patente de Estados Unidos N.º: 5.766.910, Patente de Estados Unidos N.º: 5.834.251, Patente de Estados Unidos N.º: 5.910.570, Patente de Estados Unidos N.º: 5.849.904, Patente de Estados Unidos N.º: 5.955.347, Patente de Estados Unidos N.º: 5.962.294, Patente de Estados Unidos N.º: 5.135.854, Patente de Estados Unidos N.º: 4.935.349, Patente de Estados Unidos N.º: 5.707.828 y Patente de Estados Unidos N.º: 5.047.335. Pueden obtenerse sistemas de expresión de levadura apropiados de fuentes tales como la Colección Americana de Cultivos Tipo, Rockville, MD. En el mercado están disponibles vectores de una diversidad de fuentes.

Listado de secuencias

Las SEC ID N.º: 1-6 pueden encontrarse en la Solicitud de Patente de Estados Unidos N.º: 09/892.591

SEC ID N.º: 7

Cebador: regiones de alta homología dentro de 1,6 manosiltransferasas 5'-atggcgaaggcagatggcagt-3'

SEC ID N.º: 8

Cebador: regiones de alta homología dentro de 1,6 manosiltransferasas 5'-ttagtccttccaacttccttc-3'

SEC ID N.º: 9

Cebador interno: 5'-actgccatctgccttcgccat-3'

SEC ID N.º: 10

Cebador interno: 5'-GTAATACGACTCACTATAGGGC-3' T7

SEC ID N.º: 11

Cebador interno: 5'-AATTAACCCTCACTAAAGGG-3' T3

SEC ID N.º: 12

Cebador: atgcccgtgg ggggcctgtt gccgctcttc agtagc

SEC ID N.º: 13

Cebador: tcatttctct ttgccatcaa tttccttctt ctgttcacgg

SEC ID N.º: 14

Cebador: ggcgcgccga ctcctccaag ctgctcagcg gggtcctgtt ccac

SEC ID N.º: 15

Cebador: ccttaattaa tcatttctct ttgccatcaa tttccttctt ctgttcacgg

SEC ID N.º: 16

Cebador: ggcgagctcg gcctacccgg ccaaggctga gatcatttgt ccagcttcaga

SEC ID N.º: 17

Cebador: gcccacgtcg acggatccgt taaacatcg attggagagg ctgacaccgc tacta

SEC ID N.º: 18

Cebador: cgggatccac tagtatttaa atcatatgtg cgagtgtaca actcttccca catgg

SEC ID N.º: 19

Cebador: ggacgcgtcg acggcctacc cggccgtacg aggaatttct cggatgactc ttttc

SEC ID N.º: 20

Cebador. cgggatccct cgagagatct tttttgtaga aatgtcttgg tgcct

SEC ID N.º: 21

Cebador: ggacatgcat gcactagtgc ggccgccacg tgatagttgt tcaattgatt gaaataggga caa

SEC ID N.º: 22

Cebador: ccttgctagc ttaattaacc gcggcacgtc cgacggcggc ccacgggtcc ca

SEC ID N.º: 23

Cebador: ggacatgcat gcggatcct taagagccgg cagcttgcaa attaaagcct tcgagcgtcc c

SEC ID N.º: 24

Cebador: gaaccacgtc gacggccatt gcggccaaaa ccttttttcc tattcaaaca caaggcattg c

SEC ID N.º: 25

Cebador: ctccaatact agtcgaagat tatcttctac ggtgcctgga ctc

SEC ID N.º: 26

Cebador: tggaaggttt aaacaaagct agagtaaaa tagatatagc gagattagag aatg

SEC ID N.º: 27

Cebador: aagaattcgg ctggaaggcc ttgtaccttg atgtagttcc cgttttcatc

SEC ID N.º: 28

Cebador: gcccaagccg gccttaaggg atctcctgat gactgactca ctgataataa aaatacgg

SEC ID N.º: 29

Cebador: gggcgcgta tttaaatacta gtggatctat cgaatctaaa tgtaagttaa aatctctaa

SEC ID N.º: 30

Cebador: ggccgcctgc agatttaaat gaattcgg cgcgccttaat

SEC ID N.º: 31

Cebador: taaggcgcgc cgaattcatt taaatctgca gggc

SEC ID N.º: 32

Cebador: 5'-tggcaggcgcgcctcagtcagcgctctcg-3'

SEC ID N.º: 33

Cebador: 5'-aggttaatta agtgctaattccagctagg-3'

SEC ID N.º: 34

Cebador para gen *OCH1* de *K. lactis*: ccagaagaat tcaattytgy cartgg

SEC ID N.º: 35

Cebador para gen *OCH1* de *K. lactis*: cagtgaaaat acctggnccn gtcca

SEC ID N.º: 36

Cebador para gen *MNN1* de *K. lactis*: tgccatcttt taggtccagg cccgttc

SEC ID N.º: 37

Cebador para gen *MNN1* de *K. lactis*: gatcccacga cgcatcgtat ttctttc

SEC ID N.º: 38

Secuencia de ADN del segmento de 302 pb del gen de *KlOCH1* teórico:

gcccttcagtgaaaatacctggcccggtccagttcataatatcggtaccatctgtatttttggcggttttctttgttgatgttt
gtaatttttgttgaacttctttttatcccicatgttgacattataatcatctgcaatgtctttaatacttcagc
atcatctaaaggaatgctgcttttaacatttgccacgctctccaatgttgttgcggtgatatttgtgatcaattcgcgcaataa
tggatggccagattttgattgtattgtccactgacaaaattgaattctctggaagggc

SEC ID N.º: 39

Traducción de gen *KlOCHl* teórico (excluyendo cebadores):

TIQSKSGHPLLRELITNITATTLESVANVKSSIPLDDAEVLKDIADDYNVNM
RDKKKFNKNYKHQQKKTAKNTDGTDIMN

SEC ID N.º: 40

Secuencia de ADN del segmento de 405 pb del gen *K1MNN1* teórico:

cccagcgtgccattaccgtatttgccgccgtttgaaatactcaatattcatgatggttgtaaggcgtttttatcattcgcgat
ataatatgccatcttttaggtccaggcccgttctcttagctatctttggtgtctgtgctaccgtgatatggtacct
attctttttccagtctaatctgaagatggcagatttgaaaaaggtagcaacttcaaggtatctttcacaagaaccgtcgttat
cagaacttatgtcaaatgtgaagatcaagcctattgaagaaacccccggtttcgccattggagttgattccagatatcgaaa
tatcgactagaaagaaatacgatgcgtcgtgggatctgttgttccgtggtagaaaatataaatcgttcaacgattatgat

SEC ID N.º: 41

Secuencia de ADN del gen *OCH1* de *K. lactis*:

atggggttaccaaagatttcaagaagaacgaggtacattattgtcattgtgctgatactgtacttattgttttctgtgcaatg
gaatactgcgaaagtgaatcaccatttctataacagcattggcacggtgcttcccagtacagctcgcgtggatcacttga
acttgaaaaacttggacttagcaggtacgagcaataacggtgatcatttgatggatctacgagttcaattggctagtcaat
tcccctacgattctcgagtacccatccccaaaaaggtatggcagacctggaagattgatcccagttcaaagtcacaggtt
tcttccatttcaaaatgccagaatgattggaaacatttcagtgcatccgaggaaccgccatatcaataccaattaatcaca
gatgatcaaatgataccacttctagagcagctatatggtggggtcccacaagtgataaaggcttttgaatccttgccactt
ccaattcttaaagcagacttttcagatacttgatcctttatgcaagaggtggtatatattctgacatggatacgttcccatta
aagccattgtcgtcatggccatcgacttctcagtcctacttttctagtttaaagaatccacaaaggtatagaaattccttgga
caaccttgaaacgctagaagcttcagaacctggctttgtcattggtatcgaggctgatccggatagaagcgattgggca
gagtggtacgccaggagaatacaattctgtcagtggacaatacaatcaaaatctggccatccattattgcgcgaattgat
cacaaatatcaccgcaacaacattggagagcgtggcaaatgttaaaagcagcattcctttagatgatgctgaagtattaa
aagacattgcagatgattataatgtcaacatgagggataaaaagaagttcaacaaaaattacaaacatcaacaaaagaa
aaccgccaaaaatacagatggtaccgatattatgaactggactggtccaggtatttttcagatgttatttccagtatctta
ataacgttatccagaagaatgatgacattttaatttcaatgataatcttaatgttatcaacaaacatggatccaaacatgata
caactatgagattctataaagacattgttaaaaatttacaaaacgacaaaccctcattgttctggggattcttttcattgatga

cagagcctattctagtggacgacatcatggtacttccgattacttctttctcaccaggtatcagaacaatgggcgctaaag
aagacaacgacgagatggcatttgttaagcatatttttgaaggaagttggaaagactga

SEC ID N.º: 42

Traducción de gen *OCH1* de *K. lactis* teórico:

MGLPKISRRTRYIIVIVLILYLLFSVQWNTAKVNHHFYNSIGTVLPSTARVD
HLNLKNLDLAGTSNNGDHLMDLRVQLASQFPYDSRVPIPKKVWQTWKID
PSSKSQVSSISKCQNDWKHFSASEEPPYQYQLITDDQMIPLLEQLYGGVPQ
VIKAFESLPLPILKADFFRYLILYARGGIYSDMDTFPLKPLSSWPSTSQSYFS
SLKNPQRYRNSLDNLETLEASEPGFVIGIEADPDRSDWAEWYARRIQFCQW
TIQSKSGHPLLRELITNITATTLESVANVKSSIPLDDAEVLKDIADDYNVNM
RDKKKFNKNYKHQQKKTAKNTDGTDIMNWTGPGIFSDVIFQYLNNVIQK
NDDILIFNDNLNVINKHGSKHDTTMRFYKDIVKNLQNDKPSLFWGFFSLMT
EPILVDDIMVLPITSFSPGIRTMGAKEDNDEMAFVKHIFEGSWKDZ

SEC ID N.º: 43

Secuencia de ADN del gen *MNN1* de *K. lactis*:

atgatggttgtaaggcgttttttatcagcttcgcgatataatatgccatcttttaggtccaggcccgttctcttagctatctttg
gtgtctgtgctaccgtgatatggtacctattcttttttccagtctaatctgaagatggcagatttgaaaaaggtag
caacttcaaggtatctttcacaagaaccgtcgttatcagaacttatgtcaaatgtgaagatcaagcctattgaagaaaccc
cggtttcgccattggagttgattccagatatcgaaatatcgactagaaagaaatacgatgcgtcgtgggatctgttgttcc
gtggtagaaaatataaatcgttcaacgattatgatcttcatacgaaatgtgagttttatttccagaatttatacaatttgaacg
aggattggaccaataatattcggacgttcactttcgatattaacgatgtagacacgtctacgaaaattgacgctcttaaag
attccgatggggttcaattggtggacgagaaggctatacgtttatacaagagaacgcataacgttgccttggctacgga
aaggttacgtctttatgataaatgttttgtcaatagtccaggttcaaacccattgaaaatggatcacctttcagatcgaaca
agaagagtaagactacggctttggatgacgaagtcactgggaaccgtgatacttttaccaagacgaagaaaacttcgtt
cttaagcgatatggacacgagtagtttccagaagtacgatcaatgggatttcgaacatagaatgttccccatgatcccata
tttcgaggaacacaatttcaccaacgtgatgcctatttcaccggctcaaacggtggggaaccttacctcaagggaaatt
cccggtattagatccaaaatccggtgaattgttacgtgtagagacttcagatatgataaatcgaaatcgctttggaagaa
ctggaatgatatgtcctctgcttctggtaaacgtggtattatcttggctgctggcgacggccaagtggaccaatgcatcc
gtcttattgctacgttgagagctcaaggaaacgctctacctattcaaattatccacaacaaccaattgaatgagaaatctgt
gaaactgttatcggaggccgctaaatctaccgaattctcatccggtagagctcaatctctttggttagtgaatgtgggccc
cacgttggaatcttcaatgaagagcaattttgggagatttaagaataagtggttgtcagttattttcaacacttttgaagaatt
tatattcatagatacagatgccatctcctacattaatatggctgattatttcaacttcaaggagtacaaatctactggaacact
cttctttaaggataggtctttggcaattggaactgaacagaaatgtggtcctttgttcgaaactcttgaaccaagaattcttg
aaatgtactatttcaatactttacctatgatcaatggtgattacgtggaacagcaatgtatgggcatgctcaccccagagg
aaaaagtttacaaacgtttctttgaagttggtcatcaacacaacttggaaagtggattattggccatcaacaaaaacgaac
acatcatgggattggttactgcaacagtcttaaatatcgcaccaaaggtcggaggttgcggttggggtgacaaagagttt
ttctggcttggtttgttggttgctggccaacgctactcgatctatgatatagatgcaagtgcaattggtgttcctcaacagaa
gcaatctatcgctaacggagacgaatttgatgaatataggatttgttctttacaagtggcacatacttcatacgacggacat
ttactatggataaatggtggctctcagtactgtaagaaaccagagacttttgaaggtgattggaccaacattaaggagctt
cgtgaatcgtattctgatgataaagaaaaggctctgaaggcttatagtgatacagttaaggtggaagcagcaatcgtgc
cagattccagaagtaatggttggggtagagacgatcaaagatgtaaaggctacttctggtgcggcaaatttacttcaaa

gctgaaaccgtatacttataacacggtggtaactaaaggtgatttgatccgtttcggagacgaggaaatcgaaagtatct
ccaagattaataagatctggaatgatgctattattccagacggagcttaa

SEC ID N.º: 44

Traducción de gen *MNN1* de *K. lactis* teórico:

MMVVRRFLSASRYNMPSFRSRPVLLAIFGVCATVIWYLFFFQSNLKMADL
KKVATSRYLSQEPSLSELMSNVKIKPIEETPVSPLELIPDIEISTRKKYDASW
DLLFRGRKYKSFNDYDLHTKCEFYFQNLYNLNEDWTNNIRTFTFDINDVD
TSTKIDALKDSDGVQLVDEKAIRLYKRTHNVALATERLRLYDKCFVNSPG
SNPLKMDHLFRSNKKSKTTALDDEVTGNRDTFTKTKKTSFLSDMDTSSFQ
KYDQWDFEHRMFPMIPYFEEHNFTNVMPIFTGSNGGEPLPQGKFPVLDPKS
GELLRVETFRYDKSKSLWKNWNDMSSASGKRGIILAAGDGQVDQCIRLIA
TLRAQGNALPIQIIHNNQLNEKSVKLLSEAAKSTEFSSGRAQSLWLVNVGP
TLESSMKSNFGRFKNKWLSVIFNTFEEFIFIDTDAISYINMADYFNFKEYKST
GTLFFKDRSLAIGTEQKCGPLFETLEPRILEMYYFNTLPMINGDYVEQQCM
GMLTPEEKVYKRFFEVGHQHNLESGLLAINKNEHIMGLVTATVLNIAPKV
GGCGWGDKEFFWLGLLVAGQRYSIYDIDASAIGVPQQKQSIANGDEFDEY
RICSLQVAHTSYDGHLLWINGGSQYCKKPETFEGDWTNIKELRESYSDDKE
KALKAYSDTVKVEAAIVPDSRSNGWGRDDQRCKGYFWCGKFTSKLKPYT
YNTVVTKGDLIRFGDEEIESISKINKIWNDAIIPDGA

Referencias

Aebi, M., J. Gassenbuber y cols. (1996). "Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae." Glycobiology 6 (4): 439-444.

Altmann, F., E. Staudacher y cols. (1999). "Insect cells as hosts for the expression of recombinant glycoproteins." Glycoconjugate Journal 16 (2): 109-123.

Andersen, D. C. y C. F. Goochee (1994). "The effect of cell-culture conditions on the oligosaccharide structures of secreted glycoproteins." Current Opinion in Biotechnology 5: 546-549.

Bardor, M., L. Faye y cols. (1999). "Analysis of the N-glycosylation of recombinant glycoproteins produced in transgenic plants." Trends in Plant Science 4 (9): 376-380.

Bretthauer, R. K. y F. J. Castellino (1999). "Glycosylation of Pichia pastoris-derived proteins." Biotechnology and Applied Biochemistry 30: 193-200.

Burda, P. y M. Aebi (1999). "The dolichol pathway of N-linked glycosylation." Biochimica Et Biophysica Acta-General Subjects 1426 (2): 239-257.

Chiba, Y., M. Suzuki y cols. (1998). "Production of human compatible high mannose-type (Man(5)GlcNAc(2)) sugar chains in Saccharomyces cerevisiae." Journal of Biological Chemistry 273(41): 26298-26304.

Cole, E. S., E. Higgins y cols. (1994). "Glycosylation Patterns of Human Proteins Expressed in Transgenic Goat Milk." Journal of Cellular Biochemistry: 265-265.

Davies y cols. Biotechnol Bioeng. 20 agosto de 2001; 74 (4): 288-294. (Expression of GnTIII in a Recombinant Anti-CD20 CHO Production Cell Line: Expression of Antibodies with Altered Glycoforms Leadstoan Increase in ADCC Through Higher Affinity for FcgRIII).

Dente, L., U. Ruther y cols. (1988). "Expression of Human Alpha-1-Acid Glycoprotein Genes in Cultured-Cells and in Transgenic Mice." Genes & Development 2 (2): 259-266.

Huffaker, T. C. y P. W. Robbins (1983). "Yeast Mutants Deficient in Protein Glycosylation." Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 80 (24): 7466-7470.

Jarvis, D. L., Z. S. Kawar y cols. (1998). "Engineering N-glycosylation pathways in the baculovirus-insect cell system." Current Opinion in Biotechnology 9 (5): 528-533.

Kimura, T., N. Kitamoto y cols. (1997). "A novel yeast gene, RHK1, is involved in the synthesis of the cell wall receptor for the HM-1 killer toxin that inhibits beta-1,3-glucan synthesis." Molecular & General Genetics 254 (2): 139-147.

Kimura, T., T. Komiyama y cols. (1999). "N-glycosylation is involved in the sensitivity of Saccharomyces cerevisiae to HM-1 killer toxin secreted from Hansenula mrakii IFO 0895." Applied Microbiology and Biotechnology 51 (2): 176-184.

Malissard, M., S. Zeng y cols. (2000)."Expression of functional soluble formsofhuman beta-1,4-galactosyltransferase I, alpha-2,6-sialyltransferase, and alpha-1,3-fucosyltransferase VI in the methylotrophic yeast Pichia pastoris." Biochemical and Biophysical Research Communications 267 (1): 169-173.

Maras, M. y R. Contreras (1994). Methods of Modifying Carbohydrate Moieties. United States, Alko Group Ltd., Helsinki, Finlandia.

Maras, M., A. De Bruyn y cols. (1999). "In vivo synthesis of complex N-glycans by expression of human N-acetylglucosaminyltransferase I in the filamentous fungus Trichoderma reesei." Febs Letters 452 (3): 365-370.

Maras, M., X. Saelens y cols. (1997). "In vitro conversion of the carbohydrate moiety of fungal glycoproteins to mammalian-type oligosaccharides - Evidence for N-acetylglucosaminyltransferase-I-accepting glycans from Trichoderma reesei." European Journal of Biochemistry 249 (3): 701-707.

Martinet, W., M. Maras y cols. (1998). "Modification of the protein glycosylation pathway in the methylotrophic yeast Pichia pastoris." Biotechnology Letters 20 (12): 1171-1177.

McGarvey, P. B., J. Hammond y cols. (1995). "Expression of the Rabies Virus Glycoprotein in Transgenic Tomatoes." Bio-Technology 13 (13): 1484-1487.

Moens, S. y J. Vanderleyden (1997). "Glycoproteins in prokaryotes." Archives of Microbiology 168(3): 169-175.

Nakanishishindo, Y., K. Nakayama y cols. (1993). "Structure of the N-Linked Oligosaccharides That Show the

Complete Loss of Alpha-1,6-Polymannose Outer Chain From Och1, Och1 Mnn1, and Och1 Mnn1 Alg3 Mutants of Saccharomyces- Cerevisiae." Journal of Biological Chemistry 268 (35): 26338-26345.

Raju, T. S., J. B. Briggs y cols. (2000). "Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics." Glycobiology 10 (5): 477-486.

Sharma, C. B., R. Knauer y cols. (2001). "Biosynthesis of lipid-linked oligosaccharides in yeast: the ALG3 gene encodes the Dol-P-Man : Man(5)GlcNAc(2)-PP-Dol mannosyltransferase." Biological Chemistry 382 (2): 321-328.

Staub, J. M., B. Garcia y cols. (2000). "High-yield production of a human therapeutic protein in tobacco chloroplasts." Nature Biotechnology 18 (3): 333-338.

Takeuchi, M. (1997). "Trial for molecular breeding of yeast for the production of glycoprotein therapeutics." Trends in Glycoscience and Glycotechnology 9: S29-S35.

Umana y cols., Nat Biotechnol. Feb 1999a (17) 176-180. (Engineered glycoforms of an antineuroblastoma IgG1 with optimized antibodydependent cellular cytotoxic activity)

Umana y cols., Biotechnol Bioeng. diciembre de 1999b 5; 65 (5): 542-549. (Regulated Overexpression of glycosyltransferase). Verostek, M. F., P. H. Atkinson y cols. (1993). "Glycoprotein-Biosynthesis in the Alg3 Saccharomyces-Cerevisiae Mutant .1. Role of Glucose in the Initial Glycosylation of Invertase in the Endoplasmic-Reticulum." Journal of Biological Chemistry 268 (16): 12095-12103.

Verostek, M. F., P. H. Atkinson y cols. (1993). "Glycoprotein-Biosynthesis in the Alg3 Saccharomyces-Cerevisiae Mutant .2. Structure of Novel Man6-10glcnac2 Processing Intermediates On Secreted Invertase." Journal of Biological Chemistry 268 (16): 12104-12115.

Weikert, S., D. Papac y cols. (1999). "Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins." Nature Biotechnology 17 (11): 1116-1121.

Wemer, R. G., W. Noe y cols. (1998). "Appropriate mammalian expression systems for biopharmaceuticals." Arzne-imittel-Forschung-Drug Research 48 (8): 870-880.

Yang, M. y M. Butler (2000). "Effects of ammonia on CHO cell growth erythropoietin production, and glycosylation." Biotechnology and Bioengineering 68 (4): 370-380.

Zufferey, R., R. Knauer y cols. (1995). "Stt3, a Highly Conserved Protein Required for Yeast Oligosaccharyl Transferase-Activityin-Vivo." EMBO Journal 14 (20): 4949-4960.

Listado de secuencias

<110> GlycoFi, Inc.

<120> BIBLIOTECA DE ADN COMBINATORIA PARA PRODUCIR N-GLUCANOS MODIFICADOS EN EUCARIOTAS INFERIORES

<130> L2576 EP/1 S3

<140> PCT/US04/005244

<141> 20-02-2004

<150> EP 04 71 3437.4

<151> 20-02-2004

<150> 10/371.877

<151> 20-02-2003

<150> 09/892.591

<151> 27-06-2001

<150> 60/214.358

<151> 28-06-2000

<150> 60/215.638

<151> 30-06-2000

<150> 60/279.997

<151> 30-03-2001

<160> 47

<170> PatentIn ver. 3.5

<210> 1

<211> 21

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 1

atggcgaagg cagatggcag t 21

<210> 2

<211> 21

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 2

ttagtccttc caacttcctt c 21

<210> 3

<211> 26

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<220>

<221> VARIANTE

<222> (9), (12), (18), (24)

<223> /reemplazar="a"

/reemplazar="t"

/reemplazar="c"

/reemplazar="g"

<400> 3

taytggmgng tngarcynga yatnaa 26

<210> 4

<211> 20

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<220>

<221> base modificada

<222> (6)

<223> a, t, c o g

<220>

<221> base modificada

<222> (12)

<223> a, t, c o g

<400> 4

gcrtcncccc anckytcrta 20

<210> 5

<211> 4

<212> PRT

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: tetrapéptido señal ilustrativo"

<400> 5

His Asp Glu Leu
1

<210> 6

<211> 4

<212> PRT

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: tetrapéptido señal ilustrativo"

<400> 6

Lys Asp Glu Leu
1

<210> 7

<211> 21

<212> ADN

<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 7
atggcgaagg cagatggcag t 21

<210> 8
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 8
ttagtccttc caacttcctt c 21

<210> 9
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 9
actgccatct gccttcgcca t 21

<210> 10
<211> 22
<212> ADN
<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 10

gtaatacgac tcactatagg gc 22

<210> 11

<211> 20

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 11

aattaaccct cactaaaggg 20

<210> 12

<211> 36

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 12

atgcccgtgg ggggcctgtt gccgctcttc agtagc 36

<210> 13

<211> 40

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 13

tcatttctct ttgccatcaa tttccttctt ctgttcacgg 40

<210> 14

<211> 44

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 14

ggcgcgccga ctcctccaag ctgctcagcg gggtcctgtt ccac 44

<210> 15

<211> 50

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 15

ccttaattaa tcatttctct ttgccatcaa tttccttctt ctgttcacgg 50

<210> 16

<211> 51

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 16
ggcgagctcg gcctacccgg ccaaggctga gatcatttgt ccagcttcag a 51

<210> 17
<211> 55
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 17
gcccacgtcg acggatccgt ttaaacatcg attggagagg ctgacaccgc tacta 55

<210> 18
<211> 54
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 18
cgggatccac tagtatttaa atcatatgtc gagtgtacaa ctcttcccac atgg 54

<210> 19
<211> 55
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 19

ggacgcgtcg acggcctacc cggccgtacg aggaatttct cggatgactc ttttc 55

<210> 20

<211> 45

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 20

cgggatccct cgagagatct tttttgtaga aatgtcttgg tgcct 45

<210> 21

<211> 63

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 21

ggacatgcat gcactagtgc ggccgccacg tgatagttgt tcaattgatt gaaataggga caa 60
63

<210> 22

<211> 52

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 22

ccttgctagc ttaattaacc gcggcacgtc cgacggcggc ccacgggtcc ca 52

<210> 23

<211> 61

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 23

ggacatgcat gcggatccct taagagccgg cagcttgcaa attaaagcct tcgagcgtcc 60
c 61

<210> 24

<211> 61

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 24

gaaccacgtc gacggccatt gcggccaaaa ccttttttcc tattcaaaca caaggcattg 60
c 61

<210> 25

<211> 43

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 25

ctccaatact agtcgaagat tatcttctac ggtgcctgga ctc 43

<210> 26

<211> 53

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 26

tggaaggttt aaacaaagct agagtaaaat agatatagcg agattagaga atg 53

<210> 27

<211> 50

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 27

aagaattcgg ctggaaggcc ttgtaccttg atgtagttcc cgttttcatc 50

<210> 28

<211> 58

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 28

gcccaagccg gccttaaggg atctcctgat gactgactca ctgataataa aaatacgg 58

<210> 29

<211> 59

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 29

gggcgcgtat ttaaatacta gtggatctat cgaatctaaa tgtaagttaa aatctctaa 59

<210> 30

<211> 39

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 30

ggccgcctgc agatttaaat gaattcggcg cgccttaat 39

<210> 31

<211> 34

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 31
taaggcgcgc cgaattcatt taaatctgca gggc 34

<210> 32
<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 32
tggcaggcgc gcctcagtca gcgctctcg 29

<210> 33
<211> 29
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 33
aggttaatta agtgctaatt ccagctagg 29

<210> 34
<211> 26
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 34
ccagaagaat tcaattytgy cartgg 26

<210> 35
<211> 25
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<220>
<221> VARIANTE
<222> (17), (20)
<223> /reemplazar="a" /reemplazar="t" /reemplazar="c" /reemplazar="g"

<400> 35
cagtgaaaat acctggnccn gtcca 25

<210> 36
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<221> Fuente
<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 36
tgccatcttt taggtccagg cccgttc 27

<210> 37
<211> 27
<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<400> 37

gatcccacga cgcatcgtat ttctttc 27

<210> 38

<211> 302

<212> ADN

<213> Kluyveromyces lactis

<400> 38

```
gcccttcagt gaaaatacct ggcccggtcc agttcataat atcggtacca tctgtatttt 60
tggcggtttt cttttgttga tgtttgtaat ttttgttgaa cttctttta tccctcatgt 120
tgacattata atcatctgca atgtctttta atacttcagc atcatctaaa ggaatgctgc 180
ttttaacatt tgccacgctc tccaatgttg ttgcggtgat atttgtgatc aattcgcgca 240
ataatggatg gccagatttt gattgtattg tccactgaca aaattgaatt ctctggaagg 300
gc                                                                302
```

<210> 39

<211> 80

<212> PRT

<213> Kluyveromyces lactis

<400> 39

```
Thr Ile Gln Ser Lys Ser Gly His Pro Leu Leu Arg Glu Leu Ile Thr
  1               5                  10                  15

Asn Ile Thr Ala Thr Thr Leu Glu Ser Val Ala Asn Val Lys Ser Ser
             20                  25                  30

Ile Pro Leu Asp Asp Ala Glu Val Leu Lys Asp Ile Ala Asp Asp Tyr
         35                  40                  45

Asn Val Asn Met Arg Asp Lys Lys Lys Phe Asn Lys Asn Tyr Lys His
     50                  55                  60

Gln Gln Lys Lys Thr Ala Lys Asn Thr Asp Gly Thr Asp Ile Met Asn
 65                  70                  75                  80
```

<210> 40

<211> 404

<212> ADN

<213> Kluyveromyces lactis

<400> 40

```
cccagcgtgc cattaccgta tttgccgccg tttgaaatac tcaatattca tgatggttgt 60
aaggcgtttt ttatcattcg cgatataata tgccatcttt taggtccagg cccgttctct 120
tagctatctt tggtgtctgt gctaccgtga tatggtacct attctttttc cagtctaatc 180
tgaagatggc agatttgaaa aaggtagcaa cttcaaggta tctttcacaa gaaccgtcgt 240
tatcagaact tatgtcaaat gtgaagatca agcctattga agaaaccccg gtttcgccat 300
tggagttgat tccagatatc gaaatatcga ctagaaagaa atacgatgcg tcgtgggatc 360
tgttgttccg tggtagaaaa tataaatcgt tcaacgatta tgat                   404
```

<210> 41

<211> 1363

<212> ADN

<213> Kluyveromyces lactis

<400> 41

```
atggggttac caaagatttc aagaagaacg aggtacatta ttgtcattgt gctgatactg 60
tacttattgt tttctgtgca atggaatact gcgaaagtga atcaccattt ctataacagc 120
attggcacgg tgcttcccag tacagctcgc gtggatcact tgaacttgaa aaacttggac 180
ttagcaggta cgagcaataa cggtgatcat ttgatggatc tacgagttca attggctagt 240
caattcccct acgattctcg agtacccatc cccaaaaagg tatggcagac ctggaagatt 300
gatcccagtt caaagtcaca ggtttcttcc atttcaaaat gccagaatga ttggaaacat 360
ttcagtgcat ccgaggaacc gccatatcaa taccaattaa tcacagatga tcaaatgata 420
ccacttctag agcagctata tggtggggtc ccacaagtga taaaggcttt tgaatccttg 480
ccacttccaa ttcttaaagc agacttttтc agatacttga tcctttatgc aagaggtggt 540
atatattctg acatggatac gttcccatta aagccattgt cgtcatggcc atcgacttct 600
cagtcctact tttctagttt aaagaatcca caaaggtata gaaattcctt ggacaacctt 660
gaaacgctag aagcttcaga acctggcttt gtcattggta tcgaggctga tccggataga 720
agcgattggg cagagtggta cgccaggaga atacaattct gtcagtggac aatacaatca 780

aaatctggcc atccattatt gcgcgaattg atcacaaata tcaccgcaac aacattggag 840
agcgtggcaa atgttaaaag cagcattcct ttagatgatg ctgaagtatt aaaagacatt 900
gcagatgatt ataatgtcaa catgagggat aaaaagaagt tcaacaaaaa ttacaaacat 960
caacaaaaga aaaccgccaa aaatacagat ggtaccgata ttatgaactg gactggtcca 1020
ggtatttttt cagatgttat tttccagtat cttaataacg ttatccagaa gaatgatgac 1080
attttaattt tcaatgataa tcttaatgtt atcaacaaac atggatccaa acatgataca 1140
actatgagat tctataaaga cattgttaaa aatttacaaa acgacaaacc cctcattgtt 1200
ctggggattc ttttcattga tgacagagcc tattctagtg gacgacatca tggtacttcc 1260
gattacttct ttctcaccag gtatcagaac aatgggcgct aaagaagaca acgacgagat 1320
ggcatttgtt aagcatattt ttgaaggaag ttggaaagac tga                   1363
```

<210> 42

<211> 453

<212> PRT

<213> Kluyveromyces lactis

<400> 42

```
Met Gly Leu Pro Lys Ile Ser Arg Arg Thr Arg Tyr Ile Ile Val Ile
 1               5                  10                  15

Val Leu Ile Leu Tyr Leu Leu Phe Ser Val Gln Trp Asn Thr Ala Lys
            20                  25                  30

Val Asn His His Phe Tyr Asn Ser Ile Gly Thr Val Leu Pro Ser Thr
        35                  40                  45

Ala Arg Val Asp His Leu Asn Leu Lys Asn Leu Asp Leu Ala Gly Thr
    50                  55                  60

Ser Asn Asn Gly Asp His Leu Met Asp Leu Arg Val Gln Leu Ala Ser
65                  70                  75                  80

Gln Phe Pro Tyr Asp Ser Arg Val Pro Ile Pro Lys Lys Val Trp Gln
                85                  90                  95

Thr Trp Lys Ile Asp Pro Ser Ser Lys Ser Gln Val Ser Ser Ile Ser
            100                 105                 110

Lys Cys Gln Asn Asp Trp Lys His Phe Ser Ala Ser Glu Glu Pro Pro
        115                 120                 125

Tyr Gln Tyr Gln Leu Ile Thr Asp Asp Gln Met Ile Pro Leu Leu Glu
    130                 135                 140

Gln Leu Tyr Gly Gly Val Pro Gln Val Ile Lys Ala Phe Glu Ser Leu
145                 150                 155                 160

Pro Leu Pro Ile Leu Lys Ala Asp Phe Phe Arg Tyr Leu Ile Leu Tyr
                165                 170                 175

Ala Arg Gly Gly Ile Tyr Ser Asp Met Asp Thr Phe Pro Leu Lys Pro
            180                 185                 190

Leu Ser Ser Trp Pro Ser Thr Ser Gln Ser Tyr Phe Ser Ser Leu Lys
        195                 200                 205

Asn Pro Gln Arg Tyr Arg Asn Ser Leu Asp Asn Leu Glu Thr Leu Glu
    210                 215                 220

Ala Ser Glu Pro Gly Phe Val Ile Gly Ile Glu Ala Asp Pro Asp Arg
225                 230                 235                 240

Ser Asp Trp Ala Glu Trp Tyr Ala Arg Arg Ile Gln Phe Cys Gln Trp
                245                 250                 255

Thr Ile Gln Ser Lys Ser Gly His Pro Leu Leu Arg Glu Leu Ile Thr
            260                 265                 270
```

```
Asn Ile Thr Ala Thr Thr Leu Glu Ser Val Ala Asn Val Lys Ser Ser
        275                 280                 285

Ile Pro Leu Asp Asp Ala Glu Val Leu Lys Asp Ile Ala Asp Asp Tyr
    290                 295                 300

Asn Val Asn Met Arg Asp Lys Lys Lys Phe Asn Lys Asn Tyr Lys His
305                 310                 315                 320

Gln Gln Lys Lys Thr Ala Lys Asn Thr Asp Gly Thr Asp Ile Met Asn
                325                 330                 335

Trp Thr Gly Pro Gly Ile Phe Ser Asp Val Ile Phe Gln Tyr Leu Asn
            340                 345                 350

Asn Val Ile Gln Lys Asn Asp Asp Ile Leu Ile Phe Asn Asp Asn Leu
        355                 360                 365

Asn Val Ile Asn Lys His Gly Ser Lys His Asp Thr Thr Met Arg Phe
    370                 375                 380

Tyr Lys Asp Ile Val Lys Asn Leu Gln Asn Asp Lys Pro Ser Leu Phe
385                 390                 395                 400

Trp Gly Phe Phe Ser Leu Met Thr Glu Pro Ile Leu Val Asp Asp Ile
                405                 410                 415

Met Val Leu Pro Ile Thr Ser Phe Ser Pro Gly Ile Arg Thr Met Gly
            420                 425                 430

Ala Lys Glu Asp Asn Asp Glu Met Ala Phe Val Lys His Ile Phe Glu
        435                 440                 445

Gly Ser Trp Lys Asp
    450
```

<210> 43

<211> 2241

<212> ADN

<213> Kluyveromyces lactis

<400> 43

atgatggttg taaggcgttt tttatcagct tcgcgatata atatgccatc ttttaggtcc 60
aggcccgttc tcttagctat ctttggtgtc tgtgctaccg tgatatggta cctattcttt 120
ttccagtcta atctgaagat ggcagatttg aaaaaggtag caacttcaag gtatctttca 180
caagaaccgt cgttatcaga acttatgtca aatgtgaaga tcaagcctat tgaagaaacc 240
ccggtttcgc cattggagtt gattccagat atcgaaatat cgactagaaa gaaatacgat 300
gcgtcgtggg atctgttgtt ccgtggtaga aaatataaat cgttcaacga ttatgatctt 360
catacgaaat gtgagtttta tttccagaat ttatacaatt tgaacgagga ttggaccaat 420
aatattcgga cgttcacttt cgatattaac gatgtagaca cgtctacgaa aattgacgct 480
cttaaagatt ccgatggggt tcaattggtg gacgagaagg ctatacgttt atacaagaga 540
acgcataacg ttgccttggc tacggaaagg ttacgtcttt atgataaatg ttttgtcaat 600
agtccaggtt caaacccatt gaaaatggat caccttttca gatcgaacaa gaagagtaag 660
actacggctt tggatgacga agtcactggg aaccgtgata cttttaccaa gacgaagaaa 720
acttcgttct taagcgatat ggacacgagt agtttccaga agtacgatca atgggatttc 780
gaacatagaa tgttccccat gatcccatat ttcgaggaac acaatttcac caacgtgatg 840
cctattttca ccggctcaaa cggtggggaa cctttacctc aagggaaatt cccggtatta 900
gatccaaaat ccggtgaatt gttacgtgta gagactttca gatatgataa atcgaaatcg 960
ctttggaaga actggaatga tatgtcctct gcttctggta aacgtggtat tatcttggct 1020
gctggcgacg gccaagtgga ccaatgcatc cgtcttattg ctacgttgag agctcaagga 1080
aacgctctac ctattcaaat tatccacaac aaccaattga atgagaaatc tgtgaaactg 1140
ttatcggagg ccgctaaatc taccgaattc tcatccggta gagctcaatc tctttggtta 1200
gtgaatgtgg gccccacgtt ggaatcttca atgaagagca attttgggag atttaagaat 1260
aagtggttgt cagttatttt caacactttt gaagaattta tattcataga tacagatgcc 1320
atctcctaca ttaatatggc tgattatttc aacttcaagg agtacaaatc tactggaaca 1380
ctcttcttta aggataggtc tttggcaatt ggaactgaac agaaatgtgg tcctttgttc 1440
gaaactcttg aaccaagaat tcttgaaatg tactatttca atactttacc tatgatcaat 1500
ggtgattacg tggaacagca atgtatgggc atgctcaccc cagaggaaaa agtttacaaa 1560

cgtttctttg aagttggtca tcaacacaac ttggaaagtg gattattggc catcaacaaa 1620
aacgaacaca tcatgggatt ggttactgca acagtcttaa atatcgcacc aaaggtcgga 1680
ggttgcggtt ggggtgacaa agagtttttc tggcttggtt tgttggttgc tggccaacgc 1740
tactcgatct atgatataga tgcaagtgca attggtgttc ctcaacagaa gcaatctatc 1800
gctaacggag acgaatttga tgaatatagg atttgttctt tacaagtggc acatacttca 1860
tacgacggac atttactatg gataaatggt ggctctcagt actgtaagaa accagagact 1920
tttgaaggtg attggaccaa cattaaggag cttcgtgaat cgtattctga tgataaagaa 1980
aaggctctga aggcttatag tgatacagtt aaggtggaag cagcaatcgt gccagattcc 2040
agaagtaatg gttggggtag agacgatcaa agatgtaaag gctacttctg gtgcggcaaa 2100
tttacttcaa agctgaaacc gtatacttat aacacggtgg taactaaagg tgatttgatc 2160
cgtttcggag acgaggaaat cgaaagtatc tccaagatta ataagatctg gaatgatgct 2220
attattccag acggagctta a 2241

<210> 44

<211> 747

<212> PRT

<213> Kluyveromyces lactis

<400> 44

```
Met Met Val Val Arg Arg Phe Leu Ser Ala Ser Arg Tyr Asn Met Pro
 1               5                  10                  15

Ser Phe Arg Ser Arg Pro Val Leu Leu Ala Ile Phe Gly Val Cys Ala
            20                  25                  30

Thr Val Ile Trp Tyr Leu Phe Phe Phe Gln Ser Asn Leu Lys Met Ala
        35                  40                  45

Asp Leu Lys Lys Val Ala Thr Ser Arg Tyr Leu Ser Gln Glu Pro Ser
    50                  55                  60

Leu Ser Glu Leu Met Ser Asn Val Lys Ile Lys Pro Ile Glu Glu Thr
65                  70                  75                  80

Pro Val Ser Pro Leu Glu Leu Ile Pro Asp Ile Glu Ile Ser Thr Arg
                85                  90                  95

Lys Lys Tyr Asp Ala Ser Trp Asp Leu Leu Phe Arg Gly Arg Lys Tyr
            100                 105                 110

Lys Ser Phe Asn Asp Tyr Asp Leu His Thr Lys Cys Glu Phe Tyr Phe
        115                 120                 125

Gln Asn Leu Tyr Asn Leu Asn Glu Asp Trp Thr Asn Asn Ile Arg Thr
    130                 135                 140

Phe Thr Phe Asp Ile Asn Asp Val Asp Thr Ser Thr Lys Ile Asp Ala
145                 150                 155                 160

Leu Lys Asp Ser Asp Gly Val Gln Leu Val Asp Glu Lys Ala Ile Arg
                165                 170                 175

Leu Tyr Lys Arg Thr His Asn Val Ala Leu Ala Thr Glu Arg Leu Arg
            180                 185                 190

Leu Tyr Asp Lys Cys Phe Val Asn Ser Pro Gly Ser Asn Pro Leu Lys
        195                 200                 205

Met Asp His Leu Phe Arg Ser Asn Lys Lys Ser Lys Thr Thr Ala Leu
    210                 215                 220

Asp Asp Glu Val Thr Gly Asn Arg Asp Thr Phe Thr Lys Thr Lys Lys
225                 230                 235                 240

Thr Ser Phe Leu Ser Asp Met Asp Thr Ser Ser Phe Gln Lys Tyr Asp
                245                 250                 255

Gln Trp Asp Phe Glu His Arg Met Phe Pro Met Ile Pro Tyr Phe Glu
```

```
            260                 265                 270
Glu His Asn Phe Thr Asn Val Met Pro Ile Phe Thr Gly Ser Asn Gly
        275                 280                 285
Gly Glu Pro Leu Pro Gln Gly Lys Phe Pro Val Leu Asp Pro Lys Ser
    290                 295                 300
Gly Glu Leu Leu Arg Val Glu Thr Phe Arg Tyr Asp Lys Ser Lys Ser
305                 310                 315                 320
Leu Trp Lys Asn Trp Asn Asp Met Ser Ser Ala Ser Gly Lys Arg Gly
                325                 330                 335
Ile Ile Leu Ala Ala Gly Asp Gly Gln Val Asp Gln Cys Ile Arg Leu
            340                 345                 350
Ile Ala Thr Leu Arg Ala Gln Gly Asn Ala Leu Pro Ile Gln Ile Ile
        355                 360                 365
His Asn Asn Gln Leu Asn Glu Lys Ser Val Lys Leu Leu Ser Glu Ala
    370                 375                 380
Ala Lys Ser Thr Glu Phe Ser Ser Gly Arg Ala Gln Ser Leu Trp Leu
385                 390                 395                 400
Val Asn Val Gly Pro Thr Leu Glu Ser Ser Met Lys Ser Asn Phe Gly
                405                 410                 415
Arg Phe Lys Asn Lys Trp Leu Ser Val Ile Phe Asn Thr Phe Glu Glu
            420                 425                 430
Phe Ile Phe Ile Asp Thr Asp Ala Ile Ser Tyr Ile Asn Met Ala Asp
        435                 440                 445
Tyr Phe Asn Phe Lys Glu Tyr Lys Ser Thr Gly Thr Leu Phe Phe Lys
    450                 455                 460
Asp Arg Ser Leu Ala Ile Gly Thr Glu Gln Lys Cys Gly Pro Leu Phe
465                 470                 475                 480
Glu Thr Leu Glu Pro Arg Ile Leu Glu Met Tyr Tyr Phe Asn Thr Leu
                485                 490                 495
Pro Met Ile Asn Gly Asp Tyr Val Glu Gln Gln Cys Met Gly Met Leu
            500                 505                 510
Thr Pro Glu Glu Lys Val Tyr Lys Arg Phe Phe Glu Val Gly His Gln
        515                 520                 525
His Asn Leu Glu Ser Gly Leu Leu Ala Ile Asn Lys Asn Glu His Ile
    530                 535                 540
Met Gly Leu Val Thr Ala Thr Val Leu Asn Ile Ala Pro Lys Val Gly
545                 550                 555                 560
Gly Cys Gly Trp Gly Asp Lys Glu Phe Phe Trp Leu Gly Leu Leu Val
                565                 570                 575
Ala Gly Gln Arg Tyr Ser Ile Tyr Asp Ile Asp Ala Ser Ala Ile Gly
            580                 585                 590
Val Pro Gln Gln Lys Gln Ser Ile Ala Asn Gly Asp Glu Phe Asp Glu
        595                 600                 605
Tyr Arg Ile Cys Ser Leu Gln Val Ala His Tyr Thr Ser Tyr Asp Gly
    610                 615                 620
His Leu Leu Trp Ile Asn Gly Gly Ser Gln Tyr Cys Lys Lys Pro Glu
```

625 630 635 640

Thr Phe Glu Gly Asp Trp Thr Asn Ile Lys Glu Leu Arg Glu Ser Tyr
645 650 655

Ser Asp Asp Lys Glu Lys Ala Leu Lys Ala Tyr Ser Asp Tyr Val Lys
660 665 670

Val Glu Ala Ala Ile Val Pro Asp Ser Arg Ser Asn Gly Trp Gly Arg
675 680 685

Asp Asp Gln Arg Cys Lys Gly Tyr Phe Trp Cys Gly Lys Phe Thr Ser
690 695 700

Lys Leu Lys Pro Tyr Thr Tyr Asn Thr Val Val Thr Lys Gly Asp Leu
705 710 715 720

Ile Arg Phe Gly Asp Glu Glu Ile Glu Ser Ile Ser Lys Ile Asn Lys
725 730 735

Ile Trp Asn Asp Ala Ile Ile Pro Asp Gly Ala
740 745

<210> 45

<211> 1968

<212> ADN

<213> Mus musculus

<220>

<221> CDS

<222> (1)..(1965)

<400> 45

```
atg ccc gtg ggg ggc ctg ttg ccg ctc ttc agt agc cct ggg ggc ggc   48
Met Pro Val Gly Gly Leu Leu Pro Leu Phe Ser Ser Pro Gly Gly Gly
  1               5                  10                  15

ggc ctg ggc agt ggc ctg ggc ggg ggg ctt ggc ggc ggg agg aag ggg   96
Gly Leu Gly Ser Gly Leu Gly Gly Gly Leu Gly Gly Gly Arg Lys Gly
            20                  25                  30

tct ggc ccc gct gcc ttc cgc ctc acc gag aag ttc gtg ctg ctg ctg   144
Ser Gly Pro Ala Ala Phe Arg Leu Thr Glu Lys Phe Val Leu Leu Leu
        35                  40                  45

gtg ttc agc gcc ttc atc acg ctc tgc ttc ggg gca atc ttc ttc ctg   192
Val Phe Ser Ala Phe Ile Thr Leu Cys Phe Gly Ala Ile Phe Phe Leu
    50                  55                  60

cct gac tcc tcc aag ctg ctc agc ggg gtc ctg ttc cac tcc aac cct   240
Pro Asp Ser Ser Lys Leu Leu Ser Gly Val Leu Phe His Ser Asn Pro
65                  70                  75                  80

gcc ttg cag ccg ccg gcg gag cac aag ccc ggg ctc ggg gcg cgt gcg   288
Ala Leu Gln Pro Pro Ala Glu His Lys Pro Gly Leu Gly Ala Arg Ala
                85                  90                  95

gag gat gcc gcc gag ggg aga gtc cgg cac cgc gag gaa ggc gcg cct   336
Glu Asp Ala Ala Glu Gly Arg Val Arg His Arg Glu Glu Gly Ala Pro
            100                 105                 110

ggg gac cct gga gct gga ctg gaa gac aac tta gcc agg atc cgc gaa   384
Gly Asp Pro Gly Ala Gly Leu Glu Asp Asn Leu Ala Arg Ile Arg Glu
        115                 120                 125

aac cac gag cgg gct ctc agg gaa gcc aag gag acc ctg cag aag ctg   432
Asn His Glu Arg Ala Leu Arg Glu Ala Lys Glu Thr Leu Gln Lys Leu
    130                 135                 140
```

```
ccg gag gag atc caa aga gac att ctg ctg gag aag gaa aag gtg gcc   480
Pro Glu Glu Ile Gln Arg Asp Ile Leu Leu Glu Lys Glu Lys Val Ala
145                 150                 155                 160

cag gac cag ctg cgt gac aag gat ctg ttt agg ggc ttg ccc aag gtg   528
Gln Asp Gln Leu Arg Asp Lys Asp Leu Phe Arg Gly Leu Pro Lys Val
                165                 170                 175

gac ttc ctg ccc ccc gtc ggg gta gag aac cgg gag ccc gct gac gcc   576
Asp Phe Leu Pro Pro Val Gly Val Glu Asn Arg Glu Pro Ala Asp Ala
            180                 185                 190

acc atc cgt gag aag agg gca aag atc aaa gag atg atg acc cat gct   624
Thr Ile Arg Glu Lys Arg Ala Lys Ile Lys Glu Met Met Thr His Ala
        195                 200                 205

tgg aat aat tat aaa cgc tat gcg tgg ggc ttg aac gaa ctg aaa cct   672
Trp Asn Asn Tyr Lys Arg Tyr Ala Trp Gly Leu Asn Glu Leu Lys Pro
    210                 215                 220

ata tca aaa gaa ggc cat tca agc agt ttg ttt ggc aac atc aaa gga   720
Ile Ser Lys Glu Gly His Ser Ser Ser Leu Phe Gly Asn Ile Lys Gly
225                 230                 235                 240

gct aca ata gta gat gcc ctg gat acc ctt ttc att atg ggc atg aag   768
Ala Thr Ile Val Asp Ala Leu Asp Thr Leu Phe Ile Met Gly Met Lys
                245                 250                 255

act gaa ttt caa gaa gct aaa tcg tgg att aaa aaa tat tta gat ttt   816
Thr Glu Phe Gln Glu Ala Lys Ser Trp Ile Lys Lys Tyr Leu Asp Phe
            260                 265                 270

aat gtg aat gct gaa gtt tct gtt ttt gaa gtc aac ata cgc ttc gtc   864
Asn Val Asn Ala Glu Val Ser Val Phe Glu Val Asn Ile Arg Phe Val
        275                 280                 285

ggt gga ctg ctg tca gcc tac tat ttg tcc gga gag gag ata ttt cga   912
Gly Gly Leu Leu Ser Ala Tyr Tyr Leu Ser Gly Glu Glu Ile Phe Arg
    290                 295                 300

aag aaa gca gtg gaa ctt ggg gta aaa ttg cta cct gca ttt cat act   960
Lys Lys Ala Val Glu Leu Gly Val Lys Leu Leu Pro Ala Phe His Thr
305                 310                 315                 320

ccc tct gga ata cct tgg gca ttg ctg aat atg aaa agt ggg atc ggg  1008
Pro Ser Gly Ile Pro Trp Ala Leu Leu Asn Met Lys Ser Gly Ile Gly
                325                 330                 335

cgg aac tgg ccc tgg gcc tct gga ggc agc agt atc ctg gcc gaa ttt  1056
Arg Asn Trp Pro Trp Ala Ser Gly Gly Ser Ser Ile Leu Ala Glu Phe
            340                 345                 350

gga act ctg cat tta gag ttt atg cac ttg tcc cac tta tca gga gac  1104
Gly Thr Leu His Leu Glu Phe Met His Leu Ser His Leu Ser Gly Asp
        355                 360                 365

cca gtc ttt gcc gaa aag gtt atg aaa att cga aca gtg ttg aac aaa  1152
Pro Val Phe Ala Glu Lys Val Met Lys Ile Arg Thr Val Leu Asn Lys
    370                 375                 380

ctg gac aaa cca gaa ggc ctt tat cct aac tat ctg aac ccc agt agt  1200
Leu Asp Lys Pro Glu Gly Leu Tyr Pro Asn Tyr Leu Asn Pro Ser Ser
385                 390                 395                 400

gga cag tgg ggt caa cat cat gtg tcg gtt gga gga ctt gga gac agc  1248
Gly Gln Trp Gly Gln His His Val Ser Val Gly Gly Leu Gly Asp Ser
                405                 410                 415
```

```
ttt tat gaa tat ttg ctt aag gcg tgg tta atg tct gac aag aca gat   1296
Phe Tyr Glu Tyr Leu Leu Lys Ala Trp Leu Met Ser Asp Lys Thr Asp
            420                 425                 430

ctc gaa gcc aag aag atg tat ttt gat gct gtt cag gcc atc gag act   1344
Leu Glu Ala Lys Lys Met Tyr Phe Asp Ala Val Gln Ala Ile Glu Thr
        435                 440                 445

cac ttg atc cgc aag tca agt ggg gga cta acg tac atc gca gag tgg   1392
His Leu Ile Arg Lys Ser Ser Gly Gly Leu Thr Tyr Ile Ala Glu Trp
    450                 455                 460

aag ggg ggc ctc ctg gaa cac aag atg ggc cac ctg acg tgc ttt gca   1440
Lys Gly Gly Leu Leu Glu His Lys Met Gly His Leu Thr Cys Phe Ala
465                 470                 475                 480

gga ggc atg ttt gca ctt ggg gca gat gga gct ccg gaa gcc cgg gcc   1488
Gly Gly Met Phe Ala Leu Gly Ala Asp Gly Ala Pro Glu Ala Arg Ala
                485                 490                 495

caa cac tac ctt gaa ctc gga gct gaa att gcc cgc act tgt cat gaa   1536
Gln His Tyr Leu Glu Leu Gly Ala Glu Ile Ala Arg Thr Cys His Glu
            500                 505                 510

tct tat aat cgt aca tat gtg aag ttg gga ccg gaa gcg ttt cga ttt   1584
Ser Tyr Asn Arg Thr Tyr Val Lys Leu Gly Pro Glu Ala Phe Arg Phe
        515                 520                 525

gat ggc ggt gtg gaa gct att gcc acg agg caa aat gaa aag tat tac   1632
Asp Gly Gly Val Glu Ala Ile Ala Thr Arg Gln Asn Glu Lys Tyr Tyr
    530                 535                 540

atc tta cgg ccc gag gtc atc gag aca tac atg tac atg tgg cga ctg   1680
Ile Leu Arg Pro Glu Val Ile Glu Thr Tyr Met Tyr Met Trp Arg Leu
545                 550                 555                 560

act cac gac ccc aag tac agg acc tgg gcc tgg gaa gcc gtg gag gct   1728
Thr His Asp Pro Lys Tyr Arg Thr Trp Ala Trp Glu Ala Val Glu Ala
                565                 570                 575

cta gaa agt cac tgc aga gtg aac gga ggc tac tca ggc tta cgg gat   1776
Leu Glu Ser His Cys Arg Val Asn Gly Gly Tyr Ser Gly Leu Arg Asp
            580                 585                 590

gtt tac att gcc cgt gag agt tat gac gat gtc cag caa agt ttc ttc   1824
Val Tyr Ile Ala Arg Glu Ser Tyr Asp Asp Val Gln Gln Ser Phe Phe
        595                 600                 605

ctg gca gag aca ctg aag tat ttg tac ttg ata ttt tcc gat gat gac   1872
Leu Ala Glu Thr Leu Lys Tyr Leu Tyr Leu Ile Phe Ser Asp Asp Asp
    610                 615                 620

ctt ctt cca cta gaa cac tgg atc ttc aac acc gag gct cat cct ttc   1920
Leu Leu Pro Leu Glu His Trp Ile Phe Asn Thr Glu Ala His Pro Phe
625                 630                 635                 640

cct ata ctc cgt gaa cag aag aag gaa att gat ggc aaa gag aaa tga   1968
Pro Ile Leu Arg Glu Gln Lys Lys Glu Ile Asp Gly Lys Glu Lys
                645                 650                 655
```

<210> 46

<211> 655

<212> PRT

<213> Mus musculus

<400> 46

Met Pro Val Gly Gly Leu Leu Pro Leu Phe Ser Ser Pro Gly Gly Gly
1 5 10 15

Gly Leu Gly Ser Gly Leu Gly Gly Gly Leu Gly Gly Gly Arg Lys Gly
20 25 30

Ser Gly Pro Ala Ala Phe Arg Leu Thr Glu Lys Phe Val Leu Leu Leu
35 40 45

Val Phe Ser Ala Phe Ile Thr Leu Cys Phe Gly Ala Ile Phe Phe Leu
50 55 60

Pro Asp Ser Ser Lys Leu Leu Ser Gly Val Leu Phe His Ser Asn Pro
65 70 75 80

Ala Leu Gln Pro Pro Ala Glu His Lys Pro Gly Leu Gly Ala Arg Ala
85 90 95

Glu Asp Ala Ala Glu Gly Arg Val Arg His Arg Glu Glu Gly Ala Pro
100 105 110

Gly Asp Pro Gly Ala Gly Leu Glu Asp Asn Leu Ala Arg Ile Arg Glu
115 120 125

Asn His Glu Arg Ala Leu Arg Glu Ala Lys Glu Thr Leu Gln Lys Leu
130 135 140

Pro Glu Glu Ile Gln Arg Asp Ile Leu Leu Glu Lys Glu Lys Val Ala
145 150 155 160

Gln Asp Gln Leu Arg Asp Lys Asp Leu Phe Arg Gly Leu Pro Lys Val
165 170 175

Asp Phe Leu Pro Pro Val Gly Val Glu Asn Arg Glu Pro Ala Asp Ala
180 185 190

Thr Ile Arg Glu Lys Arg Ala Lys Ile Lys Glu Met Met Thr His Ala
195 200 205

Trp Asn Asn Tyr Lys Arg Tyr Ala Trp Gly Leu Asn Glu Leu Lys Pro
210 215 220

Ile Ser Lys Glu Gly His Ser Ser Ser Leu Phe Gly Asn Ile Lys Gly
225 230 235 240

Ala Thr Ile Val Asp Ala Leu Asp Thr Leu Phe Ile Met Gly Met Lys
245 250 255

Thr Glu Phe Gln Glu Ala Lys Ser Trp Ile Lys Lys Tyr Leu Asp Phe
260 265 270

Asn Val Asn Ala Glu Val Ser Val Phe Glu Val Asn Ile Arg Phe Val
275 280 285

Gly Gly Leu Leu Ser Ala Tyr Tyr Leu Ser Gly Glu Glu Ile Phe Arg
290 295 300

Lys Lys Ala Val Glu Leu Gly Val Lys Leu Leu Pro Ala Phe His Thr
305 310 315 320

Pro Ser Gly Ile Pro Trp Ala Leu Leu Asn Met Lys Ser Gly Ile Gly
325 330 335

Arg Asn Trp Pro Trp Ala Ser Gly Gly Ser Ser Ile Leu Ala Glu Phe
340 345 350

Gly Thr Leu His Leu Glu Phe Met His Leu Ser His Leu Ser Gly Asp
355 360 365

Pro Val Phe Ala Glu Lys Val Met Lys Ile Arg Thr Val Leu Asn Lys
370 375 380

Leu Asp Lys Pro Glu Gly Leu Tyr Pro Asn Tyr Leu Asn Pro Ser Ser
385 390 395 400

Gly Gln Trp Gly Gln His His Val Ser Val Gly Gly Leu Gly Asp Ser
405 410 415

Phe Tyr Glu Tyr Leu Leu Lys Ala Trp Leu Met Ser Asp Lys Thr Asp
420 425 430

Leu Glu Ala Lys Lys Met Tyr Phe Asp Ala Val Gln Ala Ile Glu Thr
435 440 445

His Leu Ile Arg Lys Ser Ser Gly Gly Leu Thr Tyr Ile Ala Glu Trp
450 455 460

Lys Gly Gly Leu Leu Glu His Lys Met Gly His Leu Thr Cys Phe Ala
465 470 475 480

Gly Gly Met Phe Ala Leu Gly Ala Asp Gly Ala Pro Glu Ala Arg Ala
485 490 495

Gln His Tyr Leu Glu Leu Gly Ala Glu Ile Ala Arg Thr Cys His Glu
500 505 510

Ser Tyr Asn Arg Thr Tyr Val Lys Leu Gly Pro Glu Ala Phe Arg Phe
515 520 525

Asp Gly Gly Val Glu Ala Ile Ala Thr Arg Gln Asn Glu Lys Tyr Tyr
530 535 540

Ile Leu Arg Pro Glu Val Ile Glu Thr Tyr Met Tyr Met Trp Arg Leu
545 550 555 560

Thr His Asp Pro Lys Tyr Arg Thr Trp Ala Trp Glu Ala Val Glu Ala
565 570 575

Leu Glu Ser His Cys Arg Val Asn Gly Gly Tyr Ser Gly Leu Arg Asp
580 585 590

Val Tyr Ile Ala Arg Glu Ser Tyr Asp Asp Val Gln Gln Ser Phe Phe
595 600 605

Leu Ala Glu Thr Leu Lys Tyr Leu Tyr Leu Ile Phe Ser Asp Asp Asp
610 615 620

Leu Leu Pro Leu Glu His Trp Ile Phe Asn Thr Glu Ala His Pro Phe
625 630 635 640

Pro Ile Leu Arg Glu Gln Lys Lys Glu Ile Asp Gly Lys Glu Lys
645 650 655

<210> 47

<211> 26

<212> ADN

<213> Secuencia artificial

<220>

<221> Fuente

<223> /nota="Descripción de secuencia artificial: Cebador"

<220>

<221> VARIANTE

<222> (9), (12), (18)

<223> /reemplazar="a" /reemplazar="t" /reemplazar="c" /reemplazar="g"

<400> 47

taytggmgng tngarcynga yathaa 26

REIVINDICACIONES

1. Una célula huésped de un hongo unicelular o filamentoso que no presenta actividad α1,6 manosiltransferasa con respecto al N-glucano en una glucoproteína, comprendiendo dicha célula huésped un ácido nucleico que codifica una proteína de fusión que comprende el dominio catalítico de una α-1,2-manosidasa fusionada en el extremo N-terminal con un péptido señal de dirección celular SEC12, VAN1 o MNN10 para la producción de una estructura de carbohidrato de $Man_5GlcNAc_2$, en donde $Man_5GlcNAc_2$ se produce como un sustrato productivo para GnTI *in vivo* dentro de la célula huésped a un rendimiento de al menos el 30 por ciento en moles.

2. La célula huésped de la reivindicación 1, en donde la célula huésped es capaz de producir *N*-glucanos de los que más del 30 % en moles comprenden seis o menos residuos de manosa.

3. La célula huésped de la reivindicación 1 o 2, en donde la célula huésped es capaz de producir *N*-glucanos de los que más del 30 % en moles comprenden tres o menos residuos de manosa.

4. La célula huésped de una cualquiera de las reivindicaciones 1 a 3, en donde la célula huésped es capaz de producir *N*-glucanos que comprenden uno o más azúcares seleccionados del grupo constituido por GlcNAc, galactosa, ácido siálico y fucosa.

5. La célula huésped de una cualquiera de las reivindicaciones 1 a 4, en donde la célula huésped es capaz de producir *N*-glucanos con al menos una rama oligosacarídica que comprende la estructura NeuNAcGalGlcNAcMan.

6. La célula huésped de una cualquiera de las reivindicaciones 1 a 5, en donde la célula huésped está seleccionada del grupo que consiste en *Pichia pastoris, Pichia finlandica, Pichia trehalophila, Pichia koclamae, Pichia membranaefaciens, Pichia opuntiae, Pichia thermotolerans, Pichia salictaria, Pichia guercuum, Pichia pijperi, Pichia stiptis, Pichia methanolica, Pichia* sp., *Saccharomyces cerevisiae, Saccharomyces* sp., *Hansenula polymorpha, Kluyveromyces* sp., *Kluyveromyces lactis, Candida albicans, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Trichoderma reesei, Chrysosporium lucknowense, Fusarium* sp., *Fusarium gramineum, Fusarium venenatum* y *Neurospora crassa.*

7. La célula huésped de la reivindicación 6, en donde el huésped es deficiente en la actividad de una o más enzimas seleccionadas del grupo que consiste en manosiltransferasas y fosfomanosiltransferasas.

8. La célula huésped de la reivindicación 7, en donde el huésped no expresa una enzima seleccionada del grupo que consiste en 1,6-manosiltransferasa, 1,3-manosiltransferasa y 1,2-manosiltransferasa.

9. La célula huésped de la reivindicación 8, en donde el huésped es un mutante *och1* de *P. pastoris*.

10. La célula huésped de una cualquiera de las reivindicaciones 1 a 9, en donde la célula huésped expresa adicionalmente actividades transportadoras de GnTI y UDP-GlcNAc.

11. La célula huésped de una cualquiera de las reivindicaciones 1 a 10, en la que la α-1,2-manosidasa se deriva de ratón, ser humano, *Lepidoptera, Aspergillus nidulans* o *Bacillus* sp., C. *elegans, D. melanogaster, P. citrinum* o *X. laevis*.

12. La célula huésped de una cualquiera de las reivindicaciones 1 a 11, en donde la célula huésped incluye adicionalmente un ácido nucleico que codifica un dominio catalítico que codifica una actividad glucosidasa, manosidasa o glucosiltransferasa derivada de un miembro del grupo que consiste en GnTI, GnTII, GnTIII, GnTIV, GnT V, GnT VI, GalT, Fucosiltransferasa y Sialiltransferasa y en donde el dominio catalítico tiene un pH óptimo dentro de 1,4 unidades de pH del pH óptimo medio de otras enzimas representativas en el orgánulo en el que se localiza la enzima, o tiene actividad óptima a un pH entre 5,1 y 8,0.

13. La célula huésped de la reivindicación 12, en la que la manosidasa está seleccionada del grupo que consiste en manosidasa IA de C. *elegans*, manosidasa IB de C. *elegans*, manosidasa IA de *D. melanogaster*, manosidasa IB de *H. sapiens*, manosidasa I de *P. citrinum*, manosidasa IA de ratón, manosidasa IB de ratón, manosidasa IA de *A. nidulans*, manosidasa IB de *A. nidulans*, manosidasa IC de *A. nidulans*, manosidasa II de ratón, manosidasa II de *C. elegans*, manosidasa II de *H. sapiens* y manosidasa III.

14. La célula huésped de la reivindicación 12 o 13, en la que la molécula de ácido nucleico codifica una o más enzimas seleccionadas del grupo que consiste en UDP-GlcNAc transferasa, UDP-galactosiltransferasa, GDP-fucosiltransferasa, CMP-sialiltransferasa, transportador de UP-GlcNAc, transportador de UDP-galactosa, transportador de GDP-fucosa, transportador de CMP-ácido siálico y nucleótido difosfatasas.

15. La célula huésped de la reivindicación 12, en donde la célula huésped expresa además actividades del transportador de GnTI y de UDP-GlcNAc.

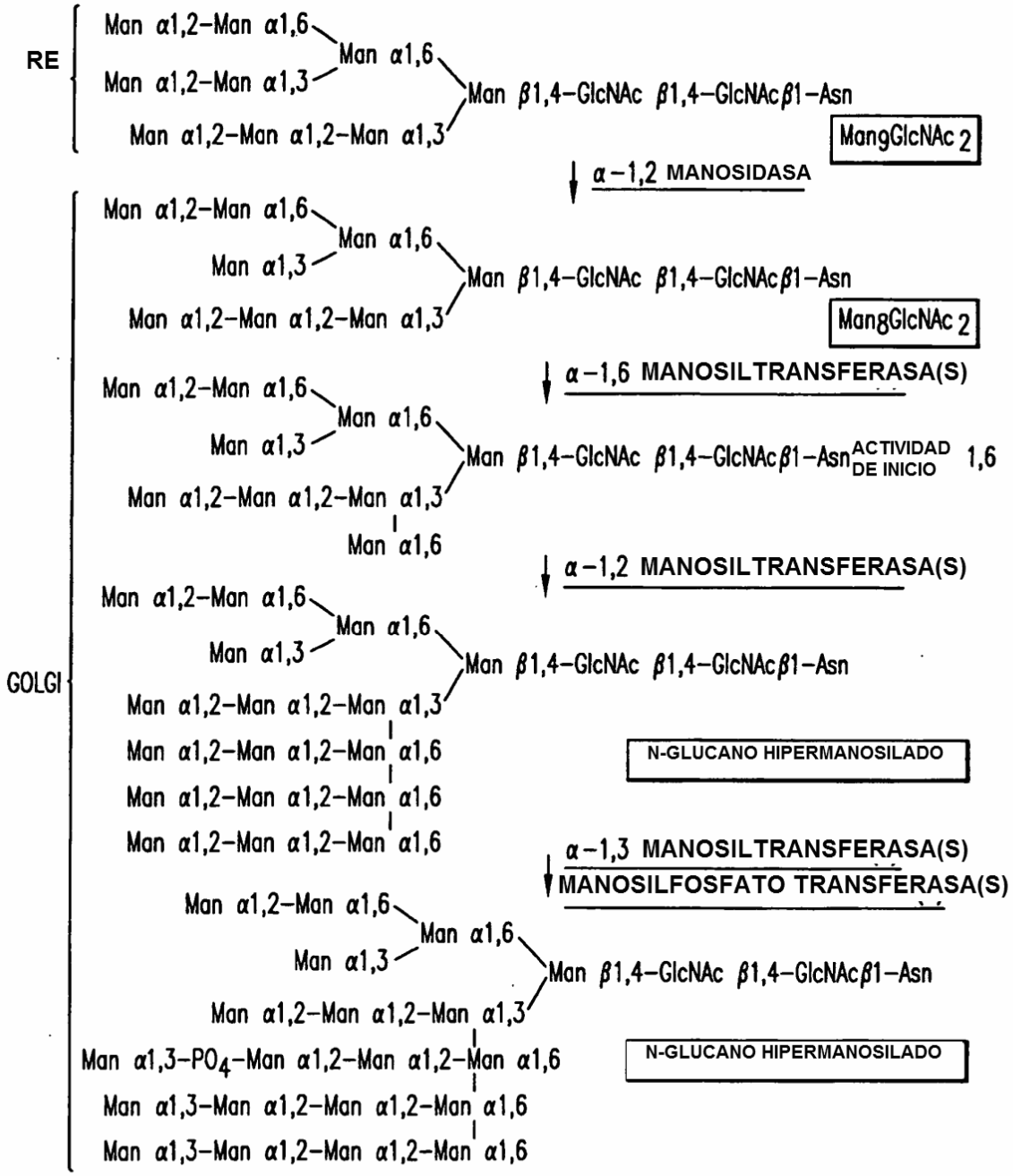

RE
Man α1,2–Man α1,6
Man α1,6
Man α1,2–Man α1,3
Man β1,4–GlcNAc β1,4–GlcNAcβ1–Asn
Man α1,2–Man α1,2–Man α1,3
ManGlcNAc 2
α–1,2 MANOSIDASA
Man8GlcNAc 2
α–1,6 MANOSILTRANSFERASA(S)
ACTIVIDAD DE INICIO
1,6
α–1,2 MANOSILTRANSFERASA(S)
GOLGI
N-GLUCANO HIPERMANOSILADO
α–1,3 MANOSILTRANSFERASA(S)
MANOSILFOSFATO TRANSFERASA(S)
Man α1,3–PO4–Man α1,2–Man α1,2–Man α1,6
Man α1,3–Man α1,2–Man α1,2–Man α1,6

FIG.1A

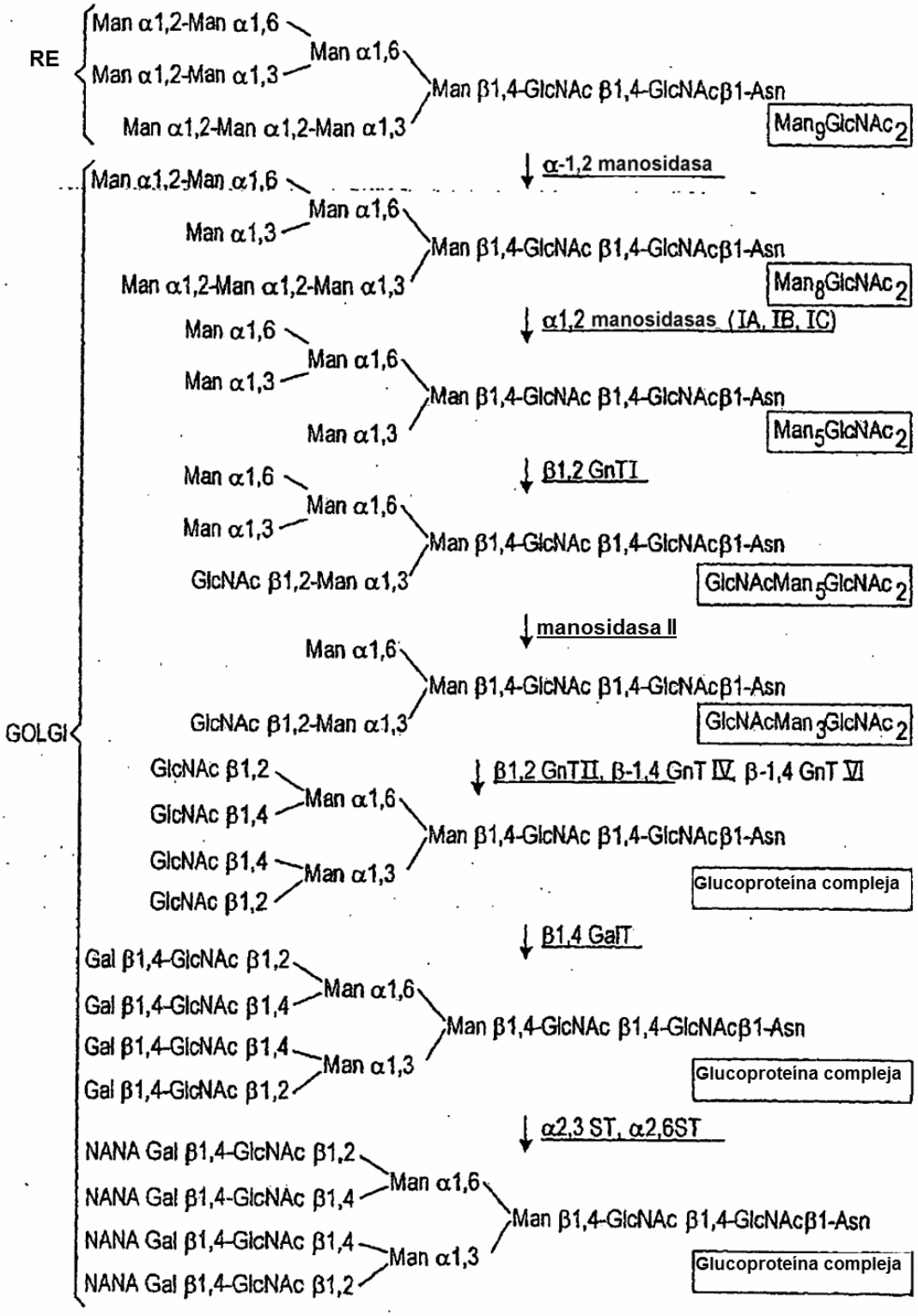

RE
Man α1,2-Man α1,6
Man α1,2-Man α1,3
Man α1,6
Man α1,2-Man α1,2-Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
Man9GlcNAc2
α-1,2 manosidasa
Man α1,2-Man α1,6
Man α1,3
Man α1,6
Man α1,2-Man α1,2-Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
Man8GlcNAc2
α1,2 manosidasas (IA, IB, IC)
Man α1,6
Man α1,3
Man α1,6
Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
Man5GlcNAc2
β1,2 GnTI
Man α1,6
Man α1,3
Man α1,6
GlcNAc β1,2-Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
GlcNAcMan5GlcNAc2
manosidasa II
Man α1,6
GlcNAc β1,2-Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
GlcNAcMan3GlcNAc2
GOLGI
β1,2 GnTII, β-1,4 GnT IV, β-1,4 GnT VI
GlcNAc β1,2
GlcNAc β1,4
GlcNAc β1,4
GlcNAc β1,2
Man α1,6
Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
Glucoproteína compleja
β1,4 GalT
Gal β1,4-GlcNAc β1,2
Gal β1,4-GlcNAc β1,4
Gal β1,4-GlcNAc β1,4
Gal β1,4-GlcNAc β1,2
Man α1,6
Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
Glucoproteína compleja
α2,3 ST, α2,6ST
NANA Gal β1,4-GlcNAc β1,2
NANA Gal β1,4-GlcNAc β1,4
NANA Gal β1,4-GlcNAc β1,4
NANA Gal β1,4-GlcNAc β1,2
Man α1,6
Man α1,3
Man β1,4-GlcNAc β1,4-GlcNAcβ1-Asn
Glucoproteína compleja

FIG. 1B

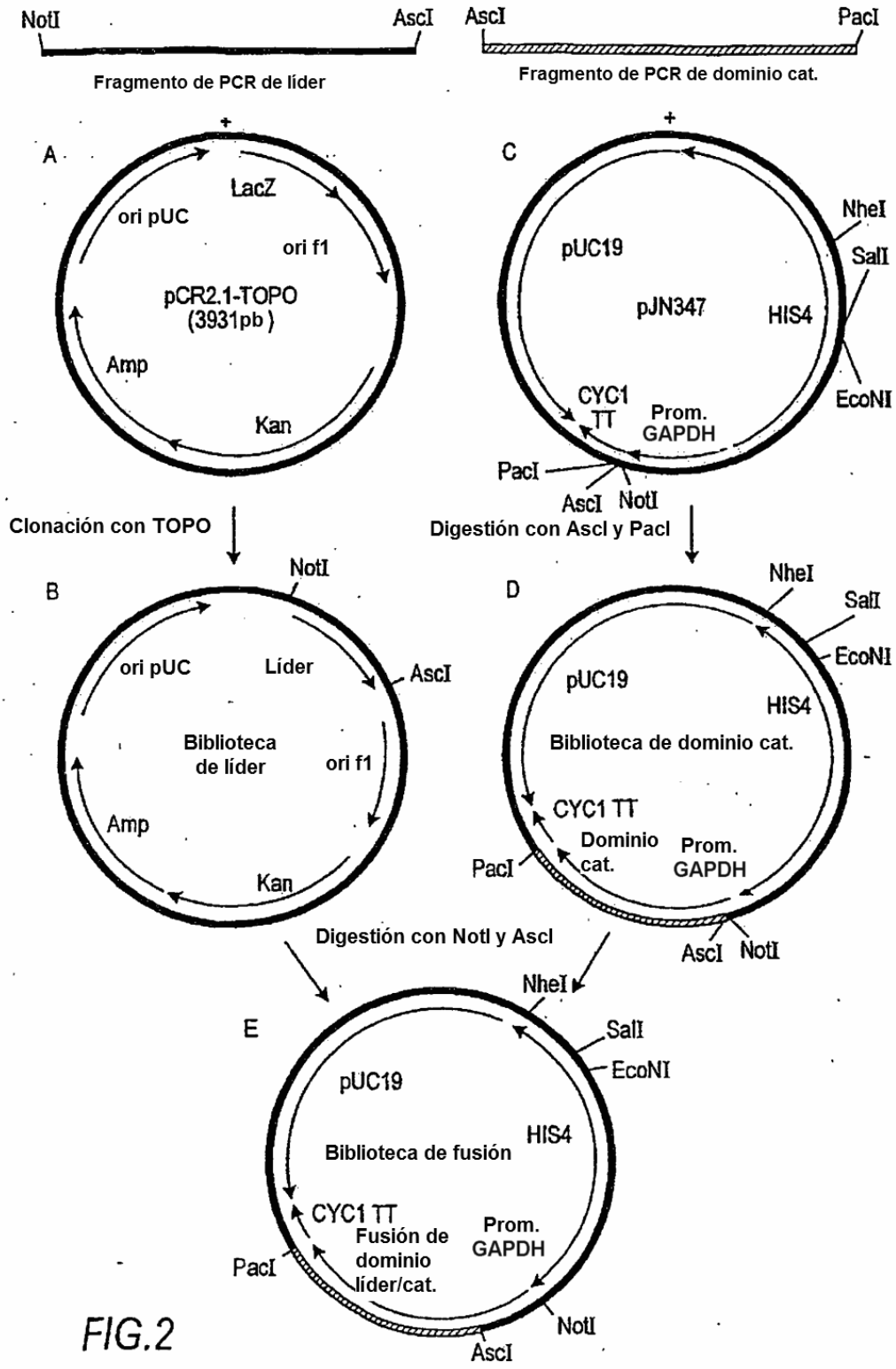

NotI
AscI
Fragmento de PCR de líder
AscI
PacI
Fragmento de PCR de dominio cat.
A
LacZ
ori pUC
ori f1
pCR2.1-TOPO
(3931pb)
Amp
Kan
C
pUC19
pJN347
HIS4
NheI
SalI
EcoNI
CYC1
TT
Prom.
GAPDH
PacI
AscI
NotI
Clonación con TOPO
Digestión con AscI y PacI
B
NotI
ori pUC
Líder
AscI
Biblioteca
de líder
ori f1
Amp
Kan
D
NheI
SalI
EcoNI
pUC19
HIS4
Biblioteca de dominio cat.
CYC1 TT
Dominio
cat.
Prom.
GAPDH
PacI
AscI
NotI
Digestión con NotI y AscI
E
NheI
SalI
EcoNI
pUC19
HIS4
Biblioteca de fusión
CYC1 TT
Fusión de
dominio
líder/cat.
Prom.
GAPDH
PacI
NotI
AscI

FIG.2

Fase abierta de lectura de alfa-1,2-manosida IA de *M. musculus.* Los dominios transmembrana y catalítico están destacados en negrita respectivamente. La secuencia de los cebadores usada para generar los truncamientos N-terminales están destacados por subrayado y el inicio de cada fragmento de proteína respectivo está indicado por una flecha.

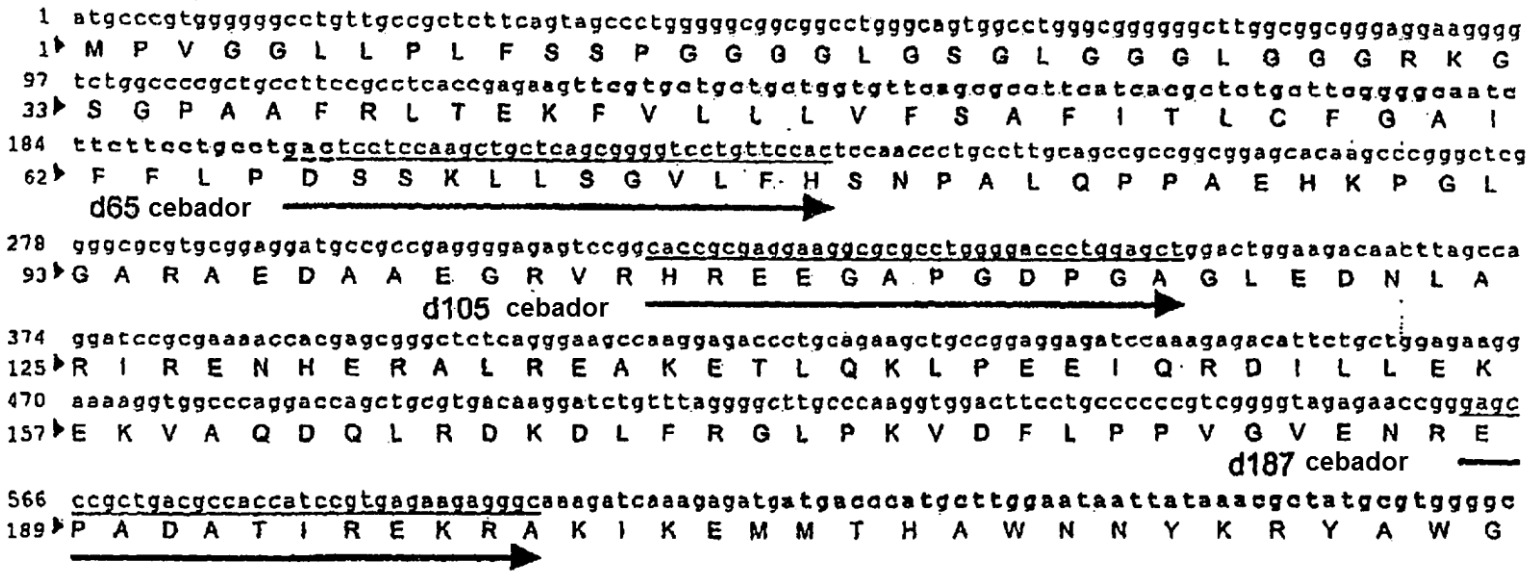


FIG. 3

ES 2 415 955 T3

```
655 ttgaacgaactgaaacctatatcaaagaaggccattcaagcagtttgtttggcaacatcaaaggagctacaatagtagatg
219▶ L N E L K P I S K E G H S S S L F G N I K G A T I V D
737 ccctggatacccttttcattattgggcatgaagactgaatttcaagaagctaaatcgtggattaaaaatatttagatttttaa
246▶A L D T L F I M G M K T E F Q E A K S W I K K Y L D F N
819 tgtgaatgctgaagtttctgtgtttttgaagtcaacatacgcttcgtggtggactgctgtcagcctactatttgtccggagag
273▶ V N A E V S V F E V N I R F V G G L L S A Y Y L S G E
901 gagatatttcgaaagaaagcagtggaaacttgtgggtaaaattgctacctgcatttcatactcctgtggaatacctgggcat
301▶E I F R K K A V E L G V K L L P A F H T P S G I P W A
983 tgctgaatatgaaaagtgggatcgggcggaactggcctctggaggcagcagtatcctgtggccgaatttggaactct
328▶L L N M K S G I G R N W P W A S G G S S I L A E F G T L
1065 gcatttagagtttatgtcactgtccactttatcaggagacccagtctttgccgaaaaggttatgaaaattcgaacagtgttg
355▶ H L E F M H H L S H L S G D P V F A E K V M K I R T V L
1147 aacaaactggacaaaccagaaggcctttatcctaactatatctgaaccccagtagtggacagtgggtcaacatcatgtctggg
383▶N K L D K P E G L Y P N Y L N P S S G Q W G Q H H V S
1229 ttggaggacttggagacagctttttatgaatatttgcttaaggcgtggttaatgtctgacaagacagatctcgaagccaagaa
410▶V G G L G D S F Y E Y L L K A W L M S D K T D L E A K K
1311 gatgtatttttgatgctgttcaggccatcgagactcactgatccgcaagtcaagtgggggcaatcaagtacatcgcagagtgg
437▶ M Y F D A V Q A I E T H L I R K S S G G L T Y I A E W
1393 aaggggggcctcctggaacacaaagatgggccacctgacgtgctttgcaggaggcatgtttgcacttgggggcagatggagctc
465▶ K G G L L E H K M G H L T C F A G G M F A L G A D G A
1475 cggaagccgggccaacactacctgaacttgaagctgaaattgccgcacttgtcatgaatcttataatcgtacatatgt
492▶P E A R A Q H Y L E L G A E I A R T C H E S Y N R T Y V
1557 gaagttgggaccggaagcgtttcgatttgatggcggtgtggaagctattgccacgaggcaaaatgaaaagtattacatctta
519▶ K L G P E A F R F D G G V E A I A T R Q N E K Y Y I L
1639 cggcccgaggtcatcgagacatacatgtacatgtggcgatgactcacgacccaagtacaggacctgggcctgggaagccg
547▶R P E V I E T Y M Y W R L T H D P K Y R T W A W E A
1721 tggaggctctagaaagtcactgcagagtgaacggaggctactcaggcttacgggatgtttacattgccgtgagagttatga
574▶V E A L E S H C R V N G G Y S G L R D V Y I A R E S Y D
1803 cgatgtccagcaaagtttctcctgggcagagacactgaagtatttgtacttgatatttcacgatgatgacctttctcaacta
601▶ D V Q Q S F F L A E T L K Y L Y L I F S D D D L L P L
1885 gaacactggatcttcaaacacgaggctcatcctttcccatactccctgtgaacagaaggaaattgatggcaaagagaaatga
629▶E H W I F N T E A H P F P I L R E Q K K E I D G K E K
```

FIG. 3 CONTINUACIÓN

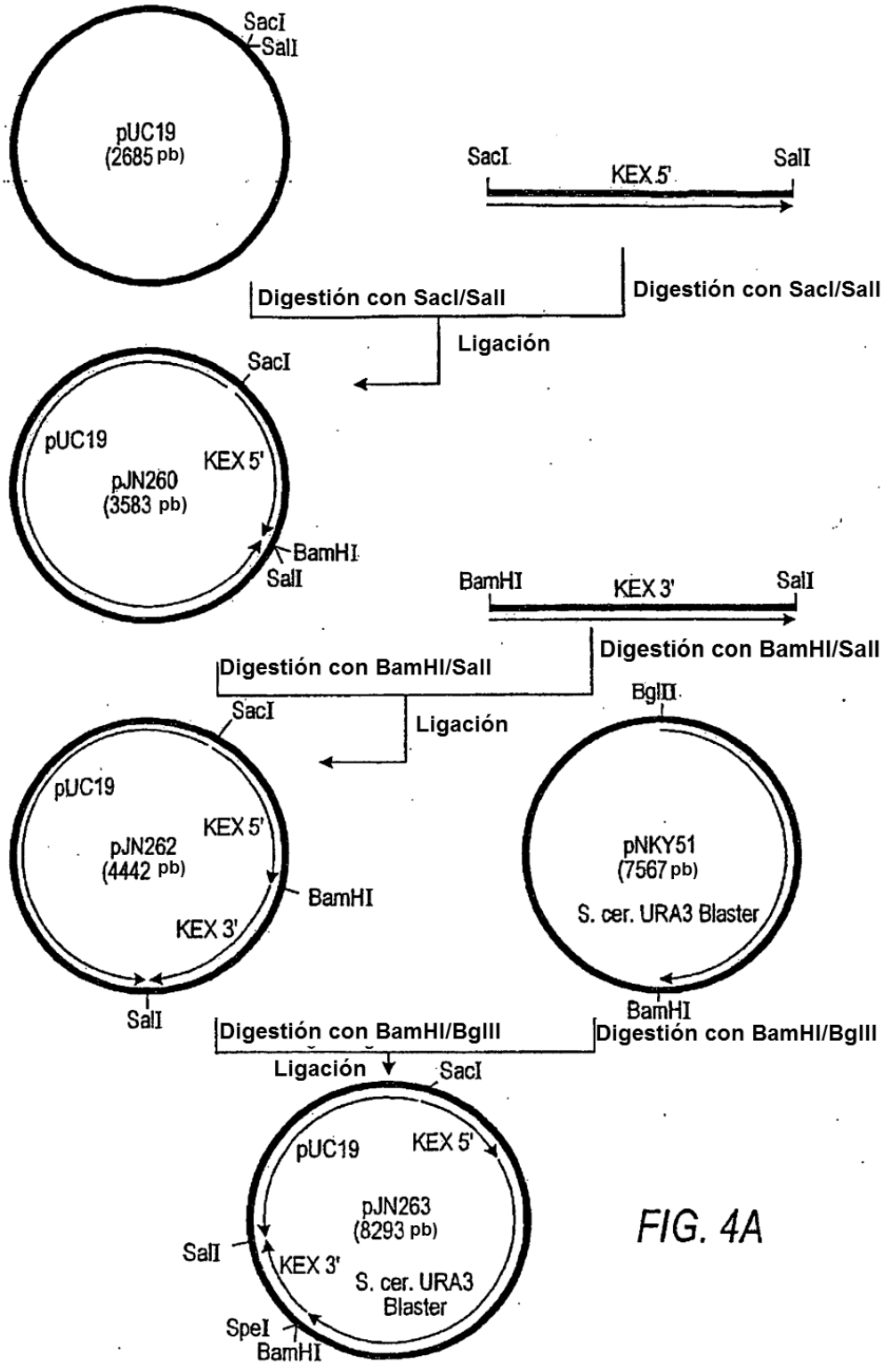

SacI
SalI
pUC19
(2685 pb)
SacI
KEX 5'
SalI
Digestión con SacI/SalI
Digestión con SacI/SalI
Ligación
SacI
pUC19
KEX 5'
pJN260
(3583 pb)
BamHI
SalI
BamHI
KEX 3'
SalI
Digestión con BamHI/SalI
Digestión con BamHI/SalI
Ligación
SacI
BglII
pUC19
KEX 5'
pJN262
(4442 pb)
BamHI
KEX 3'
SalI
pNKY51
(7567 pb)
S. cer. URA3 Blaster
BamHI
Digestión con BamHI/BglII
Digestión con BamHI/BglII
Ligación
SacI
pUC19
KEX 5'
pJN263
(8293 pb)
SalI
KEX 3'
S. cer. URA3
Blaster
SpeI
BamHI

FIG. 4A

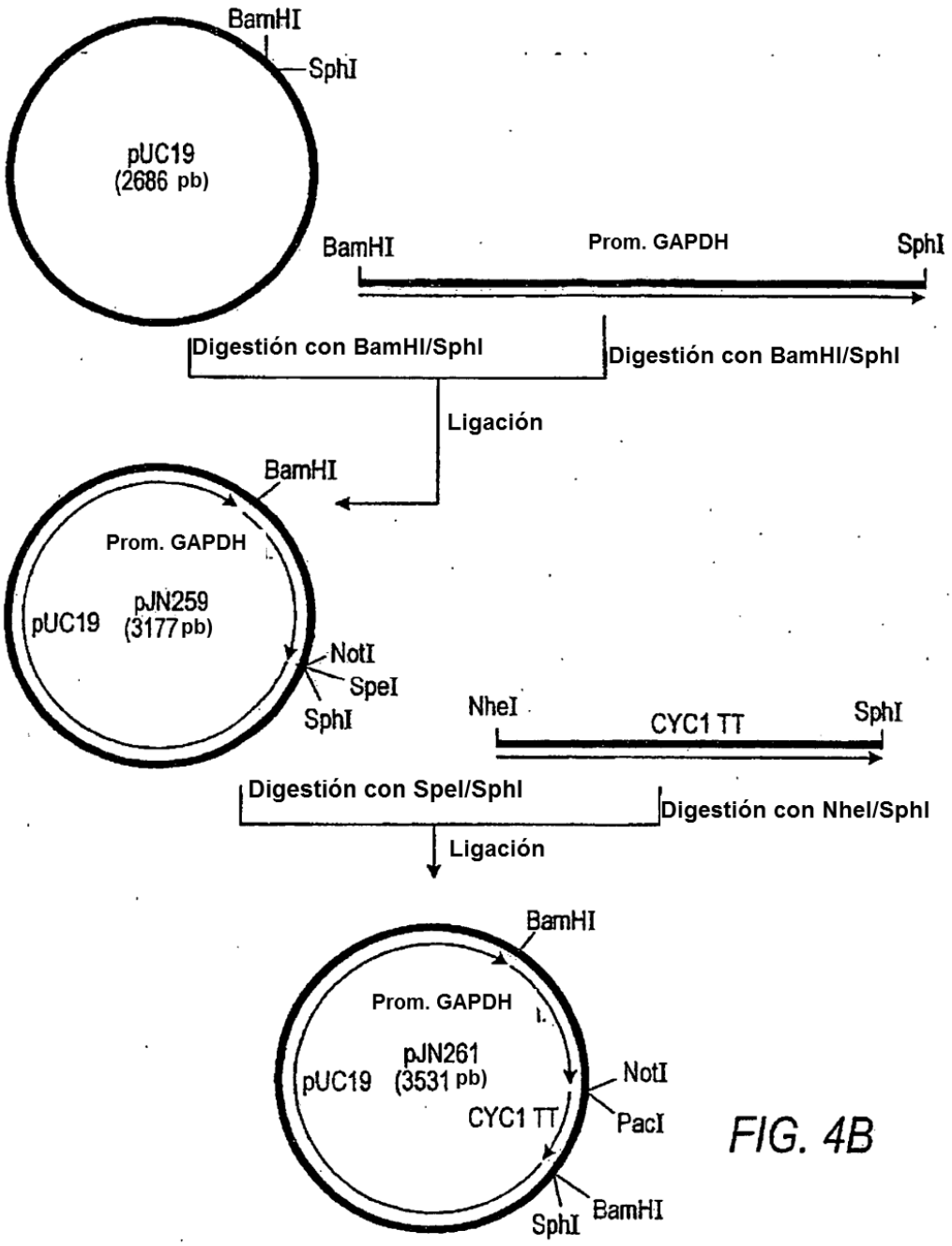

BamHI
SphI
pUC19
(2686 pb)
BamHI
Prom. GAPDH
SphI
Digestión con BamHI/SphI
Digestión con BamHI/SphI
Ligación
BamHI
Prom. GAPDH
pUC19
pJN259
(3177 pb)
NotI
SpeI
SphI
NheI
CYC1 TT
SphI
Digestión con SpeI/SphI
Digestión con NheI/SphI
Ligación
BamHI
Prom. GAPDH
pUC19
pJN261
(3531 pb)
NotI
PacI
CYC1 TT
SphI
BamHI

FIG. 4B

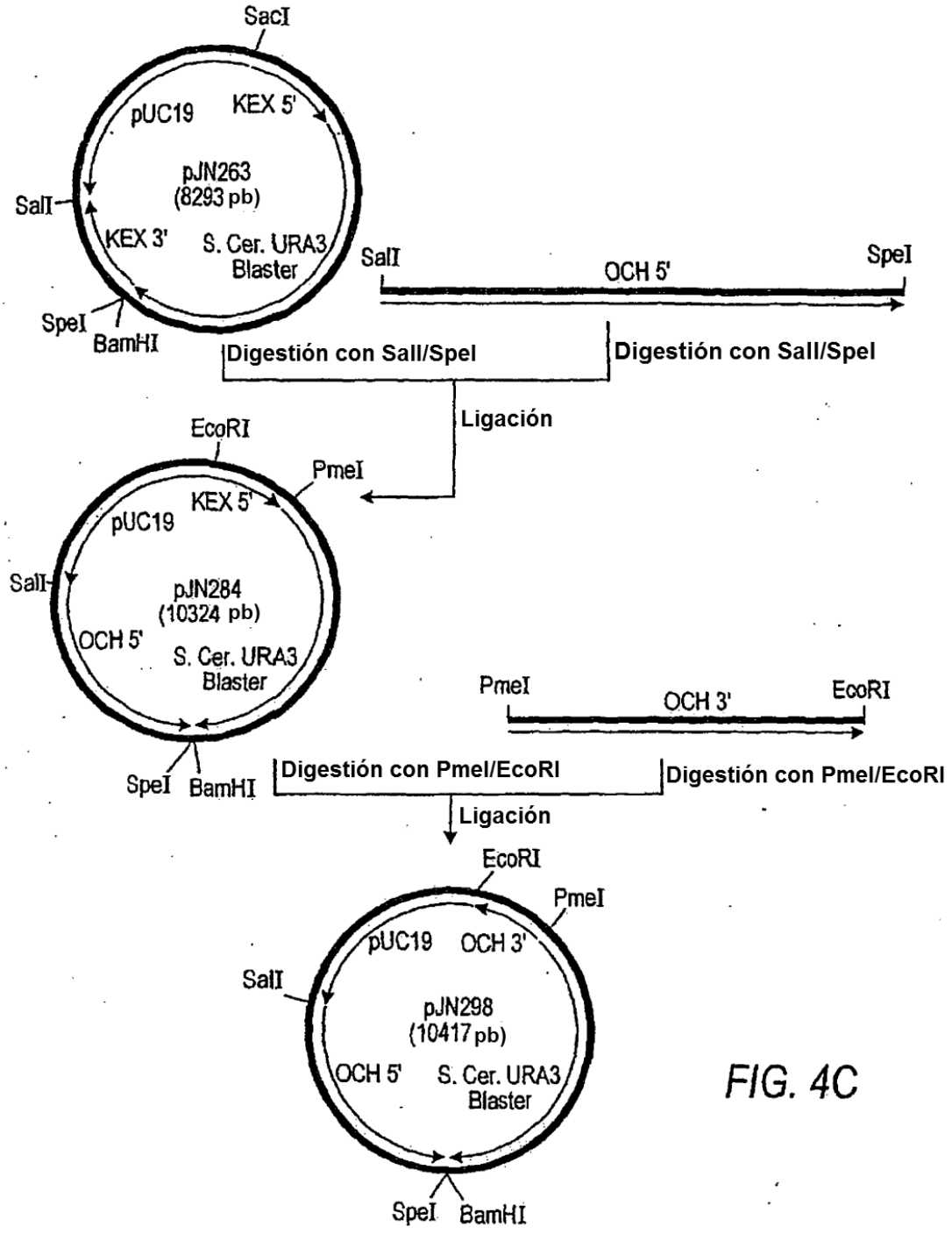

SacI
pUC19
KEX 5'
pJN263
(8293 pb)
SalI
KEX 3'
S. Cer. URA3
Blaster
SpeI
BamHI
SalI
OCH 5'
SpeI
Digestión con SalI/SpeI
Digestión con SalI/SpeI
Ligación
EcoRI
PmeI
KEX 5'
pUC19
pJN284
(10324 pb)
SalI
OCH 5'
S. Cer. URA3
Blaster
SpeI
BamHI
PmeI
OCH 3'
EcoRI
Digestión con PmeI/EcoRI
Digestión con PmeI/EcoRI
Ligación
EcoRI
PmeI
pUC19
OCH 3'
SalI
pJN298
(10417 pb)
OCH 5'
S. Cer. URA3
Blaster
SpeI
BamHI

FIG. 4C

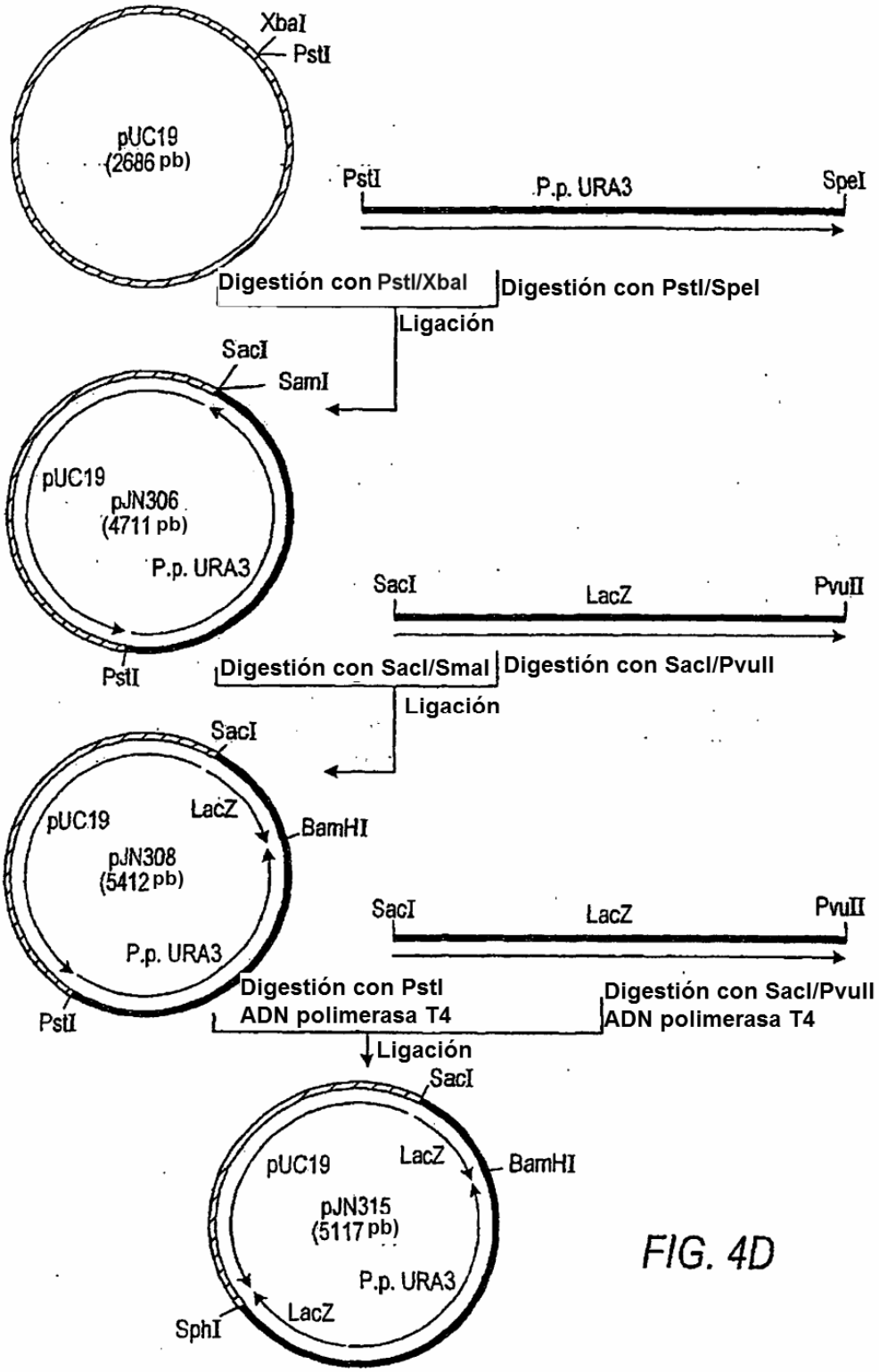

XbaI
PstI
pUC19
(2686 pb)
PstI
P.p. URA3
SpeI
Digestión con PstI/XbaI
Digestión con PstI/SpeI
Ligación
SacI
SamI
pUC19
pJN306
(4711 pb)
P.p. URA3
PstI
SacI
LacZ
PvuII
Digestión con SacI/SmaI
Digestión con SacI/PvuII
Ligación
SacI
pUC19
LacZ
BamHI
pJN308
(5412 pb)
P.p. URA3
PstI
SacI
LacZ
PvuII
Digestión con PstI
ADN polimerasa T4
Digestión con SacI/PvuII
ADN polimerasa T4
Ligación
SacI
pUC19
LacZ
BamHI
pJN315
(5117 pb)
P.p. URA3
LacZ
SphI

FIG. 4D

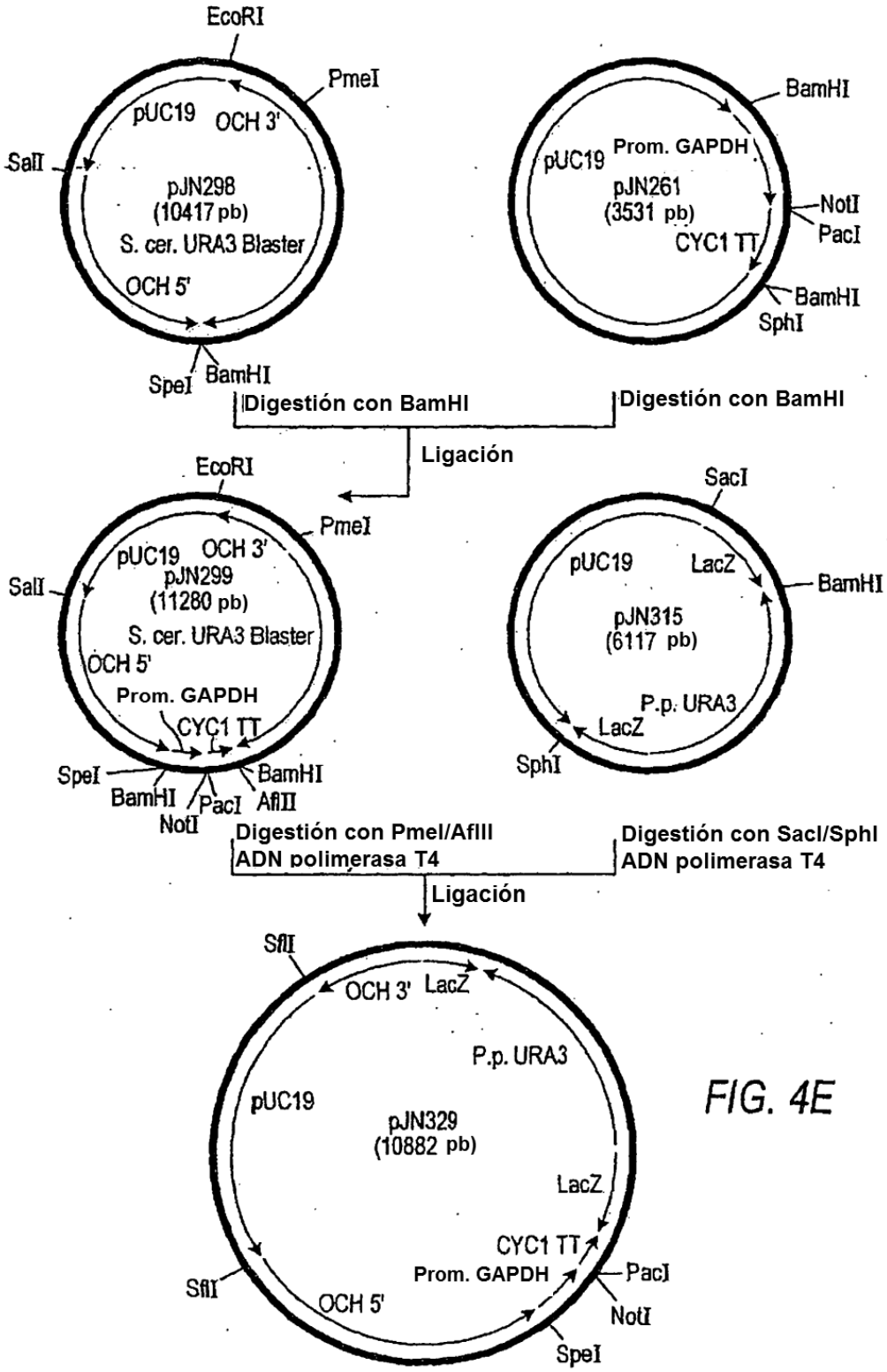

EcoRI
PmeI
pUC19
OCH 3'
SalI
pJN298
(10417 pb)
S. cer. URA3 Blaster
OCH 5'
SpeI
BamHI
BamHI
pUC19
Prom. GAPDH
pJN261
(3531 pb)
CYC1 TT
NotI
PacI
BamHI
SphI
Digestión con BamHI
Digestión con BamHI
Ligación
EcoRI
PmeI
SalI
pUC19
OCH 3'
pJN299
(11280 pb)
S. cer. URA3 Blaster
OCH 5'
Prom. GAPDH
CYC1 TT
SpeI
BamHI
BamHI
PacI
AflII
NotI
SacI
pUC19
LacZ
BamHI
pJN315
(6117 pb)
P.p. URA3
LacZ
SphI
Digestión con PmeI/AflII
ADN polimerasa T4
Digestión con SacI/SphI
ADN polimerasa T4
Ligación
SfiI
OCH 3'
LacZ
P.p. URA3
pUC19
pJN329
(10882 pb)
LacZ
CYC1 TT
Prom. GAPDH
PacI
NotI
SpeI
SfiI
OCH 5'

FIG. 4E

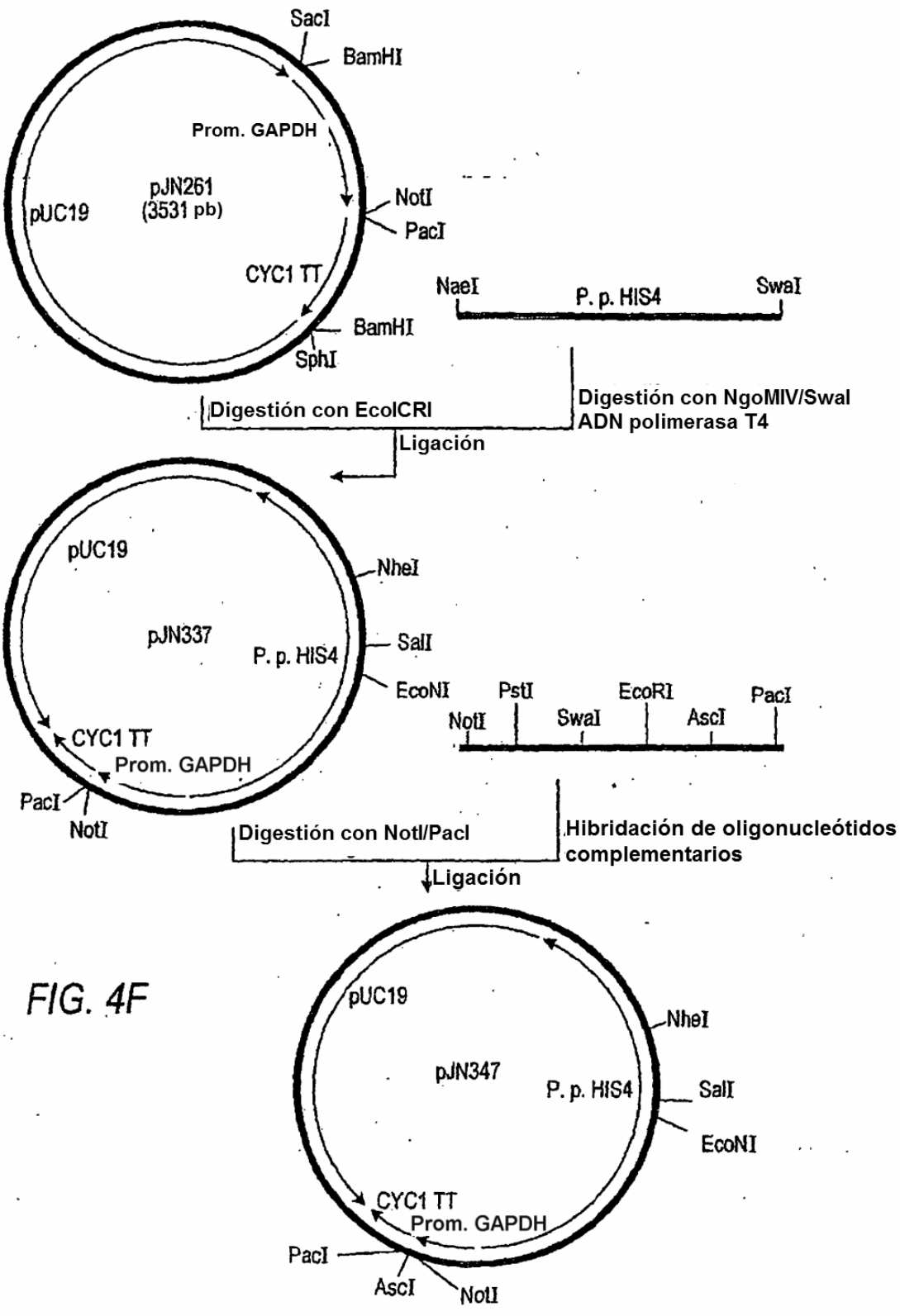

SacI
BamHI
Prom. GAPDH
pJN261
(3531 pb)
pUC19
NotI
PacI
CYC1 TT
BamHI
SphI
NaeI
P. p. HIS4
SwaI
Digestión con EcoICRI
Digestión con NgoMIV/SwaI
ADN polimerasa T4
Ligación
pUC19
NheI
pJN337
P. p. HIS4
SalI
EcoNI
PstI
EcoRI
PacI
NotI
SwaI
AscI
CYC1 TT
Prom. GAPDH
PacI
NotI
Digestión con NotI/PacI
Hibridación de oligonucleótidos
complementarios
Ligación
pUC19
NheI
pJN347
P. p. HIS4
SalI
EcoNI
CYC1 TT
Prom. GAPDH
PacI
AscI
NotI

FIG. 4F

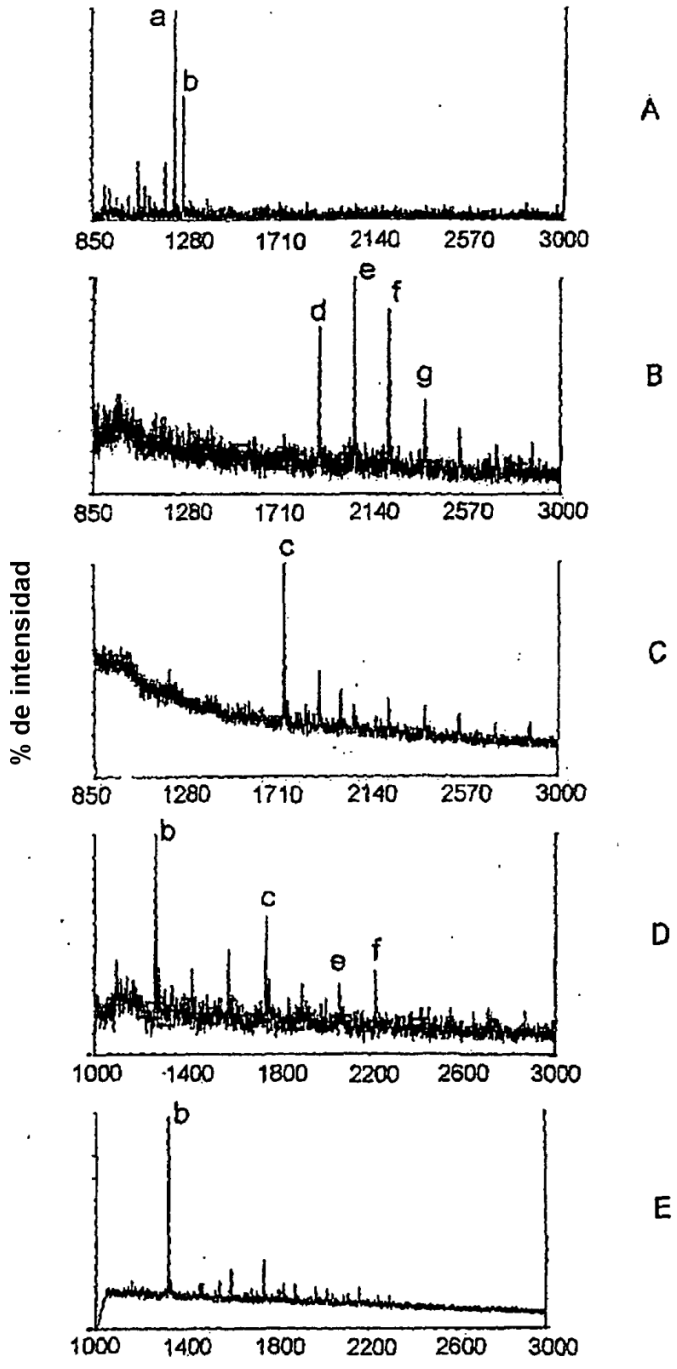
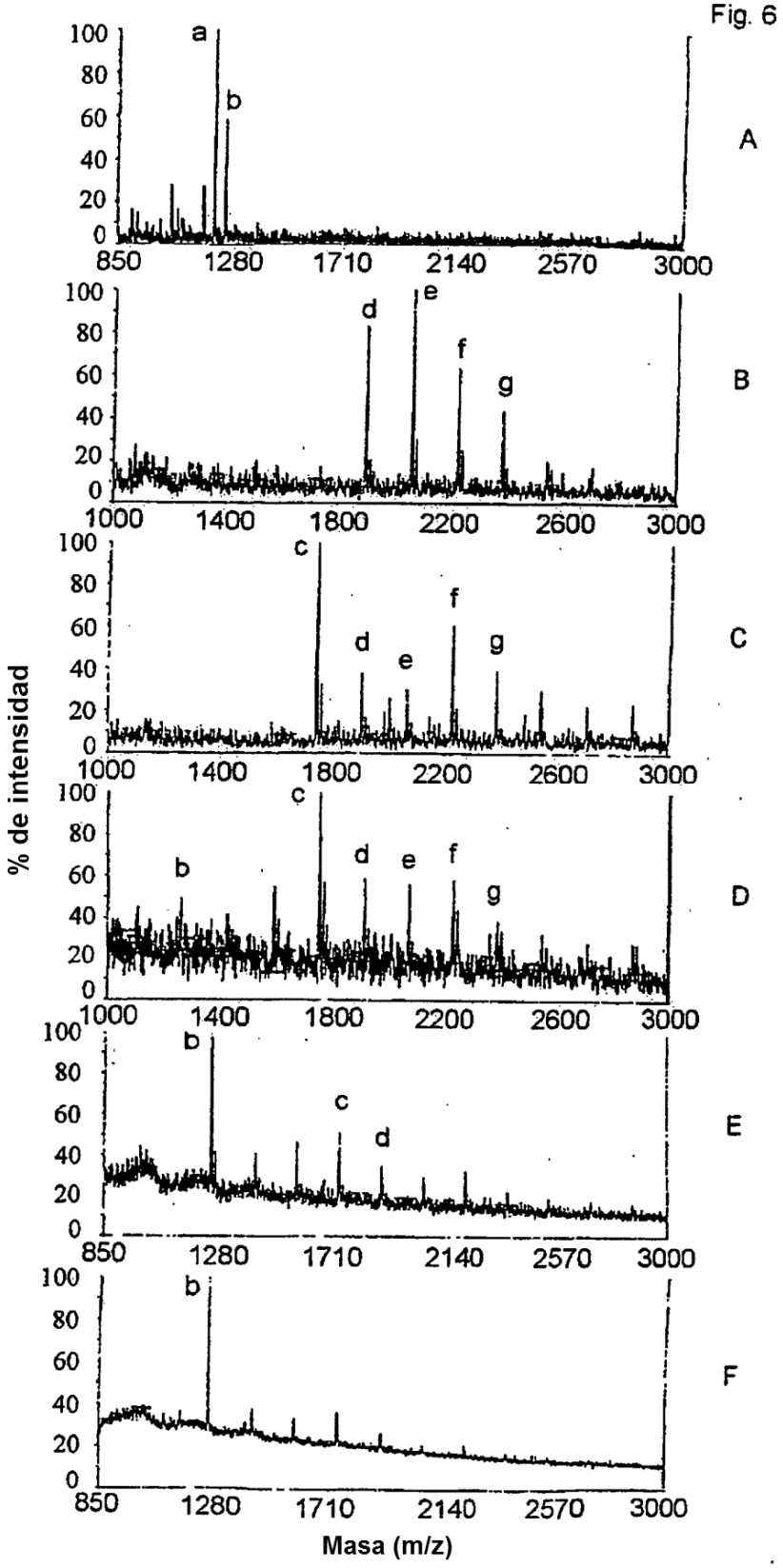

a
b
A
850
1280
1710
2140
2570
3000
e
f
d
g
B
850
1280
1710
2140
2570
3000
c
C
% de intensidad
850
1280
1710
2140
2570
3000
b
c
e
f
D
1000
1400
1800
2200
2600
3000
b
E
1000
1400
1800
2200
2600
3000
Masa (m/z)

Fig. 5

Fig. 6


% de intensidad
Masa (m/z)
A
B
C
D
E
F
a
b
c
d
e
f
g
100
80
60
40
20
0
850
1280
1710
2140
2570
3000
1000
1400
1800
2200
2600

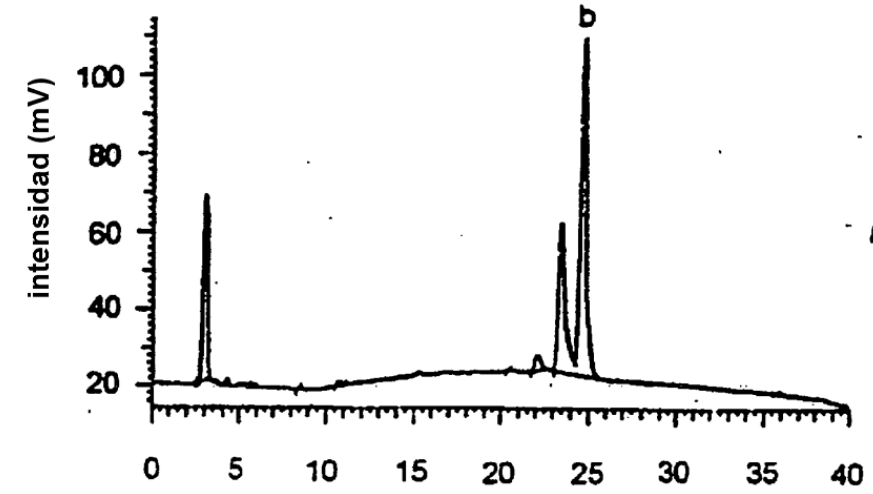


FIG. 7A

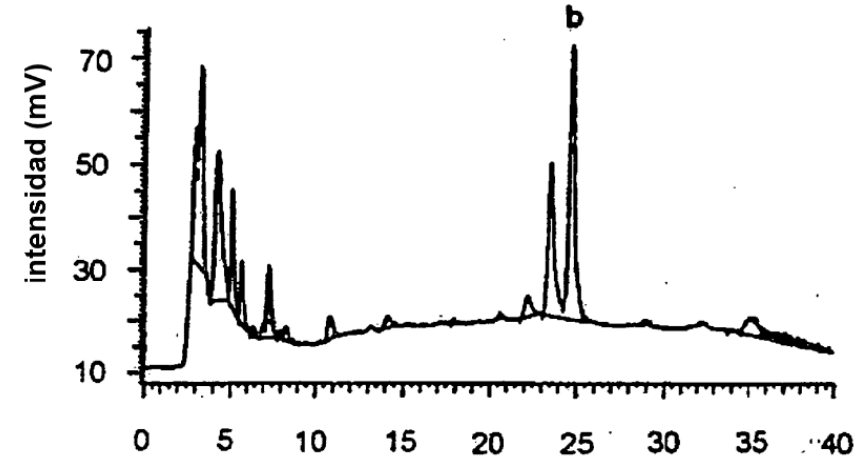


FIG. 7B

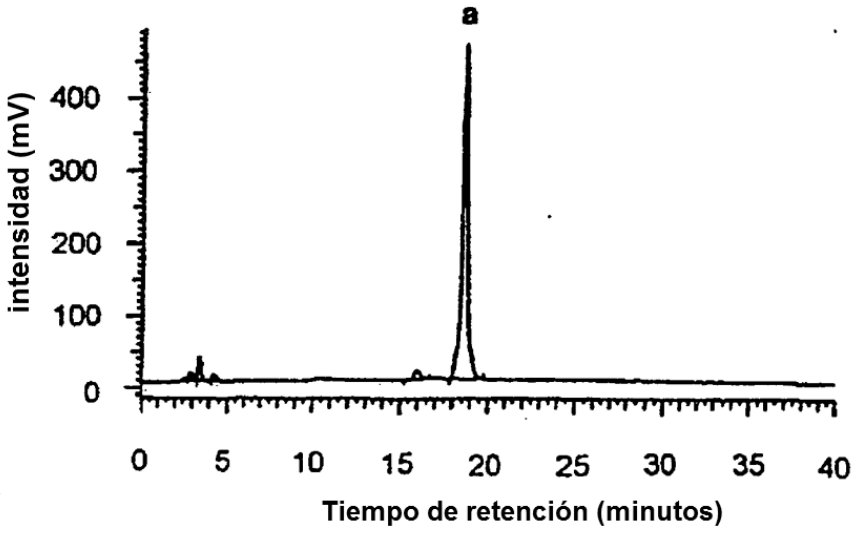


FIG. 7C

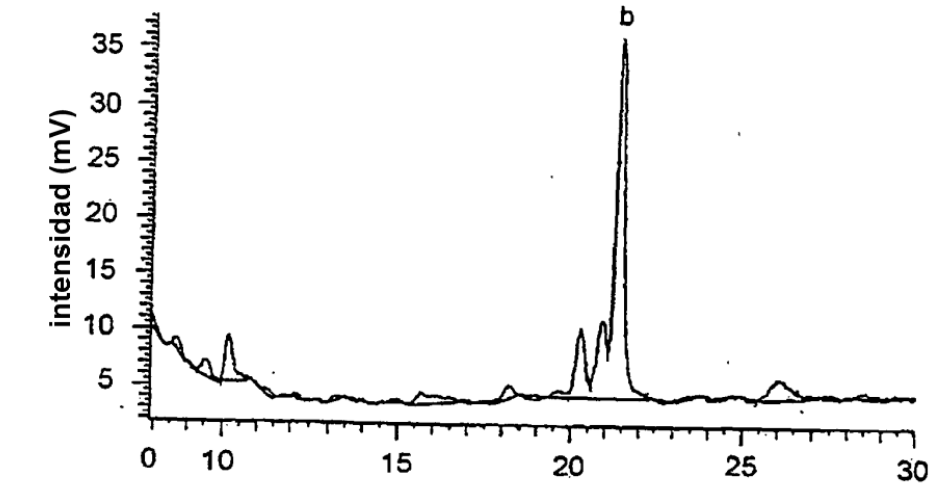


FIG. 8A

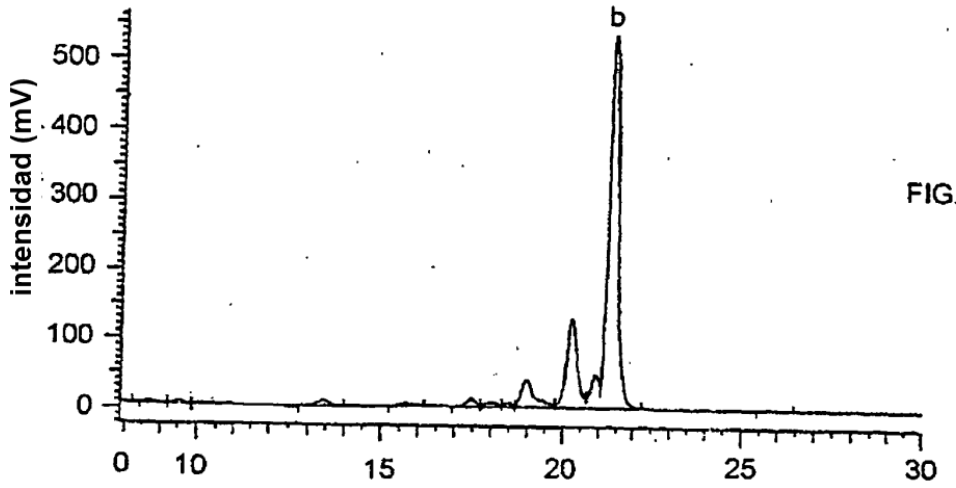


FIG. 8B

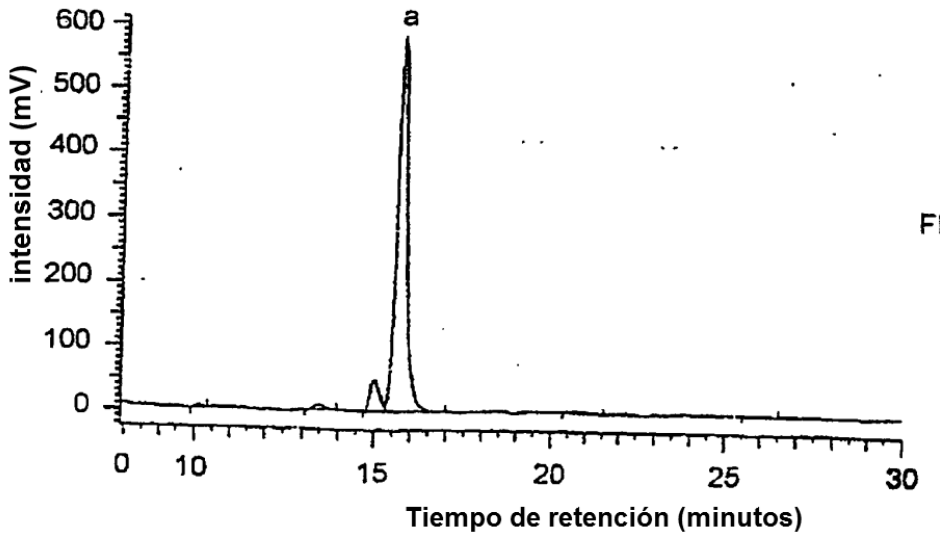


FIG. 8C

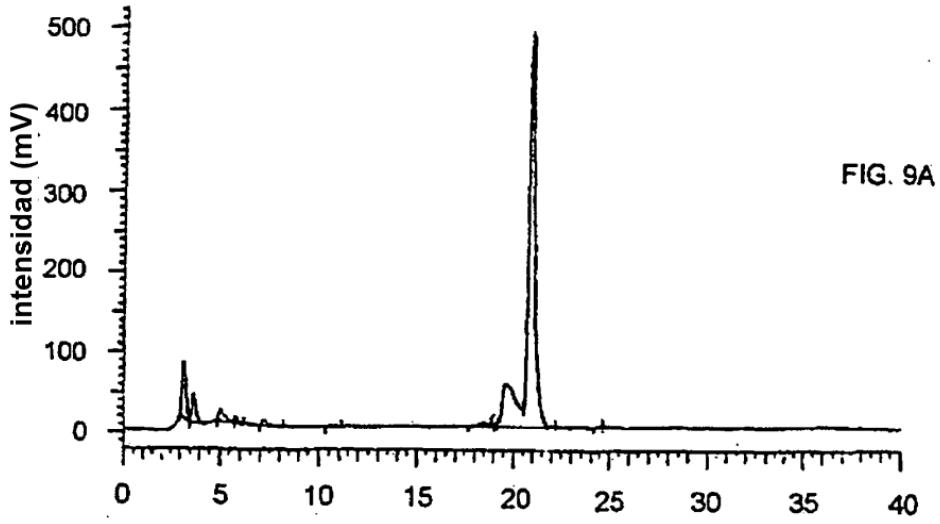


FIG. 9A

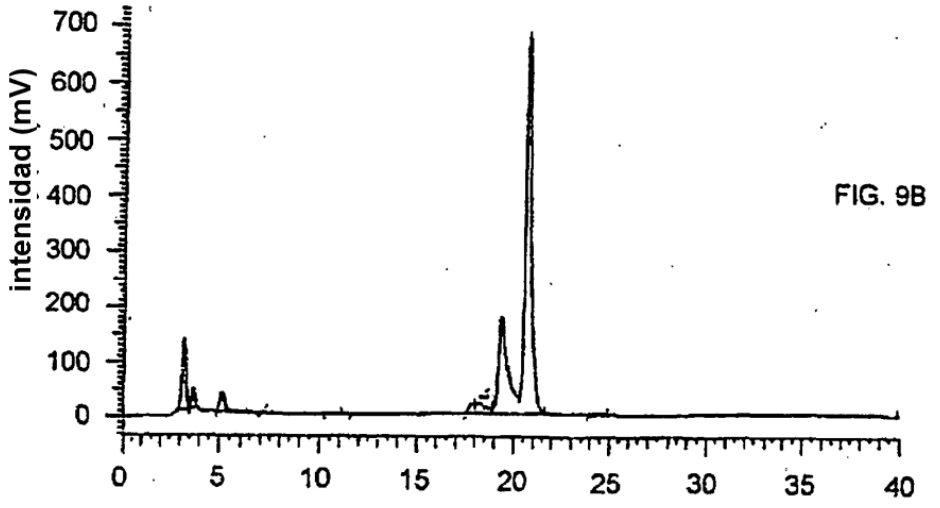


FIG. 9B

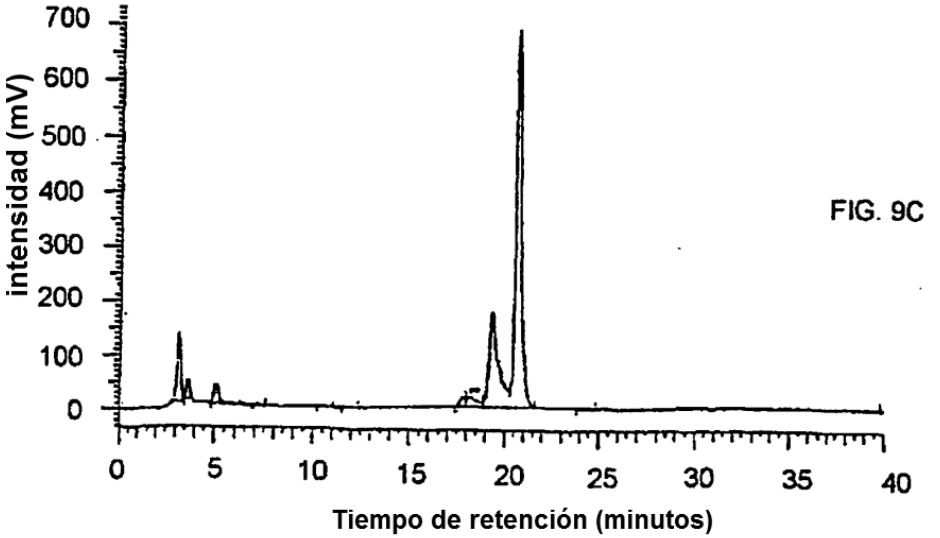


FIG. 9C

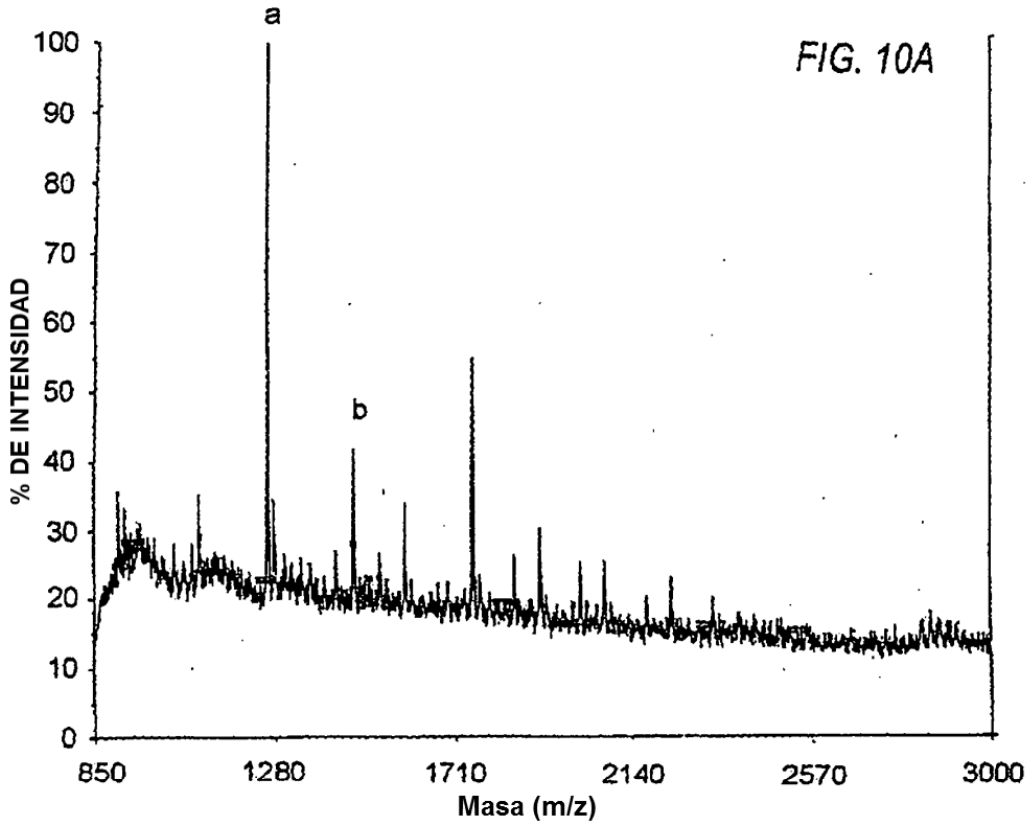


FIG. 10A

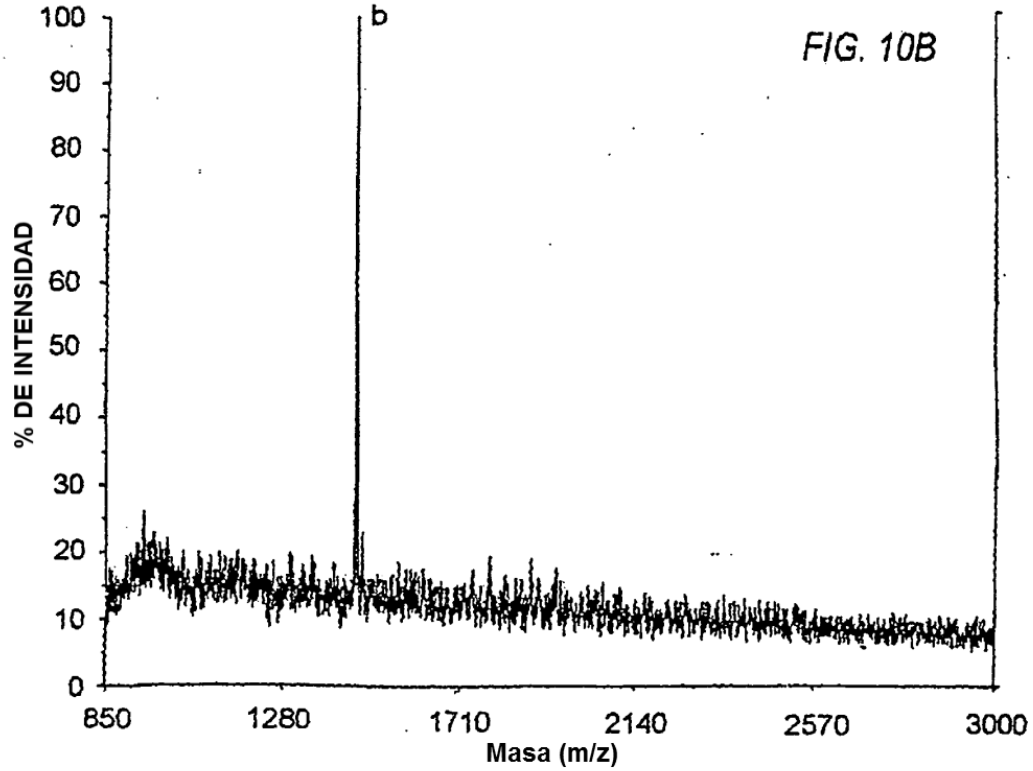


FIG. 10B

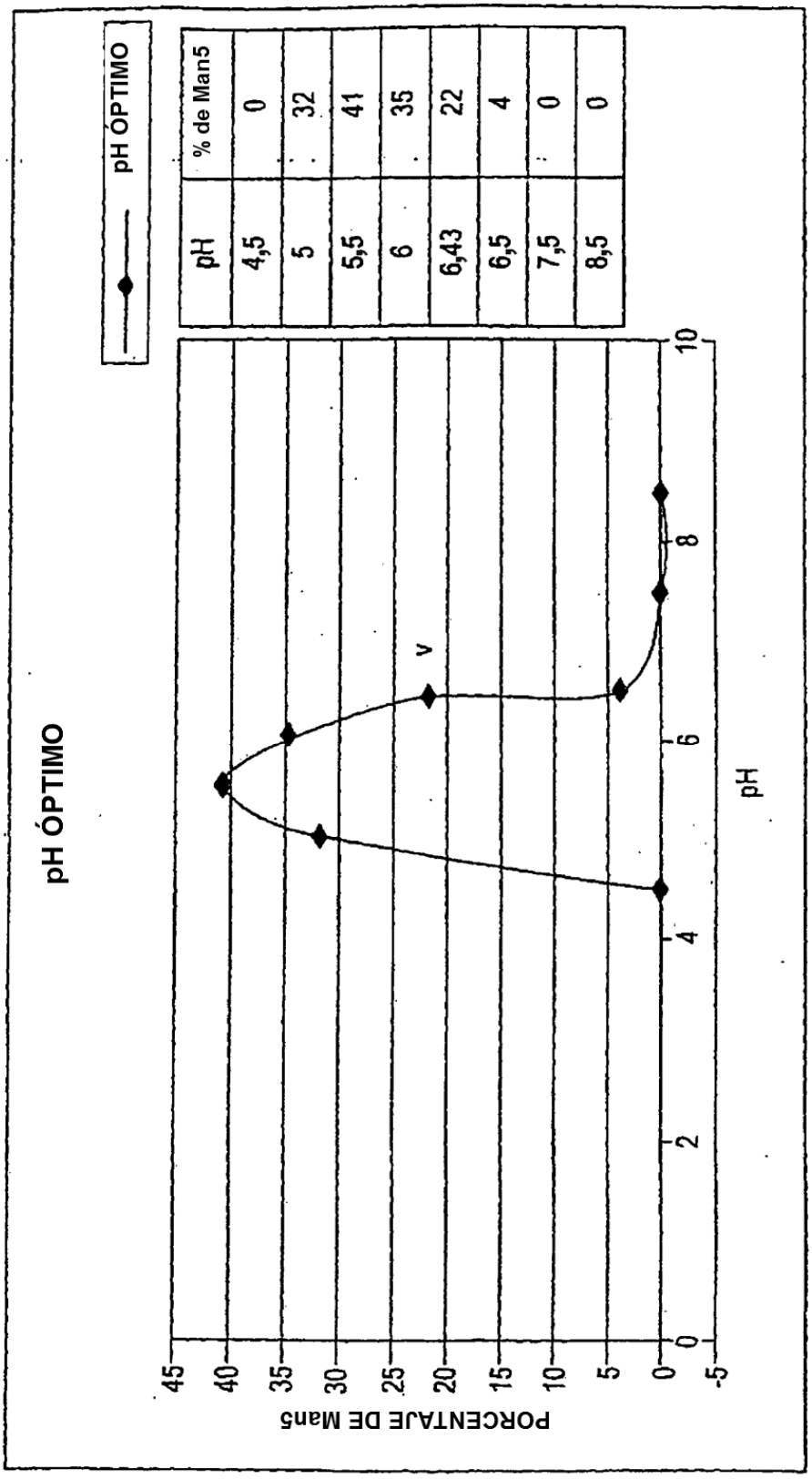
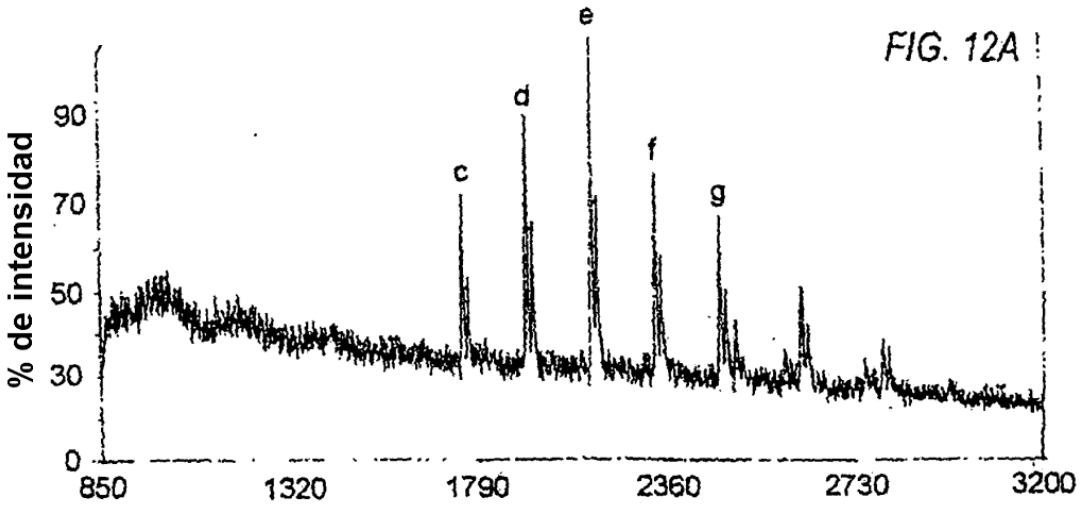
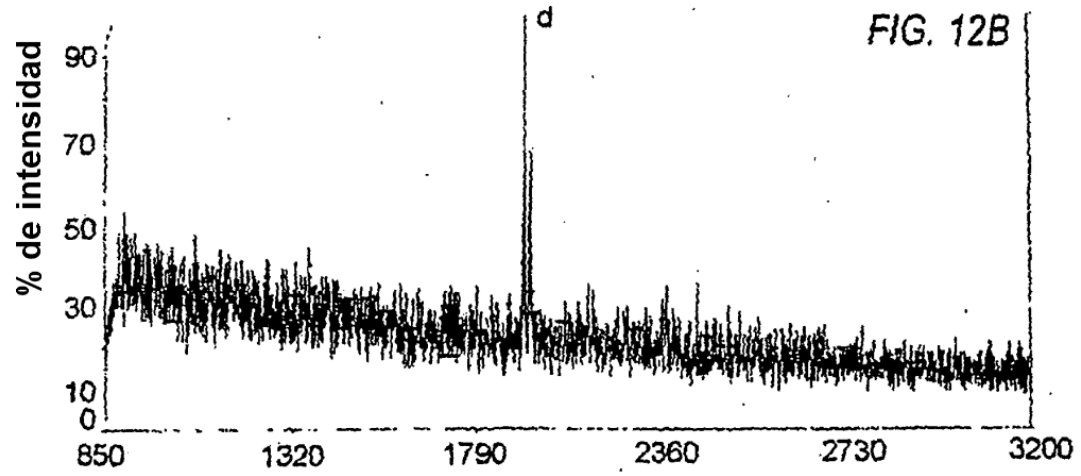
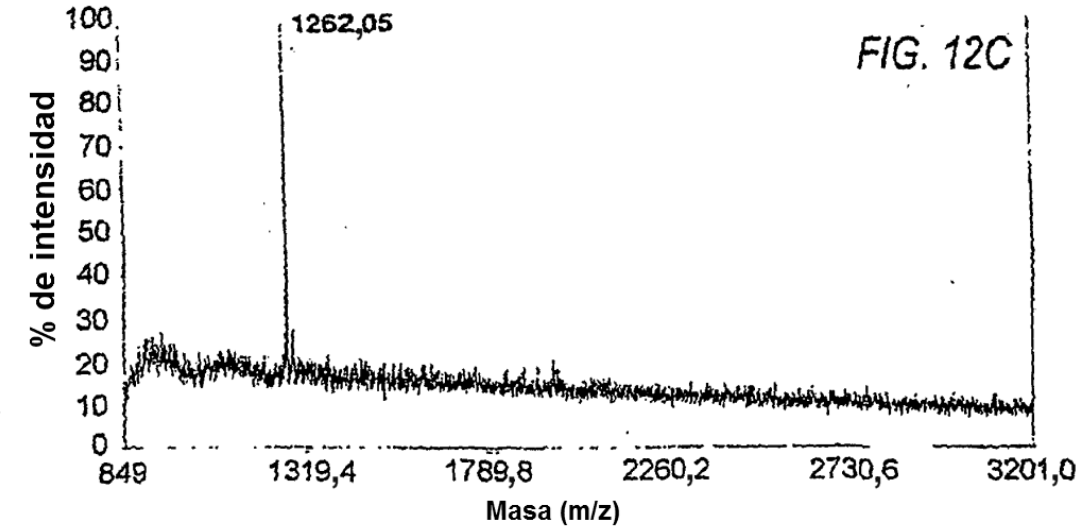

pH ÓPTIMO
pH ÓPTIMO
pH
% de Man5
4,5 0
5 32
5,5 41
6 35
6,43 22
6,5 4
7,5 0
8,5 0
PORCENTAJE DE Man5
pH
v

FIG. 11


FIG. 12A
% de intensidad
90
70
50
30
0
c
d
e
f
g
850
1320
1790
2360
2730
3200


FIG. 12B
% de intensidad
90
70
50
30
10
0
d
850
1320
1790
2360
2730
3200


FIG. 12C
% de intensidad
100
90
80
70
60
50
40
30
20
10
0
1262,05
849
1319,4
1789,8
2260,2
2730,6
3201,0
Masa (m/z)

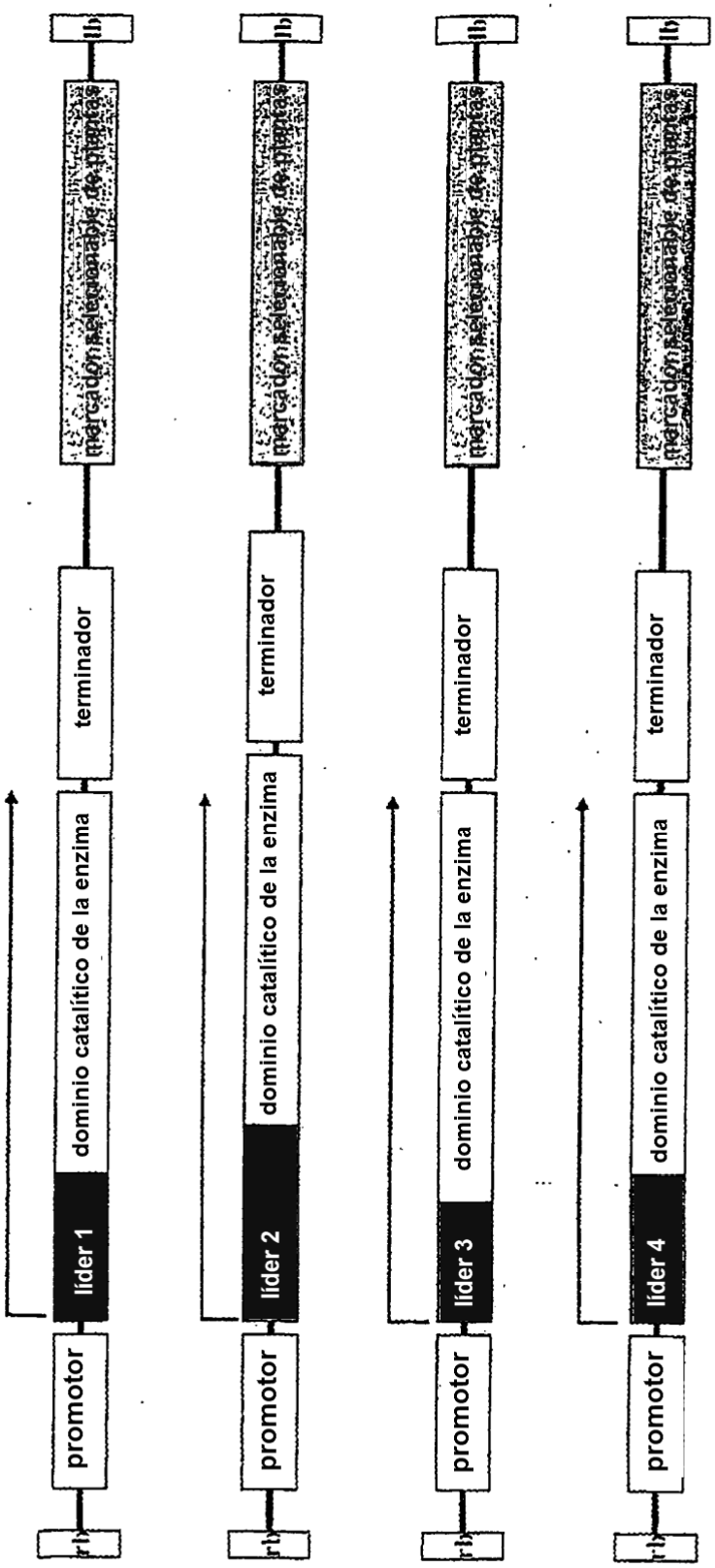

promotor
líder 1
dominio catalítico de la enzima
terminador
marcador seleccionable de plantas
promotor
líder 2
dominio catalítico de la enzima
terminador
marcador seleccionable de plantas
promotor
líder 3
dominio catalítico de la enzima
terminador
marcador seleccionable de plantas
promotor
líder 4
dominio catalítico de la enzima
terminador
marcador seleccionable de plantas

FIG.13