



19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 335 345**

51 Int. Cl.:  
**C12P 21/00** (2006.01)  
**C12N 9/10** (2006.01)  
**C12N 5/10** (2006.01)  
**C12N 15/81** (2006.01)  
**C12R 1/645** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **04713412 .7**  
96 Fecha de presentación : **20.02.2004**  
97 Número de publicación de la solicitud: **1599595**  
97 Fecha de publicación de la solicitud: **30.11.2005**

54 Título: **Expresión de N-acetilglucosaminil transferasa III en eucariotas inferiores.**

30 Prioridad: **20.02.2003 US 371877**  
**07.10.2003 US 680963**

45 Fecha de publicación de la mención BOPI:  
**25.03.2010**

45 Fecha de la publicación del folleto de la patente:  
**25.03.2010**

73 Titular/es: **GlycoFi, Inc.**  
**21 Lafayette Street, Suite 200**  
**Lebanon, New Hampshire 03766, US**

72 Inventor/es: **Bobrowicz, Piotr;**  
**Hamilton, Stephen R.;**  
**Gerngross, Tillman U.;**  
**Wildt, Stefan;**  
**Choi, Byung-Kwon;**  
**Nett, Juergen Hermann y**  
**Davidson, Robert C.**

74 Agente: **Carpintero López, Mario**

ES 2 335 345 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

## DESCRIPCIÓN

Expresión de *N*-acetilglucosaminil transferasa III en eucariotas inferiores.

5 **Campo de la invención**

La presente invención se refiere a procedimientos y composiciones por los que se pueden modificar genéticamente células huésped de levadura o de hongo unicelular o multicelular filamentosos 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, enseres humanos o animales.

**Antecedentes de la invención***Rutas de Glucosilación en Seres Humanos y Eucariotas Inferiores*

Después de que el ADN se haya transcrito y traducido en una proteína, el procesamiento post-traduccional 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 disponibles (nucleótido azúcar), de tal forma 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 está expresando la proteína particular. Las bacterias típicamente no glucosilan proteínas y, si lo hacen, solamente de un modo muy inespecífico (Moeny y Vanderleyden (1997) Arch Microbiol. 168(3): 169-175). Los eucariotas inferiores tales como hongos filamentosos y levaduras añaden principalmente azúcares manosa y manosilfosfato. El glucano resultante se conoce como un glucano de tipo "alto contenido de manosa" o un manano. Las células vegetales y las células de insecto (tales como células Sf9) glucosilan proteínas de otro modo. Por el contrario, en eucariotas superiores tales como seres humanos, la cadena lateral del oligosacárido naciente se puede cortar para eliminar varios restos de manosa y prolongar con restos de azúcar adicionales que típicamente no se encuentran en los *N*-glucanos de eucariotas inferiores. Véase, por ejemplo, Bretthauer, y col. (1999) Biotechnology and Applied Biochemistry 30: 193-200; Martinet, y col. (1998) Biotechnology Letters 20: 1171-1177; Weikert, y col. (1999) Nature Biotechnology, 17: 1116-1121; M. Malissard, y col. (2000) Biochemical and Biophysical Research Communications 267: 169-173; Jarvis, y col., (1998) Current Opinion in Biotechnology 9: 528-533; y Takeuchi (1997) Trends in Glycoscience and Glycotechnology 9: S29-S35.

La síntesis de una estructura de oligosacárido de tipo mamífero comienza con un conjunto de reacciones secuenciales en el curso de las cuales se añaden y se eliminan 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 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 se muestra en la Figura 1A.

Las etapas tempranas de la glucosilación humana se pueden dividir en al menos dos fases diferentes: (i) los oligosacáridos Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub> ligados a lípidos se ensamblan por un conjunto secuencial de reacciones en la membrana del retículo endoplásmico (RE) (Figura 13) y (ii) la transferencia de este oligosacárido del ancla lipídica dolícol pirofosfato a la proteína sintetizada *de novo*. El sitio de la transferencia específica se define por un resto de asparagina (Asn) en la secuencia Asn-Xaa-Ser/Thr (SEC ID N°: 1 y 2) donde Xaa puede ser cualquier aminoácido excepto prolina (Gavel y von Heijne (1990) Protein Eng. 3: 433-42). Tiene lugar un procesamiento adicional por glucosidasas y manosidasas en el RE antes de que la glucoproteína naciente se transfiera al aparato de Golgi temprano, donde se eliminan restos de manosa adicionales por alfa ( $\alpha$ )-1,2-manosidasas específicas de Golgi. El procesamiento continúa cuando la proteína avanza a través del Golgi. En el Golgi medial, varias enzimas modificadoras, que incluyen *N*-acetilglucosaminil transferasas (GnTI, GnTII, GnTIII, GnTIV y GnTV), manosidasa II y fucosiltransferasas, añaden y eliminan restos de azúcar específicos. Finalmente, en el Golgi *trans*, las galactosiltransferasas (GalT) y sialiltransferasas (ST) produce una estructura de glucoproteína 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 las glucoproteínas sus características humanas. La estructura de un *N*-glucano humano típico se muestra en la Figura 1B. Véase también las Figuras 14 y 15 para etapas implicadas en el procesamiento de *N*-glucano de tipo mamífero.

En prácticamente todos los eucariotas, las glucoproteínas se obtienen de un precursor común de oligosacárido ligado a lípido Glc<sub>3</sub>Man<sub>9</sub>GlcNAc<sub>2</sub>-dolícol-pirofosfato. Dentro del retículo endoplásmico, la síntesis y el procesamiento de oligosacáridos unidos a dolícol pirofosfato son idénticos en todos los eucariotas conocidos. Sin embargo, el procesamiento adicional del oligosacárido central por células fúngicas, por ejemplo, levaduras, una vez que se ha transferido a un péptido que abandona el RE y que entra en el Golgi, difiere significativamente del de seres humanos cuando se mueve a lo largo de la ruta secretora e implica la adición de varios azúcares de manosa.

En levaduras, estas etapas se catalizan por manosiltransferasas que residen en el Golgi, como Ochlp, Mntl p y Mnnlp, que añaden secuencialmente azúcares de manosa al oligosacárido central. La estructura resultante es indeseable para la producción de proteínas de tipo humano y, por tanto, es deseable reducir o eliminar la actividad de

manosiltransferasa. Se ha demostrado que los mutantes de *S. cerevisiae*, deficientes en actividad de manosiltransferasa (por ejemplo, mutantes *och11* o *mnn9*) son no mortales y presentan un contenido reducido de manosa en el oligosacárido de glucoproteínas de levadura. Puede que también se tengan que eliminar otras enzimas de procesamiento de oligosacáridos, tales como manosilfosfato transferasa, dependiendo del patrón de glucosilación endógeno particular del huésped.

#### Precursores de Nucleótido Azúcar

Los *N*-glucanos de glucoproteínas animales incluyen típicamente galactosa, fucosa y ácido siálico terminal. Estos azúcares no se encuentran en glucoproteínas producidas en levaduras y hongos filamentosos. En seres humanos, toda la variedad de precursores de nucleótido azúcar (por ejemplo, UDP-*N*-acetilglucosamina, UDP-*N*-acetilgalactosamina, CMP-ácido-*N*-acetilneuramínico, UDP-galactosa, GDP-fucosa, etc.) se sintetizan en el citosol y se transportan al Golgi, donde se unen al oligosacárido central por 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 producen típicamente un producto secundario que es un nucleósido difosfato o monofosfato. Mientras que los monofosfatos se pueden exportar directamente en intercambio por azúcares de nucleósido trifosfato por un mecanismo de cotransporte bidireccional, los difosfonucleósidos (por ejemplo, GDP) se tienen que escindir por fosfatasas (por ejemplo, GDPasa) para producir nucleósido monofosfato y fosfato inorgánico antes de ser exportados. Esta reacción es importante para la glucosilación eficaz; por ejemplo, se ha observado que la GDPasa de *Saccharomyces cerevisiae* (*S. cerevisiae*) es necesaria para la manosilación. Sin embargo, esa GDPasa tiene una actividad reducida el 90% con respecto a UDP (Berninsone y col. (1994) *J. Biol. Chem.* 269(1): 207-211). Los eucariotas inferiores típicamente carecen de actividad de difosfatasa específica de UDP en el Golgi ya que no utilizan precursores de UDP-azúcar para síntesis de glucoproteína basada en Golgi. Se ha observado que *Schizosaccharomyces pombe*, una levadura de la que se ha observado que añade restos de galactosa a polisacáridos de pared celular (de UDP-galactosa) tiene actividad de UDPasa específica, indicando el requisito potencial de tal enzima (Berninsone y col. (1994) *J. Biol. Chem.* 269(1): 207-211). Se sabe que el UDP es un inhibidor potente de glucosiltransferasas y la eliminación de este producto secundario de la glucosilación puede ser importante para evitar la inhibición por glucosiltransferasa en el lumen del Golgi (Khatara y col. (1974) *Eur. J. Biochem.* 44: 537-560). Véase Berninsone y col. (1995) *J. Biol. Chem.* 270(24): 14564-14567; Beaudet y col. (1998) *ABC Transporters: Biochemical, Cellular, and Molecular Aspects* 292: 397-413.

#### Procesamiento Secuencial de *N*-glucanos por Actividades Enzimáticas Compartimentalizadas

Las transferasas de azúcar y glucosidasas (por ejemplo, manosidasas) revisten la superficie interna (luminal) del RE y del aparato de Golgi y proporcionan de este modo una superficie "catalítica" que permite el procesamiento secuencial de glucoproteínas cuando avanzan a través del RE y la red de Golgi. Los múltiples compartimentos del Golgi *cis*, medial y *trans* y la Red de Golgi *trans* (TGN) proporcionan las diferentes localizaciones en las que puede tener lugar la secuencia ordenada de reacciones de glucosilación. Cuando una glucoproteína avanza desde síntesis en el RE hasta maduración completa en el Golgi tardío o TGN, se expone secuencialmente a diferentes glucosidasas, manosidasas y glucosiltransferasas de tal forma que se puede sintetizar una estructura específica de hidrato de carbono. Se ha dedicado mucho trabajo para poner de manifiesto el mecanismo exacto por el que estas enzimas se retienen y anclan a su orgánulo respectivo. La imagen resultante es compleja, pero las pruebas sugieren que la región de tallo, la región transmembrana y cola citoplasmática, individualmente o de forma concertada, dirigen enzimas a la membrana de orgánulos individuales y de este modo localizan el dominio catalítico asociado en ese locus (véase, por ejemplo, Gleeson (1998) *Histochem. Cell Biol.* 109: 517-532).

En algunos casos, se ha observado que estas interacciones específicas funcionan entre especies. Por ejemplo, se demostró que el dominio transmembrana de  $\alpha$ 2,6-ST de ratas, una enzima que se conoce que se localiza en el Golgi *trans* del animal, también localiza un gen indicador (invertasa) en el Golgi de levadura (Schwientek y col. (1995) *J. Biol. Chem.* 270(10): 5483-9). Sin embargo, el mismo dominio transmembrana como parte de una  $\alpha$ 2,6-ST de longitud completa se retuvo en el RE y no se transportó adicionalmente al Golgi de levadura (Krezdorn y col. (1994) *Eur. J. Biochem.* 220(3): 809-17). Una GalT de longitud completa de seres humanos no se sintetizó de forma uniforme en levaduras, a pesar de niveles de transcripción demostrablemente elevados. Por el contrario, la región transmembrana de la misma GalT humana fusionada con un indicador de invertasa fue capaz de dirigir la localización al 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 citoplasmática, una región transmembrana y ocho aminoácidos de la región de tallo al dominio catalítico de GalT humana es suficiente para la localización en Golgi de una GalT activa. Parece que otras galactosiltransferasas dependen de interacciones con enzimas que residen en orgánulos particulares debido a que después de la retirada de su región transmembrana, todavía son capaces de localizarse de forma apropiada.

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  $\alpha$ -1,2-manosidasas (Eades y Hintz (2000) *Gene* 255(1): 25-34), no añade GlcNAc a Man<sub>5</sub>GlcNAc<sub>2</sub> cuando se transforma con el gen de GnTI de conejo, a pesar de un

nivel globalmente alto de actividad de GnTI (Kalsner y col. (1995) *Glycoconj. J.* 12(3): 360-370). La GnTI, aunque se expresa activamente, se puede localizar incorrectamente, de tal forma que la enzima no está en contacto con sus dos sustratos: UDP-GlcNAc y un sustrato de Man<sub>5</sub>GlcNAc<sub>2</sub> productivo (no todas las estructuras de Man<sub>5</sub>GlcNAc<sub>2</sub> son productivas; véase a continuación). Alternativamente, el organismo huésped puede no proporcionar un nivel adecuado de UDP-GlcNAc en el Golgi o la enzima se puede localizar de forma apropiada pero, sin embargo, de forma inactiva en su nuevo entorno. Además, las estructuras de Man<sub>5</sub>GlcNAc<sub>2</sub> presentes en la célula huésped pueden diferir en estructura de Man<sub>5</sub>GlcNAc<sub>2</sub> encontrado en mamíferos. Maras y colaboradores observaron que aproximadamente un tercio de los *N*-glucanos de celobiohidrolasa I (CBHI) obtenidos de *T. reesei* se pueden cortar hasta Man<sub>5</sub>GlcNAc<sub>2</sub> por 1,2 manosidasa de *A. saitoi* *in vitro*. Menos del 1% de esos *N*-glucanos, sin embargo, podrían servir como un sustrato productivo para GnTI. Maras y col. (1997) *Eur. J. Biochem.* 249: 701-707. La mera presencia de Man<sub>5</sub>GlcNAc<sub>2</sub>, por lo tanto, no garantiza que se pueda conseguir el procesamiento *in vivo* adicional de Man<sub>5</sub>GlcNAc<sub>2</sub>. Se requiere formación de una estructura de Man<sub>5</sub>GlcNAc<sub>2</sub> productiva reactiva con GnTI. Aunque Man<sub>5</sub>GlcNAc<sub>2</sub> se podría producir en la célula (aproximadamente el 27% en mol), solamente una pequeña fracción se podría convertir en Man<sub>5</sub>GlcNAc<sub>2</sub> (menos de aproximadamente el 5%, véase Chiba y col., documento WO 01/14522). Shinkawa y col. (2003), *J. Biol. Chem.* 278(5): 3466-3473, usaron células Lec 10, una línea celular de CHO que sobre-expresa GnT III, para producir una IgG anti-CD20 con un perfil de oligosacárido del 74% de GlcNAc de bisección y el 100% de fucosa. En comparación con el mAb anti-CD20 producido comercialmente Rituxan que tiene el 0% de GlcNAc de bisección y el 94% de fucosa, la IgG producida por Lec 10 mostró una mejora de varias veces de la actividad de citotoxicidad celular dependiente de anticuerpos (ADCC). Davies y col. (2001), *Biotechnology and Bioengineering* 74(4): 288-294, describen la clonación y co-expresión del gen que codifica la enzima de glucosilación de rata  $\beta$ 1-4-*N*-acetilglucosaminiltransferasa III (GnT III) en una línea celular de ovario de hámster chino (CHO) de producción recombinante que expresa un anticuerpo IgG1 quimérico de ratón/ser humano anti-CD20. El análisis de HPLC mostró que la enzima tenía añadidos restos de *N*-acetilglucosamina (GlcNAc) de bisección en la mayoría (del 48% al 71%) de los oligosacáridos ligados a N aislados de preparaciones de anticuerpos purificados de las líneas celulares. El documento WO 02/079255 se refiere a anticuerpos recombinantes co-expresados con GnT III correspondientes a Davies y col. (2001), es decir, que describen que para la línea celular positiva a GnT III, se demostró que la mayoría de las glucoformas (48-71%) contenían un resto de GlcNAc bisechado, reflejando los efectos catalíticos *in vivo* de GnT III sobre las glucoformas de preparaciones de anticuerpos purificados.

Hasta la fecha, no existe un modo fiable de predecir si una glucosiltransferasa o manosidasa expresada de forma heteróloga particular en un eucariota inferior se (1) traducirá suficientemente, (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 eucariotas inferiores, sería deseable un esquema sistemático para conseguir la función catalítica deseada y la retención apropiada de enzimas en ausencia de herramientas predictivas, que actualmente no están disponibles.

#### *Producción de Glucoproteínas Terapéuticas*

Un número significativo de proteínas aisladas de seres humanos o animales se modifican después de la traducción, siendo la glucosilación una de las modificaciones más significativas. Un número estimado del 70% de todas las proteínas terapéuticas están glucosiladas y, por tanto, actualmente dependen de un sistema de producción (es decir, célula huésped) que sea capaz de glucosilar de un modo similar a seres humanos. Varios estudios han demostrado que la glucosilación desempeña un papel importante al determinar (1) la inmunogenicidad, (2) las propiedades farmacocinéticas, (3) el tránsito y (4) la eficacia de proteínas terapéuticas. Por tanto, no es sorprendente que se haya dirigido esfuerzos sustanciales por la industria farmacéutica a desarrollar procedimientos para obtener glucoproteínas que sean tan "humanoides" o "de tipo humano" como sea posible. Hasta la fecha, la mayoría de las glucoproteínas se preparan en un sistema huésped de mamífero. Esto puede implicar la modificación genética de tales células de mamífero para mejorar el grado de sialilación (es decir, adición terminal de ácido siálico) de proteínas expresadas por las células, que se sabe que mejora las propiedades farmacocinéticas de tales proteínas. Alternativamente, se puede mejorar el grado de sialilación por adición *in vitro* de tales azúcares usando glucosiltransferasas conocidas y sus nucleótido azúcares respectivos (por ejemplo, 2,3-sialiltransferasa y CMP-ácido siálico).

Aunque la mayoría de los eucariotas superiores realizan reacciones de glucosilación que son similares a las observadas en seres humanos, las proteínas humanas recombinantes expresadas en los sistemas huésped que se han mencionado anteriormente difieren invariablemente de su equivalente humano "natural" (Raju y col. (2000) *Glycobiology* 10(5): 477-486). Por tanto, se ha dirigido un trabajo de desarrollo extenso para encontrar modos para mejorar el "carácter humano" de proteínas preparadas en estos sistemas de expresión. Esto incluye la optimización de condiciones de fermentación y la modificación genética de huéspedes de expresión de proteína introduciendo genes que codifican enzimas implicadas en la formación de glucoformas de tipo humano. Goochee y col. (1999) *Biotechnology* 9(12): 1347-55; Andersen y Goochee (1994) *Curr Opin Biotechnol.* 5(5): 546-49; Werner y col. (1998) *Arzneimittelforschung.* 48(8): 870-80; Weikert y col. (1999) *Nat Biotechnol.* 17(11): 1116-21; Yang y Butler (2000) *Biotech. Bioeng.* 68: 370-80, Los problemas inherentes asociados con todos los sistemas de expresión de mamífero no se han resuelto.

*Producción de Glucoproteínas Usando Microorganismos Eucariotas*

Aunque la estructura de oligosacárido central transferida a una proteína en el retículo endoplásmico básicamente es idéntica en mamíferos y eucariotas inferiores, se han observado 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 de ventajas por lo demás destacadas con respecto a sistemas de expresión de mamíferos.

Las glucoproteínas terapéuticas producidas en un huésped microorganismo tal como levadura utilizando la ruta de glucosilación del huésped endógena difieren estructuralmente de las producidas en células de mamífero y muestran típicamente una eficacia terapéutica en gran medida reducida. Tales glucoproteínas típicamente son inmunogénicas en seres humanos y muestran una semivida (y, por tanto, bioactividad) reducida *in vivo* después de la administración (Takeuchi (1997) Trends in Glycoscience and Glycotechnology 9: S29-S35). Los receptores específicos en seres humanos y animales (es decir, receptores de manosa en macrófagos) pueden reconocer restos de manosa terminales y promover el aclaramiento rápido de la glucoproteína extraña del torrente sanguíneo. Los efectos adversos adicionales pueden incluir cambios en el plegamiento, solubilidad, susceptibilidad a proteasas, tránsito, transporte, compartimentalización, secreción, reconocimiento por otras proteínas o factores, antigenicidad o alergenicidad de la proteína.

Tanto las levaduras como los hongos filamentosos se han usado de forma exitosa para la producción de proteínas recombinantes, tanto intracelulares como secretadas (Cereghino y Cregg (2000) FEMS Microbiology Reviews 24(1): 45-66; Harkki y col. (1989) BioTechnology 7(6): 596; Berka y col. (1992) Abstr. Papers Amer. Chem. Soc. 203: 121-BIOT; Svetina y col. (2000) J. Biotechnol. 76(2-3): 245-251). Diversas levaduras, tales como *K. lactis*, *Pichia pastoris*, *Pichia methanolic* y *Hansenula polymorpha*, han desempeñado papeles particularmente importantes como sistemas de expresión eucariotas debido a que son capaces de crecer hasta altas densidades celulares y secretar grandes cantidades de proteína recombinante. Asimismo, los hongos filamentosos, tales como *Aspergillus niger*, *Fusarium sp.*, *Neurospora crassa* y otros se han usado para producir de forma eficaz glucoproteínas a escala industrial. Sin embargo, como se ha señalado anteriormente, las glucoproteínas expresadas en cualquiera de estos microorganismos eucariotas difieren sustancialmente en la estructura de *N*-glucano 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 que en seres 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 (compárese Figuras 1A y 1B). Sin embargo, el procesamiento posterior del oligosacárido central difiere significativamente en levaduras e implica la adición de varios azúcares de manosa. Esta etapa está catalizada por manosiltransferasas que residen en el Golgi (por ejemplo, *OCH1*, *MNT1*, *MNN1*, etc.), que añaden secuencialmente azúcares de manosa al oligosacárido central. La estructura resultante es indeseable para la producción de proteínas humanoides y, por tanto, es deseable para reducir o eliminar la actividad de manosiltransferasa. Se ha demostrado que los mutantes de *S. cerevisiae* deficientes en actividad de manosiltransferasa (por ejemplo, mutantes *och1* o *mnn9*) son no mortales y presentan un contenido de manosa reducido en el oligosacárido de glucoproteínas de levadura. Puede que también se tengan que eliminar otras enzimas de procesamiento de oligosacáridos, tales como manosilfosfato transferasas, dependiendo del patrón de glucosilación endógeno particular del huésped. Después de reducir reacciones de glucosilación endógenas indeseadas, se tiene que incluir la formación de *N*-glucanos complejos en el sistema huésped. Esto requiere la expresión estable de varias enzimas y transportadores de nucleótido-azúcar. Además, se tienen que localizar estas enzimas de tal forma que se garantice un procesamiento secuencial de la estructura de glucosilación que está madurando.

Se han realizados varios esfuerzos para modificar las rutas de glucosilación de microorganismos eucariotas para proporcionar glucoproteínas más adecuadas para el uso como agentes terapéuticos de mamífero. Por ejemplo, se han clonado y expresado varias glucosiltransferasas por separado en *S. cerevisiae* (GAI1, GnTI), *Aspergillus nidulans* (GnTI) y otros hongos (Yoshida y col. (1999) Glycobiology 9(1): 53-8, Kalsner y col. (1995) Glycoconj. J. 12(3): 360-370). Sin embargo, no se han obtenido *N*-glucanos que se parezcan a los preparados 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, la familia de *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 reacciones de glucosilación endógenas. Muchos de estos genes se han deletado individualmente dando lugar a organismos viables que tienen perfiles de glucosilación alterados. Se muestran ejemplos en la Tabla 1.

# ES 2 335 345 T3

TABLA 1

| Ejemplos de cepas de levadura que tienen manosilación alterada |                                       |                              |                                      |   |
|--|---------------------------------------|------------------------------|--------------------------------------|---|
| Cepa   | <i>N</i> -glucano<br>(tipo silvestre) | Mutación                     | <i>N</i> -glucano<br>(mutante)       | Referencia  |
| <i>S. pombe</i>  | Man <sub>9</sub> GlcNAc <sub>2</sub>  | OCH1                         | Man <sub>8</sub> GlcNAc <sub>2</sub> | Yoko-o y col.<br>(2001) FEBS<br>Lett. 489(1): 75-<br>80                         |
| <i>S. cerevisiae</i>   | Man <sub>9</sub> GlcNAc <sub>2</sub>  | OCH1/MNN1                    | Man <sub>8</sub> GlcNAc <sub>2</sub> | Nakanishi-Shindo<br>y col. (1993) J.<br>Biol. Chem.<br>268(35): 26338-<br>26345 |
| <i>S. cerevisiae</i>   | Man <sub>9</sub> GlcNAc <sub>2</sub>  | OCH1/MNN1/MNN4               | Man <sub>8</sub> GlcNAc <sub>2</sub> | Chiba y col.<br>(1998) J. Biol.<br>Chem. 273,<br>26298-26304                    |
| <i>P. pastoris</i>   | Hiperglucosilado                      | OCH1 (delección<br>completa) | No<br>hiperglucosilado               | Welfide,<br>Publicación de<br>Solicitud<br>Japonesa N° 8-<br>336387             |
| <i>P. pastoris</i>   | Man <sub>8</sub> GlcNAc <sub>2</sub>  | OCH1 (alteración)            | Man <sub>8</sub> GlcNAc <sub>2</sub> | Contreras y col.<br>documento<br>WO02/00856 A2                                  |

La Publicación de Solicitud de Patente Japonesa N° 8-336387 describe la delección de un homólogo de *OCH1* en *Pichia pastoris*. En *S. cerevisiae*, *OCH1* codifica 1,6-manosiltransferasa, que añade una manosa a la estructura de glucano Man<sub>8</sub>GlcNAc<sub>2</sub> para producir Man<sub>9</sub>GlcNAc<sub>2</sub>. La estructura Man<sub>9</sub>GlcNAc<sub>2</sub>, que contiene tres restos de 1,6 manosa, después es 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 pueden tener típicamente 30-40 restos de manosa por *N*-glucano. Debido a que Och1p inicia la transferencia de 1,6 manosa al núcleo Man<sub>8</sub>GlcNAc<sub>2</sub>, se denomina con frecuencia la “1,6 manosiltransferasa de inicio” para distinguir la misma de otras 1,6 manosiltransferasas que actúan posteriormente en el Golgi. En una cepa mutante de *och1 mnn1 mnn4* de *S. cerevisiae*, las proteínas glucosiladas con Man<sub>8</sub>GlcNAc<sub>2</sub> se acumulan y no tiene lugar la hipermanosilación. Sin embargo, Man<sub>8</sub>GlcNAc<sub>2</sub> no es un sustrato para glucosiltransferasas de mamífero, tales como UDP-GlcNAc transferasa I, y por consiguiente, el uso de esa cepa mutante, por sí misma, no es útil para producir proteínas de tipo mamífero, es decir, las que tienen patrones de glucosilación complejos o híbridos.

Se pueden cortar estructuras de Man<sub>8</sub>GlcNAc<sub>2</sub> hasta una isómero de Man<sub>5</sub>GlcNAc<sub>2</sub> en *S. cerevisiae* (aunque todavía se tiene que demostrar un corte de alta eficacia mayor del 50% *in vivo*) introduciendo una manosidasa fúngica de *A. saitoi* en el retículo endoplásmico (RE). Los defectos de esta estrategia son dos: (1) no está claro si las estructuras Man<sub>5</sub>GlcNAc<sub>2</sub> formadas se forman de hecho *in vivo* (en lugar de haberse secretado y modificado adicionalmente por

## ES 2 335 345 T3

manosidasas en el exterior de la célula); y (2) no está claro si cualquier estructura de  $\text{Man}_5\text{GlcNAc}_2$  formada, si de hecho se ha formado *in vivo*, es la isoforma correcta para ser un sustrato productivo para modificación posterior de *N*-glucano por GlcNAc transferasa I (Maras y col. (1997) Eur. J. Biochem. 249: 701-707).

- 5 Con el objetivo de proporcionar una glucoproteína más de tipo humano obtenida de un huésped fúngico, la Patente de Estados Unidos N° 5.834.251 describe un procedimiento para producir una glucoproteína híbrida obtenida de *Trichoderma reesei*. Un *N*-glucano híbrido tiene solamente restos de manosa en el brazo  $\text{Man}\alpha 1-6$  de la estructura de manosa central y una o dos antenas complejas en el brazo  $\text{Man}\alpha 1-3$ . Aunque esta estructura tiene utilidad, el procedimiento tiene la desventaja de que se tienen que realizar numerosas etapas enzimática *in vitro*, lo que es costoso y requiere 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 de Manosidasa Intracelular Implicada en el Corte de N-glucano*

- 15 Se requiere actividad de alfa-1,2-manosidasa para el corte de  $\text{Man}_8\text{GlcNAc}_2$  para formar  $\text{Man}_5\text{GlcNAc}_2$ , que es un intermedio principal para la formación de *N*-glucano complejo en mamíferos. El trabajo previo ha demostrado que se puede expresar  $\alpha$ -1,2-manosidasa murina, fúngica y humana truncada en la levadura metilótrofa *P. pastoris* y presenta actividad de corte de  $\text{Man}_8\text{GlcNAc}_2$  hasta  $\text{Man}_5\text{GlcNAc}_2$  (Lal y col. (1998) Glycobiology 8(10): 981-95; Tremblay y col. (1998) Glycobiology 8 (6): 585-95, Callewaert y col. (2001) FEBS Lett. 503(2-3): 173-8). Sin embargo, hasta la fecha no existen informes que demuestran el alto nivel de corte *in vivo* de  $\text{Man}_8\text{GlcNAc}_2$  hasta  $\text{Man}_5\text{GlcNAc}_2$  en una glucoproteína secretada de *P. pastoris*.

- Además, la mera presencia de una  $\alpha$ -1,2-manosidasa en la célula no garantiza, por sí misma, el corte intracelular apropiado de  $\text{Man}_8\text{GlcNAc}_2$  hasta  $\text{Man}_5\text{GlcNAc}_2$ . (Véase, por ejemplo, Contreras y col. documento WO 02/00856 A2, en el que se localiza una manosidasa marcada con HDEL de *T. reesei* principalmente en el RE y se co-expresa con una proteína indicadora de hemaglutinina (HA) de gripe en la que prácticamente no se pudo detectar  $\text{Man}_5\text{GlcNAc}_2$ . Véase también Chiba y col. (1998) J. Biol. Chem. 273 (41): 26298-26304, en el que una fusión quimérica de  $\alpha$ -1,2-manosidasa/dominio transmembrana de Ochlp localizada en el RE, Golgi temprano y citosol de *S. cerevisiae*, no tenía actividad de corte de manosidasa). Por consiguiente, la mera localización de una manosidasa en el RE o Golgi es insuficiente para garantizar la actividad de la respectiva enzima en ese orgánulo dirigido. (Véase también Martinet y col. (1998) Biotech. Letters 20(12): 1171-1177, que muestra que una  $\alpha$ -1,2-manosidasa de *T. reesei*, aunque se localice intracelularmente, aumentó más que disminuyó el alcance de manosilación). Hasta la fecha no existen informes que demuestren la localización intracelular de una  $\alpha$ -1,2-manosidasa heteróloga activa en hongos o levaduras usando una secuencia de localización transmembrana.

- 35 Aunque es útil modificar cepas que son capaces de producir  $\text{Man}_5\text{GlcNAc}_2$  como la estructura de *N*-glucano principal, cualquier intento de modificar adicionalmente estas estructuras de precursor de alto contenido de manosa para parecerse más estrechamente a glucanos humanos requiere etapas adicionales *in vivo* o *in vitro*. Se describen procedimientos para humanizar adicionalmente glucanos de fuentes fúngicas y de levadura *in vitro* en la Patente de Estados Unidos N° 5.834.251 (anteriormente). Si se tiene que humanizar  $\text{Man}_5\text{GlcNAc}_2$  adicionalmente *in vivo*, se tiene que garantizar que las estructuras de  $\text{Man}_5\text{GlcNAc}_2$  generadas, de hecho, se generen intracelularmente y no sean el producto de actividad de manosidasa en el medio. La formación de *N*-glucano complejo en levaduras u hongos requerirá que se generen altos niveles de  $\text{Man}_5\text{GlcNAc}_2$  dentro de la célula debido a que solamente los glucanos  $\text{Man}_5\text{GlcNAc}_2$  intracelulares se pueden procesar adicionalmente hasta *N*-glucanos híbridos y complejos *in vivo*. Además, se tiene que demostrar que la mayoría de las estructuras de  $\text{Man}_5\text{GlcNAc}_2$  generadas de hecho son un sustrato para GnTI y por tanto, permiten la formación de *N*-glucanos híbridos y complejos.

- Por consiguiente, existe la necesidad de procedimientos para producir glucoproteínas caracterizadas por un alto contenido intracelular de  $\text{Man}_5\text{GlcNAc}_2$  que se pueda procesar adicionalmente a estructuras de glucoproteína de tipo humano en células huésped eucariotas no humanas y, particularmente, en levaduras y hongos filamentosos.

### *N-Acetilglucosaminiltransferasas*

- 55 Las *N*-acetilglucosaminiltransferasas ("GnT") pertenecen a otra clase de enzimas de glucosilación que modifican oligosacáridos ligados a N en la ruta secretora. Tales glucosiltransferasas catalizan la transferencia de monosacárido de donadores de nucleótido azúcar específicos en una posición de hidroxilo particular de un monosacárido en una cadena de glucano creciente en uno de dos posibles enlaces anoméricos ( $\alpha$  o  $\beta$ ). Dennis y col. (1999) Bioessays 21 (5): 412-21. Las GnT específicas añaden *N*-acetilglucosamina ("GlcNAc") al brazo  $\text{Man}\alpha 1,6$  o al brazo  $\text{Man}\alpha 1,3$  de un sustrato de *N*-glucano (por ejemplo,  $\text{Man}_5\text{GlcNAc}_2$  ("núcleo de manosa-5") y  $\text{Man}_3\text{GlcNAc}_2$  (una "estructura central interna")). El producto de reacción (por ejemplo,  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  o  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ ) después se puede modificar hasta estructuras de oligosacáridos ligados a N bi-, tri- y tetra-antennarias. El documento WO 02/00879 describe un procedimiento para producir glucoproteínas que tienen estructuras de hidrato de carbono "de tipo humano" que comprenden la etapa de dirigir enzimas tales como GnT I-VI a sitios *in vivo* de glucosilación tales como el RE/red de Golgi. El documento WO 02/00879 describe adicionalmente ejemplos particulares de células huésped tales como *Pichia pastoris* que son deficientes en OCH1 y que expresan enzimas de glucosilación tales como GnT I, GnT II y manosidasas. Choi y col. (2003) PNAS 100 (9): 5022-27 describen el aislamiento de cepas modificadas genéticamente de *Pichia pastoris* que producen glucoproteínas que tienen *N*-glucanos que tienen una estructura de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  con un alto rendimiento.

La *N*-acetilglucosaminiltransferasa III (“GnTIII”) es una enzima que cataliza la adición de una GlcNAc en la manosa central del núcleo de trimanosa (Man $\alpha$ 1,6 (Man $\alpha$ 1,3) Man $\beta$ 1,4 - GlcNAc $\beta$ 1,4 - GlcNAc $\beta$ 1,4 - Asn) de un oligosacárido ligado a N. La adición por GnTIII de una GlcNAc de bisección a un sustrato aceptor (por ejemplo, núcleo de trimanosa) produce un denominado *N*-glucano bisechado. Por ejemplo, la adición por GnTIII de una GlcNAc de bisección a la estructura de GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> puede producir un *N*-glucano bisechado, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>. De forma similar, la adición por GnTIII de una GlcNAc de bisección a una estructura de GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> produce otro *N*-glucano bisechado, GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>. Esta última estructura se ha implicado en una mayor citotoxicidad celular dependiente de anticuerpos (ADCC). Umana y col. (1999) Nat. Biotechnol. 17(2): 176-80. Se pueden formar otros *N*-glucanos bisechados por la acción de GnTIII. Por ejemplo, GlcNAcMan<sub>4</sub>GlcNAc<sub>2</sub> se puede convertir en GlcNAc<sub>2</sub>Man<sub>4</sub>GlcNAc<sub>2</sub> bisechado, Man<sub>5</sub>GlcNAc<sub>2</sub> se puede convertir en GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> bisechado y GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> se puede convertir en GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> bisechado. Véase, por ejemplo, Narasimhan (1982) J. Biol. Chem. 257: 10235-42. Hasta ahora, la actividad de GnTIII se ha demostrado solamente en células de mamífero.

La re-modificación de glucoformas de inmunoglobulinas expresadas por células de mamífero es una tarea tediosa y compleja. Especialmente en el caso de GnTIII, donde la sobre-expresión de esta enzima se ha implicado en la inhibición del crecimiento, se han tenido que emplear procedimientos que implican expresión génica regulada (inducible) para producir inmunoglobulinas con *N*-glucanos bisechados. Umana y col. (1999) Biotechnol Bioeng. 65(5): 542-9; Umana y col. (1999) Nat. Biotechnol. 17(2): 176-80; Umana y col. documento WO 03/011878; Patente de Estados Unidos N° 6.602.684. Tal efecto de inhibición del crecimiento complica la capacidad de co-expresar la proteína diana y GnTIII y puede impartir un límite superior a la sobre-expresión de GnTIII. Patente de Estados Unidos N° 6.602.684. Puede ser necesaria la optimización cuidadosa de los niveles de expresión de GnTIII. *Id.* Por lo tanto, lo que se necesita es un sistema de producción de proteína que utilice la capacidad inherente de títulos de producto robustos tales como los producidos en células huésped eucariotas inferiores (por ejemplo, levaduras y hongos filamentosos), que sea capaz de producir *N*-glucanos bisechados en proteínas, especialmente proteínas terapéuticas, expresadas en estas células. Sin embargo, como se ha descrito anteriormente, el desarrollo de las células huésped eucariotas inferiores usadas en tal sistema de producción de proteína requiere que las rutas de glucosilación endógenas de las células huésped se modifiquen adicionalmente.

### 30 Sumario de la invención

Se han desarrollado células huésped y líneas celulares que tienen rutas de glucosilación modificadas genéticamente que permiten que realicen 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 producen glucoproteínas más similares, si no sustancialmente idénticas, a sus equivalentes de mamífero, por ejemplo, humanos. Las células huésped de la invención, es decir, una célula huésped de levadura u hongo unicelular o multicelular filamentosos desarrollado en cultivo, se modifican para producir una glucoproteína que tiene una estructura de Man<sub>3</sub>GlcNAc<sub>2</sub> o de Man<sub>5</sub>GlcNAc<sub>2</sub> en la que se introduce un ácido nucleico que codifica una actividad de *N*-acetilglucosaminiltransferasa III en la célula huésped en la que dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnT III $\Delta$ 32 de ratón o GnT III $\Delta$ 86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*, de tal forma que más del 10% en mol de una glucoproteína recombinante producida en la célula huésped comprende una estructura de *N*-glucano bisechado. Este resultado se consigue usando una combinación de modificación y/o selección de cepas que, por ejemplo, no expresan enzimas que crean las estructuras indeseables características de las glucoproteínas fúngicas y que expresan *N*-acetilglucosaminiltransferasa III y, por ejemplo, expresan enzimas heterólogas adicionales capaces de producir una glucoproteína “de tipo humano”.

Por tanto, la presente invención proporciona un procedimiento de producción para una glucoproteína recombinante que comprende un glucano bisechado usando células huésped de levadura u hongo unicelular o multicelular filamentosos que tienen un patrón de glucosilación diferente de seres humanos y que se modifican para producir una glucoproteína que tiene una estructura de Man<sub>3</sub>GlcNAc<sub>2</sub> o de Man<sub>5</sub>GlcNAc<sub>2</sub> y que comprenden una actividad de *N*-acetilglucosaminiltransferasa III, de tal forma que las células huésped se parecen más estrechamente a estructuras de hidratos de carbono encontradas en proteínas de mamífero, por ejemplo, humanas. Este procedimiento permite obtener una célula huésped modificada que se puede usar 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 especialista en el campo de la expresión de proteínas. Se crean o seleccionan células huésped con oligosacáridos modificados. Para la producción de proteínas terapéuticas, este procedimiento se puede adaptar para modificar líneas celulares en las que se puede obtener cualquier estructura de glucosilación deseada.

Por consiguiente, en una realización, la invención proporciona procedimientos para preparar una glucoproteína de tipo humano en una célula huésped de levadura u hongo unicelular o multicelular filamentosos por introducción en la célula de una actividad de *N*-acetilglucosaminiltransferasa III donde dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnT III $\Delta$ 32 de ratón o GnT III $\Delta$ 86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*. En una realización preferida, dicha actividad de *N*-acetilglucosaminiltransferasa III se expresa en la célula y, en una realización aún más preferida, esta expresión da como resultado la producción de *N*-glucanos que comprenden estructuras bisechadas de GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> o GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>. En otra realización preferida, dicha actividad de *N*-acetilglucosa-

miniltransferasa III es sustancialmente intracelular. En otra realización preferida de la invención, la glucoproteína que incluye los *N*-glucanos con estructuras bisecadas se aísla de la célula huésped eucariota inferior. En una realización aún más preferida, la glucoproteína producida en la célula huésped es una proteína terapéutica.

5 En otro aspecto, la invención proporciona una célula huésped que incluye tanto dicha actividad de *N*-acetilglucosaminiltransferasa III como una actividad de *N*-acetilglucosaminiltransferasa II. En una realización preferida, la célula huésped que incluye dicha actividad de *N*-acetilglucosaminiltransferasa III produce *N*-glucanos o glucoproteínas que comprenden estructuras de GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> o GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc que son capaces de reaccionar con esta actividad. La actividad produce un glucano bisecado. La célula huésped de la invención, por  
10 tanto, incluye un *N*-glucano con un glucano bisecado. En una realización preferida, el *N*-glucano incluye más del 10% en mol del glucano bisecado. En algunas realizaciones, la célula huésped incluye un *N*-glucano que comprende estructuras bisecadas de GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> o GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>. En una realización preferida, la célula huésped incluye una estructura central de Man<sub>5</sub>GlcNAc<sub>2</sub> o una estructura central de Man<sub>3</sub>GlcNAc<sub>2</sub> que se modifica por una GlcNAc de bisección. En una realización aún más preferida, la célula produce más del 10%  
15 en mol de la estructura modificada.

En otra realización de la invención, la célula huésped contiene una actividad de *N*-acetilglucosaminiltransferasa I además de dicha actividad de *N*-acetilglucosaminiltransferasa III. En una realización preferida, las actividades son sustancialmente intracelulares. En otra realización preferida, la célula produce *N*-glucanos que comprenden  
20 GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> que son capaces de reaccionar con dicha actividad de GnTIII. Dicha actividad de GnTIII de la célula produce un glucano bisecado.

En otra realización, la célula huésped de la invención contiene tanto dicha actividad de *N*-acetilglucosaminiltransferasa III como una actividad de manosidasa II. En una realización preferida, la célula huésped contiene adicionalmente  
25 una actividad de *N*-acetilglucosaminiltransferasa I. En otra realización preferida, la célula huésped contiene además una actividad de *N*-acetilglucosaminiltransferasa II. En otra realización preferida, la célula huésped contiene además tanto una actividad de *N*-acetilglucosaminiltransferasa I como una actividad de *N*-acetilglucosaminiltransferasa II.

30 En otra realización, la célula huésped de la invención es deficiente en una actividad de *OCHI* manosiltransferasa. Tal célula puede, por ejemplo, ser deficiente en una actividad de Dol-P-Man:Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol manosiltransferasa. En otra realización más, la célula huésped de la invención puede comprender además una actividad de  $\alpha$ -1,2-manosidasa I. En otra realización, la célula huésped puede comprender además un transportador de UDP-GlcNAc.

35 La presente invención también proporciona glucoproteínas recombinantes que se preparan mediante los procedimientos de la invención. La glucoproteína recombinante de la invención incluye una GlcNAc de bisección en una estructura central de Man<sub>5</sub>GlcNAc<sub>2</sub>, Man<sub>4</sub>GlcNAc<sub>2</sub> o Man<sub>3</sub>GlcNAc<sub>2</sub> y se produce en una célula huésped como se define en este documento. En otra realización, la glucoproteína incluye una GlcNAc de bisección unida a una estructura central seleccionada entre el grupo constituido por: GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>,  
40 GlcNAcMan<sub>4</sub>GlcNAc<sub>2</sub> y GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> se produce en una célula huésped eucariota inferior. En una realización preferida, más del 10% en mol de las estructuras centrales de la glucoproteína de la invención se modifican por la GlcNAc de bisección.

En otro aspecto, la invención proporciona composiciones farmacéuticas que contienen las glucoproteínas recombinantes de tipo humano de la invención producidas en una célula huésped de la invención. En este documento se describen adicionalmente vectores que codifican proteínas que tienen actividad de *N*-acetilglucosaminiltransferasa III y que contienen secuencias de péptido de dirección unidas. Las proteínas codificadas por los vectores se localizan en una célula huésped eucariota inferior de tal forma que producen *N*-glucanos que tienen estructuras bisecadas. La invención también se define por las reivindicaciones.  
50

#### Breve descripción de los dibujos

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

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

La Figura 2 ilustra la construcción de una genoteca de ADN combinatoria de construcciones de fusión. La Figura 2A ilustra la inserción de un fragmento peptídico de dirección en pCR2.1 -TOPO (Invitrogen, Carlsbad, CA). La Figura 2B muestra la sub-biblioteca de péptido de dirección generada que tiene sitios de restricción *NotI*-*AscI*. La  
60 Figura 2C ilustra 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 ilustra una construcción de fusión particular generada a partir de la sub-biblioteca de péptido de dirección y la sub-biblioteca de dominio catalítico.

65 La Figura 3 ilustra la secuencia de ácido nucleico de fase de lectura abierta de  $\alpha$ -1,2-manosidasa IA de *M. musculus* (SEC ID N°: 48) y la secuencia polipeptídica codificada (SEC ID N°: 49). Las secuencias de los cebadores de PCR usados para generar truncamientos *N*-terminales están subrayadas.

## ES 2 335 345 T3

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

Las Figuras 5A-5E muestran análisis de MALDI-TOF que demuestra producción del dominio de kringle 3 (K3) de glucoproteínas de plasminógeno humano que tienen  $\text{Man}_5\text{GlcNAc}_2$  como la estructura de *N*-glucano predominante en *P. pastoris*. La Figura 5A ilustra el glucano  $\text{Man}_5\text{GlcNAc}_2$  convencional [a] (Glyko, Novato, CA) y  $\text{Man}_5\text{GlcNAc}_2 + \text{Na}^+$  [b]. La Figura 5B muestra glucanos liberados por PNGasa de tipo silvestre de K3. Los *N*-glucanos mostrados son los siguientes:  $\text{Man}_9\text{GlcNAc}_2$  [d];  $\text{Man}_{10}\text{GlcNAc}_2$  [e];  $\text{Man}_{11}\text{GlcNAc}_2$  [f];  $\text{Man}_{12}\text{GlcNAc}_2$  [g]. La Figura 5C ilustra la delección de *och1* que da como resultado la producción de  $\text{Man}_8\text{GlcNAc}_2$  [c] como el *N*-glucano predominante. Las Figuras 5D y 5E muestran la producción de  $\text{Man}_5\text{GlcNAc}_2$  [b] después del corte *in vivo* de  $\text{Man}_8\text{GlcNAc}_2$  con una  $\alpha$ -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  $\text{Man}_5\text{GlcNAc}_2$  [b].

Las Figuras 6A-6F muestran análisis de MALDI-TOF que demuestra producción de glucoproteínas IFN- $\beta$  que tienen  $\text{Man}_5\text{GlcNAc}_2$  como la estructura de *N*-glucano predominante en *P. pastoris*. La Figura 6A muestra  $\text{Man}_5\text{GlcNAc}_2$  convencional [a] y  $\text{Man}_5\text{GlcNAc}_2 + \text{Na}^+$  [b] como el patrón (Glyko, Novato, CA). La Figura 6B muestra glucanos liberados por PNGasa - de tipo silvestre de IFN- $\beta$ . La Figura 6C ilustra el knock-out de *och1* que produce  $\text{Man}_8\text{GlcNAc}_2$  [c];  $\text{Man}_9\text{GlcNAc}_2$  [d];  $\text{Man}_{10}\text{GlcNAc}_2$  [e];  $\text{Man}_{11}\text{GlcNAc}_2$  [f];  $\text{Man}_{12}\text{GlcNAc}_2$  [g]; y ninguna producción de  $\text{Man}_5\text{GlcNAc}_2$  [b]. La Figura 6D muestra una cantidad relativamente pequeña de  $\text{Man}_5\text{GlcNAc}_2$  [b] entre otros *N*-glucanos intermedios  $\text{Man}_8\text{GlcNAc}_2$  [c] a  $\text{Man}_{12}\text{GlcNAc}_2$  [g]. La Figura 6E muestra una cantidad significativa de  $\text{Man}_5\text{GlcNAc}_2$  [b] con respecto a los demás glucanos  $\text{Man}_8\text{GlcNAc}_2$  [c] y  $\text{Man}_9\text{GlcNAc}_2$  [d] producidos por pGC5 (*MNS1(m)* de *Saccharomyces*/manosidasa IB  $\Delta 99$  de ratón). La Figura 6F muestra la producción predominante de  $\text{Man}_5\text{GlcNAc}_2$  [b] en la glucoproteína secretada IFN- $\beta$  por pFB8 (*SEC12(m)* de *Saccharomyces*/manosidasa IA  $\Delta 187$  de ratón). El *N*-glucano se indica por un pico con una masa (*m/z*) de 1254 coherente con su identificación como  $\text{Man}_5\text{GlcNAc}_2$  [b].

La Figura 7 muestra un cromatograma líquido de alto rendimiento para: (A) patrón de  $\text{Man}_9\text{GlcNAc}_2$  marcado con 2-AB (control negativo); (B) sobrenadante de medio de *P. pastoris*,  $\Delta och1$  transformado con manosidasa de pFB8, que demuestra una ausencia de actividad de manosidasa extracelular en el sobrenadante; y (C) patrón de  $\text{Man}_9\text{GlcNAc}_2$  marcado con 2-AB después de la exposición a manosidasa de *T. reesei* (control positivo).

La Figura 8 muestra un cromatograma líquido de alto rendimiento para: (A) patrón de  $\text{Man}_9\text{GlcNAc}_2$  marcado con 2-AB (control negativo); (B) sobrenadante de medio de *P. pastoris*,  $\Delta och1$  transformado con manosidasa de pGC5, que demuestra una ausencia de actividad de manosidasa extracelular en el sobrenadante; y (C) patrón de  $\text{Man}_9\text{GlcNAc}_2$  marcado con 2-AB después de exposición a manosidasa de *T. reesei* (control positivo).

La Figura 9 muestra un cromatograma líquido de alto rendimiento para: (A) patrón de  $\text{Man}_9\text{GlcNAc}_2$  marcado con 2-AB (control negativo); (B) sobrenadante de medio de *P. pastoris*,  $\Delta och1$  transformado con manosidasa de pBC 18-5, que demuestra ausencia de actividad de manosidasa extracelular en el sobrenadante; y (C) sobrenadante de medio de *P. pastoris*,  $\Delta och1$  transformado 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  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  en *P. pastoris*. La Figura 10A ilustra una cepa de *P. pastoris* (YSH-3) con una GnTI humana pero sin el transportador de UDP-GlcNAc dando como resultado cierta producción de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [b] pero una producción predominante de  $\text{Man}_5\text{GlcNAc}_2$  [a]. La Figura 10B ilustra la adición de transportador de UDP-GlcNAc de *K. lactis* en una cepa (PBP-3) con la GnTI humana, que dio como resultado la producción predominante de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [b]. El único pico prominente de masa (*m/z*) en 1457 es coherente con su identificación como  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [b] como se muestra en la Figura 10B.

La Figura 11 muestra un óptimo de pH de una enzima manosidasa heteróloga codificada por pBB27-2 (*MNN10(s)* de *Saccharomyces*/manosidasa IB  $\Delta 31$  de *C. elegans*) expresada en *P. pastoris*.

Las Figuras 12A-12C muestran 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 de tipo de alto contenido de manosa. La Figura 12B muestra los *N*-glucanos liberados de células a las que se ha delecionado *och1 mnn1*, poniendo de manifiesto un pico de masa (*m/z*) distinto en 1908 coherente con su identificación como  $\text{Man}_9\text{GlcNAc}_2$  [d]. La Figura 12C muestra los *N*-glucanos liberados de células a las que se ha delecionado *och1 mnn1* después de una digestión *in vitro* con  $\alpha$ -1,2-manosidasa correspondientes a un pico coherente con  $\text{Man}_5\text{GlcNAc}_2$ .

La Figura 13 es una vista esquemática de la estructura del oligosacárido ligado a dolcilil pirofosfato.

La Figura 14 es una vista esquemática de la generación de *N*-glucanos  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  de células huésped fúngicas que son deficientes en actividades de *alg3*, *alg9* o *alg12*.

La Figura 15 es una vista esquemática de reacciones de procesamiento requeridas para producir estructuras de oligosacárido de tipo mamífero en una célula huésped fúngica con un genotipo de *alg3*, *och1*.

## ES 2 335 345 T3

La Figura 16 muestra Comparaciones de Secuencia de *Alg3* de *S. cerevisiae* (Blast) (SEC ID N°: 9-20, respectivamente, en orden de aparición).

La Figura 17 muestra Secuencias de *ALG3* (SEC ID N°: 21) y *Alg3p* (SEC ID N°: 22) de *S. cerevisiae*.

La Figura 18 muestra Secuencias de *ALG3* (SEC ID N°: 23) y *Alg3p* (SEC ID N°: 24) de *P. pastoris*.

La Figura 19 muestra Comparaciones de Secuencia de *ALG3* de *P. pastoris* (Blast) (SEC ID N°: 23-31, respectivamente, en orden de aparición).

La Figura 20 muestra Secuencias de *ALG3* (SEC ID N°: 33) y *Alg3p* (SEC ID N°: 34) de *K. lactis*.

La Figura 21 muestra Comparaciones de Secuencia de *ALG3* de *K. lactis* (Blast) (SEC ID N°: 35-40, respectivamente, en orden de aparición).

La Figura 22 muestra un modelo de una inmunoglobulina IgG. La cadena pesada y la cadena ligera se pueden subdividir, basándose en estructura secundaria y terciaria similar, en dominios. Las dos cadenas pesadas (dominios  $V_H$ ,  $C_H1$ ,  $C_H2$  y  $C_H3$ ) se unen mediante tres puentes disulfuro. Las cadenas ligeras (dominios  $V_L$  y  $C_L$ ) se unen por otro puente disulfuro a la porción  $C_H1$  de la cadena pesada y, junto con los fragmentos  $C_H1$  y  $V_H$ , establecen la región Fab. Los antígenos se unen en la porción terminal de la región Fab. Las funciones de efector, tales como unión a Receptor Fc-gamma, se han localizado en el dominio  $C_H2$ , justo cadena abajo de la región bisagra y están influidas por la *N*-glucosilación de asparagina 297 en la cadena pesada.

La Figura 23 es una revisión esquemática de un vector de expresión de IgG1 modular.

La Figura 24 muestra Secuencias de Ácido Nucleico (SEC ID N°: 45) y de Aminoácidos (SEC ID N°: 46) de *GnTIII* de *M. musculus*.

La Figura 25 (parte superior) es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 producida en un **YSH-1** de *P. pastoris* que presenta un pico predominante en 1461 m/z correspondiente a la masa de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d]; la Figura 25 (parte inferior) muestra un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 producida en un **YSH-1** de *P. pastoris* transformado con manosidasa  $\text{II}\Delta 74$  de *D. melanogaster*/MNN2(s) de *S. cerevisiae* que muestra un pico predominante en 1140 m/z correspondiente a la masa de  $\text{GlcNAcMan}_3\text{GlcNAc}_2$  [b] y otros picos que se corresponden a  $\text{GlcNAcMan}_4\text{GlcNAc}_2$  [c] en 1303 m/z y  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d] en 1465 m/z. Esta cepa se denominó **YSH-37**.

La Figura 26 (parte superior) es el análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 producida en **YSH-1** de *P. pastoris* como se muestra en la Figura 25 (parte superior); la Figura 26 (parte inferior) es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 expresada en células **YSH-1** de *P. pastoris* transformadas con una construcción **pVA53** (MNN2(s) de *S. cerevisiae*/mGnTIII). El pico en 1463 m/z se corresponde a la masa de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d] y el pico en 1666 m/z se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$  [a].

La Figura 27 (parte superior) es el análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 producida en **YSH-1** de *P. pastoris* como se muestra en la Figura 25 (parte superior); la Figura 27 (parte inferior) es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 expresada en células **YSH-1** de *P. pastoris* transformadas con una construcción **pVA55** (MNN2(s) de *S. cerevisiae*/mGnTIII). El pico en 1463 m/z se corresponde a la masa de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d] y el pico en 1667 m/z se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$  [a].

Figura 28 (parte superior) es el análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 producida en **YSH-1** de *P. pastoris* como se muestra en la Figura 25 (parte superior); la Figura 28 (parte inferior) es un análisis de EM de MALDI-TOF de *N*-glucanos aislado de una glucoproteína kringle 3 expresada en células **YSH-1** de *P. pastoris* transformadas con una construcción **pVB51** (GNT1(s) de *K. lactis*/mGnTIII). El pico predominante en 1463 m/z se corresponde a la masa de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d] y se observa un segundo pico en 1726 m/z [e], que no se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ .

La Figura 29 es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 expresada en células **YSH-44** de *P. pastoris*. El pico predominante en 1356 m/z se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  [x].

La Figura 30 es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 expresada en células **YSH-44** de *P. pastoris* transformadas con una construcción **pVA53** (MNN2(s) de *S. cerevisiae*/mGnTIII). El pico en 1340 m/z se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  [x] y el pico en 1542 m/z se corresponde a la masa de  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$  [y].

## ES 2 335 345 T3

La Figura 31 es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 expresada en células **PPB6-5** de *P. pastoris*. El pico predominante en 1340 m/z se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  [x].

5 La Figura 32 es un análisis de EM de MALDI-TOF de *N*-glucanos aislados de una glucoproteína kringle 3 expresada en células **PPB6-5** de *P. pastoris* transformadas con una construcción **pVA53** (*MNN2(s)* de *S. cerevisiae/mGnTIII*). El pico en 1340 m/z se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  [x] y el pico en 1543 m/z se corresponde a la masa de  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$  [y].

10 La Figura 33 muestra un cromatograma líquido de alto rendimiento, que demuestra una ausencia de actividad de GnTIII extracelular (**pVA53**) en el sobrenadante. El *N*-glucano  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  purificado de K3 expresado en una cepa **PPB-3** se añadió a: BMMY (A); UDP-GlcNAc 1 mM (Sigma Chemical Co., St. Louis, MO) en BMMY (B); el sobrenadante de **YSH-44** transformado con **pVA53** [**YSH-57**] (C); y el sobrenadante de **YSH-57** + UDP-GlcNAc 1 mM (D).

15 La Figura 34 muestra un cromatograma líquido de alto rendimiento, que muestra una ausencia de actividad de GnTIII extracelular (**pVA53**) en el sobrenadante. El *N*-glucano  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  purificado de K3 expresado en una cepa **YSH-44** se añadió a: BMMY (A); UDP-GlcNAc 1 mM (Sigma Chemical Co., St. Louis, MO) en BMMY (B); y el sobrenadante de **YSH-44** transformado con pVA53 [**YSH-57**] (C).

20 La Figura 35 es un diagrama esquemático que compara las rutas de glucosilación normal en seres humanos y *P. pastoris* (Panel A) con una ruta de *N*-glucosilación humanizada modificada en eucariotas inferiores (Panel B). La ruta modificada representa la construcción de la cepa de *P. pastoris* **PPB6-5**, que, después de la modificación con GnTIII, se convierte en la cepa de *P. pastoris* **PPB38**.

25 La Figura 36 es un diagrama esquemático que muestra la glucoforma secretada de forma predominante producida por cada una de las cepas de *P. pastoris* designadas y la modificación génica usada para modificar cada una de las cepas.

30 La Figura 37 es una representación estructural de la transferencia de una GlcNAc al intermedio de oligosacárido,  $\text{GlcNAcMan}_3\text{GlcNAc}_2$ , producido en glucoproteínas en una célula huésped eucariota inferior, cuando se cataliza por GnTIII.

35 La Figura 38 es una representación estructural de la transferencia de una GlcNAc al intermedio de oligosacárido,  $\text{GlcNAcMan}_3\text{GlcNAc}_2$ , producido en glucoproteínas en una célula huésped de eucariota inferior, cuando se cataliza por GnTII, y la transferencia posterior de una GlcNAc al producto de esa reacción,  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ , cuando se cataliza por GnTIII.

### Descripción detallada de la invención

40 A menos que se defina de otro modo en este documento, los términos científicos y técnicos usados en relación con la presente invención deben tener los significados que se comprenden de forma común por los especialistas habituales en la técnica. Además, a menos que se requiera de otro modo por el contexto, los términos en singular deben incluir los plurales y los términos en plural deberán incluir el singular. Los procedimientos y las técnicas de la presente invención generalmente se realizan 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 de proteína y ácido nucleico e hibridación descritas en este documentos son las bien conocidas y usadas de forma común en la técnica.

50 Los procedimientos y las técnicas de la presente invención se realizan generalmente de acuerdo con procedimientos convencionales bien conocidos en la técnica y como se describen en diversas referencias generales y más específicas que se citan y analizan a lo largo de la presente memoria descriptiva a menos que se indique de otro modo. Véase, por ejemplo, Sambrook y col., Molecular Cloning: A Laboratory Manual, 2a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y. (1989); Ausubel y col., 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 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 de laboratorio y técnicas de biología molecular y celular, bioquímica de proteínas, enzimología y química medicinal y farmacéutica descritas en este documento son las que se conocen bien y se usan de forma común en la técnica.

65 Las siguientes expresiones, a menos que se indique de otro modo, se debe entender que tienen los siguientes significados:

Como se usa en este documento, la expresión "*N*-glucano" se refiere a un oligosacárido ligado a N, por ejemplo, uno que está unido por un enlace de asparagina-*N*-acetilglucosamina a un resto de asparagina de un polipéptido. Los

## ES 2 335 345 T3

*N*-glucanos tienen un núcleo de pentasacárido común de Man<sub>3</sub>GlcNAc<sub>2</sub> (“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 Man<sub>3</sub>GlcNAc<sub>2</sub> (“Man<sub>3</sub>”). La expresión “núcleo de pentamanosa” o “núcleo de Manosa-5” o “Man<sub>5</sub>” usada con respecto al *N*-glucano se refiere a la estructura Man<sub>5</sub>GlcNAc<sub>2</sub>. Los *N*-glucanos difieren con respecto al número de ramas (antenas) que comprenden azúcares periféricos (por ejemplo, GlcNAc, fucosa y ácido siálico) que se unen a la estructura central de Man<sub>3</sub>. Los *N*-glucanos se clasifican de acuerdo con sus constituyentes ramificados (por ejemplo, manosa superior, complejo o híbrido).

Un *N*-glucano de tipo “alto contenido en manosa” tiene cinco o más restos de manosa. Un *N*-glucano de tipo “complejo” típicamente tiene al menos una GlcNAc unida al brazo de 1,3 manosa y al menos una GlcNAc unida al brazo de 1,6 manosa del núcleo de trimanosa. Los *N*-glucanos complejos también pueden tener restos de galactosa (“Gal”) que están modificados opcionalmente con ácido siálico o derivados (“NeuAc”, donde “Neu” se refiere a ácido neuramínico y “Ac” se refiere a acetilo). Un *N*-glucano complejo tiene típicamente al menos una rama que termina en un oligosacárido tal como, por ejemplo: NeuNAc-; NeuAca2-6GalNAca1-; NeuAca2-3Galβ1-3GalNAca1-; NeuAca2-3/6Galβ1-GlcNAcβ1-; GlcNAca1-4Galβ1-(solamente mucinas); Fuca1-2Galβ1-(grupo sanguíneo H). Los ésteres de sulfato pueden tener lugar en restos de galactosa, GalNAc y GlcNAc y los ésteres de fosfato pueden tener lugar en restos de manosa. El NeuAc (Neu: ácido neuramínico; Ac: acetilo) puede estar O-acetilado o sustituido por NeuGI (ácido-*N*-glucolilneuramínico). Los *N*-glucanos complejos también pueden tener sustituciones intracadena que comprenden GlcNAc “de bisección” y fucosa central (“Fuc”). Un *N*-glucano “híbrido” tiene al menos una GlcNAc en el extremo del brazo de 1,3 manosa del núcleo de trimanosa y cero o más manosas en el brazo 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 desorción e ionización por láser asistida por matriz-tiempo de vuelo (MALDI-TOF).

Las abreviaturas usadas en este documento son de uso común en la técnica, véase, por ejemplo, las abreviaturas de azúcares, anteriormente. Otras abreviaturas comunes incluyen “PNGasa”, que se refiere al 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 este 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 sintética (que también son útiles y que se pueden manipular adicionalmente *in vitro* o *in vivo*) que tienen al menos cinco restos de manosa. Preferiblemente, las glucoproteínas producidas de acuerdo con la invención contienen al menos el 30% en mol, preferiblemente al menos el 40% en mol y más preferiblemente el 50, 60, 70, 80, 90 o incluso el 100% en mol del intermedio Man<sub>3</sub>GlcNAc<sub>2</sub>, al menos de forma transitoria. Esto se puede conseguir, por ejemplo, modificando una célula huésped de la invención para expresar una enzima de glucosilación “mejor”, es decir, más eficaz. Por ejemplo, se selecciona una manosidasa de tal forma que tendrá una actividad óptima en las condiciones presentes en el sitio en la célula huésped donde las proteínas se glucosilan y se introduce en la célula huésped preferiblemente dirigiendo la enzima a un orgánulo de la célula huésped en el que se desea la actividad.

El término “enzima”, cuando se usa en este documento en relación con alterar 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, quimeras, complejos y similares. Un “fragmento catalíticamente activo” de una enzima se refiere a un polipéptido que tiene un nivel detectable de actividad funcional (enzimática). La actividad de la enzima es “sustancialmente intracelular” cuando se puede medir menos del 10% de la actividad de la enzima en el exterior de la célula en comparación con la que se puede medir en células lisadas.

Una célula huésped eucariota inferior, cuando se describe en este documento en relación con perfiles de glucosilación, se refiere a cualquier célula eucariota que produce de forma normal *N*-glucanos de alto contenido de manosa y, por tanto, quiere decir que incluye algunas células animales o vegetales y más típicamente células eucariotas inferiores, que incluyen células fúngicas y de algas uni- y multicelulares. Una célula huésped de la presente invención es una levadura o un hongo unicelular o multicelular filamentosos.

Como se usa en este documento, la expresión “ruta de secreción” se refiere a la línea de ensamblaje de diversas enzimas de glucosilación a la que un precursor de oligosacárido ligado a lípido y un sustrato de *N*-glucano se exponen secuencialmente, 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. Una enzima X que actúa sobre un glucano ligado a lípido o un *N*-glucano antes que la enzima Y se dice que está o que actúa “aguas arriba” de una enzima Y; de forma similar, la enzima Y está o actúa “aguas abajo” de la enzima X.

La expresión “péptido de dirección” como se usa en este 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 sub-celulares, por ejemplo, orgánulos.

La expresión “polinucleótido” o “molécula de ácido nucleico” se refiere a una forma polimérica de nucleótidos de al menos 10 bases de longitud. La expresión incluye moléculas de ADN (por ejemplo, ADNc o ADN genómico o sin-

tético) y moléculas de ARN (por ejemplo, ARNm o ARN sintético), así como análogos de ADN o ARN que contienen análogos de nucleótidos no naturales, enlaces internucleosídicos no nativos o ambos. El ácido nucleico puede estar en cualquier conformación topológica. Por ejemplo, el ácido nucleico puede ser de cadena única, de doble cadena, de triple cadena, formar cuádruplex, tener parcialmente doble cadena, estar ramificado, formar una horquilla, ser circular o estar en una conformación de candado. La expresión incluye formas de cadena única y doble de ADN. Una molécula de ácido nucleico de esta invención puede incluir cadenas tanto con sentido como antisentido de ARN, ADNc, ADN genómico y formas sintéticas y polímeros mixtos de lo anterior. Se pueden modificar química o bioquímicamente o pueden contener bases de nucleótidos no naturales o derivatizadas, como se entenderá de forma sencilla por los especialistas en la técnica. 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, metil fosfonatos, fosfotriésteres, fosforamidatos, carbamatos, etc.), enlaces cargados (por ejemplo, fosforotioatos, fosforoditioatos, etc.), restos colgantes (por ejemplo, polipéptidos), intercaladores (por ejemplo, acridina, psolareno, etc.), quelantes, alquilantes y enlaces modificados (por ejemplo, ácidos nucleicos alfa anoméricos, etc.). También se incluyen moléculas sintéticas que imitan polinucleótidos en su capacidad de unirse a una secuencia indicada por formación de 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 de fosfato en la cadena principal de la molécula.

A menos que se indique de otro modo, un “ácido nucleico que comprende la SEC ID N°: X” se refiere a un ácido nucleico, al menos una porción del cual tiene (i) la secuencia de la SEC ID N°: X, o (ii) una secuencia complementaria a la SEC ID N°: X. La elección entre ambos se indica por el contexto. Por ejemplo, si el ácido nucleico se usa como una sonda, la elección entre ambos se indica por el requisito de que la sonda sea complementaria a 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 está asociado de forma natural. La expresión incluye un ácido nucleico o polinucleótido que (1) se ha retirado de su entorno de origen natural, (2) no está asociado con todo o una porción de un polinucleótido en el que el “polinucleótido aislado” se encuentra en la naturaleza, (3) se une operativamente a un polinucleótido al que no está unido en la naturaleza o (4) no tiene lugar en la naturaleza. La expresión “aislado” o “sustancialmente puro” también se puede usar con referencia a aislados de ADN recombinante o clonado, análogos de polinucleótidos sintetizados químicamente o análogos de polinucleótidos que se sintetizan biológicamente por sistemas heterólogos.

Sin embargo, “aislado” no requiere necesariamente que el propio ácido nucleico o polinucleótido descrito de este modo se haya eliminado físicamente de su entorno nativo. Por ejemplo, una secuencia de ácido nucleico endógeno en el genoma de un organismo se considera “aislada” en este documento si se pone una secuencia heteróloga (es decir, una secuencia que no es adyacente de forma natural a esta secuencia de ácido nucleico endógeno) adyacente a la secuencia de ácido nucleico endógeno, de tal forma que la expresión de esta secuencia de ácido nucleico endógeno está alterada. A modo de ejemplo, el promotor nativo de un gen en el genoma de una célula humana se puede sustituir (por ejemplo, por recombinación homóloga) por una secuencia de promotor no nativo, de tal forma que este gen tenga un patrón de expresión alterado. Este gen ahora se convertiría en “aislado” debido a que se ha separado de al menos algunas de las secuencias que de forma natural lo flanquean.

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

Como se usa en este documento, la frase “variante degenerada” de una secuencia de ácido nucleico de referencia incluye secuencias de ácido nucleico que se pueden traducir, 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 correspondencia máxima. La longitud de la comparación de identidad de secuencia puede ser a lo largo de un tramo de al menos aproximadamente nueve nucleótidos, habitualmente al menos aproximadamente 20 nucleótidos, más habitualmente al menos aproximadamente 24 nucleótidos, típicamente al menos aproximadamente 28 nucleótidos, más típicamente al menos aproximadamente 32 nucleótidos y preferiblemente al menos aproximadamente 36 o más nucleótidos. Existen varios algoritmos diferentes conocidos en la técnica que se pueden usar para medir la identidad de secuencia de nucleótidos. Por ejemplo, las secuencias polinucleotídicas se pueden comparar usando FASTA, Gap o Bestfit, que son programas en el Paquete de Wisconsin Versión 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA proporciona alineamientos y porcentaje de identidad de secuencia de las regiones del mejor solapamiento entre las secuencias problema y de búsqueda (Pearson (1990) *Methods Enzymol.* 183: 63-98). Por ejemplo, se puede determinar el porcentaje de iden-

## ES 2 335 345 T3

tividad de secuencia entre secuencias de ácido nucleico 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 como se proporcionan en GCG Versión 6.1.

5 La expresión “homología sustancial” o “similitud sustancial”, cuando se refiere a un ácido nucleico o fragmento del mismo, indica que, cuando se alinea óptimamente con inserciones o deleciones de nucleótidos apropiadas con otro ácido nucleico (o su cadena complementaria), existe una identidad de secuencia de nucleótidos en al menos aproximadamente el 50%, más preferiblemente el 60% de las bases de nucleótidos, habitualmente al menos aproximadamente el 70%, más habitualmente al menos aproximadamente el 80%, preferiblemente al menos aproximadamente el 90% y  
10 más preferiblemente al menos aproximadamente el 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.

15 Alternativamente, existe homología o similitud sustancial cuando un ácido nucleico o fragmento del mismo hibrida con otro ácido nucleico, con una cadena de otro ácido nucleico, o con la cadena 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 ácido nucleico se verá afectada por condiciones tales como concentración salina, temperatura, disolventes, la composición de bases de las especies hibridantes, la longitud de las regiones complementarias y el número  
20 de desapareamientos de bases de nucleótidos entre los ácidos nucleicos hibridantes, como se entenderá fácilmente por los especialistas en la técnica. El especialista habitual en la técnica conoce cómo variar esos parámetros para conseguir una rigurosidad particular de hibridación.

25 En general, la “hibridación rigurosa” se realiza a aproximadamente 25°C por debajo del punto de fusión ( $T_m$ ) térmico para el híbrido de ADN específico en un conjunto particular de condiciones. El “lavado riguroso” se realiza a temperaturas aproximadamente 5°C por debajo de 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 emparejada. Véase Sambrook y col., *anteriormente*, página 9.51. Para los propósitos de este documento, las “condiciones de alta rigurosidad” se definen para 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 al 1% a 65°C durante 8-12 horas, seguido de dos lavados en SSC 0,2X, SDS al 0,1% a 65°C durante 20 minutos. Se entenderá por el especialista que la hibridación a 65°C tendrá lugar a diferentes velocidades dependiendo de varios factores que incluyen la longitud y el porcentaje de identidad de las secuencias que están hibridando.

35 El término “mutado” cuando se aplica a secuencias de ácido nucleico significa que los nucleótidos en una secuencia de ácido nucleico se pueden insertar, delecionar o cambiar en comparación con una secuencia de ácido nucleico de referencia. Se puede realizar una única alteración en un locus (una mutación puntual) o se pueden insertar, delecionar o cambiar múltiples nucleótidos en un único locus. Además, se pueden realizar una o más alteraciones en cualquier número de loci dentro de una secuencia de ácido nucleico. Una secuencia de ácido nucleico se puede mutar mediante  
40 cualquier procedimiento conocido en la técnica incluyendo, pero sin limitación, técnicas de mutagénesis tales como “PCR con tendencia a error” (un procedimiento para realizar PCR en condiciones en las que la fidelidad de copia de la ADN polimerasa es baja, de tal forma que se obtiene una alta tasa de mutaciones puntuales a lo largo de toda la longitud del producto de PCR. Véase, por ejemplo, Leung y col. (1989) *Technique 1: 11-15* y Caldwell y Joyce (1992) *PCR Methods Applic. 2: 28-33*); y “mutagénesis dirigida a 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 y col. (1988) *Science 241: 53-57*).

45 El término “vector” como se usa en este documento tiene por objeto referirse a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al que se ha unido. Un tipo de vector es un “plásmido”, que se refiere a un bucle de ADN de doble cadena circular en el que se pueden ligar segmentos de ADN adicionales. Otros vectores incluyen cósmidos, cromosomas artificiales bacterianos (BAC) y cromosomas artificiales de levadura (YAC). Otro tipo de vector es un vector viral, en el que se pueden ligar segmentos de ADN adicionales en el genoma viral (analizado con más detalle más adelante). Ciertos vectores son capaces de replicación autónoma en una célula huésped en la que se introducen (por ejemplo, vectores que tienen un origen de replicación que funciona en la célula huésped). Otros  
55 vectores se pueden integrar en el genoma de una célula huésped después de la introducción en la célula huésped y, por lo tanto, se pueden replicar junto con el genoma del huésped. Además, ciertos vectores preferidos son capaces de dirigir la expresión de genes a los que están ligados operativamente. Tales vectores se denominan en este documento “vectores de expresión recombinantes” (o simplemente “vectores de expresión”).

60 Las secuencias de control de la expresión “unidas operativamente” se refieren a una unión en la que la secuencia de control de la expresión es contigua 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.

65 La expresión “secuencia de control de la expresión” como se usa en este documento se refiere a secuencias polinucleotídicas que son necesarias para afectar a la expresión de secuencias codificantes a las que están unidas operativamente. Las secuencias de control de la expresión son secuencias que controlan la transcripción, acontecimientos post-transcripcionales y traducción de secuencias de ácido nucleico. Las secuencias de control de la expresión incluyen secuencias apropiadas 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 citoplasmático, secuencias que mejoran la eficacia de la traducción (por ejemplo, sitios de unión a ribosoma); secuencias que mejoran la estabilidad de la proteína; y, cuando se desea, secuencias que mejoran la secreción de proteína. La naturaleza de tales secuencias de control difiere dependiendo del organismo huésped; en procariotas, tales secuencias de control incluyen generalmente un promotor, un sitio de unión a ribosoma y una secuencia de terminación de la transcripción. La expresión “secuencias de control” tiene por objeto incluir, como mínimo, todos los componentes cuya presencia sea esencial para la expresión y también puede incluir componentes adicionales cuya presencia sea ventajosa, por ejemplo, secuencias líder y secuencias de compañero de fusión.

La expresión “célula huésped recombinante” (o simplemente “célula huésped”), como se usa en este documento, tiene por objeto referirse a una célula en la que se ha introducido un ácido nucleico tal como un vector recombinante. Se debe entender que tales términos tienen por objeto referirse no solamente a la célula sujeto particular sino también a la progenie de tal célula. Debido a que pueden tener lugar ciertas modificaciones en generaciones sucesivas debido a mutación o influencias ambientales, tal progenie puede no ser, de hecho, idéntica a la célula parental, pero se incluye todavía en el alcance de la expresión “célula huésped” como se usa en este documento. Una célula huésped recombinante puede ser una célula aislada o una línea celular desarrollada en cultivo o puede ser una célula que reside en un tejido u organismos vivo.

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

El término “polipéptido” como se usa en este documento incluye proteínas tanto de origen natural como las que no son de origen 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, cada uno de los cuales tiene una o más actividades distintas.

El término “proteína aislada” o “polipéptido aislado” es una proteína o polipéptido que, debido a su origen o fuente de derivación (1) no está asociado con componentes asociados de forma natural que lo acompañan en su estado nativo, (2) cuando existe en una pureza no encontrada en la naturaleza, donde la pureza se puede estimar 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 tiene lugar en la naturaleza (por ejemplo, es un fragmento de un polipéptido encontrado en la naturaleza o incluye análogos o derivados de aminoácidos no encontrados en la naturaleza o enlaces diferentes de los enlaces peptídicos convencionales). Por tanto, un polipéptido que se sintetiza químicamente o se sintetiza en un sistema celular diferente de la célula de la que se origina de forma natural estará “aislado” de sus componentes asociados de forma natural. Un polipéptido o una proteína también se puede convertir en sustancialmente libre de componentes asociados de forma natural por aislamiento, usando técnicas de purificación de proteína bien conocidas en la técnica. Como se define, por tanto, “aislado” no requiere necesariamente que la proteína, polipéptido, péptido u oligopéptido descrito de este modo se haya retirado físicamente de su entorno nativo.

La expresión “fragmento polipeptídico” como se usa en este documento se refiere a un polipéptido que tiene una delección 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 tienen típicamente al menos 5, 6, 7, 8, 9 ó 10 aminoácidos de longitud, preferiblemente al menos 12, 14, 16 ó 18 aminoácidos de longitud, más preferiblemente al menos 20 aminoácidos de longitud, más preferiblemente al menos 25, 30, 35, 40 ó 45 aminoácidos, aún más preferiblemente al menos 50 ó 60 aminoácidos de longitud y aún más preferiblemente 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, marcado, por ejemplo, con radionúclidos, y diversas modificaciones enzimáticas, como se entenderá de forma sencilla por los especialistas en la técnica. Una diversidad de procedimientos para marcar polipéptidos y de sustituyentes o marcadores útiles para tales fines se conocen bien en la técnica e incluyen isótopos radioactivos tales como <sup>125</sup>I, <sup>32</sup>P, <sup>35</sup>S y <sup>3</sup>H, ligandos que se une a antiligandos marcados (por ejemplo, anticuerpos), fluoróforos, agentes quimioluminiscentes, enzimas y antiligandos que pueden servir como miembros de pares de unión específicos para un ligando marcado. La elección del marcador depende de la sensibilidad requerida, la facilidad de conjugación con el cebador, requisitos de estabilidad e instrumentación disponible. Los procedimientos para marcar polipéptidos se conocen bien en la técnica. Véase Ausubel y col., *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, y Suplementos hasta 2002).

Un “mutante polipeptídico” o “muteína” se refiere a un polipéptido cuya secuencia contiene una inserción, duplicación, delección, 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 único aminoácido en una posición se ha cambiado a otro aminoácido, una o más inserciones y/o delecciones, en las que uno o más aminoácidos se insertan o delecionan, respectivamente, en la secuencia de la

## ES 2 335 345 T3

proteína de origen natural y/o truncamientos de la secuencia de aminoácidos en cualquiera o en ambos de los extremos amino o carboxi. Una muteína puede tener la misma actividad biológica, pero preferiblemente, tiene una diferente, en comparación con la proteína de origen natural.

5 Una muteína tiene una homología de secuencia global de al menos el 70% con su equivalente de tipo silvestre. Se prefieren incluso más las muteínas que tienen una homología de secuencia global del 80%, 85% o del 90% con la proteína de tipo silvestre. En una realización incluso más preferida, una muteína muestra una identidad de secuencia del 95%, aún más preferiblemente el 97%, aún más preferiblemente el 98% y aún más preferiblemente el 99% de identidad de secuencia global. La homología de secuencia se puede medir mediante cualquier algoritmo de análisis de  
10 secuencia común, tal como Gap o Bestfit.

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

Como se usa en este 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  $\alpha$ -,  $\alpha$ -disustituídos, aminoácidos de *N*-alquilo u otros aminoácidos no convencionales también pueden ser componentes adecuados para polipéptidos de la presente invención. Los ejemplos de aminoácidos no convencionales incluyen: 4-hidroxi-prolina,  $\gamma$ -carboxi-glutamato,  $\epsilon$ -*N-N-N*-trimetilisina,  $\epsilon$ -*N*-acetilisina, *O*-fosfo-serina, *N*-acetilserina, *N*-formilmetionina, 3-metilhistidina, 5-hidroxisilisina, *N*-metilarginina y otros aminoácidos similares e iminoácidos (por ejemplo, 4-hidroxi-prolina). En la notación de polipéptidos usada en este  
20 documento, la dirección de la izquierda es la dirección amino terminal y la dirección de 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. Alternativamente, una proteína tiene homología con una segunda proteína si las dos proteínas tienen secuencias de aminoácidos “similares”. (Por tanto, la expresión “proteínas homólogas” se define para querer decir que las dos proteínas tienen secuencias de aminoácidos similares). Una proteína homóloga preferida es una que muestra el 60% de homología de secuencia con la proteína de tipo silvestre, se prefiere más una homología de secuencia del 70%. Se prefieren aún más proteínas homólogas que muestran una homología de secuencia del 80%, 85% o del 90% con la  
30 proteína de tipo silvestre. En una realización aún más preferida, una proteína homóloga muestra una identidad de secuencia del 95%, 97%, 98% o del 99%. Como se usa en este documento, la homología entre dos regiones de secuencia de aminoácidos (especialmente con respecto a similitudes estructurales predichas) se interpreta que implica similitud en función.

Cuando “homólogo” se usa con 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 aminoácidos conservativa” es una en la que un resto aminoacídico se sustituye por otro resto aminoacídico que tiene una cadena lateral (grupo R) con propiedades químicas similares (por ejemplo, carga o hidrofobicidad). En general, una sustitución de aminoácidos conservativa no cambiará sustancialmente las propiedades funcionales de una proteína. En los casos  
40 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 se puede ajustar al alza para corregir la naturaleza conservativa de la sustitución. Los medios para realizar este ajuste se conocen bien por los especialistas en la técnica (véase, por ejemplo, Pearson (1990) *Methods Enzymol.* 183: 63-98).

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 típicamente usando software de análisis de secuencia. Véase, por ejemplo, el Programa de Software de Análisis de Secuencia del Genetics Computer Group (GCG), Universidad de Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. El software de análisis de proteína hace coincidir secuencias similares usando la 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 se pueden usar con parámetros por defecto para determinar la homología de secuencia o la identidad de secuencia entre polipéptidos estrechamente 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 y col. (1990) *J. Mol. Biol.* 215: 403-410; Gish y States (1993) *Nature Genet.* 3: 266-272; Madden y col. (1996) *Meth.*

## ES 2 335 345 T3

Enzymol. 266: 131-141; Altschul y col. (1997) Nucleic Acids Res. 25: 3389-3402; Zhang y Madden (1997) Genome Res. 7: 649-656), especialmente blastp o tblastn (Altschul y col., 1997). Los parámetros preferidos para BLASTp son: Valor de expectativa: 10 (defecto); Filtro: seg (defecto); Coste de abrir un hueco: 11 (defecto); Coste de extender un hueco: 1 (defecto); Alineamientos máx.: 100 (defecto); Tamaño de palabra: 11 (defecto); N° de descripciones: 100 (defecto); Matriz de Penalización: BLOWSUM62.

La longitud de secuencias polipeptídicas comparadas para homología generalmente será de al menos aproximadamente 16 restos aminoacídicos, habitualmente al menos aproximadamente 20 restos, más habitualmente al menos aproximadamente 24 restos, típicamente al menos aproximadamente 28 restos y preferiblemente más de aproximadamente 35 restos. Cuando se explora una base de datos que contiene secuencias de un gran número de diferentes organismos, es preferible comparar secuencias de aminoácidos. La búsqueda en bases de datos usando secuencias de aminoácidos se puede medir por algoritmos diferentes de blastp conocidos en la técnica. Por ejemplo, las secuencias polipeptídicas se pueden comparar usando FASTA, un programa en GCG Versión 6.1. FASTA proporciona alineamientos y porcentaje de identidad de secuencia de las regiones del mejor solapamiento entre las secuencias problema y de búsqueda (véase Pearson (1990) Methods Enzymol. 183: 63-98). Por ejemplo, el porcentaje de identidad de secuencia entre secuencias de aminoácidos se puede determinar 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.

El término “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 debido a que se pueden construir 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 uno polipéptido de interés, más preferiblemente al menos 20 ó 30 aminoácidos, incluso más preferiblemente al menos 40, 50 ó 60 aminoácidos, aún más preferiblemente al menos 75, 100 ó 125 aminoácidos. Las proteínas de fusión se pueden producir recombinantemente construyendo una secuencia de ácido nucleico que codifica el polipéptido o un fragmento del mismo en fase con una secuencia de ácido nucleico que codifica una proteína o péptido diferente y después expresando la proteína de fusión. Alternativamente, una proteína de fusión se puede producir químicamente reticulando el polipéptido o un fragmento del mismo con otra proteína.

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

El término “dominio” como se usa en este documento se refiere a una estructura de una biomolécula que contribuye a una función conocida o sospechada de la biomolécula. Los dominios pueden tener la misma extensión que regiones o porciones de los mismos; los dominios también pueden incluir regiones distintas, no contiguas, de una biomolécula. Los ejemplos de dominio de proteína incluyen, pero sin limitación, un dominio de Ig, un dominio extracelular, un dominio transmembrana y un dominio citoplasmático.

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

A menos que se defina de otro modo, todos los términos técnicos y científicos usados en este documento tienen el mismo significado como se entiende de forma común por el especialista habitual en la técnica a la que se refiere esta invención. Los procedimientos y materiales ilustrativos se describen más adelante, aunque también se pueden usar procedimientos y materiales similares o equivalentes a los descritos en este documento en la práctica de la presente invención y serán evidentes para los especialistas en la técnica. En caso de conflicto, dominará la presente memoria descriptiva, incluyendo las definiciones. Los materiales, procedimientos y ejemplos son solamente ilustrativos y no tienen por objeto ser limitantes.

A lo largo de esta memoria descriptiva y las reivindicaciones, la palabra “comprender” o variaciones tales como “comprende” o “que comprende” se entenderá que 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.

### 55 *Procedimientos Para Producir Glucoproteínas de tipo Humano en Células Huésped Eucariotas Inferiores*

La invención proporciona procedimientos para producir una glucoproteína recombinante que comprende un glucano bisecado y que tiene glucosilación de tipo humano en una célula huésped de levadura o unicelular o multicelular modificada para producir una glucoproteína que tiene una estructura central de Man<sub>3</sub>GlcNAc<sub>2</sub> o de Man<sub>5</sub>GlcNAc<sub>2</sub> y que comprende una actividad de *N*-acetilglucosaminiltransferasa III, donde dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnT IIIΔ32 de ratón o GnT IIIΔ86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*. Como se describe con más detalle más adelante, una célula huésped eucariota que no expresa de forma natural, o que está modificada para no expresar, una o más enzimas implicadas en la producción de estructuras de alto contenido de manosa, se selecciona como una célula huésped de partida. Tal célula huésped seleccionada se modifica para expresar una o más enzimas u otros factores requeridos para producir glucoproteínas de tipo humano. Una cepa huésped deseada se puede modificar en una enzima o más de una enzima al mismo tiempo. Además, una molécula de ácido nucleico que codifica una o más enzimas o actividades se puede usar para modificar una cepa huésped de la invención. Preferiblemente, se crea una biblioteca

de moléculas de ácido nucleico que codifican enzimas potencialmente útiles (por ejemplo, enzimas quiméricas que comprenden un fragmento enzimático catalíticamente activo ligado en fase con una secuencia de dirección subcelular heteróloga) (por ejemplo, por ligación de sub-bibliotecas que comprenden fragmentos enzimáticos y secuencias de dirección subcelulares) y una cepa que tiene una o más enzimas con actividades óptimas o que produce las glucoproteínas más de “tipo humano” se puede seleccionar transformando células huésped diana con uno o más miembros de la biblioteca.

En particular, los procedimientos descritos en este documento permiten obtener, *in vivo*, estructuras de  $\text{Man}_5\text{GlcNAc}_2$  con alto rendimiento, al menos de forma transitoria, con el fin de modificar las mismas adicionalmente para producir *N*-glucanos complejos. Un esquema exitoso para obtener estructuras de  $\text{Man}_5\text{GlcNAc}_2$  adecuadas con rendimientos apropiados en una célula huésped, tal como un organismo eucariota inferior, implica generalmente dos estrategias paralelas: (1) reducir estructuras de alto contenido de manosa preparadas por actividades de manosiltransferasa endógenas, si las hay y (2) eliminar 1,2- $\alpha$ -manosa por manosidasas para producir altos niveles de estructuras de  $\text{Man}_5\text{GlcNAc}_2$  adecuadas que se pueden hacer reaccionar adicionalmente dentro de la célula huésped para formar glucoforamas complejas, de tipo humano.

Por consiguiente, una primera etapa implica la selección o creación de una célula huésped eucariota, por ejemplo, un eucariota inferior, capaz de producir una estructura precursora específica de  $\text{Man}_5\text{GlcNAc}_2$  que es capaz de aceptar GlcNAc *in vivo* por la acción de una GlcNAc transferasa I (“GnTI”). En una realización, el procedimiento implica preparar o usar una célula huésped eucariota no humana empobrecida en una actividad de 1,6 manosiltransferasa con respecto al *N*-glucano en una glucoproteína. Preferiblemente, la célula huésped está empobrecida en una actividad de 1,6 manosiltransferasa de inicio (véase más adelante). Tal célula huésped carecerá de una o más enzimas implicadas en la producción de estructuras de alto contenido de manosa que son indeseables para producir glucoproteínas de tipo humano.

Después se introducen una o más actividades enzimáticas en tal célula huésped para producir *N*-glucanos dentro de la célula huésped caracterizada por que tener al menos el 30% en mol de las estructuras de hidrato de carbono  $\text{Man}_5\text{GlcNAc}_2$  (“ $\text{Man}_5$ ”). Las estructuras de  $\text{Man}_5\text{GlcNAc}_2$  son necesarias para la formación de *N*-glucano complejo: se tiene que formar  $\text{Man}_5\text{GlcNAc}_2$  *in vivo* con un alto rendimiento (por ejemplo, en más del 30%), al menos de forma transitoria, ya que las reacciones posteriores de glucosilación de tipo mamífero y humano requieren  $\text{Man}_5\text{GlcNAc}_2$  o un derivado del mismo.

Esta etapa también requiere la formación de una estructura isomérica particular de  $\text{Man}_5\text{GlcNAc}_2$  dentro de la célula con un alto rendimiento. Aunque las estructuras de  $\text{Man}_5\text{GlcNAc}_2$  son necesarias para la formación de *N*-glucano complejo, su presencia no es de ninguna manera suficiente. Esto es debido a que  $\text{Man}_5\text{GlcNAc}_2$  puede aparecer en diferentes formas isoméricas, que pueden servir o no como un sustrato para GlcNAc transferasa I. Ya que la mayoría de las reacciones de glucosilación no están completas, una proteína glucosilada particular contiene generalmente una variedad de diferentes estructuras de hidratos de carbono (es decir, glucoforamas) sobre su superficie. Por tanto, la mera presencia de cantidades traza (es decir, menos del 5%) de una estructura particular,  $\text{Man}_5\text{GlcNAc}_2$  es de poca relevancia práctica para producir glucoproteínas de tipo mamífero o humanas. Es la formación de un intermedio de  $\text{Man}_5\text{GlcNAc}_2$  que acepta GlcNAc transferasa I (Figura 1B) con un alto rendimiento (es decir, por encima del 30%) lo que se requiere. 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).

Por consiguiente, algo o todo del  $\text{Man}_5\text{GlcNAc}_2$  producido por la célula huésped seleccionada tiene que ser un sustrato productivo para actividades enzimáticas junto con una ruta de glucosilación de mamífero, por ejemplo, puede servir como un sustrato para una actividad de GlcNAc transferasa I *in vivo*, formando de este modo el intermedio de *N*-glucano de tipo humano  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  en la célula huésped. Preferiblemente, al menos el 10%, más preferiblemente al menos el 30% y mucho más preferiblemente el 50% o más del intermedio de  $\text{Man}_5\text{GlcNAc}_2$  producido en una célula huésped es un sustrato productivo para GnTI *in vivo*. Se entiende, por ejemplo, que si  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  se produce en el 10% y  $\text{Man}_5\text{GlcNAc}_2$  se produce en el 25% en una proteína diana, la cantidad total del  $\text{Man}_5\text{GlcNAc}_2$  producido de forma transitoria es del 35% debido a que  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  es un producto de  $\text{Man}_5\text{GlcNAc}_2$ .

El especialista habitual en la técnica puede seleccionar células huésped de la naturaleza, por ejemplo, hongos existentes u otros eucariotas inferiores que producen niveles significativos de  $\text{Man}_5\text{GlcNAc}_2$  *in vivo*. Sin embargo, hasta ahora, no se ha demostrado que ningún eucariota inferior proporcione tales estructuras *in vivo* en más del 1,8% de los *N*-glucanos totales (véase, por ejemplo, Maras y col. (1997) Eur. J. Biochem. 249: 701-707). Alternativamente, tales células huésped se pueden modificar genéticamente para producir la estructura de  $\text{Man}_5\text{GlcNAc}_2$  *in vivo*. Se pueden usar procedimientos tales como los descritos en la Patente de Estados Unidos N° 5.595.900 para identificar la ausencia o presencia de glucosiltransferasas particulares, manosidasas y transportadores de nucleótido azúcar en una célula huésped diana u organismo de interés.

#### *Inactivación de Enzimas de Glucosilación de Células Huésped Indeseables*

Los procedimientos de la invención se refieren a la modificación de células huésped de levadura o de hongo unicelular o multicelular filamentoso para producir glucoproteínas recombinantes que tienen una estructura de  $\text{Man}_5\text{GlcNAc}_2$  o de  $\text{Man}_5\text{GlcNAc}_2$  y que comprenden estructuras de *N*-glucano bisecado de tipo humano que se consiguen introduciendo en las células huésped una actividad de *N*-acetilglucosaminiltransferasa III, donde dicha actividad es de una

## ES 2 335 345 T3

enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnT IIIΔ32 de ratón o GnT IIIΔ86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*. En una realización preferida, los procedimientos se refieren a preparar células huésped de levadura o de hongo filamentoso unicelular en las que los precursores de oligosacáridos están enriquecidos en Man<sub>5</sub>GlcNAc<sub>2</sub>. Preferiblemente, se usa una célula huésped de levadura o de hongo unicelular o multicelular filamentoso que no expresa una o más enzimas implicadas en la producción de estructuras de alto contenido de manosa. Tal célula huésped se puede encontrar en la naturaleza o se puede crear, por ejemplo, partiendo de o produciéndose a partir de uno de muchos de tales mutantes ya descritos en levaduras. Por tanto, dependiendo de la célula huésped seleccionada, uno o varios genes que codifican enzimas que se conoce que son características de reacciones de glucosilación no humana se tendrán que deletar. Tales genes y sus proteínas correspondientes se han caracterizado de forma considerable en varios eucariotas inferiores (por ejemplo, *S. cerevisiae*, *T. reesei*, *A. nidulans*, etc.), proporcionando de este modo una lista de glucosiltransferasas conocidas en eucariotas inferiores, sus actividades y su respectiva secuencia genética. Es probable que estos genes se seleccionen del grupo de manosiltransferasas, por ejemplo, 1,3 manosiltransferasas (por ejemplo, *MNN1* en *S. cerevisiae*) (Graham, 1991), 1,2 manosiltransferasas (por ejemplo, la familia *KTR/KRE* 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 deletado individualmente dando lugar a fenotipos viables con perfiles de glucosilación alterados. Los ejemplos se muestran en la Tabla 1.

Las células huésped preferidas de la invención, como se describen en este documento para ilustrar las etapas de manipulación requeridas, son mutantes sin hipermanosilación (*och1*) de *Pichia pastoris* o *K. lactis*. Al igual que otros eucariotas inferiores, *P. pastoris* procesa estructuras de Man<sub>9</sub>GlcNAc<sub>2</sub> en el RE con una α-1,2-manosidasa para producir Man<sub>8</sub>GlcNAc<sub>2</sub> (Figura 1A). Por la acción de varias manosiltransferasas, esta estructura se convierte después en estructuras hipermanosiladas (Man<sub>>9</sub>GlcNAc<sub>2</sub>), también conocidas como mananos (Figura 35A). Además, se ha observado que *P. pastoris* es capaz de añadir grupos fosfato no terminales, por la acción de manosilfosfato transferasas, a la estructura de hidrato de carbono. Esto difiere de las reacciones realizadas en células de mamífero, que implican la eliminación en lugar de la adición de azúcares de manosa (Figura 35A). Es de importancia particular eliminar la capacidad de la célula huésped eucariota, por ejemplo, hongo, de hipermanosilar una estructura de Man<sub>8</sub>GlcNAc<sub>2</sub> existente. Esto se puede conseguir seleccionando una célula huésped que no hipermanosile o modificando genéticamente tal célula.

Los genes que están implicados en el proceso de hipermanosilación se han identificado, por ejemplo, en *Pichia pastoris* y creando mutaciones en estos genes, se puede reducir la producción de glucoformas "indeseables". Tales genes se pueden identificar por homología con manosiltransferasas existentes o sus reguladores (por ejemplo, *OCH1*, *MNN4*, *MNN6*, *MNN1*) encontrados en otros eucariotas inferiores tales como *C. albicans*, *Pichia angusta* o *S. cerevisiae* o mutageneizando la cepa huésped y seleccionando un fenotipo de glucosilación con manosilación reducida. Basándose en homologías frente a manosiltransferasas y manosilfosfato transferasas conocidas, se pueden diseñar cebadores de PCR (cuyos ejemplos se muestran en la Tabla 2) o usar genes o fragmentos génicos que codifican tales enzimas como sondas para identificar homólogos en genotecas de ADN de la diana o un organismo relacionado. Alternativamente, se puede identificar un homólogo funcional que tiene actividad de manosiltransferasa por su capacidad de complementar fenotipos de glucosilación particulares en organismos relacionados.

TABLA 2

*Cebadores de PCR*

| Cebador A de PCR   | Cebador B de PCR  | Gen o genes diana en<br><i>P. pastoris</i> | Homólogos   |
|--|---|--|---|
| <b>ATGGCGAAGGC</b><br><b>AGATGGCAGT</b><br>(SEC ID N°: 3)  | <b>TTAGTCCTTCCAA</b><br><b>CTTCCTTC</b> (SEC<br>ID N°: 4) | 1,6-manosiltransferasa                     | <i>OCH1 S. cerevisiae</i> ,<br><i>Pichia albicans</i> |
| <b>TAYTGGMGNGTN</b><br><b>GARCYNGAYATH</b><br><b>AA</b> (SEC ID N°: 5)   | <b>GCRTCNCCCCAN</b><br><b>CKYTCRTA</b> (SEC<br>ID N°: 6)  | 1,2<br>manosiltransferasas                 | Familia de <i>KTR/KRE</i><br><i>S. cerevisiae</i>     |
| 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 los genes que codifican la actividad de 1,6-manosiltransferasa en *P. pastoris*, por ejemplo, se realizarían las siguientes etapas: los mutantes de *OCHI* de *S. cerevisiae* son sensibles a temperatura y crecen lentamente a temperaturas elevadas. Por tanto, se pueden identificar homólogos funcionales de *OCHI* en *P. pastoris* complementando un mutante de *OCHI* de *S. cerevisiae* con una genoteca de ADN o ADNc de *P. pastoris*. Están disponibles mutantes de *S. cerevisiae*, por ejemplo, en la Universidad de Stanford y están disponibles en el mercado en ResGen, Invitrogen Corp. (Carlsbad, CA). Los mutantes que presentan un fenotipo de crecimiento normal a temperatura elevada, después de haberse transformado con una genoteca de ADN de *P. pastoris*, probablemente llevarán un homólogo de *OCHI* de *P. pastoris*. Tal genoteca se puede crear 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 un 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 y Hieter (1989) Genetics 122: 19-27) y pFL44S, un plásmido de alto número de copias (2  $\mu$ ) basado en un pUC19 modificado que contiene el marcador URA3 (Bonneaud y col. (1991) Yeast 7: 609-615). Tales vectores se usan de forma común por investigadores académicos y están disponibles vectores similares en varios proveedores diferentes (por ejemplo, Invitrogen (Carlsbad, CA); Pharmacia (Piscataway, NJ); New England Biolabs (Beverly, MA)). Los ejemplos adicionales incluyen pYES/GS, plásmido de expresión de levadura basado en origen de replicación 2  $\mu$  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 por transformación la genoteca de ADN 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 *OCHI* en *P. pastoris* usando mutagénesis *in vivo* y/o técnicas de recombinación bien conocidas por los especialistas en la técnica.

Alternativamente, si toda la secuencia genómica de una célula huésped particular, por ejemplo, un hongo, de interés se conoce, se pueden identificar tales genes simplemente buscando en bases de datos de ADN disponibles públicamente, que están disponibles en varias fuentes, tales como NCBI, Swissprot. Por ejemplo, buscando una secuencia genómica dada o base de datos con secuencias de un gen de 1,6 manosiltransferasa conocido (por ejemplo, *OCHI* de *S. cerevisiae*), se pueden identificar genes de alta homología en tal genoma de célula huésped que pueden (pero no necesariamente) codificar proteínas que tienen actividad de 1,6-manosiltransferasa. Sin embargo, únicamente la homología de secuencia de ácido nucleico no es suficiente para probar que se ha identificado y aislado un homólogo que codifica una enzima que tiene la misma actividad. Por ejemplo, hasta la fecha no existen datos para mostrar que una delección de *OCHI* en *P. pastoris* elimina la actividad de 1,6-manosiltransferasa de inicio crucial (Martinet y col. (1998) Biotech. Letters 20(12): 1171-1177; Contreras y col. documento WO 02/00856 A2). Por tanto, ningún dato prueba que el homólogo del gen *OCHI* de *P. pastoris* realmente codifica esa función. Esa demostración se proporciona por primera vez en este documento.

Se han identificado homólogos para varias manosiltransferasas de *S. cerevisiae* en *P. pastoris* usando estas estrategias. Los genes homólogos tienen con frecuencia funciones similares a genes implicados en la manosilación de proteínas en *S. cerevisiae* y, por tanto, su delección se puede usar para manipular el patrón de glucosilación en *P. pastoris* o, por analogía, en cualquier otra célula huésped, por ejemplo, células de hongo, vegetales, de insecto o animales, con rutas de glucosilación similares.

La creación de knock-outs génicos, una vez que se ha determinado una secuencia de gen diana dada, es una técnica bien establecida en la técnica que se puede realizar por el especialista habitual en la técnica (véase, por ejemplo, Rothstein (1991) Methods in Enzymology 194: 281). La elección de un organismo huésped se puede ver influida por la disponibilidad de buenas técnicas de transformación y alteración génica.

Si se tienen que anular varias manosiltransferasas, el procedimiento desarrollado por Alani y Kleckner (1987) Genetics 116: 541-545, por ejemplo, permite el uso repetido de un marcador de selección, por ejemplo, el marcador URA3 en levadura, para eliminar secuencialmente toda actividad de manosiltransferasa endógena indeseable. Esta técnica se ha refinado por otros pero implica básicamente el uso de dos secuencias de ADN repetidas, que flanquean un marcador de selección contrario. Por ejemplo: *URA3* se puede usar como un marcador para garantizar la sección de un transformante que tiene integrada una construcción. Flanqueando el marcador *URA3* con repeticiones directas, se pueden seleccionar en primer lugar transformantes que tengan integrada la construcción y, por tanto, que tienen alterado el gen diana. Después del aislamiento de los transformantes, y su caracterización, se puede contraseleccionar en una segunda ronda los que sean resistentes a ácido 5-fluoroótico (5-FOA). Las colonias que son capaces de sobrevivir en placas que contienen 5-FOA han perdido de nuevo el marcador *URA3* por un acontecimiento de entrecruzamiento que implica las repeticiones que se han mencionado anteriormente. Por tanto, esta estrategia permite el uso repetido del mismo marcador y facilita la alteración de múltiples genes sin requerir marcadores adicionales. También se pueden usar técnicas similares para la eliminación secuencial de genes adaptados para el uso en otras células huésped eucariotas con otros marcadores de selección y de contraselección.

La eliminación de manosiltransferasas específicas, tales como 1,6 manosiltransferasa (*OCHI*) o manosilfosfato transferasas (*MNN6*, o genes que complementan mutantes *Ibd*) o reguladores (*MNN4*) en *P. pastoris* permite crear cepas modificadas de este organismo que sintetiza principalmente Man<sub>8</sub>GlcNAc<sub>2</sub> y que se puede usar para modificar

adicionalmente el patrón de glucosilación para parecerse a estructuras de glucoforma más complejas, por ejemplo, las producidas en células de mamífero, por ejemplo, humanas. Una realización preferida de este procedimiento utiliza secuencias de ADN 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.

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

#### Modificación o Selección de Huéspedes que Tienen Actividad de $\alpha$ -1,6 Manosiltransferasa de Inicio Disminuida

En una realización preferida, el procedimiento de la invención implica preparar o usar una célula huésped de levadura u hongo unicelular o multicelular filamentos que esté disminuida o empobrecida en la actividad de una  $\alpha$ -1,6-manosiltransferasa de inicio, es decir, una enzima específica de inicio que inicia la manosilación de cadena externa en el brazo  $\alpha$ -1,3 de la estructura central de  $\text{Man}_3\text{GlcNAc}_2$ . En *S. cerevisiae*, esta enzima está codificada por el gen *OCH1*. La alteración del gen *OCH1* en *S. cerevisiae* da como resultado un fenotipo en el que los azúcares ligados a N carecen completamente de la cadena externa de poli-manosa. Las estrategias anteriores para obtener glucosilación de tipo mamífero en cepas fúngicas han requerido la inactivación de *OCH1* (véase, por ejemplo, Chiba y col. (1998) J. Biol. Chem. 273: 26298-304). La alteración de la actividad de  $\alpha$ -1,6-manosiltransferasa de inicio en una célula huésped de la invención puede ser opcional, sin embargo (dependiendo de la célula huésped seleccionada), ya que la enzima Och1p requiere un  $\text{Man}_8\text{GlcNAc}_2$  intacto para el inicio de cadena externa de manosa eficaz. Por tanto, las células huésped seleccionadas o producidas de acuerdo con esta invención que acumulan oligosacáridos que tienen siete o menos restos de manosa pueden producir *N*-glucanos hipoglucosilados que probablemente serán malos sustratos para Och1p (véase, por ejemplo, Nakayama y col. (1997) FEBS Lett. 412(3): 547-50).

El gen *OCH1* se clonó de *P. pastoris* (Ejemplo 1) y *K. lactis* (Ejemplo 9), como se ha descrito. Las secuencias de ácido nucleico y aminoácidos del gen *OCH1* de *K. lactis* se exponen en las SEC ID N°: 7 y 8. Mediante el uso de cebadores específicos de genes, se preparó una construcción de cada clon para delecionar el gen *OCH1* del genoma de *P. pastoris* y *K. lactis* (Ejemplos 1 y 9, respectivamente). De este modo se obtuvieron células huésped empobrecidas en actividad de  $\alpha$ -1,6-manosiltransferasa de inicio y modificadas para producir *N*-glucanos que tienen una estructura de hidrato de carbono  $\text{Man}_5\text{GlcNAc}_2$  (véase, por ejemplo, Figuras 5, 6 y 12; Ejemplos 4 y 9).

Por tanto, en este documento se describe además una molécula de ácido nucleico aislado que tiene una secuencia de ácido nucleico que comprende o que está constituida por al menos cuarenta y cinco, preferiblemente al menos 50, más preferiblemente al menos 60 y mucho más preferiblemente 75 o más restos de nucleótidos del gen *OCH1* de *K. lactis* (SEC ID N°: 7) y homólogos, variantes y derivados de la misma. En este documento también se describen moléculas de ácido nucleico que hibridan en condiciones rigurosas con las moléculas de ácido nucleico que se han descrito anteriormente. De forma similar, se proporcionan polipéptidos aislados (incluyendo mutéfnas, variantes alélicas, fragmentos, derivados y análogos) codificados por las moléculas de ácido nucleico descritas en este documento. También se describen vectores, que incluyen vectores de expresión, que comprenden las anteriores moléculas de ácido nucleico como se describe adicionalmente en este documento. De forma similar, se proporcionan células huésped transformadas con las moléculas de ácido nucleico o vectores descritos en este documento.

En este documento se describen adicionalmente procedimientos para preparar o usar una célula huésped eucariota no humana disminuida o empobrecida en una actividad del gen *alg* (es decir, actividades de *alg* incluyendo actividades enzimáticas equivalentes en células huésped no fúngicas) e introducir en la célula huésped al menos una actividad de glucosidasa. En una realización específica, la actividad de glucosidasa se introduce provocando la expresión de una o más actividades de manosidasa dentro de la célula huésped, por ejemplo, por activación de una actividad de manosidasa o por expresión de una molécula de ácido nucleico de una actividad de manosidasa en la célula huésped.

En otra realización, el procedimiento implica preparar o usar una célula huésped disminuida o empobrecida en la actividad de una o más enzimas que transfieren un resto de azúcar al brazo 1,6 de precursores de oligosacáridos ligados a lípidos (Figura 13). Una célula huésped de la invención se selecciona o se modifica introduciendo una mutación en uno o más de los genes que codifican una enzima que transfiere un resto de azúcar (por ejemplo, manosila) al brazo 1,6 de un precursor de oligosacárido ligado a lípido. El resto de azúcar es más preferiblemente manosa, es preferiblemente una glucosa, GlcNAc, galactosa, ácido siálico, fucosa o un resto de fosfato de GlcNAc. En una realización preferida, la actividad de una o más enzimas que manosilan el brazo 1,6 de precursores de oligosacáridos ligados a lípidos está disminuida o empobrecida. El procedimiento puede comprender además la etapa de introducir en la célula huésped al menos una actividad de glucosidasa (véase a continuación).

## ES 2 335 345 T3

En otra realización más, la invención proporciona un procedimiento para producir una glucoproteína recombinante de tipo humano en un huésped de levadura u hongo unicelular o multicelular filamentoso, donde la glucoproteína comprende un *N*-glucano bisecado que tiene al menos dos GlcNAc unidas a una estructura central de trimanosa.

5

En cada anterior realización, el procedimiento se refiere a preparar una célula huésped de levadura u hongo filamentoso unicelular o multicelular en la que los precursores de oligosacárido unidos a lípido están enriquecidos en estructuras de  $\text{Man}_X\text{GlcNAc}_2$ , donde X es 3, 4 ó 5 (Figura 14). Estas estructuras se transfieren en el RE de la célula huésped a cadenas polipeptídicas nacientes por oligosacaril-transferasa y después se pueden procesar por tratamiento con glucosidasas (por ejemplo,  $\alpha$ -manosidasas) y glucosiltransferasas (por ejemplo, GnT1) para producir *N*-glucanos que tienen estructuras centrales de  $\text{GlcNAcMan}_X\text{GlcNAc}_2$ , donde X es 3, 4 ó 5 y es preferiblemente 3 (Figuras 14 y 15). Como se muestra en la Figura 14, los *N*-glucanos que tienen una estructura central de  $\text{GlcNAcMan}_X\text{GlcNAc}_2$  donde X es mayor de 3 se pueden convertir en  $\text{GlcNAcMan}_3\text{GlcNAc}_2$ , por ejemplo, por tratamiento con una actividad de  $\alpha$ -1,3 y/o  $\alpha$ -1,2-1,3 manosidasa, cuando sea aplicable.

15

El procesamiento adicional de  $\text{GlcNAcMan}_3\text{GlcNAc}_2$  por tratamiento con glucosiltransferasas (por ejemplo, GnTII) produce estructuras centrales de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  que después se pueden modificar, como se desee, por ejemplo, por tratamiento *ex vivo* o por expresión heteróloga en la célula huésped de un conjunto de enzimas de glucosilación, que incluyen glucosiltransferasas, transportadores de azúcar y manosidasas (véase a continuación), para convertirse en *N*-glucanos de tipo humano. Las glucoproteínas de tipo humano recombinantes preferidas que se pueden producir de acuerdo con la invención incluyen las que comprenden *N*-glucanos bisecados que tienen siete o menos o tres o menos restos de manosa; comprenden uno o más azúcares seleccionados entre el grupo constituido por galactosa, GlcNAc, ácido siálico y fucosa; y comprenden al menos una rama de oligosacárido que comprende la estructura NeuNAc-Gal-GlcNAc-Man.

25

En una realización, la célula huésped tiene actividad disminuida o empobrecida de Dol-P-Man: $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol Manosiltransferasa, que es una actividad implicada en la primera etapa de manosilación de  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol a  $\text{Man}_6\text{GlcNAc}_2$ -PP-Dol en el lado luminal del RE (por ejemplo, *ALG3* Figura 13; Figura 14). En *S. cerevisiae*, esta enzima está codificada por el gen *ALG3*. Como se ha descrito anteriormente, las células de *S. cerevisiae* que alojan una mutación de *alg3-1* parcial acumulan  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol y las células que tienen una delección en *alg3* parecen transferir estructuras de  $\text{Man}_5\text{GlcNAc}_2$  a cadenas polipeptídicas nacientes dentro del RE. Por consiguiente, en esta realización, las células huésped acumularán *N*-glucanos enriquecidos en estructuras de  $\text{Man}_5\text{GlcNAc}_2$  que después se pueden convertir en  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  por tratamiento con glucosidasas (por ejemplo, con actividades de  $\alpha$ -1,2 manosidasa,  $\alpha$ -1,3 manosidasa o  $\alpha$ -1,2-1,3 manosidasa) y actividades de glucosiltransferasa (por ejemplo, GnTI, GnTII) (Figura 14; Figura 35B).

35

Como se describe en el Ejemplo 10, se diseñaron cebadores degenerados basándose en un alineamiento de secuencias de proteína Alg3 de *S. cerevisiae*, *D. melanogaster* y seres humanos (*H. sapiens*) (Figuras 16 y 17) y se usaron para amplificar un producto de ADN genómico de *P. pastoris*. El producto de PCR resultante se usó como una sonda para identificar y aislar un clon genómico de *P. pastoris* que comprende una fase de lectura abierta (ORF) que codifica una proteína que tiene una identidad de secuencia global del 35% y una similitud de secuencia del 53% con el gen *ALG3* de *S. cerevisiae* (Figuras 18 y 19). Este gen de *P. pastoris* se denomina en este documento "*PpALG3*". El gen *ALG3* se identificó de forma similar y se aisló de *K. lactis* (Ejemplo 10; Figuras 20 y 21).

45

Por lo tanto, en este documento se describe además una molécula de ácido nucleico aislado que tiene una secuencia de ácido nucleico que comprende o que está constituida por al menos cuarenta y cinco, preferiblemente al menos 50, más preferiblemente al menos 60 y mucho más preferiblemente 75 o más restos de nucleótidos del gen *ALG3* de *P. pastoris* (Figura 18) y el gen *ALG3* de *K. lactis* (Figura 20) y homólogos, variantes y derivados de las mismas. También se describen en este documento moléculas de ácido nucleico que hibridan en condiciones rigurosas con las moléculas de ácido nucleico que se han descrito anteriormente. De forma similar, se proporcionan polipéptidos aislados (incluyendo muteínas, variantes alélicas, fragmentos, derivados y análogos) codificados por las moléculas de ácido nucleico que se describen adicionalmente en este documento (los productos génicos de *ALG3* de *P. pastoris* y *K. lactis* se muestran en las Figuras 18 y 20). Además, también se proporcionan vectores, incluyendo vectores de expresión, que comprenden una molécula de ácido nucleico también descrita en este documento, como se describe adicionalmente en este documento.

55

Usando cebadores específicos de genes, se preparó una construcción para delecionar el gen *PpALG3* en el genoma de *P. pastoris* (Ejemplo 10). Esta cepa se usó para generar una célula huésped empobrecida en actividad de Dol-P-Man: $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol Manosiltransferasa y produce precursores de  $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol ligados a lípidos que se transfieren a cadenas polipeptídicas nacientes para producir *N*-glucanos que tienen una estructura de hidrato de carbono  $\text{Man}_5\text{GlcNAc}_2$ .

60

Como se describe en el Ejemplo 11, tal célula huésped se puede modificar por expresión de manosidasas apropiadas para producir *N*-glucanos que tienen la estructura de hidrato de carbono central  $\text{Man}_3\text{GlcNAc}_2$  deseada. La expresión de GnT en la célula huésped (por ejemplo, dirigiendo una molécula de ácido nucleico o una biblioteca de moléculas de ácido nucleico como se describe más adelante) permite que la célula huésped modificada produzca *N*-glucanos que tienen una o dos estructuras de GlcNAc unidas a cada brazo de la estructura central  $\text{Man}_3$  (es decir,  $\text{GlcNAc}_1\text{Man}_3\text{GlcNAc}_2$ ,  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  o  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ ; véase la Figura 15). Estas estructuras se

65

pueden procesar adicionalmente usando los procedimientos de la invención para producir *N*-glucanos de tipo humano en proteínas que entran en la ruta de secreción de la célula huésped.

En una realización preferida, el procedimiento de la invención implica preparar o usar una célula huésped de levadura u hongo unicelular o multicelular filamentoso que esté tanto (a) disminuida o empobrecida en la actividad de un gen *alg* o en una o más actividades que manosilan *N*-glucanos en el brazo  $\alpha$ -1,6 de la estructura de hidrato de carbono central Man<sub>3</sub>GlcNAc<sub>2</sub> ("Man3"); como (b) disminuida o empobrecida en la actividad de una  $\alpha$ -1,6-manosiltransferasa de inicio, es decir, una enzima específica de inicio que inicia la manosilación de cadena externa (en el brazo  $\alpha$ -1,3 de la estructura central Man3). En *S. cerevisiae*, esta enzima está codificada por el gen *OCHI*. La alteración del gen *ochl* en *S. cerevisiae* da como resultado un fenotipo en el que los azúcares ligados a N carecen completamente de la cadena externa de poli-manosa. Las estrategias previas para obtener glucosilación de tipo mamífero en cepas fúngicas han requerido la inactivación de *OCHI* (véase, por ejemplo, Chiba y col. (1998) J. Biol. Chem. 273: 26298-304). La alteración de la actividad de  $\alpha$ -1,6-manosiltransferasa de inicio en una célula huésped de la invención es opcional, sin embargo (dependiendo de la célula huésped seleccionada), ya que la enzima *Ochlp* requiere Man<sub>3</sub>GlcNAc intacto para el inicio de cadena externa de manosa eficaz. Por tanto, las células huésped seleccionadas o producidas de acuerdo con esta invención, que acumulan oligosacáridos ligados a lípido que tienen siete o menos restos de manosa, después de la transferencia producirán *N*-glucanos hipoglucosilados que probablemente serán malos sustratos para *Ochlp* (véase, por ejemplo, Nakayama y col. (1997) FEBS Lett. 412(3): 547-50).

#### 20 Modificación o Selección de Huéspedes que Tienen Actividad de *N* Acetilglucosaminiltransferasa III

La invención proporciona un procedimiento para producir una glucoproteína de tipo humano recombinante en una célula huésped de levadura u hongo unicelular o multicelular filamentoso expresando una actividad de *N*-acetilglucosaminiltransferasa III, donde dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnT III $\Delta$ 32 de ratón o GnT III $\Delta$ 86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*. En una realización, una célula huésped (por ejemplo, *P. pastoris*) se modifica para producir más *N*-glucanos de tipo humano, por ejemplo, por activación de dicha actividad de *N*-acetilglucosaminiltransferasa III o por expresión de una molécula de ácido nucleico de dicha actividad de *N*-acetilglucosaminiltransferasa III. Usando técnicas bien conocidas en la técnica, se diseñan cebadores específicos de genes para complementar las regiones homólogas de un gen de GnTIII, preferiblemente un gen de GnTIII de mamífero (por ejemplo, GnTIII de ratón) (Figura 24), para los que ya están disponibles secuencias en la técnica (por ejemplo, Número de Acceso de Genbank L39373) y se amplifican por PCR.

En una realización, la invención proporciona un procedimiento para producir una glucoproteína de tipo humano recombinante en una célula huésped de levadura u hongo unicelular o multicelular filamentoso (por ejemplo, *P. pastoris*), donde la glucoproteína comprende un *N*-glucano bisecado que muestra una GlcNAc de bisección en una estructura central de trimanosa o trimanosilo (Man<sub>3</sub>GlcNAc<sub>2</sub>). En esta realización, GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> (que se puede producir haciendo reaccionar un núcleo de trimanosa con la actividad de *N*-acetilglucosaminiltransferasa I ("GnTI")) pero que se produce típicamente cortando GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> por una actividad de  $\alpha$ -1,3/ $\alpha$ -1,6-manosidasa, tal como Manosidasa II (Hamilton y col. (2003) Science 301: 1244-46)) se hace reaccionar con dicha actividad de *N*-acetilglucosaminiltransferasa III para producir un GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecado. Por consiguiente, la invención proporciona una actividad de GnTIII que transfiere  $\beta$ -1,4 GlcNAc a sustratos que son capaces de aceptar la GlcNAc de bisección en células huésped de levadura u hongos unicelulares o multicelulares filamentosos.

En otra realización, la invención proporciona un procedimiento para producir una glucoproteína de tipo humano recombinante en una célula huésped de levadura u hongo unicelular o multicelular filamentoso (por ejemplo, *P. pastoris*), donde la glucoproteína comprende un *N*-glucano bisecado que muestra una GlcNAc de bisección en una estructura central de trimanosa o trimanosilo (Man<sub>3</sub>GlcNAc<sub>2</sub>) que tiene al menos dos GlcNAc unidas al núcleo de trimanosa. En esta realización, Man<sub>3</sub>GlcNAc<sub>2</sub> se hace reaccionar con una actividad de GnTI y después con una actividad de *N*-acetilglucosaminiltransferasa II ("GnTII") y dicha actividad de GnTIII (en cualquier orden) para producir un GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecado (Figura 38). Se debe entender que la estructura central de trimanosilo bisecada de esta realización también puede contener un grupo manosilo adicional en lugar de un resto de GlcNAc. Por ejemplo, GlcNAcMan<sub>4</sub>GlcNAc<sub>2</sub> se puede hacer reaccionar con dicha actividad de GnTIII para producir un GlcNAc<sub>2</sub>Man<sub>4</sub>GlcNAc<sub>2</sub> bisecado.

La invención también proporciona un procedimiento para producir una glucoproteína recombinante más de tipo humano en una célula huésped de levadura u hongo filamento unicelular o multicelular (por ejemplo, *P. pastoris*), donde la glucoproteína producida comprende un *N*-glucano bisecado que tiene al menos dos GlcNAc unidas a una estructura central de pentamanosa (Man<sub>5</sub>GlcNAc<sub>2</sub>) y que muestra un *N*-glucano bisecado. Por consiguiente, en esta realización, una estructura central de pentamanosa (Man<sub>5</sub>GlcNAc<sub>2</sub>) se hace reaccionar con dicha actividad de GnTIII para producir una estructura de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> y GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> bisecada.

En una realización alternativa, una estructura central de pentamanosa producida por la mutación de genes *och1* y *alg3* se hace reaccionar con  $\alpha$ 1,2-manosidasa, GnTI, GnTII y dichas actividades de GnTIII y UDP-GlcNAc para producir un glucano GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecado (Figura 35B). En otra realización, una estructura de central de pentamanosa se hace reaccionar con GnTI y dicha actividad de GnTIII (en cualquier orden o en combinación) para producir una estructura de GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> bisecada (Figura 37).

## ES 2 335 345 T3

En una realización más preferida, al usar el procedimiento de genoteca de ADN combinatoria como se describe más adelante, una construcción **pVA53** que comprende el líder *MNN2(s)* de *S. cerevisiae* (Nº de Acceso de GenBank NP\_009571) fusionado con un dominio de GnTIII catalíticamente activo de ratón (GnTIIIΔ32) se expresa en una cepa de *P. pastoris* **YSH-1** (Ejemplo 13), produciendo de este modo *N*-glucanos que tienen una estructura de GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> bisecada (Ejemplo 20). La Figura 26 (parte inferior) presenta el espectro de MALDI-TOF de *N*-glucanos liberados de una proteína kringle 3 expresada en la cepa que se ha mencionado anteriormente, que se denomina **PBP26** (Figura 36), que muestra un pico predominante en 1666 m/z [a], que se corresponde a GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> bisecado. (Para comparación, la Figura 26 (parte superior) presenta el espectro de MALDI-TOF de *N*-glucanos liberados de una proteína kringle 3 expresada en la cepa **YSH-1** que carece de la construcción **pVA53**. El pico predominante en 1461 m/z [d] se corresponde al glucano no modificado: GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>). Por consiguiente, en una realización, un huésped de la presente invención se caracteriza por su capacidad de producir, al menos de forma transitoria, *N*-glucanos bisecados que muestran al menos el 50% en mol de un GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> o al menos el 50% en mol de una estructura de GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que tiene una GlcNAc de bisección. El porcentaje en mol de los glucanos es con referencia al porcentaje de glucanos neutros totales como se detecta por MALDI-TOF. Se entiende que si, por ejemplo, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que tiene una GlcNAc de bisección se produce en el 20% y GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> se produce en el 25% en una proteína diana, la cantidad total de GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> producido de forma transitoria que tiene una GlcNAc de bisección es del 45%, debido a que GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> es un producto de un GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que tiene una GlcNAc de bisección que se hecho reaccionar adicionalmente con GnTII.

De forma similar, en otra realización, una construcción **pVA55** que comprende el líder *MNN2(1)* de *S. cerevisiae* (Nº de Acceso de GenBank NP\_009571) fusionado con un dominio de GnTIII catalíticamente activo de ratón (GnTIIIΔ32) se expresa en una cepa de *P. pastoris* (**YSH-1**) produciendo de este modo *N*-glucanos GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> y una estructura GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> de *N*-glucano bisecado. Como se muestra en la Figura 27 (parte inferior), estas estructuras se corresponden a picos en 1463 m/z y 1667 m/z, respectivamente. (Para comparación, la Figura 27 (parte superior) presenta el espectro de MALDI-TOF de *N*-glucanos liberados de una proteína kringle 3 expresada en la cepa **YSH-1** que carece de la construcción **pVA53**. El pico predominante se corresponde a GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> no modificada en 1461 m/z [d]. Por consiguiente, en otra realización, un huésped de la presente invención se caracteriza por su capacidad de producir, al menos de forma transitoria, *N*-glucanos que muestran al menos el 20% en mol de un GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> o al menos el 20% en mol de una estructura de GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que tiene una GlcNAc de bisección.

En una realización aún más preferida, una construcción **pVA53** que comprende el líder *MNN2(s)* de *S. cerevisiae* (Nº de Acceso de GenBank NP\_009571) fusionado con un dominio de GnTIII catalíticamente activo de ratón (GnTIIIΔ32) se expresa en una cepa de *P. pastoris* **YSH-44** (Ejemplo 15) produciendo de este modo *N*-glucanos que tienen una estructura de GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecado (Ejemplo 20). La Figura 30 presenta el espectro de MALDI-TOF de *N*-glucanos liberados de una proteína kringle 3 expresada en la cepa que se ha mencionado anteriormente denominada **YSH-57**, que muestra un pico predominante en 1542 m/z [y], que se corresponde al glucano bisecado GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>. (Para comparación, la Figura 29 presenta el espectro de MALDI-TOF de *N*-glucanos liberados de una proteína kringle 3 expresada en la cepa **YSH-44** que carece de la construcción **pVA53**. El pico predominante en 1356 m/z [x] en la Figura 29 se corresponde al glucano no modificado: GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>). Por consiguiente, en una realización, un huésped de la presente invención se caracteriza por su capacidad de producir, al menos transitoriamente, *N*-glucanos que presentan al menos el 80% en mol de una estructura de GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que tiene una GlcNAc de bisección. El porcentaje en mol de los glucanos es con referencia al porcentaje de glucanos neutros totales como se detecta por MALDI-TOF.

Como alternativa, en otra realización, una construcción **pVA53** que comprende el líder *MNN2(s)* de *S. cerevisiae* (Nº de Acceso de GenBank NP\_009571) fusionado con un dominio de GnTIII catalíticamente activo de ratón (GnTIIIΔ32) se expresa en una cepa de *P. pastoris* (**PBP6-5**) (Ejemplo 11) produciendo de este modo *N*-glucanos que tienen un GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> y una estructura de GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecado. Como se muestra en la Figura 32, estas estructuras se corresponden a picos en 1340 m/z y 1543 m/z, respectivamente. Por consiguiente, en otra realización, un huésped de la presente invención se caracteriza por su capacidad de producir, al menos transitoriamente, *N*-glucanos que muestran al menos el 20% en mol de una estructura de GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que tienen una GlcNAc de bisección en una célula huésped mutante de *alg3*.

La invención proporciona procedimientos para producir una glucoproteína de tipo humano recombinante que comprende un glucano bisecado en una levadura u hongo unicelular o multicelular filamentoso, donde la glucoproteína comprende una estructura central de Man<sub>5</sub>GlcNAc<sub>2</sub> o una estructura central de Man<sub>3</sub>GlcNAc<sub>2</sub> y donde la estructura central está modificada adicionalmente por dos o más GlcNAc. En la invención, más del 80% de las estructuras centrales están modificadas de este modo. En una realización altamente preferida, una de los GlcNAc es una GlcNAc de bisección.

En otro aspecto de la invención, una genoteca de ácido nucleico combinatoria que codifica al menos un dominio catalítico de GnTIII se usa para expresar una actividad de GnTIII en una célula huésped de levadura u hongo unicelular o multicelular filamentoso (Ejemplo 18). Preferiblemente, una biblioteca descrita adicionalmente en este documento comprende una sub-biblioteca de secuencias líder fusionadas en fase con una única molécula de ácido nucleico o una sub-biblioteca de moléculas de ácido nucleico que comprenden secuencias de GnTIII, una o más de la cuales codifican un dominio catalítico que tiene actividad de GnTIII en la célula huésped. Alternativamente, una única molécula de

ácido nucleico o una sub-biblioteca de moléculas de ácido nucleico que comprenden secuencias líder se fusiona en fase con una sub-biblioteca de moléculas de ácido nucleico que comprenden secuencias de GnTIII, una o más de las cuales codifican un dominio catalítico que tiene actividad de GnTIII en la célula huésped. (Véase más adelante). La expresión de estas y otras de tales bibliotecas combinatorias se realiza en una célula huésped que expresa una glucoproteína diana cuyas estructuras de *N*-glucano se analizan para determinar si y cuánta GnTIII se expresa. Una amplia variedad de enzimas de GnTIII catalíticamente activa se puede producir en una célula huésped usando los procedimientos y las bibliotecas descritas en este documento. Es este aspecto de los procedimientos descritos en este documento lo que permite a un especialista crear y delimitar entre enzimas de GnTIII que tienen poca o ninguna actividad y las enzimas que se expresa de forma activa y que producen niveles predominantes de un intermedio de oligosacárido bisechado deseado tal como  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ ,  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$  o  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  en las células huésped.

Como se describe adicionalmente más adelante, la dirección apropiada de una enzima responsable de una etapa dada en la ruta de glucosilación a la localización subcelular apropiada y la suficiencia de la actividad de la enzima al pH particular de esa localización subcelular son factores importantes en la producción de glucoproteínas que tienen *N*-glucanos con las estructuras deseadas. El uso de bibliotecas combinatorias de proteínas de fusión para generar diversas poblaciones de quimeras enzimáticas y la exploración de estas bibliotecas en células transformadas proporciona un procedimiento potente para identificar cepas huésped con la actividad de interés en la localización apropiada. En realizaciones preferidas de la invención, la actividad de la enzima se localiza de tal forma que una glucoproteína que contiene *N*-glucano expresada en la célula es capaz de reaccionar con la actividad durante el proceso de secreción.

No todas las combinaciones de líder/dominios catalíticos producen, sin embargo, actividades enzimáticas deseadas. Se crea una amplia diversidad de combinaciones de líder/dominio catalítico, solamente unas pocas de las cuales pueden ser útiles para producir los intermedios deseados en este momento. Sin embargo, también se describen en este documento incluso las combinaciones que no muestran en este momento una actividad enzimática deseada en la célula huésped ilustrada. La Figura 28 (parte inferior) muestra una construcción **pVB51** que comprende el líder GNT(s) de *K. lactis* (Nº de Acceso GenBank AF106080) fusionado con un dominio de GnTIII catalíticamente activo de ratón (GnTIII  $\Delta 32$ ) expresado en una cepa de *P. pastoris* **YSH-1**, que no muestra de forma sencilla actividad de GnTIII. (Para comparación, la Figura 28 (parte superior) presenta el espectro de MALDI-TOF de *N*-glucanos liberados de una proteína kringle 3 expresada en la cepa **YSH-1** que carece de la construcción **pVA53**. El pico predominante se corresponde a  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  no modificado en 1461 m/z). Se observa el pico predominante en la Figura 28 (parte inferior) en 1463 m/z, que se correlaciona con la masa de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$ . También se observa un segundo pico en 1726 m/z, que no se correlaciona con la masa de  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ . Se considera que esta y otras de tales combinaciones pueden ser útiles, con o sin ligeras modificaciones usando técnicas bien conocidas en la técnica, cuando se expresan, por ejemplo, en otras células huésped incluyendo las que se han modificado para producir glucoformas de tipo humano.

El uso de bibliotecas combinatorias para generar diversas poblaciones de quimeras enzimáticas y la exploración de estas bibliotecas en células transformadas permite adicionalmente que se identifiquen cepas en las que la actividad de la enzima sea sustancialmente intracelular. El siguiente Ejemplo 6 proporciona un ejemplo de condiciones de ensayo útiles para medir la actividad extracelular de  $\alpha$ -1,2-manosidasa. Los Ejemplos 22 y 23 también proporcionan ejemplos de ensayos para actividad de glucosiltransferasa (GnTIII) en el medio. Véase también la Tabla 9, a continuación y Choi y col. (2003) Proc. Natl. Acad. Sci. U.S.A. 100(9): 5022-27. Para los propósitos de la invención, una actividad enzimática es sustancialmente intracelular cuando menos del 10% de la actividad enzimática se puede medir en el medio extracelular.

Como se describe en los Ejemplos 11, 12, 13, 14, 15 y 19-21, una célula huésped se puede modificar por la expresión de glucosiltransferasas apropiadas (por ejemplo, *N*-acetilglucosaminiltransferasa) para producir *N*-glucanos que tengan las estructuras de hidrato de carbono deseadas (por ejemplo,  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ ,  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ ). La expresión de GnT en la célula huésped (por ejemplo, dirigiendo una molécula de ácido nucleico o una biblioteca de moléculas de ácido nucleico como se describe más adelante y en Choi y col. (2003) Proc. Natl. Acad. Sci. U.S.A. 100(9): 5022-27 y el documento WO 02/00879) permite que la célula huésped modificada produzca *N*-glucanos que tengan la  $\text{GlcNAc}$  de bisección en la manosa central. Estas estructuras se pueden procesar adicionalmente usando los procedimientos de la invención para producir *N*-glucanos de tipo humano en proteínas que entran en la ruta de secreción de la célula huésped.

En una realización más preferida, la co-expresión de transportador o transportadores y transferasa o transferasas de UDP-azúcar apropiados protegerá los restos  $\alpha$ -1,6 y  $\alpha$ -1,3 terminales así como la manosa central con  $\text{GlcNAc}$ , dando como resultado el precursor del complejo de tipo mamífero (por ejemplo,  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ ) y *N*-glucosilación híbrida. Estas cadenas de oligosacáridos ligados a N unidos a péptidos sirven entonces como un precursor para modificación adicional hasta una estructura de oligosacárido de tipo mamífero. La expresión posterior de galactosiltransferasas y la modificación genética de la capacidad de transferir ácido siálico a los extremos (véase la Figura 1B) producirá una estructura de *N*-glucano de tipo mamífero (por ejemplo, de tipo humano).

#### Células Huésped de la Invención

Una célula huésped de la invención es una célula de levadura o de hongo unicelular o multicelular filamentosos, por ejemplo, levadura, un hongo unicelular y multicelular o filamentosos. Sin embargo, una amplia diversidad de células huésped se consideran útiles en los procedimientos de la invención. Las células vegetales o las células de insecto, por

ejemplo, se pueden modificar para expresar una glucoproteína de tipo humano de acuerdo con la invención. Asimismo, una diversidad de células huésped no humanas, de mamífero, se pueden alterar para expresar glucoproteínas más de tipo humano o alteradas de otro usando los procedimientos de la invención. Como entenderá un especialista en la técnica, se puede usar cualquier célula huésped eucariota, (incluyendo una célula humana) junto con una biblioteca descrita en este documento para expresar una o más proteínas quiméricas que se dirigen a una localización subcelular, por ejemplo, orgánulo, en la célula huésped donde la actividad de la proteína está modificada y, preferiblemente, está aumentada. Tal proteína es preferiblemente -pero no tiene porque ser necesariamente- una enzima implicada en la glucosilación de proteína, como se ilustra en este documento. Se considera que se puede dirigir cualquier secuencia codificante de proteína y seleccionar para actividad modificada en una célula huésped eucariota usando los procedimientos descritos en este documento.

Los eucariotas inferiores que son capaces producir glucoproteínas que tienen el *N*-glucano Man<sub>5</sub>GlcNAc<sub>2</sub> unido son particularmente útiles debido a que (a) carecen 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 incluso más de tipo humano, por ejemplo, por la acción de GlcNAc transferasa I (Figura 1B;  $\beta$ 1,2 GnTI) para formar GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>. Se obtiene un rendimiento de más del 30% en mol, más preferiblemente un rendimiento del 50, 60, 70, 80, 90 ó incluso del 100% en mol, de las glucoproteínas con *N*-glucanos que tienen una estructura de Man<sub>5</sub>GlcNAc<sub>2</sub>. En una realización preferida, se demuestra que más del 50% de la estructura de Man<sub>5</sub>GlcNAc<sub>2</sub> es un sustrato para una actividad de GnTI y puede servir como tal 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 reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* y *Neurospora crassa*.

En cada anterior realización, el procedimiento se refiere a preparar una célula huésped de levadura u hongo unicelular o multicelular filamentoso en la que los precursores de oligosacáridos están enriquecidos en Man<sub>5</sub>GlcNAc<sub>2</sub>. Estas estructuras son deseables debido a que se pueden procesar después por tratamiento *in vitro*, por ejemplo, usando el procedimiento de Maras y Contreras, Patente de Estados Unidos N° 5.834.251. Sin embargo, en una realización preferida, los precursores enriquecidos en Man<sub>5</sub>GlcNAc<sub>2</sub> se procesan por al menos una reacción de glucosilación adicional *in vivo* -con glucosidasas (por ejemplo,  $\alpha$ -manosidasas) y glucosiltransferasas (por ejemplo, GnTI)- para producir *N*-glucanos de tipo humano. Los precursores de oligosacáridos enriquecidos en Man<sub>5</sub>GlcNAc<sub>2</sub>, por ejemplo, se procesan preferiblemente a los que tienen estructuras centrales de GlcNAcMan<sub>x</sub>GlcNAc<sub>2</sub>, donde X es 3, 4 ó 5 y es preferiblemente 3. Los *N*-glucanos que tienen una estructura central de GlcNAcMan<sub>x</sub>GlcNAc<sub>2</sub> donde X es mayor de 3 se pueden convertir en GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, por ejemplo, por tratamiento con una actividad de  $\alpha$ -1,3 y/o  $\alpha$ -1,6 manosidasa, cuando sea aplicable. El procesamiento adicional de GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> por tratamiento con glucosiltransferasas (por ejemplo, GnTII) produce estructuras centrales de GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> que después se pueden modificar, cuando se desee, por ejemplo, por tratamiento *ex vivo* o por expresión heteróloga en la célula huésped de enzimas de glucosilación adicionales, que incluyen glucosiltransferasas, transportadores de azúcar y manosidasas (véase más adelante), para convertirse en *N*-glucanos de tipo humano.

Las glucoproteínas de tipo humano recombinantes preferidas que comprenden un glucano bisecado que se pueden producir 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 entre el grupo constituido por galactosa, GlcNAc, ácido siálico y fucosa.

Otra célula huésped no humana preferida de la invención es un hongo unicelular o filamentoso, que está disminuido o empobrecido en la actividad de una o más actividades del gen *alg* (incluyendo una actividad enzimática que es un homólogo o equivalente a una actividad de *alg*). Otra célula huésped preferida de la invención está disminuida o empobrecida en la actividad de una o más enzimas (diferentes de actividades de *alg*) que manosilan el brazo  $\alpha$ -1,6 de una estructura de oligosacárido ligado a lípido.

Aunque se prefieren células huésped de levadura o unicelulares o multicelulares, una amplia diversidad de células huésped que tienen las propiedades que se han mencionado anteriormente se consideran útiles en los procedimientos de la invención. Las células vegetales, por ejemplo, se pueden modificar para expresar una glucoproteína de tipo humano de acuerdo con la invención. Asimismo, una diversidad de células huésped de mamífero, no humanas, se pueden alterar para expresar glucoproteínas más de tipo humano usando los procedimientos de la invención. Se puede modificar una célula huésped apropiada o se puede usar uno de los muchos de tales mutantes que ya se han descrito en levaduras. Una célula huésped preferida de la invención, como se ilustra en este documento, es un mutante sin hipermanosilación (*OCH1*) en *Pichia pastoris*, que se ha modificado adicionalmente para delecionar el gen *alg3*.

La invención proporciona adicionalmente una célula huésped de levadura u hongo unicelular o multicelular filamentoso capaz de producir glucoproteínas recombinantes que tienen *N*-glucanos bisecados, tales como GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>, GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> y preferiblemente GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecados. En la invención, la célula huésped de levadura u hongo unicelular o multicelular filamentoso compren-

de una actividad de GnTIII donde dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III de GnTIIIΔ32 de ratón o GnTIIIΔ86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*. En una realización más preferida, las células huésped comprenden además una o más actividades seleccionadas entre: GnTI, GnTII, GnTIV y GnTV. Las células huésped preferidas expresan GnTI, GnTII y dicha actividad de GnTIII. Otras células huésped preferidas expresan adicionalmente GnTIV y/o GnTV. Aún más preferiblemente, la una o más actividades de GnT de las células huésped son sustancialmente intracelulares.

Por tanto, en realizaciones preferidas de la invención, las células huésped de la invención que comprenden la una o más actividades de GnT producen *N*-glucanos bisecados que comprenden estructuras, que incluyen, pero sin limitación, GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>4</sub>GlcNAc<sub>2</sub> o GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>, que son capaces de reaccionar con dicha actividad de enzima GnTIII para producir *N*-glucanos bisecados de forma correspondiente. Las actividades de enzima convierten de este modo las glucoproteínas que contienen estos *N*-glucanos en formas con propiedades nuevas y más deseables. Debido a que se entiende actualmente que GnTIII inhibe la actividad de GnT adicional en células de mamífero, el especialista debe entender que la reacción de glucosilación secuencial puede tener o no importancia. Sin embargo, la presente invención considera la adición de GnTI y dicha actividad de GnTIII en cualquier orden o de manera conjunta. También se debe entender que otras actividades enzimáticas dentro de la célula, tales como, por ejemplo, una o más actividades de manosidasa deseadas (por ejemplo, una 1,2 manosidasa, Manosidasa I, Manosidasa II) pueden actuar de forma concertada con las actividades de GnT para generar otras glucoproteínas de tipo humano más de interés (véase la Figura 1B).

En una realización preferida, una manosidasa II o un fragmento catalíticamente activo de la misma se introduce en la célula huésped para cortar los brazos que contienen α1,3 y α1,6 manosa de una estructura central de pentamano bisecada tal como GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>. Los glucanos resultantes (por ejemplo, GlcNAc<sub>2</sub>Man<sub>4</sub>GlcNAc<sub>2</sub> y GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecados) son sustratos preferidos para la modificación de *N*-glucano de tipo humano posterior.

En otra realización de la invención, las células huésped comprenden una estructura central de Man<sub>5</sub>GlcNAc<sub>2</sub> o una estructura central de Man<sub>3</sub>GlcNAc<sub>2</sub> modificada por dos o más GlcNAc. Se debe entender que cualquier estructura central puede incluir modificaciones adicionales además de la modificación por GlcNAc. Preferiblemente, el 10% o más de las estructuras centrales están modificadas por GlcNAc. Más preferiblemente, el 20%, 30%, 40%, 50%, 60%, 70%, 80% o incluso más de las estructuras centrales contienen la modificación de GlcNAc.

### 35 Formación de *N*-glucanos complejos

La formación de la síntesis de *N*-glucano complejo es un procedimiento secuencial por el que se eliminan restos de azúcar específicos y se unen a la estructura de oligosacárido central. En eucariotas superiores, esto se consigue exponiendo el sustrato de forma secuencial a diversas enzimas de procesamiento. Estas enzimas realizan reacciones específicas dependiendo de su localización particular dentro de toda la cascada de procesamiento. Esta "línea de ensamblaje" consiste en RE, Golgi temprano, medial y tardío y la red de Golgi *trans*, todos con su entorno de procesamiento específico. Para re-crear el procesamiento de glucoproteínas humanas en el Golgi y RE de eucariotas inferiores, se tienen que expresar numerosas enzimas (por ejemplo, glucosiltransferasas, glucosidasas, fosfatasas y transportadores) y se tienen que dirigir específicamente a estos orgánulos y, preferiblemente, en una localización de tal forma que funcionen de forma más eficaz en relación a su entorno así como en relación a otras enzimas en la ruta.

Debido a que un objetivo de los procedimientos descritos en este documento es conseguir una cepa de producción de proteína robusta que sea capaz de funcionar 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, uno o más genes que codifican enzimas que se conoce que son características de reacciones de glucosilación no humanas preferiblemente se delecionan. La cepa celular modificada se transforma con una variedad de diferentes genes que codifican actividades deseadas y estos genes se transforman de un modo estable, garantizando de este modo que se mantenga la actividad deseada a lo largo del proceso de fermentación.

Cualquier combinación de las siguientes actividades enzimáticas se puede modificar de forma única o de forma múltiple en el huésped usando los procedimientos de la invención: sialiltransferasas, manosidasas, fucosiltransferasas, galactosiltransferasas, GlcNAc transferasas, transportadores específicos del RE y Golgi (por ejemplo, transportadores de cotransporte unidireccional y bidireccional 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 CMP-ácido-*N*-acetilneuramínico. Preferiblemente, se introducen actividades enzimáticas en una o más moléculas de ácido nucleico (véase también más adelante). Las moléculas de ácido nucleico se pueden introducir de forma única o de forma múltiple, por ejemplo, en el contexto de una biblioteca de ácido nucleico tal como una biblioteca combinatoria descrita en este documento. Sin embargo, se tiene que entender que se pueden introducir actividades enzimáticas únicas o múltiples en una célula huésped de cualquier modo, 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 descrita en este documento.

*Expresión de Glucosiltransferasas para Producir N-glicanos Complejos*

Con la información de secuencia de ADN, el especialista puede clonar moléculas de ADN que codifican actividades de GnT (por ejemplo, Ejemplo 3, 8, 11, 15 y 18). Usando técnicas convencionales bien conocidas por los especialistas en la técnica, se pueden insertar moléculas de ácido nucleico que codifican GnTI, II, III, IV o V (o que codifican fragmentos catalíticamente activos de los mismos) en vectores de expresión apropiados bajo el 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, una célula huésped fúngica, tal como *Pichia sp.*, *Kluyveromyces sp.* y *Aspergillus sp.*, como se describe en este documento, de tal forma que una o más de estas enzimas de GnT de mamífero se pueden expresar de forma activa en una célula huésped de elección para la producción de una glucoproteína compleja de tipo humano (por ejemplo, Ejemplos 8, 20 y 21).

Varias glucosiltransferasas individuales se han clonado y expresado en *S. cerevisiae* (GalT, GnTI), *Aspergillus nidulans* (GnTI) y otros hongos, sin embargo, sin demostrar el resultado deseado de "humanización" en el patrón de glucosilación de los organismos (Yoshida y col. (1999) *Glycobiology* 9(1): 53-8; Kalsner y col. (1995) *Glycoconj. J.* 12(3): 360-370). Se especuló que la estructura de hidrato de carbono requerida para aceptar azúcares por la acción de tales glucosiltransferasas no estaba presente en cantidades suficientes, lo que contribuyó muy probablemente a la ausencia de la formación de *N*-glucano complejo.

Un procedimiento preferido de la invención proporciona la expresión funcional de una actividad de GnT, tal como GnTI, GnTII y GnTIII de esta invención en el aparato de Golgi temprano, medial o tardío, y garantiza un suministro suficiente de UDP-GlcNAc (por ejemplo, por expresión de un transportador de UDP-GlcNAc; véase los siguientes Ejemplos).

25 *Procedimientos para Proporcionar Precursores de Nucleótido Azúcar al Aparato de Golgi*

Para que una glucosiltransferasa funcione satisfactoriamente en el Golgi, la enzima requiere una concentración suficiente de un nucleótido azúcar apropiado, que es el donador de alta energía del resto de azúcar añadido a una glucoproteína naciente. En seres humanos, toda la variedad de precursores de nucleótido azúcar (por ejemplo, UDP-*N*-acetilglucosamina, UDP-*N*-acetilgalactosamina, CMP-ácido-*N*-acetilneuramínico, UDP-galactosa, etc.) se sintetiza generalmente en el citosol y se transporta al Golgi, donde se une al oligosacárido central por glucosiltransferasas.

Para repetir este procedimiento en células huésped no humanas, tales como eucariotas inferiores, se tienen que expresar transportadores específicos de nucleósido azúcar en el Golgi para garantizar niveles adecuados de precursores de nucleósido azúcar (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). Se pueden proporcionar nucleótido azúcar a los compartimentos apropiados, por ejemplo, expresando en el microorganismo huésped un gen exógeno que codifica un transportador de nucleótido azúcar. La elección de una enzima transportadora está influida por la naturaleza de la glucosiltransferasa exógena que se está usando. 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 CMP-ácido siálico.

La proteína transportadora añadida transporta un nucleótido azúcar desde el citosol al aparato de Golgi, donde el nucleótido azúcar se puede hacer reaccionar por la glucosiltransferasa, por ejemplo, para alargar un *N*-glucano. La reacción libera un nucleósido difosfato o monofosfato, por ejemplo, UDP, GDP o CMP. Los nucleósidos monofosfato se pueden exportar directamente desde el Golgi en intercambio por azúcares de nucleósido trifosfato por un mecanismo de cotransporte bidireccional. La acumulación de un nucleósido difosfato, sin embargo, inhibe la actividad adicional de una glucosiltransferasa. Ya que esta reacción parece ser importante para la glucosilación eficaz, frecuentemente es deseable proporcionar una copia expresada de un gen que codifica una nucleótido difosfatasa. La difosfatasa (específica para UDP o GDP cuando sea apropiado) hidroliza el difosfonucleósido para producir nucleósido monofosfato y un fosfato inorgánico.

Las enzimas transportadoras adecuadas, que son típicamente de origen de mamífero, se describen más adelante. Tales enzimas se pueden introducir en una célula huésped seleccionada usando los procedimientos de la invención.

En otro ejemplo, la  $\alpha$  2,3- o  $\alpha$  2,6-sialiltransferasa protege restos de galactosa con ácido siálico en el Golgi *trans* y TGN de seres humanos conduciendo a una forma madura de la glucoproteína (Figura 1B). Para volver a introducir esta etapa de procesamiento en una levadura o un hongo modificado metabólicamente se requerirá (1) actividad de  $\alpha$  2,3- o  $\alpha$  2,6-sialiltransferasa y (2) un suministro suficiente de CMP-ácido-*N*-acetilneuramínico, en el Golgi tardío de levadura. Para obtener una actividad de  $\alpha$  2,3-sialiltransferasa suficiente en el Golgi tardío, por ejemplo, el dominio catalítico de una sialiltransferasa conocida (por ejemplo, de seres humanos) se tiene que dirigir al Golgi tardío en hongos (véase anteriormente). Asimismo, se tienen que insertar transportadores para permitir el transporte de CMP-ácido-*N*-acetilneuramínico al Golgi tardío. Actualmente no existe indicación de que los hongos sinteticen o incluso que puedan transportar cantidades suficientes de CMP-ácido-*N*-acetilneuramínico al Golgi. Por consiguiente, para garantizar el suministro adecuado de sustrato para las glucosiltransferasas correspondientes, se tiene que insertar metabólicamente la producción de CMP-ácido siálico en el hongo.

*UDP-N-acetilglucosamina*

El ADNc de transportador de UDP-*N*-acetilglucosamina humano, que se reconoció por una búsqueda de homología en la base de datos de etiquetas de secuencia expresada (dbEST), se ha clonado (Ishida (1999) J. Biochem. 126(1): 68-77). El transportador de membrana de Golgi de mamífero para UDP-*N*-acetilglucosamina se clonó por corrección fenotípica con ADNc de células de riñón caninas (MDCK) de un mutante de *Kluyveromyces lactis* caracterizado recientemente deficiente en transporte de Golgi del nucleótido azúcar anterior (Guillen y col. (1998) Proc. Natl. Acad. Sci. USA 95(14): 7888-7892). Los resultados demuestran que el gen de transportador de UDP-GlcNAc de Golgi de mamífero tiene toda la información necesaria para que la proteína se exprese y se dirija funcionalmente al aparato de Golgi de levadura y que dos proteínas con secuencias de aminoácidos muy diferentes pueden transportar el mismo soluto dentro de la misma membrana de Golgi (Guillen y col. (1998) Proc. Natl. Acad. Sci. USA 95(14): 7888-7892).

Por consiguiente, se puede incorporar la expresión de un transportador de UDP-GlcNAc en una célula huésped mediante una construcción de ácido nucleico que puede contener, por ejemplo: (1) una región por la que la construcción transformada se mantiene en la célula (por ejemplo, origen de replicación o una región que media en la integración cromosómica), (2) un gen marcador que permite la selección de células que se han transformado, incluyendo marcadores de contraselección y reciclables tales como *ura3* o *T-urf13* (Soderholm y col. (2001) Biotechniques 31(2): 306-10) u otros marcadores de selección bien caracterizados (por ejemplo, *his4*, *bla*, *Sh ble* etc.), (3) un gen o fragmento del mismo que codifica un transportador de UDP-GlcNAc funcional (por ejemplo, de *K. lactis*, (Abeijon, (1996) Proc. Natl. Acad. Sci. U.S.A. 93: 5963-5968) o de *H. sapiens* (Ishida y col. (1996) J. Biochem. (Tokio) 120(6): 1074-8) y (4) un promotor que activa la expresión de la biblioteca de construcción de fusión de dominio de localización/catalítico que se ha mencionado anteriormente.

*GDP-Fucosa*

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

*UDP-Galactosa*

Dos genes heterólogos, *gmal2(+)* que codifica alfa 1,2-galactosiltransferasa (alfa 1,2 GalT) de *Schizosaccharomyces pombe* y (hUGT2) que codifica el transportador de UDP-galactosa (UDP-Gal) humano, se han expresado funcionalmente en *S. cerevisiae* para examinar las condiciones intracelulares requeridas para galactosilación. La correlación entre la galactosilación de la proteína y la actividad de transporte de UDP-galactosa indicó que un suministró exógeno de transportador de UDP-Gal, en lugar de alfa 1,2 GalT, desempeñó un papel clave para la galactosilación eficaz en *S. cerevisiae* (Kainuma (1999) Glycobiology 9(2): 133-141). Asimismo, se clonó un transportador de UDP-galactosa de *S. pombe* (Segawa (1999) FEBS Letters 451(3): 295-298).

*CMP-Ácido-N-acetilneuramínico (CMP-Ácido Siálico)*

El transportador de CMP-ácido siálico humano (hCST) se ha clonado y expresado en células CHO Lec 8 (Aoki y col. (1999) J. Biochem. (Tokio) 126(5): 940-50; Eckhardt y col. (1997) Eur. J. Biochem. 248(1): 187-92). La expresión funcional del transportador de CMP-ácido siálico murino se consiguió en *Saccharomyces cerevisiae* (Berninsone y col. (1997) J. Biol. Chem. 272(19): 12616-9). El ácido siálico se ha encontrado algunos hongos, sin embargo, no está claro si el sistema huésped elegido será capaz de suministrar niveles suficientes de CMP-ácido siálico. El ácido siálico se puede suministrar en el medio o, alternativamente, también se pueden integrar rutas fúngicas implicadas en la síntesis de ácido siálico en el genoma huésped.

*Expresión de Difosfatasa*

Cuando se transfieren azúcares a una glucoproteína, se libera un nucleósido difosfato o monofosfato de los precursores de nucleótido azúcar. Aunque los monofosfatos se pueden exportar directamente en sustitución por azúcares de trifosfato de nucleósido por un mecanismo de cotransporte bidireccional, los difosfonucleósidos (por ejemplo, GDP) se tienen que escindir por fosfatasa (por ejemplo, GDPasa) para producir nucleósidos monofosfato y fosfato inorgánico antes de ser exportados. Esta reacción parece ser importante para la glucosilación eficaz, ya que se ha observado que la GDPasa de *S. cerevisiae* es necesaria para la manosilación. Sin embargo, la enzima tiene solamente el 10% de la actividad frente a UDP (Berninsone y col. (1994) J. Biol. Chem. 269(1): 207-211). Los eucariotas inferiores con frecuencia no tienen actividad de difosfatasa específica de UDP en el Golgi ya que no utilizan precursores de UDP-azúcar para la síntesis de glucoproteína en el Golgi. Se observó que *Schizosaccharomyces pombe*, una levadura que añade restos de galactosa a polisacáridos de pared celular (de UDP-galactosa), tiene actividad UDPasa específica, sugiriendo adicionalmente el requisito de tal enzima (Berninsone y col. (1994) J. Biol. Chem. 269(1): 207-211). Se conoce que el UDP es un inhibidor potente de glucosiltransferasas y la eliminación de este producto secundario de la glucosilación es importante para evitar la inhibición de glucosiltransferasa en el lumen del Golgi (Khatara y col. (1974) Eur. J. Biochem. 44: 537-560).

*Procedimientos para Alterar N-Glucanos en un Huésped Expresando una Actividad Enzimática Dirigida de una Molécula de Ácido Nucleico*

La presente invención proporciona un procedimiento para producir una glucoproteína de tipo humano recombinante que comprende un glucano bisecado en una célula huésped de levadura u hongo unicelular o multicelular filamentoso 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 hidrato de carbono  $\text{Man}_5\text{GlcNAc}_2$  e introducir en la célula un ácido nucleico que codifica una actividad de *N*-acetilglucosaminiltransferasa III donde dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnTIII $\Delta$ 32 de ratón o GnTIII $\Delta$ 86 de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*. En una realización preferida, una molécula de ácido nucleico que codifica una o más actividades de manosidasa implicadas en la producción de  $\text{Man}_5\text{GlcNAc}_2$  a partir de  $\text{Man}_8\text{GlcNAc}_2$  o  $\text{Man}_9\text{GlcNAc}_2$  se introduce en el huésped. Adicionalmente se proporcionan en este documento 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 codifica una o más enzimas o actividades de glucosilación. Las actividades de enzima preferidas se seleccionan entre el grupo constituido por 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 difosfatasa. En una realización particularmente preferida, el huésped se selecciona o se modifica para expresar dos o más actividades enzimáticas en las que el producto de una actividad aumenta los niveles de sustrato de otra actividad, por ejemplo, una glucosiltransferasa y un transportador de azúcar correspondiente, por ejemplo, actividades de GnTI y de transportador de UDP-GlcNAc. En otra realización preferida, el huésped se selecciona o se modifica para expresar una actividad para eliminar productos que pueden inhibir las 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 normalmente no está ligado a o asociado 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 se obtiene preferiblemente de un miembro del grupo constituido por: proteínas unidas a membrana del RE o Golgi, señales de recuperación, proteínas de membrana de Tipo II, proteínas de membrana de Tipo I, transportadores de nucleótido azúcar transmembrana, manosidasas, sialiltransferasas, glucosidasas, manosiltransferasas y fosfomanosiltransferasas.

El componente de dominio catalítico de la construcción o proteína de fusión se obtiene preferiblemente de una glucosidasa, manosidasa o actividad de glucosiltransferasa obtenida de un miembro del grupo constituido por GnTI, GnTII, GnTIII, GnTIV, GnTV, GnTVI, GaIT, Fucosiltransferasa y Sialiltransferasa. El dominio catalítico tiene preferiblemente un óptimo de pH dentro de 1,4 unidades de pH del óptimo de pH promedio de otras enzimas representativas en el orgánulo en el que se localiza la enzima o tiene una actividad óptima a un pH entre 5,1 y 8,0. En una realización preferida, el dominio catalítico codifica una manosidasa seleccionada entre el grupo constituido por 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, manosidasa lix y manosidasa III de *H. sapiens*.

50

*Selección de una Enzima de Glucosilación: Óptimos de pH y Localización Subcelular*

En una realización de la invención, una glucoproteína de tipo humano recombinante que comprende un glucano bisecado se prepara de forma eficaz en una célula huésped de levadura u hongo unicelular o multicelular filamentoso introduciendo en un compartimento subcelular de la célula una enzima de glucosilación seleccionada para tener un óptimo de pH similar a los óptimos de pH de otras enzimas en el compartimento subcelular dirigido. Por ejemplo, la mayoría de la enzimas que son activas en el RE y aparato de Golgi de *S. cerevisiae* tienen de óptimos de pH que están entre aproximadamente 6,5 y 7,5 (véase la Tabla 3). Debido a que la glucosilación de proteínas es un proceso altamente evolucionado y eficaz, el pH interno del RE y Golgi probablemente también está en el intervalo de aproximadamente 6-8. Todas las estrategias previas para reducir la manosilación por la acción de manosidasas recombinantes en células fúngicas, sin embargo, han introducido enzimas que tienen un óptimo de pH de aproximadamente pH 5,0 (Martinet y col. (1998) Biotech. Letters 20(12): 1171-1177 y Chiba y col. (1998) J. Biol. Chem. 273(41): 26298-26304). 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, a saber, el RE y Golgi temprano, para la producción *in vivo* eficaz de  $\text{Man}_5\text{GlcNAc}_2$  en *N*-glucanos.

65

## ES 2 335 345 T3

Por consiguiente, una realización preferida de esta invención dirige una enzima de glucosilación seleccionada (o dominio catalítico de la misma), por ejemplo, una  $\alpha$ -manosidasa, a una localización subcelular en la célula huésped (por ejemplo, un orgánulo) donde el óptimo de pH de la enzima o dominio está dentro de 1,4 unidades de pH del óptimo de pH promedio de otras enzimas marcadoras representativas localizadas en el mismo o en los mismos orgánulos. El óptimo de pH de la enzima a dirigir a un orgánulo específico se debe hacer coincidir con el óptimo de pH de otras enzimas encontradas en el mismo orgánulo para maximizar la actividad por unidad de enzima obtenida. La Tabla 3 resume la actividad de manosidasas de diversas fuentes y sus respectivos óptimos de pH. La Tabla 4 resume sus localizaciones subcelulares típicas.

TABLA 3

| <b>Manosidasas y su óptimo de pH.</b>     |  |                 |  |  |
|---|--|-----------------|--|--|
| Fuente                                    | Enzima   | Óptimo de<br>pH | Referencia   |  |
| <i>Aspergillus saitoi</i>                 | $\alpha$ -1,2-manosidasa                             | 5,0             | Ichishima y col. (1999) <i>Biochem. J.</i> 339 (Pt 3): 589-597   |  |
| <i>Trichoderma reesei</i>                 | $\alpha$ -1,2-manosidasa                             | 5,0             | Maras y col. (2000) <i>J. Biotechnol.</i> 77(2-3): 255-263       |  |
| <i>Penicillium citrinum</i>               | $\alpha$ -D-1,2-manosidasa                           | 5,0             | Yoshida y col. (1993) <i>Biochem. J.</i> 290 (Pt 2): 349-354     |  |
| <i>C. elegans</i>                         | $\alpha$ -1,2-manosidasa                             | 5,5             | véase la <b>Figura 11</b>  |  |
| <i>Aspergillus nidulans</i>               | $\alpha$ -1,2-manosidasa                             | 6,0             | Eades y Hintz (2000) <i>Gene</i> 255 (1): 25-34                  |  |
| <i>Homo sapiens</i> IA (Golgi)            | $\alpha$ -1,2-manosidasa                             | 6,0             |  |  |
| <i>Homo sapiens</i> IB (Golgi)            | $\alpha$ -1,2-manosidasa                             | 6,0             |  |  |
| Células de insecto de <i>Lepidopteran</i> | $\alpha$ -1,2-Man <sub>6</sub> -manosidasa de Tipo I |                 | Ren y col. (1995) <i>Biochem.</i> 34(8): 2489-2495               |  |
| <i>Homo sapiens</i>                       | $\alpha$ -D-manosidasa                               | 6,0             | Chandrasekaran y col. (1984) <i>Cancer Res.</i> 44(9): 4059-68   |  |
| <i>Xanthomonas manihotis</i>              | $\alpha$ -1,2,3-manosidasa                           | 6,0             | Patente de Estados Unidos N° 6.300.113                           |  |
| Ratón IB ( <i>Golgi</i> )                 | $\alpha$ -1,2-manosidasa                             | 6,5             | Schneikert y Herscovics (1994) <i>Glycobiology.</i> 4(4): 445-50 |  |
| <i>Bacillus sp.</i> (secretada)           | $\alpha$ -D-1,2-manosidasa                           | 7,0             | Maruyama y col. (1994) <i>Carbohydrate Res.</i> 251: 89-98       |  |

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

En una realización más preferida, la enzima de glucosilación dirigida es una manosidasa, glucosiltransferasa o una glucosidasa. En una realización especialmente preferida, la actividad de manosidasa se dirige al RE o Golgi *cis*, donde tienen lugar las reacciones tempranas de glucosilación. Aunque este procedimiento es útil para producir una

## ES 2 335 345 T3

glucoproteína de tipo humano en una célula huésped no humana, se entenderá que el procedimiento también es útil de forma más general para modificar perfiles de hidratos de carbono de una glucoproteína en cualquier célula huésped eucariota, incluyendo células huésped humanas.

5 Las secuencias de dirección que median en la retención de proteínas en ciertos orgánulos de la ruta secretora de la célula huésped se conocen bien y se describen en la bibliografía científica y bases de datos públicas, como se analiza con más detalle más adelante con respecto a bibliotecas para la selección de secuencias de dirección y enzimas dirigidas. Tales secuencias de dirección subcelulares se pueden usar 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 actividad mejorada u óptima basándose en óptimos de pH o la presencia de otros factores estimuladores.

15 Cuando se intenta cortar estructuras de alto contenido de manosa para producir  $\text{Man}_5\text{GlcNAc}_2$  en el RE o el aparato de Golgi de una célula huésped tal como *S. cerevisiae*, por ejemplo, se puede elegir cualquier enzima o combinación de enzimas que (1) tenga un óptimo de pH suficientemente próximo (es decir, entre pH 5,2 y pH 7,8), y (2) sabe que genera, sola o de forma concertada, la estructura de  $\text{Man}_5\text{GlcNAc}_2$  isomérica específica requerida para aceptar la adición posterior de GlcNAc por GnTI. Cualquier enzima o combinación de enzimas que se demuestra que genera una estructura que se puede convertir en  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  por GnTI *in vitro* constituiría una elección apropiada. Este conocimiento se puede obtener de la bibliografía científica o experimentalmente. Por ejemplo, se puede determinar si una manosidasa potencial puede convertir  $\text{Man}_8\text{GlcNAc}_2\text{-2AB}$  (2-aminobenzamida) en  $\text{Man}_5\text{GlcNAc}_2\text{-AB}$  y después verificar que la estructura de  $\text{Man}_5\text{GlcNAc}_2\text{-2AB}$  obtenida puede servir como un sustrato para GnTI y UDP-GlcNAc para dar  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  *in vitro*. La manosidasa IA de una fuente humana o murina, por ejemplo, sería una elección apropiada (véase, por ejemplo, el Ejemplo 4). Los ejemplos descritos en este documento utilizan oligomanosa ligada a N marcada con 2-aminobenzamida seguido de análisis de HPLC para realizar esta determinación.

TABLA 4

30 **Localización celular y óptimos de pH de diversas enzimas relacionadas con glucosilación de *S. cerevisiae*.**

| Gen                     | Actividad                            | Localización | óptimo de pH | Referencia(s)  |
|-------------------------|--------------------------------------|--------------|--------------|--|
| 35<br>40<br><i>KTR1</i> | $\alpha$ -1,2-<br>manosiltransferasa | Golgi        | 7,0          | Romero y col.<br>(1997) <i>Biochem.</i><br><i>J</i> 321 (Pt.2):<br>289-295           |
| 45<br><i>MNS1</i>       | $\alpha$ -1,2-manosidasa             | RE           | 6,5          | Lipari y col.<br>(1994)<br><i>Glycobiology.</i><br><i>Oct</i> , 4(5): 697-<br>702    |
| 50<br><i>CWH41--</i>    | glucosidasa I<br>manosiltransferasa  | RE Golgi     | 6,8 7-8      | Lehele y Tanner<br>(1974) <i>Biochim.</i><br><i>Biophys. Acta</i><br>350(1): 225-235 |
| 60<br><i>KRE2</i>       | $\alpha$ -1,2-<br>manosiltransferasa | Golgi        | 6,5-9,0      | Romero y col.<br>(1997) <i>Biochem.</i><br><i>J.</i> 321 (Pt 2):<br>289-295          |

## ES 2 335 345 T3

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

Un experimento que ilustra el óptimo de pH para una enzima  $\alpha$ -1,2-manosidasa se describe en el Ejemplo 7. Una proteína de fusión quimérica BB27-2 (*MNN10* (s) de *Saccharomyces*/manosidasa IB  $\Delta$ 31 de *C. elegans*), que sale al medio, se sometió a diversos intervalos de pH para determinar la actividad óptima de la enzima. Los resultados del experimento muestran que la  $\alpha$ -1,2-manosidasa tiene un pH óptimo de aproximadamente 5,5 para su función (Figura 11).

En una realización preferida, un único gen de manosidasa clonado se expresa 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 una producción adecuada de  $\text{Man}_5\text{GlcNAc}_2$ . En los casos en los que se usan múltiples genes, las manosidasas codificadas tendrán preferiblemente todas óptimos de pH dentro del intervalo preferido de aproximadamente 5,1 a aproximadamente 8,0, o especialmente entre aproximadamente 5,5 y aproximadamente 7,5. Las actividades de manosidasa preferidas incluyen  $\alpha$ -1,2-manosidasas obtenidas 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 Glucosilación de Célula Huésped Usando una Genoteca de ADN Combinatoria

Ciertos procedimientos de la invención se realizan preferiblemente (pero no necesariamente) usando una o más bibliotecas de ácido nucleico. Una característica ilustrativa de una biblioteca de ácido nucleico combinatoria descrita en este documento es que comprende secuencias que codifican péptidos señal de dirección celular y secuencias que codifican proteínas a dirigir (por ejemplo, enzimas o dominios catalíticos de las mismas, incluyendo, pero sin limitación, las que median en la glucosilación).

Una biblioteca de ácido nucleico 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 a dirigir. Alternativamente, una biblioteca de ácido nucleico 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 a dirigir a una célula huésped. Como se describe más adelante con más detalle, una secuencia de ácido nucleico obtenida de (a) y una secuencia de ácido nucleico obtenida de (b) se ligan para producir una o más construcciones de fusión que codifican un péptido señal de dirección celular ligado funcionalmente a un dominio polipeptídico de interés. Un ejemplo de un enlace funcional es cuando el péptido señal de dirección celular se liga al dominio polipeptídico de interés en el mismo macro de lectura de traducción (“en macro”).

En una realización preferida, una genoteca de ADN combinatoria expresa una o más proteínas de fusión que comprenden péptidos señal de dirección celular ligados en macro con dominios enzimáticos catalíticos. La proteína de fusión codificada comprende preferiblemente un dominio catalítico de una enzima implicada en modificación de tipo mamífero o humano de *N*-glucanos. En una realización más preferida, el dominio catalítico se obtiene de una enzima seleccionada entre el grupo constituido por manosidasas, glucosiltransferasas y otras glucosidasas que se liga en macro con uno o más péptidos señal de dirección. El dominio enzimático puede ser exógeno y/o endógeno a la célula huésped. Un péptido señal particularmente preferido es uno que normalmente está asociado con una proteína que experimenta transporte del RE al Golgi.

La genoteca de ADN combinatoria descrita en este documento se puede usar para producir y localizar enzimas *in vivo* implicadas en modificación de *N*-glucano de tipo mamífero o humano. Las construcciones de fusión de la genoteca de ADN combinatoria se modifican de tal forma que las enzimas codificadas se localizan en el RE, Golgi o la red de Golgi *trans* de la célula huésped donde están implicadas en producir *N*-glucanos particulares en una glucoproteína de interés. La localización de enzimas que modifican *N*-glucanos de la presente invención se consigue mediante un mecanismo de anclaje o mediante interacción proteína-proteína donde el péptido de localización construido a partir de la genoteca de ADN combinatoria se localiza en un orgánulo deseado de la ruta secretora tal como el RE, Golgi o la red de Golgi *trans*.

Un ejemplo de un *N*-glucano útil, que se produce de forma eficaz y en cantidades suficientes para modificación adicional con reacciones de glucosilación (compleja) de tipo humano es  $\text{Man}_5\text{GlcNAc}_2$ . Una cantidad suficiente de  $\text{Man}_5\text{GlcNAc}_2$  se necesita en una glucoproteína de interés para procesamiento de tipo humano adicional *in vivo* (por ejemplo, más del 30% en mol). El intermedio de  $\text{Man}_5\text{GlcNAc}_2$  se puede usar como un sustrato para modificación de *N*-glucano adicional para producir  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  (Figura 1B; véase anteriormente). Por consiguiente, la genoteca de ADN combinatoria de la presente invención se puede usar para producir enzimas que producen posteriormente  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  u otros *N*-glucanos complejos deseados, en una cantidad útil.

Un aspecto adicional de las construcciones de fusión producidas usando la genoteca de ADN combinatoria de la presente invención es que permiten una actividad de corte de *N*-glucano intracelular suficiente y con frecuencia prácticamente completa en la célula huésped modificada. Las construcciones de fusión preferidas producidas por la genoteca de ADN combinatoria de la invención codifican una enzima de glucosilación, por ejemplo, una manosidasa, que se localiza de forma eficaz en un compartimento de la célula huésped intracelular y muestra de este modo muy poca y preferiblemente ninguna actividad extracelular. Se demuestra que las construcciones de fusión preferidas de la presente invención que codifican una enzima manosidasa se localizan donde los *N*-glucanos se modifican, a saber, el RE y Golgi. Las enzimas de fusión de la presente invención se dirigen a tales orgánulos particulares en la ruta secretora donde se localizan y actúan en *N*-glucanos tales como  $\text{Man}_8\text{GlcNAc}_2$  para producir  $\text{Man}_5\text{GlcNAc}_2$  en una glucoproteína de interés.

Las construcciones de fusión de GnTIII generadas a partir de una genoteca de ADN combinatoria para producir glucanos bisecados se ensayaron para determinar cualquier actividad extracelular. Un ejemplo de una construcción de fusión de GnTIII que muestra alteración *in vivo* de glucosilación de célula huésped se denomina **pVA53**. Después de transformar **YSH-1** de *P. pastoris* con la construcción de fusión **pVA53**, el sobrenadante se ensayó para detectar cualquier actividad de GnTIII *ex vivo*. La Figura 33 no muestra ningún cambio aparente en el sustrato convencional  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  en condiciones que pondrían de manifiesto actividad de GnTIII extracelular en el medio (Ejemplo 22). De forma similar, la Figura 34 no muestra actividad de GnTIII extracelular detectable en el medio en **YSH-57** de *P. pastoris* que reacciona con el sustrato  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$  (Ejemplo 23).

Las enzimas producidas por la genoteca de ADN combinatoria descrita en este documento pueden modificar *N*-glucanos en una glucoproteína de interés como se muestra para proteínas K3 o IFN- $\beta$  expresadas en *P. pastoris*, como se muestra en las Figuras 5, 6 y 25-34 (véase también Ejemplos 2, 4 y 18-23). Sin embargo, se entiende que otros tipos de glucoproteínas, sin limitación, incluyendo eritropoyetina, citocinas tales como interferón- $\alpha$ , interferón- $\beta$ , interferón- $\gamma$ , interferón- $\omega$  y granulocito-CSF, factores de coagulación tales como factor VIII, factor IX y proteína C humana, cadena  $\alpha$  de receptor de IgE soluble, IgG, fragmentos de IgG, IgM, interleucinas, urocinasa, quimasa e inhibidor de tripsina de urea, proteína de unión de IGF, factor de crecimiento epidérmico, factor liberador de hormona del crecimiento, proteína de fusión de anexina V, angiostatina, factor de crecimiento endotelial vascular 2, factor inhibidor de progenitor mielóide 1, osteoprotegerina,  $\alpha$ -1 antitripsina, ADNasa II,  $\alpha$ -feto proteínas, AAT, rhTBP-1 (oncept, también conocido como proteína de unión a TNF 1), TACI-Ig (interaccionador de ligando de ciclofilina y modulador de calcio y activador transmembrana), FSH (hormona estimuladora de folículos), GM-CSF, GLP-1 (proteína similar a glucagón 1) con y sin agonista de receptor FC de IL-1, sTNFr (enbrel, también conocido como fusión de receptor Fc TNF soluble) ATHI, rhTrombina, glucocerebrosidasa y CTLA4-Ig (Antígeno 4 asociado a Linfocitos T Citotóxicos - Ig) se pueden glucosilar de este modo.

#### *Construcción de una Genoteca de ADN combinatoria de Construcciones de Fusión*

Una genoteca de ADN combinatoria de construcciones de fusión presenta uno o más péptidos señal de dirección celular ("péptidos de dirección") obtenidos generalmente de dominios *N*-terminales de proteínas nativas (por ejemplo, realizando deleciones C-terminales). Sin embargo, algunos péptidos de dirección se obtienen del extremo C de proteínas nativas (por ejemplo, *SEC12*). Las proteínas unidas a membrana del RE o del Golgi se usan preferiblemente como una fuente para secuencias de péptidos de dirección. Estas proteínas tienen secuencias que codifican una cola citosólica (ct), un dominio transmembrana (tmd) y una región de tallo (sr) que varían en longitud. Estas regiones se pueden reconocer por alineamientos de secuencia de proteína y comparaciones con homólogos conocidos y/u otras proteínas localizadas (por ejemplo, comparando gráficos de hidrofobicidad).

Los péptidos de dirección se indican en este documento como cortos (s) medios (m) y largos (l) con respecto a las partes de una membrana de tipo II. La secuencia de péptido de dirección indicado como corto (s) se corresponde al dominio transmembrana (tmd) de la proteína unida a membrana. La secuencia de péptido de dirección indicado como largo (l) se corresponde a la longitud del dominio transmembrana (tmd) y la región de tallo (sr). La secuencia de péptido de dirección indicado como media (m) se corresponde al dominio transmembrana (tmd) y aproximadamente la mitad de la longitud de la región de tallo (sr). Las regiones de dominio catalítico se indican en este documento por el número de deleción de nucleótidos con respecto a su enzima de glucosilación de tipo silvestre.

#### *Sub-bibliotecas*

En algunos casos, una biblioteca de ácido nucleico combinatoria descrita en este documento se puede ensamblar directamente a partir de genes existentes o de tipo silvestre. Preferiblemente, la genoteca 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 novedosas que codifican dominios de proteína dirigida útiles tales como los que tienen actividades de glucosilación.

#### *Sub-bibliotecas de Dominio Catalítico 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. Los dominios catalíticos se pueden seleccionar del huésped a modificar, así como de otros organismos relacionados o no relacionados. Las enzimas de mamífero, vegetales, de insecto, de reptiles, de

## ES 2 335 345 T3

algas o fúngicas son todas útiles y se deben elegir para representar un amplio espectro de propiedades bioquímicas con respecto a temperatura y óptimos de pH. En una realización preferida, los genes se truncan para dar fragmentos, algunos de los cuales codifican los dominios catalíticos de las enzimas. Al eliminar secuencias de dirección endógenas, las enzimas después se pueden redirigir y expresar en otros loci celulares.

5 La elección de tales dominios catalíticos se puede guiar por el conocimiento del entorno particular en el que el dominio catalítico posteriormente será activo. Por ejemplo, si una enzima de glucosilación particular tiene que ser activa en el Golgi tardío, y todas las enzimas conocidas del organismo huésped en el Golgi tardío tienen un cierto óptimo de pH, o se conoce que el Golgi tardío tiene un pH particular, entonces se elige un dominio catalítico que muestre una actividad adecuada y preferiblemente máxima a ese pH, como se ha analizado anteriormente.

### *Sub-bibliotecas de Secuencia de Péptido de Dirección*

15 Otra sub-biblioteca útil incluye secuencias de ácido nucleico que codifican péptidos señal de dirección que dan como resultado la localización de una proteína en una localización particular dentro del RE, Golgi y red de Golgi *trans*. Estos péptidos de dirección se pueden seleccionar entre el organismo huésped a modificar así como otros organismos relacionados o no relacionados. Generalmente, tales secuencias se incluyen en tres categorías: (1) Secuencias *N*-terminales que codifican una cola citosólica (ct), un dominio transmembrana (tmd) y parte o todo de una región de tallo (sr), que de forma conjunta o individualmente ancla proteínas a la membrana interna (luminal) del Golgi; (2) 20 señales de recuperación que se encuentran generalmente en el extremo C tales como el tetrapéptido HDEL (SEC ID N°: 41) o KDEL (SEC ID N°: 42); y (3) regiones transmembrana de diversas proteínas, por ejemplo, transportadores de nucleótido azúcar, que se sabe que se localizan en el Golgi.

25 En el primer caso, cuando el péptido de dirección consiste en diversos elementos (ct, tmd y sr), la biblioteca se diseña de tal forma que la ct, el tmd y diversas partes de la región de tallo están representados. Por consiguiente, una realización preferida de la sub-biblioteca de secuencias de péptido de dirección incluye secuencias de ct, tmd y/o sr de proteínas unidas a membrana del RE o Golgi. En algunos casos, puede ser deseable proporcionar a la sub-biblioteca longitudes variables de secuencia de sr. Esto se puede conseguir 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 de tallo.

30 Otras fuentes útiles más de secuencias de péptido de dirección incluyen péptidos señal de recuperación, por ejemplo, los tetrapéptidos HDEL o KDEL, que se encuentran típicamente en el extremo C de proteínas que se transportan de forma retrógrada al RE o Golgi. Otras fuentes más de secuencias de péptido de dirección incluyen (a) proteínas de membrana de tipo II, (b) las enzimas enumeradas en la Tabla 3, (c) transportadores de nucleótido azúcar transmembrana que se localizan en el Golgi y (d) secuencias a las que se hace referencia en la Tabla 5.

TABLA 5

| <b>Fuentes de secuencias de dirección de compartimento útiles</b> |                      |                            |  |
|---|----------------------|----------------------------|--|
| <b>Gen o Secuencia</b>  | <b>Organismo</b>     | <b>Función</b>             | <b>Localización de producto génico</b> |
| <i>MNSI</i>   | <i>A. nidulans</i>   | $\alpha$ -1,2-manosidasa   | RE                                     |
| <i>MNSI</i>   | <i>A. niger</i>      | $\alpha$ -1,2-manosidasa   | RE                                     |
| <i>MNSI</i>   | <i>S. cerevisiae</i> | $\alpha$ -1,2-manosidasa   | RE                                     |
| <i>GLSI</i>   | <i>S. cerevisiae</i> | glucosidasa                | RE                                     |
| <i>GLSI</i>   | <i>A. niger</i>      | glucosidasa                | RE                                     |
| <i>GLSI</i>   | <i>A. nidulans</i>   | glucosidasa                | RE                                     |
| HDEL en extremo C   | Universal en hongos  | señal de recuperación      | RE                                     |
| <i>SEC 12</i>   | <i>S. cerevisiae</i> | proteína de vesícula COPII | RE/Golgi                               |
| <i>SEC 12</i>   | <i>A. niger</i>      | proteína de vesícula COPII | RE/Golgi                               |
| <i>OCH1</i>   | <i>S. cerevisiae</i> | 1,6-manosiltransferasa     | Golgi (cis)                            |
| <i>OCH1</i>   | <i>P. pastoris</i>   | 1,6-manosiltransferasa     | Golgi (cis)                            |

| <b>Fuentes de secuencias de dirección de compartimento útiles</b> |                      |                                    |  |
|---|----------------------|------------------------------------|--|
| <b>Gen o Secuencia</b>  | <b>Organismo</b>     | <b>Función</b>                     | <b>Localización de producto génico</b> |
| MNN9  | <i>S. cerevisiae</i> | complejo de 1,6-manosiltransferasa | Golgi                                  |
| MNN9  | <i>A. niger</i>      | indeterminado                      | Golgi                                  |
| VAN1  | <i>S. cerevisiae</i> | indeterminado                      | Golgi                                  |
| VAN1  | <i>A. niger</i>      | indeterminado                      | Golgi                                  |
| ANP1  | <i>S. cerevisiae</i> | indeterminado                      | Golgi                                  |
| HOCI  | <i>S. cerevisiae</i> | indeterminado                      | Golgi                                  |
| MNN10   | <i>S. cerevisiae</i> | indeterminado                      | Golgi                                  |
| MNN10   | <i>A. niger</i>      | indeterminado                      | Golgi                                  |
| MNN11   | <i>S. cerevisiae</i> | indeterminado                      | Golgi ( <i>cis</i> )                   |
| MNN11   | <i>A. niger</i>      | indeterminado                      | Golgi ( <i>cis</i> )                   |
| MNT1  | <i>S. cerevisiae</i> | 1,2-manosiltransferasa             | Golgi ( <i>cis</i> , medial)           |
| KTR1  | <i>P. pastoris</i>   | indeterminado                      | Golgi (medial)                         |
| KRE2  | <i>P. pastoris</i>   | indeterminado                      | Golgi (medial)                         |
| KTR3  | <i>P. pastoris</i>   | indeterminado                      | Golgi (medial)                         |
| MNN2  | <i>S. cerevisiae</i> | 1,2-manosiltransferasa             | Golgi (medial)                         |
| KTR1  | <i>S. cerevisiae</i> | indeterminado                      | Golgi (medial)                         |
| KTR2  | <i>S. cerevisiae</i> | indeterminado                      | Golgi (medial)                         |
| MNN1  | <i>S. cerevisiae</i> | 1,3-manosiltransferasa             | Golgi ( <i>trans</i> )                 |
| MNN6  | <i>S. cerevisiae</i> | fosfomanosiltransferasa            | Golgi ( <i>trans</i> )                 |
| 2,6 ST  | <i>H. sapiens</i>    | 2,6-sialiltransferasa              | Golgi <i>trans</i> , red               |
| UDP-GalT  | <i>S. pombe</i>      | transportador de UDP-Gal           | Golgi                                  |

En cualquier caso, se prefiere en gran medida que se seleccionen secuencias de péptido de dirección que sean apropiadas para la actividad o las actividades enzimáticas particulares para funcionar de forma óptima dentro de la secuencia de reacciones de glucosilación deseadas. Por ejemplo, al desarrollar un microorganismo modificado capaz de sialilación terminal de *N*-glucanos nacientes, un proceso que tiene lugar en el Golgi tardío en seres humanos, es deseable utilizar una sub-biblioteca de secuencias de péptido de dirección obtenida de proteínas de Golgi tardías. De forma similar, el corte de Man<sub>8</sub>GlcNAc<sub>2</sub> por una  $\alpha$ -1,2-manosidasa para dar Man<sub>5</sub>GlcNAc<sub>2</sub> es una etapa temprana en la formación de *N*-glucano complejo en seres humanos (Figura 1B). Por lo tanto, es deseable que esta reacción ocurra en RE o Golgi temprano de un microorganismo huésped modificado. 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 genoteca de ADN combinatoria) uniendo funcionalmente una o una serie de secuencias de péptido de dirección a una o una serie de secuencias que codifican dominios catalíticos. En una realización preferida, esto se consigue por la ligación en fase de una sub-biblioteca que comprende ADN que codifica secuencias de péptido de dirección (anteriormente) con una sub-biblioteca que comprende ADN que codifica enzimas de glucosilación o fragmentos catalíticamente activos de las mismas (véase más adelante).

La biblioteca resultante comprende genes sintéticos que codifican proteínas de fusión que contienen secuencia de péptido de dirección. En algunos casos es deseable proporcionar una secuencia de péptido de dirección en el extremo

N de una proteína de fusión o, en otros casos, en el extremo C. En algunos casos, las secuencias de péptido de dirección se pueden insertar en la fase de lectura abierta de una enzima, suponiendo que la estructura de la proteína de dominios plegados individuales no esté alterada. Cada tipo de proteína de fusión se construye (de un modo dirigido por pasos o semialeatorio) y las construcciones óptimas se pueden seleccionar después de la transformación de células huésped y caracterización de patrones de glucosilación en células transformadas usando procedimientos de la invención.

#### *Alteración de Glucosilación de Célula Huésped Usando Construcciones de Fusión de Bibliotecas Combinatorias*

La construcción de una genoteca de ADN combinatoria preferida se ilustra esquemáticamente en la Figura 2 y se describe en el Ejemplo 4. La construcción de fusión se puede unir operativamente a una multitud de vectores, tales como vectores de expresión bien conocidos en la técnica. Una amplia diversidad de tales construcciones de fusión se ensambló usando actividades representativas como se muestra en la Tabla 6. Las combinaciones de péptido de dirección/dominios catalíticos se pueden ensamblar para el uso en dirigir actividades de manosidasa, glucosiltransferasa y glucosidasa en el RE, Golgi y la red de Golgi *trans* de acuerdo con la invención. Sorprendentemente, el mismo dominio catalítico puede no tener ningún efecto hasta un efecto muy profundo sobre patrones de *N*-glucosilación, dependiendo del tipo del péptido de dirección usado (véase, por ejemplo, Tabla 7, Ejemplo 4).

#### *Construcciones de Fusión de Manosidasa*

Un ejemplo representativo de una construcción de fusión de manosidasa obtenida de una genoteca de ADN combinatoria de la invención es pFB8, que es un péptido de dirección *SEC 12*(m) de *Saccharomyces* truncado (988-1296 nucleótidos de *SEC12* de SwissProt P11655) ligado en fase con una delección de 187 aminoácidos *N*-terminales de una  $\alpha$ -manosidasa IA de ratón (Genbank AN 6678787). Por tanto, la nomenclatura usada en este documento se refiere al péptido de dirección/región de dominio catalítico de una enzima de glucosilación como *SEC 12*(m) de *Saccharomyces*/manosidasa IA  $\Delta$ 187 de ratón. La proteína de fusión codificada se localiza en el RE mediante la secuencia de péptido de dirección *SEC12* mientras que conserva su actividad de dominio catalítico de manosidasa y es capaz de producir *N*-glucanos *in vivo* que tienen una estructura de Man<sub>5</sub>GlcNAc<sub>2</sub> (Ejemplo 4; Figuras 6F y 7B).

La construcción de fusión pGC5, *MNS1*(m) de *Saccharomyces*/manosidasa IB  $\Delta$ 99 de ratón, es otro ejemplo de una construcción de fusión que tiene actividad de corte de manosidasa intracelular (Ejemplo 4; Figuras 5D y 8B). La construcción de fusión pDC18-5 (VAN1(s) de *Saccharomyces*/manosidasa IB  $\Delta$ 80 de *C. elegans*) es otro ejemplo de una construcción de fusión eficaz capaz de producir *N*-glucanos que tienen una estructura de Nlan<sub>5</sub>GlcNAc<sub>2</sub> *in vivo*. Al crear una genoteca de ADN combinatoria de estas y otras de tales construcciones de fusión de manosidasa de acuerdo con la invención, un especialista puede distinguir y seleccionar las construcciones que tengan actividad de corte intracelular óptima de las que tienen relativamente poca o ninguna actividad. Los procedimientos que usan bibliotecas de ADN combinatoria de la invención son ventajosos debido a que solamente unas pocas construcciones de fusión de manosidasa seleccionadas pueden producir un *N*-glucano particularmente deseado *in vivo*.

Además, la actividad de corte de manosidasa puede ser específica de una proteína particular de interés. Por tanto, se tiene que entender adicionalmente que no todas las construcciones de fusión de péptido de dirección/dominio catalítico de manosidasa pueden funcionar de forma igualmente buena para producir la glucosilación apropiada en una glucoproteína de interés. Por consiguiente, una proteína de interés se puede introducir en una célula huésped transfectada con una genoteca de ADN combinatoria para identificar una o más construcciones de fusión que expresan una actividad de manosidasa óptima para la proteína de interés. Un especialista en la técnica será capaz de producir y seleccionar una construcción o construcciones de fusión óptimas usando la estrategia de genoteca de ADN combinatoria descrita en este documento.

Además, es evidente que otras de tales construcciones de fusión que muestran dominios catalíticos de manosidasa activos localizados (o más generalmente, dominios de una enzima) se pueden preparar usando técnicas tales como las ilustradas en el Ejemplo 4 y descritas en este documento. Será un asunto de experimentación rutinaria para un especialista en la técnica preparar y usar la genoteca de ADN combinatoria de la presente invención para optimizar, por ejemplo, la producción de Man<sub>5</sub>GlcNAc<sub>2</sub> 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 genoteca de ADN combinatoria de glucosiltransferasa usando los procedimientos de la invención. Una genoteca de ADN combinatoria de secuencias obtenidas de actividades de glucosiltransferasa I (GnTI) se ensambló con péptidos de dirección y se exploró para producción eficaz en una célula huésped eucariota inferior de una estructura de *N*-glucano GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> en una glucoproteína marcadora. Se identificó una construcción de fusión que se demostró que produce GcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> (pPB104), MNN9 (s) de *Saccharomyces*/GnTI  $\Delta$ 38 humana (Ejemplo 8). Se ensambló una amplia diversidad de tales construcciones de fusión de GnTI (Ejemplo 8, Tabla 10). Otras combinaciones de péptido de dirección/dominios catalíticos de GnTI se pueden ensamblar de forma sencilla preparando una genoteca de ADN combinatoria. También es evidente para el especialista en la técnica que se pueden preparar otras de tales construcciones de fusión que muestran actividad de glucosiltransferasa como se demuestra en el Ejemplo 8. Será un asunto de experimentación rutinaria para el especialista en la técnica usar el procedimiento de genoteca de ADN combinatoria descrito en este documento para optimizar la producción

de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> usando una construcción de fusión seleccionada en un vector de expresión y una línea de célula huésped particular.

Como se ha indicado anteriormente para construcciones de fusión de manosidasas, no todas las construcciones de fusión de péptido de dirección/dominio catalítico de GnTI funcionarán igualmente bien para producir la glucosilación apropiada en una glucoproteína de interés como se describe en este documento. Sin embargo, un especialista en la técnica será capaz de producir y seleccionar una construcción o construcciones de fusión óptimas usando una estrategia de genoteca de ADN como se describe en este documento. El Ejemplo 8 ilustra una realización preferida de una genoteca de ADN combinatoria que comprende construcciones de fusión de péptidos de dirección y dominio catalítico de GnTI implicadas en producir glucoproteínas con una estructura predominantemente de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>.

#### *Uso de Múltiples Construcciones de Fusión para Alterar la Glucosilación de la Célula Huésped*

En otro ejemplo del uso de los procedimientos y las bibliotecas de la invención para alterar la glucosilación de células huésped, una cepa de *P. pastoris* con una delección de *OCH1* que expresa una proteína indicadora (K3) se transformó 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 8). En primer lugar, la construcción de fusión de manosidasa pFB8 (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) se introdujo por transformación en una cepa de *P. pastoris* que carecía de actividad de 1,6 manosiltransferasas de inicio (es decir, delección de *och1*, Ejemplo 1). En segundo lugar, se construyó **pPB103** que comprende un gen *MNN2-2* de *K. lactis* (Genbank AN AF106080) que codifica un transportador de UDP-GlcNAc para aumentar la producción adicional de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>. La adición del transportador de UDP-GlcNAc aumentó la producción de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> significativamente en la cepa de *P. pastoris* como se ilustra en la Figura 10B. En tercer lugar, se introdujo **pPB104** que comprende *MNN9* (s) de *Saccharomyces*/GnTI Δ38 humano en la cepa. Esta cepa de *P. pastoris* se denomina "**PBP-3**". (Véase la Figura 36).

Se entiende por el especialista en la técnica que las células huésped tales como las cepas de levadura que se han descrito anteriormente se pueden transformar de forma secuencial y/o co-transformar con uno o más vectores de expresión. También se debe entender que el orden de la transformación no es particularmente pertinente para producir la glucoproteína de interés. El especialista reconoce que las modificaciones rutinarias de los procedimientos descritos en este documento pueden proporcionar resultados mejorados en la producción de la glucoproteína de interés.

La importancia de usar una secuencia de péptido de dirección particular con una secuencia de dominio catalítico particular se entiende fácilmente a partir de los experimentos descritos en este documento. La genoteca de ADN combinatoria proporciona una herramienta para construir fusiones de enzima que están implicadas en modificar *N*-glucanos en una glucoproteína de interés, que es especialmente útil para producir glucoproteínas de tipo humano. (Sin embargo, cualquier fusión de enzima se puede seleccionar 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 Man<sub>5</sub>GlcNAc<sub>2</sub> como se muestra en las Figuras 5D y 5E. Esto otorga una masa molecular reducida al glucano escindido en comparación con la K3 de la cepa con delección de *OCH1* parental, como se detectó por espectrometría de masas MALDI-TOF en la Figura 5C.

De forma similar, se usó la misma estrategia para producir otra glucoproteína secretada: IFN-β que comprende predominantemente Man<sub>5</sub>GlcNAc<sub>2</sub>. El Man<sub>5</sub>GlcNAc<sub>2</sub> se retiró por digestión con PNGasa (Papac y col. (1998) *Glycobiology* 8: 445-454) y se sometió a MALDI-TOF como se muestra en las Figuras 6A-6F. Un único pico prominente en 1254 (m/z) confirma la producción de Man<sub>5</sub>GlcNAc<sub>2</sub> en IFN-β en las Figuras 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 humano) y **pPB103** (gen *MNN2-2* de *K. lactis*), el *N*-glucano híbrido. Se detectó GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> [b] por MALDI-TOF (Figura 10).

Después de identificar transformantes con un alto grado de corte de manosa, se realizaron experimentos adicionales para confirmar que la actividad de manosidasa (corte) tuvo lugar *in vivo* y no era predominantemente el resultado de actividad extracelular en el medio de cultivo (Ejemplo 6; Figuras 7-9).

Aunque la presente invención se ilustra usando un organismo huésped *P. pastoris*, se entiende por los especialistas en la técnica que se pueden alterar otras células huésped eucariotas, incluyendo otras especies de huéspedes de levadura y fúngicos, como se describe en este documento, para producir glucoproteínas de tipo humano. Las técnicas descritas en este documento para la identificación y alteración de genes de glucosilación de célula huésped indeseables, por ejemplo, *OCH1*, se entiende que se puede aplicar a estos y/u otros genes homólogos o relacionados funcionalmente en otras células huésped eucariotas tales como otras cepas de levadura y fúngicas. Como se describe en el Ejemplo 9, los genes *och1 mnn1* se deleccionaron de *K. lactis* para modificar una célula huésped que conduce a *N*-glucanos que se convierten completamente en Man<sub>5</sub>GlcNAc<sub>2</sub> por 1,2-manosidasa (Figura 12C).

El gen *MNN1* se clonó de *K. lactis* como se describe en el Ejemplo 9. Las secuencias de ácido nucleico y de aminoácidos deducida del gen *MNN1* de *K. lactis* se muestran en las SEC ID N°: 43 y 44, respectivamente. Mediante el uso de cebadores específicos de genes, se preparó una construcción para deleccionar el gen de *MNN1* del genoma de *K. lactis* (Ejemplo 9). Las células huésped empobrecidas en actividades de *och1* y *mnn1* producen *N*-glucanos que tienen una estructura de hidrato de carbono Man<sub>5</sub>GlcNAc<sub>2</sub> (véase, por ejemplo, la Figura 12B). Tales células huésped

## ES 2 335 345 T3

se pueden modificar usando adicionalmente, por ejemplo, procedimientos y bibliotecas de la invención para producir glucoproteínas de tipo mamífero o humanas.

Por tanto, en otra realización, en este documento se describe una molécula de ácido nucleico aislada que tiene una secuencia de ácido nucleico que comprende o que está constituida por al menos cuarenta y cinco, preferiblemente al menos 50, más preferiblemente al menos 60 y mucho más preferiblemente 75 o más restos de nucleótidos del gen MNN1 de *K. lactis* (SEC ID N°: 43) y homólogos, variantes y derivados de la misma. También se describen en este documento moléculas de ácido nucleico que hibridan en condiciones rigurosas con las moléculas de ácido nucleico que se han descrito anteriormente. De forma similar, se proporcionan polipéptidos aislados (incluyendo muteínas, variantes alélicas, fragmentos, derivados y análogos) codificados por las moléculas de ácido nucleico descritas en este documento. Además, también se describen en este documento vectores, incluyendo vectores de expresión, que comprenden una molécula de ácido nucleico descrita en este documento, como se describe adicionalmente en este documento. De forma similar, se proporcionan células huésped transformadas con las moléculas de ácido nucleico o vectores descritos en este documento.

Otro aspecto descrito en este documento, por tanto, se refiere a una cepa huésped eucariota no humana que expresa glucoproteínas que comprenden *N*-glucanos modificados que se parecen a los preparados por células humanas. El realizar los procedimientos de la invención en especies diferentes de células de levadura y fúngicas, por tanto, se considera y se describe adicionalmente en este documento. Se considera que se puede usar una biblioteca de ácido nucleico combinatoria descrita en este documento para seleccionar construcciones que modifican la ruta de glucosilación en cualquier sistema de célula huésped eucariota. Por ejemplo, las bibliotecas combinatorias descritas en este documento también se pueden usar en plantas, algas e insectos y en otras células huésped eucariotas, incluyendo células de mamífero y humanas, para localizar proteínas, incluyendo enzimas de glucosilación o dominios catalíticos de las mismas, en una localización deseada a lo largo de una ruta secretora de célula huésped. Preferiblemente, las enzimas de glucosilación o dominios catalíticos y similares se dirigen a una localización subcelular a lo largo de la ruta secretora de célula huésped donde son capaces de funcionar y, preferiblemente, donde se diseñan o seleccionan para funcionar de forma más eficaz.

Las células vegetales y de insecto también se pueden modificar para alterar la glucosilación de proteínas expresadas usando la biblioteca combinatoria y los procedimientos de la invención. Además, la glucosilación en células de mamífero, incluyendo células humanas, también se puede modificar usando la biblioteca combinatoria y los procedimientos de la invención. Puede ser posible, por ejemplo, optimizar una actividad enzimática particular o modificar de otro modo las proporciones relativas de diversos *N*-glucanos preparados en una célula huésped de mamífero usando la biblioteca combinatoria y los procedimientos de la invención.

Los ejemplos de modificaciones de glucosilación que se pueden realizar usando un procedimiento de acuerdo con esta realización de la invención son: (1) modificar una célula huésped eucariota para cortar restos de manosa de  $\text{Man}_8\text{GlcNAc}_2$  para producir un *N*-glucano  $\text{Man}_5\text{GlcNAc}_2$ ; (2) modificar una célula huésped eucariota para añadir un resto de *N*-acetilglucosamina (GlcNAc) a  $\text{Man}_5\text{GlcNAc}_2$  por acción de GlcNAc transferasa I; (3) modificar una célula huésped eucariota para expresar funcionalmente una enzima tal como una *N*-acetilglucosaminil Transferasa (GnTI, GnTII, GnTIII, GnTIV, GnTV, GnTVI), manosidasa II, fucosiltransferasa (FT), galactosil transferasa (GalT) o una sialiltransferasa (ST).

Al repetir el procedimiento, se pueden incorporar rutas de glucosilación cada vez más complejas en un huésped diana, tal como un microorganismo eucariota inferior. En una realización preferida, el organismo huésped se transforma dos o más veces con bibliotecas de ADN que incluyen secuencias que codifican actividades de glucosilación. La selección de fenotipos deseados se puede realizar después de cada ronda de transformación o, alternativamente, después de que hayan tenido lugar varias transformaciones. Las rutas de glucosilación complejas se pueden modificar de forma rápida de este modo.

### *Reacciones de Glucosilación Secuencial*

En una realización preferida, tales bibliotecas de péptido de dirección/dominio catalítico están diseñadas para incorporar información existente sobre la naturaleza secuencial de las reacciones de glucosilación en eucariotas superiores. Las reacciones que se sabe que tienen lugar de forma temprana en el curso del procesamiento de glucoproteína requieren la dirección de enzimas que catalizan tales reacciones a una parte temprana del Golgi o el RE. Por ejemplo, el corte de  $\text{Man}_8\text{GlcNAc}_2$  a  $\text{Man}_5\text{GlcNAc}_2$  por manosidasas es una etapa temprana en la formación de *N*-glucano complejo (Figuras 1B y 35A). Debido a que el procesamiento de proteína se inicia en el RE y después avanza a lo largo del Golgi temprano, medial y tardío, es deseable que esta reacción tenga lugar en el RE o Golgi temprano. Cuando se diseña una biblioteca para la localización de manosidasa I, por ejemplo, se intenta, por tanto, hacer coincidir las señales de dirección de RE y Golgi temprano con el dominio catalítico de manosidasa I.

### *Generación de Diversidad de Secuencia Adicional*

El procedimiento de esta realización es más eficaz cuando un ácido nucleico, por ejemplo, una biblioteca de ADN, introducida por transformación en el huésped contiene una gran diversidad de secuencias, aumentando de este modo la probabilidad de que al menos un transformante mostrará el fenotipo deseado. Las mutaciones de un solo aminoácido, por ejemplo, pueden alterar espectacularmente la actividad de enzimas de procesamiento de glucoproteínas (Romero

## ES 2 335 345 T3

y col. (2000) J. Biol. Chem. 275(15): 11071-4). Por consiguiente, antes de la transformación, una genoteca de ADN o un sub-biblioteca constituyente se puede someter a una o más técnicas para generar diversidad de secuencia adicional. Por ejemplo, una o más rondas de transposición de genes, PCR con tendencia a error, mutagénesis *in vitro* u otros procedimientos para generar diversidad de secuencia se pueden realizar para obtener una mayor diversidad de secuencias dentro de la combinación de construcciones de fusión.

### Secuencias de Control de la Expresión

Además de las secuencias de fase de lectura abierta que se han descrito anteriormente, generalmente es preferible proporcionar a cada construcción de biblioteca secuencias de control de la expresión, tales como promotores, terminadores de la transcripción, potenciadores, sitios de unión a ribosoma y otras secuencias funcionales como puede ser necesario para garantizar la transcripción y traducción eficaz de las proteínas de fusión después de la introducción por transformación de construcciones de fusión en el organismo huésped.

Los componentes de vector adecuados, por ejemplo, marcadores de selección, 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, se seleccionan como una función de qué célula huésped particular se elige. Los criterios de selección para componentes de vector adecuados para el uso 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 de Selección

También es preferible proporcionar a cada construcción al menos un marcador de selección, tal como un gen para otorgar 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 se pueden usar los genes *URA3*, *HIS4*, *SUC2*, *G418*, *BLA* o *SH BLE*. Se conoce una multitud de marcadores de selección y están disponibles para el uso en células huésped de levadura, hongos, plantas, insectos, de mamífero y otras eucariotas.

### Transformación

Después, la biblioteca de ácido nucleico se introduce por transformación en el organismo huésped. En levaduras, se puede usar 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 con 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 genoteca de ADN en el cromosoma del huésped. En una realización preferida, la integración tiene lugar por recombinación homóloga, usando técnicas bien conocidas en la técnica. Por ejemplo, se proporcionan elementos de genoteca de ADN con secuencias flanqueantes homólogas a secuencias del organismo huésped. De este modo, la integración tiene lugar en un sitio definido en el genoma del huésped, sin alteración de genes deseables o esenciales.

En una realización especialmente preferida, el ADN de genoteca se integra en el sitio de un gen indeseado en un cromosoma del huésped, realizando la alteración o delección del gen. Por ejemplo, la integración en los sitios de los genes *OCH1*, *MNN1* o *MNN4* permite la expresión de un ADN de biblioteca deseado mientras que se evita la expresión de enzimas implicadas en la hiperfosforilación de levadura de glucoproteínas. En otras realizaciones, se puede introducir el ADN de la biblioteca en el huésped por una molécula de ácido nucleico, plásmido, vector (por ejemplo, vector viral o retroviral), cromosoma, y se puede introducir como una molécula de ácido nucleico autónoma o por integración homóloga o aleatoria en el genoma del huésped. En cualquier caso, generalmente es deseable incluir dentro de cada construcción de ADN de biblioteca al menos un gen marcador de selección para permitir la selección sencilla de organismos huésped que se han transformado de forma estable. Son especialmente adecuados los genes marcadores reciclables tales como *ura3*, que se pueden seleccionar en positivo o negativo.

### Procedimientos de Exploración y Selección

Después de la transformación de la cepa huésped con la genoteca de ADN, se seleccionan los transformantes que presentan un fenotipo de glucosilación deseado. La selección se puede realizar en una única etapa o por una serie de etapas de enriquecimiento y/o empobrecimiento fenotípico usando cualquiera de una diversidad de ensayos o procedimientos de detección. La caracterización fenotípica se puede realizar manualmente o usando un equipamiento de exploración de alto rendimiento automatizado. De forma común, un microorganismo huésped presenta *N*-glucanos de proteína sobre la superficie celular, donde se localizan diversas glucoproteínas.

Se pueden explorar las células que tienen la mayor concentración de GlcNAc terminal sobre la superficie celular, por ejemplo, o las células que secretan la proteína con el mayor contenido de GlcNAc terminal. Tal exploración se puede basar en un procedimiento visual, como un procedimiento de tinción, la capacidad de unirse a anticuerpos de unión de GlcNAc terminal específicos o lectinas conjugadas con un marcador (tales lectinas están disponibles en E. Y. Laboratories Inc., San Mateo, CA), la capacidad reducida de lectinas específicas de unirse a restos de manosa terminales, la capacidad de incorporar un azúcar marcado radiactivamente *in vitro*, unión alterada a colorantes o superficies cargadas o se puede conseguir usando un dispositivo de Separación de Células Activadas por Fluorescencia (FACS) junto con una lectina o anticuerpo marcado con fluoróforo (Guillen y col. (1998) Proc. Natl. Acad. Sci. USA 95(14): 7888-7892).

Por consiguiente, las células intactas se pueden explorar para un fenotipo de glucosilación deseado exponiendo las células a una lectina o un anticuerpo que se une específicamente al *N*-glucano deseado. Está disponible en el mercado una amplia diversidad de lectinas específicas de oligosacáridos (por ejemplo, en EY Laboratories, San Mateo, CA). Alternativamente, están disponibles en el mercado anticuerpos para *N*-glucanos humanos o animales específicos o se pueden producir usando técnicas convencionales. Una lectina o anticuerpo apropiado se puede conjugar a una molécula indicadora, tal como un cromóforo, fluoróforo, radioisótopo o una enzima que tiene un sustrato cromogénico (Guillen y col., 1998. Proc. Natl. Acad. Sci. USA 95(14): 7888-7892).

Después, la exploración se puede realizar usando procedimientos analíticos tales como espectrofotometría, fluorimetría, separación de células activadas por fluorescencia o recuento por escintilación. En otros casos, puede ser necesario analizar glucoproteínas aisladas o *N*-glucanos de células transformadas. El aislamiento de proteína se puede realizar por técnicas conocidas en la técnica. En una realización preferida, una proteína indicadora se secreta al medio y se purifica por cromatografía de afinidad (por ejemplo, cromatografía de afinidad de Ni o afinidad de glutatión S-transferasa). En los casos en los que se prefiere un *N*-glucano aislado, se puede usar una enzima tal como endo- $\beta$ -*N*-acetilglucosaminidasa (Genzyme Co., Boston, MA, New England Biolabs, Beverly, MA) para escindir los *N*-glucanos de glucoproteínas. Después, se pueden analizar proteínas aisladas o *N*-glucanos por cromatografía líquida (por ejemplo, HPLC), espectrometría de masas u otros medios adecuados. La Patente de Estados Unidos N° 5.595.900 describe varios procedimientos por los que se pueden identificar células con estructuras de hidratos de carbono 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 agotar la población transformada de células que tengan fenotipos indeseados. Por ejemplo, cuando se usa el procedimiento para incorporar una actividad de manosidasa funcional en células, los transformantes deseados tendrán niveles menores de manosa en la glucoproteína celular. La exposición de la población transformada a un radioisótopo letal de manosa en el medio agota la población de transformantes que tienen el fenotipo indeseado, es decir, niveles altos de manosa incorporada (Huffaker y Robbins (1983) Proc Natl Acad Sci USA. 80 (24): 7466-70). Alternativamente, se puede usar una lectina o un anticuerpo citotóxico, dirigido contra un *N*-glucano indeseable, para agotar una población transformada de fenotipos indeseados (por ejemplo, Stanley y Siminovitch (1977) Somatic Cell Genet 3(4): 391-405). La Patente de Estados Unidos N° 5.595.900 describe varios procedimientos por los que se pueden identificar células con una estructura de hidrato de carbono extracelular deseada. La realización de forma repetida de esta estrategia permite la incorporación secuencial de glucanos más y más complejos en eucariotas inferiores.

Para detectar células huésped que tengan sobre su superficie un alto grado del intermedio de *N*-glucano de tipo humano GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, por ejemplo, se pueden seleccionar transformantes que permiten la transferencia más eficaz de GlcNAc por GlcNAc Transferasa de UDP-GlcNAc en un ensayo celular *in vitro*. Esta exploración se puede realizar cultivando células que alojan la biblioteca introducida por transformación bajo presión selectiva sobre 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 GnTII, la liberación de UDP se determina por HPLC o un ensayo ligado a enzimas para UDP. Alternativamente, se puede usar UDP-GlcNAc y GnTII marcado radiactivamente, lavar las células y después buscar la liberación de GlcNAc radiactiva por *N*-acetilglucosaminidasa. Todo esto se puede realizar manualmente o de manera automatizada mediante el uso de equipamiento de exploración de alto rendimiento. Los transformantes que liberan más UDP, en el primer ensayo, o más GlcNAc marcada radiactivamente en el segundo ensayo, se espera que tengan un mayor grado de GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> sobre su superficie y, por tanto, constituyen el fenotipo deseado. Se pueden adaptar ensayos similares para buscar los *N*-glucanos en proteínas secretadas del mismo modo.

Alternativamente, se puede usar cualquier otra exploración adecuada tal como un ensayo de unión de lectina que sea capaz de poner de manifiesto patrones de glucosilación alterados sobre la superficie de células transformadas. En este caso, la unión reducida de lectinas específicas para manosas terminales puede ser una herramienta de selección adecuada. La lectina de *Galantus nivalis* se une específicamente a  $\alpha$ -1,3 manosa terminal, que se espera que se reduzca si está presente suficiente actividad de manosidasa II en el Golgi. También se puede enriquecer en transformantes deseados realizando una etapa de separación cromatográfica que permita la retirada de células que contienen un alto contenido de manosa terminal. Esta etapa de separación se realizaría con una columna de lectina que se une específicamente a células con un alto contenido de manosa terminal (por ejemplo, lectina de *Galantus nivalis* unida a agarosa, Sigma, St. Louis, MO) con respecto a las que tienen un bajo contenido de manosa terminal.

Además, se pueden crear directamente tales construcciones de proteína de fusión, ya que información adicional son respecto a la localización de enzimas modificadoras de hidratos de carbono activas en diferentes huéspedes eucariotas inferiores que está disponible en la bibliografía científica. Por ejemplo, se conoce que  $\beta$ 1,4-GalTr humana se puede fusionar al dominio de membrana de MNT, una manosiltransferasa de *S. cerevisiae* y se localiza en el aparato de Golgi mientras que conserva su actividad catalítica (Schwientek y col. (1995) J. Biol. Chem. 270 (10): 5483-9). Si *S. cerevisiae* o un organismo relacionado es el huésped a modificar, se pueden incorporar directamente tales hallazgos en la estrategia global para obtener *N*-glucanos complejos de tal huésped. Varios de tales fragmentos génicos en *P. pastoris* se han identificado que están relacionados con glucosiltransferasas en *S. cerevisiae* y, por tanto, se podrían usar con ese propósito.

#### Sitios de Integración

Como un objetivo final de este esfuerzo de modificación genética es una cepa de producción de proteína robusta que sea capaz de funcionar bien en un procedimiento de fermentación industrial, la integración de múltiples genes en el cromosoma del huésped (por ejemplo, fúngico) implica preferiblemente una planificación cuidadosa. La cepa modificada probablemente se tiene que transformar con una variedad de diferentes genes y estos genes se tendrán que transformar de un modo estable para garantizar que la actividad deseada se mantenga a lo largo del procedimiento de fermentación. Como se describe en este documento, cualquier combinación de diversas actividades enzimáticas deseadas se puede introducir por ingeniería en el huésped de expresión de proteína fúngica, por ejemplo, sialiltransferasas, manosidasas, fucosiltransferasas, galactosiltransferasas, glucosiltransferasas, GlcNAc transferasas, transportadores específicos del RE y Golgi (por ejemplo, transportadores de cotransporte unidireccional y bidireccional 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, CMP-ácido-*N*-acetilneuramínico. Los ejemplos de procedimientos preferidos para modificar la glucosilación en una célula huésped eucariota inferior, tal como *Pichia pastoris*, se muestran en la Tabla 6.

(Tabla pasa a página siguiente)

ES 2 335 345 T3

TABLA 6

| Algunas realizaciones preferidas para modificar glucosilación en un microorganismo eucariota inferior |   |   |   |  |
|---|---|---|---|--|
| <i>Estructura Deseada</i>   | <i>Actividades Catalíticas Adecuadas</i>                                      | <i>Fuentes Adecuadas de Secuencias de Localización</i>  | <i>Deleciones Génicas Adecuadas</i>       | <i>Transportadores y/o Fosfatasas Adecuadas</i>                                  |
| Man <sub>5</sub> GlcNAc <sub>2</sub>  | α -1,2-manosidasa (murina, humana, <i>Bacillus sp.</i> , <i>A. nidulans</i> ) | MnsI (Extremo N, <i>S. cerevisiae</i> )<br>Och1 (Extremo N, <i>S. cerevisiae</i> , <i>P. pastoris</i> )<br>Ktr1 Mnn9<br>Mnt1 ( <i>S. cerevisiae</i> )<br>KDEL, HDEL (Extremo C)                                 | <i>OCH1</i><br><i>MNN4</i><br><i>MNN6</i> | Ninguno  |
| GlcNAcMan <sub>5</sub> GlcNAc <sub>2</sub>  | GlcNAc Transferasa I, (humana, murina, de rata, etc.)                         | OchI (Extremo N, <i>S. cerevisiae</i> , <i>P. pastoris</i> )<br>Extremo KTR1<br>Mnn1 (Extremo N, <i>S. cerevisiae</i> )<br>Mnt1 (Extremo N, <i>S. cerevisiae</i> )<br>GDPasa (Extremo N, <i>S. cerevisiae</i> ) | <i>OCH1</i><br><i>MNN4</i><br><i>MNN6</i> | Transportador de UDP-GlcNAc (humano, murino, <i>K. lactis</i> ), UDPasa (humana) |

ES 2 335 345 T3

| Algunas realizaciones preferidas para modificar glucosilación en un microorganismo eucariota inferior |  |   |   |  |
|---|--|---|---|--|
| <i>Estructura Deseada</i>   | <i>Actividades Catalíticas Adecuadas</i> | <i>Fuentes Adecuadas de Secuencias de Localización</i>  | <i>Deleciones Génicas Adecuadas</i>       | <i>Transportadores y/o Fosfatasas Adecuadas</i>                                  |
| GlcNAcMan <sub>3</sub> GlcNAc <sub>2</sub>  | Manosidasa II                            | Ktr1 Mnn1 (Extremo N, <i>S. cerevisiae</i> )<br>Mnt1 (Extremo N, <i>S. cerevisiae</i> )<br>Kre2/Mnt1 ( <i>S. cerevisiae</i> )<br>Kre2 ( <i>P. pastoris</i> )<br>Ktr1 ( <i>S. cerevisiae</i> )<br>Ktr1 ( <i>P. pastoris</i> )<br>Mnn1 ( <i>S. cerevisiae</i> ) | <i>OCH1</i><br><i>MNN4</i><br><i>MNN6</i> | Transportador de UDP-GlcNAc (humano, murino, <i>K. lactis</i> ), UDPasa (humana) |
| GlcNAc <sub>(2-4)</sub> -Man <sub>3</sub> GlcNAc <sub>2</sub>   | GlcNAc Transferasa II, III, IV, V        | Mnn1 (Extremo N, <i>S. cerevisiae</i> )<br>Mnt1 (Extremo N)   | <i>OCH1</i><br><i>MNN4</i><br><i>MNN6</i> | transportador de UDP-GlcNAc (humano, murino,                                     |
|   | (humana, murina)                         | <i>S. cerevisiae</i> )<br>Kre2/Mnt1 ( <i>S. cerevisiae</i> )<br>Kre2 ( <i>P. pastoris</i> )<br>Ktr1 ( <i>S. cerevisiae</i> )<br>Ktr1 ( <i>P. pastoris</i> )<br>Mnn1 ( <i>S. cerevisiae</i> )  |   | <i>K. lactis</i> ) UDPasa (humano)   |

5

10

15

20

25

30

35

40

45

50

55

60

65

ES 2 335 345 T3

| Algunas realizaciones preferidas para modificar glucosilación en un microorganismo eucariota inferior         |   |   |   |   |
|---|---|---|---|---|
| <i>Estructura Deseada</i>   | <i>Actividades Catalíticas Adecuadas</i>                    | <i>Fuentes Adecuadas de Secuencias de Localización</i>  | <i>Deleciones Génicas Adecuadas</i>       | <i>Transportadores y/o Fosfatasas Adecuadas</i>           |
| Gal <sub>(1-4)</sub> GlcNAc <sub>(2-4)</sub> -<br>Man <sub>3</sub> GlcNAc <sub>2</sub>                        | β-1,4-Galactosil transferasa (humana)                       | Mnn1 (Extremo N, <i>S. cerevisiae</i> )<br>Mnt1 (Extremo N, <i>S. cerevisiae</i> )<br>Kre2/Mnt1 ( <i>S. cerevisiae</i> )<br>Kre2 ( <i>P. pastoris</i> )<br>Ktrl ( <i>S. cerevisiae</i> )<br>Ktrl ( <i>P. pastoris</i> )<br>Mnn1 ( <i>S. cerevisiae</i> )      | <i>OCH1</i><br><i>MNN4</i><br><i>MNN6</i> | transportador de UDP-Galactosa (humano, <i>S. pombe</i> ) |
| NANA <sub>(1-4)</sub> -Gal <sub>(1-4)</sub> GlcNAc <sub>(2-4)</sub> -<br>Man <sub>3</sub> GlcNAc <sub>2</sub> | α-2,6-Sialiltransferasa (humana)<br>α-2,3-Sialiltransferasa | KTR1 MNN1 (Extremo N, <i>S. cerevisiae</i> )<br>MNT1 (Extremo N, <i>S. cerevisiae</i> )<br>Kre2/Mnt1 ( <i>S. cerevisiae</i> )<br>Kre2 ( <i>P. pastoris</i> )<br>Ktrl ( <i>S. cerevisiae</i> )<br>Ktrl ( <i>P. pastoris</i> )<br>MNN1 ( <i>S. cerevisiae</i> ) | <i>OCH1</i><br><i>MNN4</i><br><i>MNN6</i> | Transportador de CMP-ácido Siálico (humano)               |

Como una estrategia para modificar la formación de *N*-glucanos complejos en una célula huésped tal como un eucariota inferior implica tanto la eliminación como la adición de actividades de glucosiltransferasa particulares, un esquema comprensible intentará coordinar ambos requisitos. Los genes que codifican enzimas que son indeseables sirven como sitios de integración potenciales para genes que son deseables. Por ejemplo, la actividad de 1,6 manosiltransferasa es un distintivo de glucosilación en muchos eucariotas inferiores conocidos. El gen que codifica alfa-

1,6-manosiltransferasa (*OCHI*) se ha clonado de *S. cerevisiae* y mutaciones en el gen dan lugar a un fenotipo viable con manosilación reducida. El locus génico que codifica la actividad de alfa-1,6-manosiltransferasa, por lo tanto, es una diana principal para la integración de genes que codifican actividad de glucosiltransferasa. De un modo similar, se puede elegir una diversidad de otros sitios de integración cromosómicos que, basándose en un acontecimiento de alteración de gen en ese locus, se espera que: (1) mejoren la capacidad de las células de glucosilar de un modo más de tipo humano, (2) mejoren la capacidad de las células de secretar proteínas, (3) reduzcan la proteólisis de proteínas extrañas y (4) mejoren otras características del procedimiento que faciliten la purificación o el propio procedimiento de fermentación.

#### 10 *Glucoproteínas Diana*

Los procedimientos descritos en este 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 demostrado que manosa-6-fosfato dirige las proteínas al lisosoma, lo que puede ser esencial para la función apropiada de varias enzimas relacionadas con trastornos de almacenamiento lisosómico tales como enfermedad de Gaucher, de Hunter, de Hurler, de Scheie, de Fabry y de Tay-Sachs, por mencionar solamente unas pocas. Asimismo, la adición de uno o más restos de ácido siálico a una cadena lateral de glucano puede aumentar la vida de una glucoproteína terapéutica *in vivo* después de la administración. Por consiguiente, las células huésped (por ejemplo, de eucariota inferior o mamífero) se pueden modificar genéticamente para aumentar el alcance de ácido siálico terminal en glucoproteínas expresadas en las células. Alternativamente, se puede conjugar ácido siálico a la proteína de interés *in vitro* antes de la administración usando una transferasa de ácido siálico y un sustrato apropiado. Se pueden emplear cambios en la composición del medio de cultivo además de la expresión de actividades enzimáticas implicadas en glucosilación de tipo humano para producir glucoproteínas que se parecen más estrechamente a formas humanas (Weikert y col. (1999) *Nature Biotechnology* 17, 1116-1121; Werner y col. (1998) *Arzneimittelforschung* 48(8): 870-880; Andersen y Goochee (1994) *Cur. Opin. Biotechnol.* 5: 546-549; Yang y Butler (2000) *Biotechnol. Bioengin.* 68(4): 370-380). Se ha demostrado que las modificaciones de glucano específicas de anticuerpos monoclonales (por ejemplo, la adición de una GlcNAc de bisección) mejoran la citotoxicidad celular dependiente de anticuerpos (Umana y col. (1999) *Nat. Biotechnol.* 17(2): 176-80), lo que puede ser deseable para la producción de anticuerpos u otras proteínas terapéuticas.

Las proteínas terapéuticas se administran típicamente por inyección, por vía oral, por vía pulmonar u otros medios. Los ejemplos de glucoproteínas diana adecuadas que se pueden producir de acuerdo con la invención incluyen, sin limitación: eritropoyetina, citocinas tales como interferón- $\alpha$ , interferón- $\beta$  y interferón- $\gamma$ , interferón- $\omega$  y granulocito-CSF, factores de coagulación tales como factor VIII, factor IX y proteína humana C, cadena  $\alpha$  de receptor de IgE soluble, IgG, fragmentos de IgG, IgM, interleucinas, urocinasa, quimasa, inhibidor de tripsina de urea, proteína de unión de IGF, factor de crecimiento epidérmico, factor de liberación de hormona del crecimiento, proteína de fusión de anexina V, angiostatina, factor de crecimiento endotelial vascular 2, factor inhibidor de progenitor mieloides-1, osteoprotegerina,  $\alpha$ -1-antitripsina, ADNasa II y  $\alpha$ -feto proteínas, AAT, rhTBP-1 (oncept, también conocido como proteína de unión a TNF 1), TACI-Ig (interaccionador de ligando de ciclofilina y modulador de calcio y activador de transmembrana), FSH (hormona estimuladora de folículos), GM-CSF, GLP-1 (proteína similar a glucagón 1) con y sin agonista de receptor FC de IL-1, sTNFr, (enbrel, también conocido como fusión de Fc de receptor de TNF soluble), ATIII, rhThrombina, glucocerebrosidasa y CTLA4-Ig (Antígeno 4 asociado a Linfocitos T Citotóxicos-Ig).

#### 45 *Expresión de GnT- para Reforzar la Funcionalidad de Anticuerpos*

La adición de restos de *N*-acetilglucosamina a la estructura de GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub> por *N*-acetilglucosaminiltransferasas II y III produce un denominado *N*-glucano GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisechado (Figura 15). Esta estructura se ha implicado en una mayor citotoxicidad celular dependiente de anticuerpos (ADCC) (Umana y col. (1999) *Nat. Biotechnol.* 17(2): 176-80). La re-modificación de glucoformas de inmunoglobulinas expresadas por células de mamífero es una tarea tediosa y compleja. Especialmente en el caso de GnTIII, donde la sobre-expresión de esta enzima se ha implicado en la inhibición del crecimiento, se han tenido que emplear procedimientos que implican expresión génica regulada (inducible) para producir inmunoglobulinas con *N*-glucanos bisechados (Umana y col. (1999) *Biotechnol Bioeng.* 65(5): 542-9; Umana y col. (1999) *Nat. Biotechnol.* 17(2): 176-80); Umana y col., documento WO 03/011878; Patente de Estados Unidos N° 6.602.684.

Por consiguiente, en otra realización, la invención proporciona sistemas y procedimientos para producir *N*-glucanos de tipo humano que tienen *N*-acetilglucosamina (GlcNAc) de bisección en una estructura central de trimanosa o pentamanosa. Por tanto, las inmunoglobulinas que tienen *N*-glucanos bisechados se pueden producir de acuerdo con los procedimientos de la presente invención. Los sistemas y procedimientos descritos en este documento no padecerán problemas previos, por ejemplo, citotoxicidad asociada con sobre-expresión de GnTIII o ADCC, ya que las células huésped de la invención se modifican y seleccionan para ser células viables y preferiblemente robustas que producen *N*-glucanos que tienen glucoformas de tipo humano sustancialmente modificadas tales como GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>. Por tanto, la adición de una *N*-acetilglucosamina de bisección en una célula huésped de la invención tendrá un efecto insignificante sobre el fenotipo de crecimiento o la viabilidad de estas células huésped.

Además, el trabajo de otros ha mostrado que no existe una correlación lineal entre los niveles de expresión de GnTIII y el grado de ADCC. Umana y col. (1999) *Nature Biotechnol.* 17: 176-80. Por tanto, parece que es un desafío encontrar el nivel de expresión óptimo en células de mamífero y mantener el mismo a lo largo de un procedimiento de

## ES 2 335 345 T3

fermentación aprobado por la FDA. Sin embargo, en las células de la invención, tales como células fúngicas, el hallar un promotor de una fuerza apropiada para establecer un nivel de expresión de GnTIII robusto, fiable y óptimo, es una tarea comparativamente sencilla para el especialista en la técnica.

5 Una célula huésped de levadura o unicelular o multicelular capaz de producir glucoproteínas con *N*-glucanos de bisección se modifica de acuerdo con la invención introduciendo en la célula huésped la actividad de GnTIII de la invención (Ejemplo 12). La célula huésped se transforma con un ácido nucleico que codifica los dominios de GnTIII de la invención (véase, por ejemplo, la Figura 24) que tienen actividad enzimática, fusionado opcionalmente con un péptido de dirección señal de célula heterólogo (por ejemplo, usando las bibliotecas y procedimientos asociados de la invención). Las células huésped modificadas para expresar la actividad de GnTIII de la invención producirán mayores títulos de anticuerpos de lo que son capaces las células de mamífero. También producirán anticuerpos con una mayor potencia con respecto a ADCC.

15 Se ha demostrado que los anticuerpos producidos por líneas de células de mamífero transfectadas con GnTIII son tan eficaces como los anticuerpos producidos por líneas celulares no transfectadas, pero a una concentración 10-20 veces menor (Davies y col. (2001) *Biotechnol. Bioeng.* 74(4): 288-94). Un aumento de productividad del vehículo de producción de la invención con respecto a sistemas de mamífero en un factor de veinte, y un aumento de diez veces de la potencia dará como resultado una mejora de productividad neta de doscientos. Por tanto, se pueden producir altos títulos de un anticuerpo que tiene una alta potencia (por ejemplo, hasta varios órdenes de magnitud más potente de lo que se puede producir actualmente) de acuerdo con los procedimientos de la presente invención. El sistema y procedimiento es seguro y proporciona anticuerpos de alta potencia con un bajo coste en periodos de tiempo cortos. Las células huésped modificadas para expresar la actividad de GnTIII de acuerdo con la invención producen inmunoglobulinas que tienen *N*-glucanos bisecados a velocidades de al menos 50 mg/litro/día hasta al menos 500 mg/litro/día. Además, cada molécula de inmunoglobulina (Ig) (que comprende GlcNAc de bisección) es más potente que la misma molécula de Ig producida sin GlcNAc de bisección.

25 Lo siguiente son ejemplos que ilustran diversos aspectos de la invención. Estos ejemplos no se deben considerar como limitantes: los ejemplos se incluyen solamente con fines de ilustración.

### 30 Ejemplo 1

#### *Clonación y Alteración del gen OCH1 en P. pastoris*

##### 35 *Generación de un mutante de OCH1 en P. pastoris*

Una ORF de 1215 pb del gen *OCH1* de *P. pastoris* que codifica una supuesta  $\alpha$ -1,6 manosiltransferasa se amplificó a partir de ADN genómico de *P. pastoris* (cepa X-33, Invitrogen, Carlsbad, CA) usando los oligonucleótidos 5'-ATGGCGAAGGCAGATGGCAGT-3' (SEC ID N°: 3) y 5'-TTAGTCCTTCCAACCTCCTTC-3' (SEC ID N°: 4) que se diseñaron basándose en la secuencia de *OCH1* de *P. pastoris* (Publicación de Solicitud de Patente Japonesa N° 8-336387). Posteriormente, se amplificaron 2685 pb cadena arriba y 1175 pb cadena abajo de la ORF del gen *OCH1* de una genoteca de ADN genómico de *P. pastoris* (Boehm, T. y col. (1999) *Yeast* 15(7): 563-72) usando los oligonucleótidos internos 5'-ACTGCCATCTGCCTTCGCCAT-3' (SEC ID N°: 47) en el gen *OCH1* y 5'-GTAATAC GACTCACTATAGGGC-3' T7 (SEC ID N°: 48) y oligonucleótidos 5'-AATTAACCCTCACTAAAGGG-3' T3 (SEC ID N°: 4), en la cadena principal del plásmido lambda ZAP II que lleva la biblioteca (Stratagene, La Jolla, CA). El fragmento resultante de 5075 pb se clonó en el vector pCR2.1-TOPO (Invitrogen, Carlsbad, CA) y se denominó PBK9.

Después de ensamblar una construcción knockout génica que sustituyó la fase de lectura de *OCH1* con un gen de resistencia *HIS4*, *P. pastoris* se transformó y las colonias se exploraron para sensibilidad a temperatura a 37°C. Las mutantes de *OCH1* de *S. cerevisiae* son sensibles a temperatura y crecen lentamente a temperaturas elevadas. Por tanto, se pueden identificar homólogos funcionales de *OCH1* en *P. pastoris* complementando un mutante de *OCH1* de *S. cerevisiae* con una genoteca de ADN o ADNc de *P. pastoris*. Aproximadamente 20 cepas sensibles a temperatura se sometieron adicionalmente a una exploración por PCR de colonias para identificar colonias con un gen de *och1* deleciónado. Se obtuvieron varias deleciones de *och1*.

55 El pBK9.1 linealizado, que tiene una secuencia de 2,1 kb cadena arriba y una secuencia de 1,5 kb cadena abajo del casete del gen *OCH1* que lleva el gen *HIS4* de *Pichia*, se introdujo por transformación en BK1 de *P. pastoris* [GS115 (*his4* Invitrogen Corp., San Diego, CA) que lleva el gen de IFN- $\beta$  humano en el locus de *AOX1*] para anular el gen *OCH1* de tipo silvestre. La exploración inicial de transformantes se realizó usando medio de retirada de histidina seguido de siembra por duplicado para seleccionar las colonias sensibles a temperatura. Veinte de doscientas de las colonias positivas a histidina mostraron un fenotipo sensible a temperatura a 37°C. Para excluir la integración aleatoria de pBK9.1 en el genoma de *Pichia*, los 20 aislados sensibles a 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 las veinte colonias eran defectuosas en *och1* y se analizaron adicionalmente usando una transferencia de Southern y una transferencia de Western indicando la alteración de *och1* funcional por la construcción knock-out de *och1*. El ADN genómico se digirió usando dos enzimas de restricción separadas *BglIII* y *ClaI* para confirmar el knock-out de *och1* y para confirmar la integración en la fase de lectura abierta. La transferencia de Western mostró mutantes de *och1* que carecían de una banda separada producida en el tipo silvestre de GS115 en 46,2 kDa.

## ES 2 335 345 T3

### Ejemplo 2

#### *Modificación de *P. pastoris* con $\alpha$ -1,2-Manosidasa para Producir Precursores de IFN- $\beta$ que Contienen Man<sub>5</sub>GlcNAc<sub>2</sub>*

5 Se requiere una  $\alpha$ -1,2-manosidasa para el corte de Man<sub>8</sub>GlcNAc<sub>2</sub> para producir Man<sub>5</sub>GlcNAc<sub>2</sub>, un intermedio esencial para la formación de *N*-glucano complejo. Aunque la producción de un precursor de Man<sub>5</sub>GlcNAc<sub>2</sub> es esencial, no es necesariamente suficiente para la producción de glucanos híbridos y complejos debido a que el isómero específico de Man<sub>5</sub>GlcNAc<sub>2</sub> puede o no ser un sustrato para GnTI. Un mutante de *och1* de *P. pastoris* se modifica para expresar interferón  $\beta$  humano secretado bajo el control de un promotor *aox*. Se construyó una genoteca de ADN por la  
10 ligación en fase del dominio catalítico de manosidasa IB humana (una  $\alpha$ -1,2-manosidasa) con una sub-biblioteca que incluye secuencias que codifican péptidos de localización de Golgi temprano y RE. La genoteca de ADN después se introduce por transformación en el organismo huésped, dando como resultado una población genéticamente mixta en la que los transformantes individuales expresan cada uno interferón- $\beta$  así como un gen de manosidasa sintético de la biblioteca. Las colonias de transformantes individuales se cultivan y la producción de interferón se induce por adición  
15 de metanol. En estas condiciones, más del 90% de la proteína secretada es interferón- $\beta$  glucosilado.

Los sobrenadantes se purifican para retirar sales y contaminantes de bajo peso molecular por cromatografía en fase inversa de sílice C<sub>18</sub>. Los transformantes deseados que expresan  $\alpha$ -1,2-manosidasa activa apropiadamente dirigida producen interferón- $\beta$ , que incluye *N*-glucanos de la estructura Man<sub>5</sub>GlcNAc<sub>2</sub>, que tiene una masa molecular reducida  
20 en comparación con el interferón- $\beta$  de la cepa parental. El interferón- $\beta$  purificado se analiza por espectrometría de masas MALDI-TOF y se identifican las colonias que expresan la forma deseada de interferón- $\beta$ .

### Ejemplo 3

#### *Generación de una Cepa Mutante de *och1* que Expresa una $\alpha$ -1,2-Manosidasa, GnTI y GnTII para la Producción de una Glucoproteína de Tipo Humano*

La fase de lectura abierta de 1215 pb del gen de *OCH1* de *P. pastoris* así como 2685 pb cadena arriba y 1175  
30 pb cadena abajo se amplificó por PCR (véase también el documento WO 02/00879), se clonó en el vector pCR2.1-TOPO (Invitrogen) y se denominó PBk9. Para crear una cepa knock-out de *och1* que contiene múltiples marcadores auxótrofos, 100  $\mu$ g de pJN329, un plásmido que contiene un alelo mutante de *och1::URA3* flanqueado con sitios de restricción *SfiI* se digirió con *SfiI* y se usó para transformar la cepa de *P. pastoris* JC308 (Cereghino y Col. (2001) Gene 263: 159-169) por electroporación. Después de la incubación de medio definido que carece de uracilo durante  
35 10 días a temperatura ambiente, se seleccionaron 1000 colonias y se volvieron a sembrar en estrías. Se sometieron clones URA<sup>+</sup> que eran incapaces de desarrollarse a 37°C, pero que se desarrollaban a temperatura ambiente, a PCR de colonias para ensayar la integración correcta del alelo mutante *och1::URA3*. Un clon que mostró el patrón de PCR esperado se denominó YJN153. El dominio de Kringle 3 de plasminógeno humano (K3) se usó como una proteína modelo. Un plásmido marcado con Neo<sup>R</sup> que contenía el gen de K3 se introdujo por transformación en la cepa YJN153  
40 y una cepa resultante, que expresaba K3, se denominó BK64-1.

El plásmido pPB103, que contenía el gen de *MNN2-2* de *Kluyveromyces lactis* que codifica un transportador de UDP-*N*-acetilglucosamina de Golgi, se construyó clonando un fragmento como de *BglIII-HindIII* del vector pDL02 (Abeijon y col. (1996) Proc. Natl. Acad. Sci. U.S.A 93: 5963-5968) en pBLADE-SX digerido con *BglIII* y *BamHI*  
45 y con extremos romos que contenía el gen *ADE1* de *P. pastoris* (Cereghino y col. (2001) Gene 263: 159-169). Este plásmido se linealizó con *EcoNI* y se introdujo por transformación en la cepa BK64-1 por electroporación y una cepa que se había confirmado que contenía el *MNN2-2* por análisis de PCR se denominó **PBP1**.

Se generó una biblioteca de construcciones de manosidasa, que comprenden fusiones en fase de los dominios líder de varias proteínas de membrana de tipo I o tipo II de *S. cerevisiae* y *P. pastoris* fusionados con los dominios catalíticos de varios genes de  $\alpha$ -1,2-manosidasa de fuentes humanas, de ratón, de mosca, de gusano y de levadura (véase, por ejemplo, el documento WO02/00879, incorporado en este documento como referencia). Esta biblioteca se creó en un vector de integración de *HIS4* de *P. pastoris* y se exploró linealizando con *SalI*, introduciendo por transformación por electroporación en la cepa **PBP1** y analizando los glucanos liberados de la proteína indicadora K3. Una construcción  
55 activa elegida era una quimera de los nucleótidos 988-1296 (Extremo C) del gen *SEC12* de levadura fusionado con una deleción *N*-terminal del gen de  $\alpha$ -1,2-manosidasa IA de ratón (Figura 3), que carecía de los 187 nucleótidos. Una cepa de *P. pastoris* que expresa 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 genes de 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 exploró linealizando con *AatII*, introduciendo por transformación por electroporación en la cepa **PBP2** y analizando los glucanos liberados de K3. Una construcción activa elegida era una quimera de los primeros 120 pb del gen de *MNN9* de *S. cerevisiae* fusionado con una deleción del gen de GnTI humana, que carecía de los primeros 154 pb. Una cepa de *P. pastoris* que  
65 expresa esta construcción se denominó **PBP-3**. (Véase también la Figura 36).

Se generó una biblioteca de construcciones de GnTII, que comprendía fusiones en fase de la biblioteca líder con los dominios catalíticos de los genes de GnTII de fuentes humanas y de rata (documento WO 02/00879). Esta

## ES 2 335 345 T3

biblioteca se creó en un vector de integración de *P. pastoris* que contenía el gen de *NST*<sup>R</sup> que otorga resistencia al fármaco nourseotricina. Los plásmidos de biblioteca se linealizaron con *EcoRI*, se introdujeron por transformación en la cepa **RDP27** por electroporación y las cepas resultantes se exploraron por análisis de los glucanos liberados de K3 purificada.

5

### *Materiales para las Sigüientes Reacciones*

El MOPS, cacodilato sódico, cloruro de manganeso, UDP-galactosa y CMP-ácido-*N*-acetilneuramínico eran de Sigma. El ácido trifluoroacético (TFA) era de Sigma/Aldrich, Saint Louis, MO. La  $\alpha$ -2,6-sialiltransferasa de rata recombinante de *Spodoptera frugiperda* y  $\beta$ 1,4-galactosiltransferasa de leche bovina eran de Calbiochem (San Diego, CA). La *N*-glucosidasa F de proteína, manosidasas y oligosacáridos eran de Glyko (San Rafael, CA). La resina de DEAE de ToyoPearl era de TosoHaas. La resina quelante de metales "HisBind" era de Novagen (Madison, WI). Las placas de aclaramiento de lisado de 96 pocillos eran de Promega (Madison, WI). Las placas de 96 pocillos de unión de proteína eran de Millipore (Bedford, MA). Las sales y los agentes de tamponamiento eran de Sigma (St. Louis, MO). Las matrices de MALDI eran de Aldrich (Milwaukee, WI).

15

### *Purificación de Proteína*

20

Se purificó Kringle 3 usando un formato de 96 pocillos en un robot de manejo de muestras Beckman Biomek 2000 (Beckman/Coulter Ranch Cucamonga, CA). Se purificó Kringle 3 de medio de expresión usando un marcador hexa-histidina C-terminal. La purificación robótica es una adaptación del protocolo proporcionado por Novagen para su resina HisBind. En resumen, un volumen asentado de 150  $\mu$ l ( $\mu$ l) de resina se vierte en los pocillos de una placa de unión de lisado de 96 pocillos, se lava con 3 volúmenes de agua y se carga con 5 volúmenes de NiSO<sub>4</sub> 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 del drenado, las columnas se lavan con 10 volúmenes de tampón de unión y 6 volúmenes de tampón de lavado (imidazol 30 mM, NaCl 0,5 M, Tris-HCl 20 mM, pH 7,9) y la proteína se eluye con 6 volúmenes de tampón de elución (imidazol 1M, NaCl 0,5 M, Tris-HCl 20 mM, pH 7,9). Las glucoproteínas eluidas se evaporan hasta sequedad por liofilización.

25

30

### *Liberación de Glucanos Ligados a N*

35

Los glucanos se liberan y separan de las glucoproteínas por una modificación de un procedimiento descrito previamente (Papac, y col. A. J. S. (1998) *Glycobiology* 8, 445-454). Los pocillos de una placa de 96 pocillos MultiScreen IP (membrana Immobilon-P) (Millipore) se humectan con 100  $\mu$ l de metanol, se lavan con 3X150  $\mu$ l de agua y 50  $\mu$ l de tampón RCM (urea 8 M, Tris 360 mM, EDTA 3,2 mM, pH 8,6), drenando con vacío ligero después de cada adición. Las muestras de proteína secadas se disuelven en 30  $\mu$ l de tampón RCM y se transfieren a los pocillos que contienen 10  $\mu$ l de tampón RCM. Los pocillos se drenan y lavan dos veces con tampón RCM. Las proteínas se reducen por adición de 60  $\mu$ l de DTT 0,1 M en tampón RCM durante 1 h a 37°C. Los pocillos se lavan tres veces con 300  $\mu$ l de agua y se carboximetilan por adición de 60  $\mu$ l de ácido iodoacético 0,1 M durante 30 min en oscuridad a temperatura ambiente. Los pocillos se lavan de nuevo tres veces con agua y las membranas se bloquean por la adición de 100  $\mu$ l de PVP 360 al 1% en agua durante 1 h a temperatura ambiente. Los pocillos se drenan y lavan tres veces con 300  $\mu$ l de agua y se desglucosilan por la adición de 30  $\mu$ l de NH<sub>4</sub>HCO<sub>3</sub> 10 mM pH 8,3 que contiene una millitud de *N*-glucanasa (Glyko). Después de 16 horas a 37°C, la solución que contiene los glucanos se eliminó por centrifugación y se evaporó hasta sequedad.

45

50

### *Espectrometría de Masas de Desorción e Ionización por Láser Asistida por Matriz-Tiempo de Vuelo*

Los pesos moleculares de los glucanos se determinaron usando un espectrómetro de masas de MALDI-TOF Voyager DE PRO lineal (Applied Biosciences) usando extracción retrasada. Los glucanos secados de cada pocillo se disolvieron en 15  $\mu$ l de agua y 0,5  $\mu$ l se aplicaron de forma puntual sobre placas de muestra de acero inoxidable y se mezclaron con 0,5  $\mu$ l de matriz S-DHB (9 mg/ml de ácido dihidroxibenzoico, 1 mg/ml de ácido 5-metoxisalicílico en 1:1 agua/acetónitrilo TFA a 0,1%) y se dejaron secar.

55

Se generaron iones por irradiación con un láser de nitrógeno pulsado (337 nm) con un tiempo de pulso de 4 ns. El instrumento se accionó en el modo de extracción retrasada con un retraso de 125 ns y con una tensión de aceleración de 20 kV. La tensión de rejilla era 93,00%, la tensión de cable de guía era del 0,10%, la presión interna era menor de 5 X 10<sup>-7</sup> torr y el umbral de masa baja era 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. Se usó el oligosacárido Man<sub>5</sub>GlcNAc<sub>2</sub> como un patrón de peso molecular externo. Todos los espectros se generaron con el instrumento en el modo de ión positivo. La precisión de masa estimada de los espectros era del 0,5%.

65

## ES 2 335 345 T3

### Ejemplo 4

*Modificación de P. pastoris para Producir Man<sub>5</sub>GlcNAc<sub>2</sub> como la Estructura de N-Glucano Predominante usando una Genoteca de ADN combinatoria*

5 Un mutante de *och1* de *P. pastoris* (véase los Ejemplos 1 y 3) se modificó para expresar y secretar proteínas tales como el dominio de kringle 3 del plasminógeno humano (K3) bajo el control del promotor *AOXI* inducible. El dominio de Kringle 3 de plasminógeno humano (K3) se usó como una proteína modelo. Un fragmento de ADN que codifica el K3 se amplificó usando polimerasa Pfu turbo (Stratagene, La Jolla, CA) y se clonó en los sitios *EcoRI* y *XbaI* de pPICZαA (Invitrogen, Carlsbad, CA), dando como resultando un marcador 6-His C-terminal. Para mejorar la eficacia de glucosilación ligada a N de K3 (Hayes y col. 1975 J. Arch. Biochem. Biophys. 171, 651-655), Pro<sub>46</sub> se sustituyó con Ser<sub>46</sub> usando mutagénesis dirigida. El plásmido resultante se denominó pBK64. La secuencia correcta de la construcción de PCR se confirmó por secuenciación de ADN.

15 Se construyó una genoteca de ADN combinatoria por la ligación en fase de dominios catalíticos de α-1,2-manosidasa IB murina (Genbank AN 6678787) y IA (Genbank AN 6754619) con una sub-biblioteca que incluye secuencias que codifican péptidos de localización de vesícula Cop II, RE y Golgi temprano de acuerdo con la Tabla 6. La genoteca de ADN combinada se usó para generar construcciones de fusión individuales, que después se introdujeron por transformación en el organismo huésped que expresa K3, dando como resultado una población genéticamente mixta en la que los transformantes individuales expresan cada uno K3 así como un gen de fusión de señal de localización/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 del sobrenadante para eliminar sales y contaminantes de bajo peso molecular por cromatografía de afinidad de Ni. Después de la purificación por afinidad, la proteína se sometió a precipitación salina por cromatografía de exclusión por tamaño en una resina Sephadex G10 (Sigma, St. Louis, MO) y se sometió directamente a análisis de MALDI-TOF descrito más adelante o los N-glicanos se eliminaron por digestión con PNGasa como se describe más adelante (liberación de N-glicanos) y se sometieron a análisis de MALDI-TOF Miele y col. (1997) Biotechnol. Appl. Biochem. 25: 151-157.

30 Siguiendo esta estrategia, se obtuvo un conjunto diverso de transformantes; algunos no mostraron modificación de los N-glicanos en comparación con la cepa knock-out de *och1*, y otros mostraron un alto grado de corte de manosa (Figuras 5D y 5E). Los transformantes deseados que expresan α-1,2-manosidasa activa dirigida apropiadamente produjeron K3 con N-glicanos de la estructura Man<sub>5</sub>GlcNAc<sub>2</sub>. Esto otorga una masa molecular reducida a la glucoproteína en comparación con la K3 de la cepa de delección de *och1* parental, una diferencia que se detectó de forma sencilla por espectrometría de masas MALDI-TOF (Figura 5). La Tabla 7 indica los niveles de producción relativos de Man<sub>5</sub>GlcNAc<sub>2</sub>.

TABLA 7

40 *Una genoteca de ADN combinatoria representativa de secuencias de localización/dominios catalíticos que muestra niveles relativos de producción de Man<sub>5</sub>GlcNAc<sub>2</sub>*

|                      |    | Secuencias de Péptido de Dirección |                    |                    |                     |                     |             |
|----------------------|----|------------------------------------|--------------------|--------------------|---------------------|---------------------|-------------|
| Dominios Catalíticos |    | <i>MNS1</i><br>(s)                 | <i>MNS1</i><br>(m) | <i>MNS1</i><br>(l) | <i>SEC12</i><br>(s) | <i>SEC12</i><br>(m) |             |
|                      | 45 | manosidasa<br>1AΔ187 de ratón      | FB4<br>++          | FB5<br>+           | FB6<br>-            | FB7<br>++           | FB8<br>++++ |
|                      | 50 | manosidasa<br>1BΔ58 de ratón       | GB4<br>++          | GB5<br>+           | GB6<br>+            | GB7<br>++           | GB8<br>+    |
|                      | 55 | manosidasa<br>1BΔ99 de ratón       | GC4<br>-           | GC5<br>+++         | GC6<br>+            | GC7<br>+            | GC8<br>+    |
|                      | 60 | manosidasa<br>1BΔ170 de ratón      | GD4<br>-           | GD5<br>-           | GD6<br>-            | GD7<br>+            | GD8<br>+    |
|                      | 65 |                                    |                    |                    |                     |                     |             |

# ES 2 335 345 T3

TABLA 8

Otra genoteca de ADN combinatoria de secuencias de localización/dominios catalíticos que muestra niveles relativos de producción de  $\text{Man}_5\text{GlcNAc}_2$

5

|                      |  | Secuencias de Péptido de Dirección     |                           |                           |                            |                            |                            |      |
|----------------------|--|--|---------------------------|---------------------------|----------------------------|----------------------------|----------------------------|------|
| Dominios Catalíticos |  | <b>VAN1</b><br><b>(s)</b>              | <b>VAN1</b><br><b>(m)</b> | <b>VAN1</b><br><b>(l)</b> | <b>MNN10</b><br><b>(s)</b> | <b>MNN10</b><br><b>(m)</b> | <b>MNN10</b><br><b>(l)</b> |      |
|                      | 10                                     | manosidasa                             | BC18-5                    | BC19                      | BC20                       | BC27                       | BC28                       | BC29 |
|                      | 15                                     | 1B $\Delta$ 80 de<br><i>C. elegans</i> | +++++                     | ++++                      | +++                        | +++++                      | +++++                      | +++  |
| 20                   | manosidasa                             | BB18                                   | BB19                      | BB20                      | BB18                       | BB19                       | BB20                       |      |
|                      | 1B $\Delta$ 31 de<br><i>C. elegans</i> | +++++                                  | +++++                     | ++++                      | +++++                      | +++++                      | ++++                       |      |

25

Se seleccionaron péptidos de dirección de *MNS I* (SwissProt P32906) en *S. cerevisiae* (largo, medio y corto) (véase anteriormente Bibliotecas de Ácido Nucleico; Genoteca 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 de péptido de dirección eran deleciones *N*-terminales, algunas secuencias de péptido de dirección, tales como *SEC12*, eran deleciones *C*-terminales. Los dominios catalíticos usados en este experimento se seleccionaron 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 este documento, indica los niveles relativos de producción de  $\text{Man}_5\text{GlcNAc}_2$ . La notación (-) indica ninguna producción aparente de  $\text{Man}_5\text{GlcNAc}_2$ . La notación (+) indica menos del 10% de producción de  $\text{Man}_5\text{GlcNAc}_2$ . La notación (+ +) aproximadamente el 10-20% de producción de  $\text{Man}_5\text{GlcNAc}_2$ . La notación (+ + +) indica aproximadamente el 20-40% de producción de  $\text{Man}_5\text{GlcNAc}_2$ . La notación con (+ + + +) indica aproximadamente el 50% de producción de  $\text{Man}_5\text{GlcNAc}_2$ . La notación con (+ + + + +) indica más del 50% de producción de  $\text{Man}_5\text{GlcNAc}_2$ .

La Tabla 9 muestra la cantidad relativa de  $\text{Man}_5\text{GlcNAc}_2$  en K3 secretada. Se generaron seiscientos ocho (608) cepas diferentes de *P. pastoris*,  $\Delta\text{och1}$  transformando las mismas con una única construcción de una biblioteca genética combinatoria que se generó fusionando diecinueve (19) dominios catalíticos de  $\alpha$ -1,2-manosidasa con treinta y dos (32) líderes fúngicos de RE y Golgi *cis*.

45

(Tabla pasa a página siguiente)

50

55

60

65

## ES 2 335 345 T3

TABLA 9

|    | Cantidad de Man <sub>5</sub> GlcNAc <sub>2</sub> en K3 secretada<br>(% de glucanos totales)  | Número de<br>construcciones (%) |
|----|--|---------------------------------|
| 5  | N.D.*  | 19 (3,1)                        |
| 10 | 0-10%  | 341 (56,1)                      |
|    | 10-20%   | 50 (8,2)                        |
|    | 20-40%   | 75 (12,3)                       |
| 15 | 40-60%   | 72 (11,8)                       |
|    | Más del 60%  | 51 (8,4) †                      |
|    | Total  | 608 (100)                       |
| 20 | * Varias construcciones de fusión no se ensayaron debido a que los correspondientes plásmidos no se podían propagar en <i>E. coli</i> antes de la introducción por transformación en <i>P. pastoris</i> .  |                                 |
| 25 | † Los clones con el mayor grado de corte de Man <sub>5</sub> GlcNAc <sub>2</sub> (30/51) se analizaron adicionalmente para actividad de manosidasa en el sobrenadante del medio. La mayoría (28/30) mostró una actividad de manosidasa detectable en el sobrenadante (por ejemplo, <b>Figura 4B</b> ). Solamente dos construcciones presentaron niveles altos de Man <sub>5</sub> GlcNAc <sub>2</sub> aunque carecían de actividad de manosidasa en el medio (por ejemplo, <b>Figura 4C</b> ). |                                 |
| 30 |  |                                 |
| 35 |  |                                 |

La Tabla 7 muestra dos construcciones **pFB8** y **pGC5**, entre otras, que presentan Man<sub>5</sub>GlcNAc<sub>2</sub>. La Tabla 8 muestra una construcción más preferida, **pBC18-5**, una secuencia de péptido de dirección de *VAN1(s)* de *S. cerevisiae* (de SwissProt 23642) ligada en fase con una delección *N*-terminal de 80 aminoácidos de manosidasa IB de *C. elegans* (Genbank AN CAA98114) (*Van1 (s)* de *Saccharomyces*/manosidasa IB  $\Delta$ 80 de *C. elegans*). Esta construcción de fusión también produce una estructura predominante de Man<sub>5</sub>GlcNAc<sub>2</sub>, como se muestra en la Figura 5E. Se demostró que esta construcción produce más del 50% de Man<sub>5</sub>GlcNAc<sub>2</sub> (+++++).

### *Generación de una biblioteca de localización/manosidasa combinatoria*

La generación de una genoteca de ADN combinatoria de dominios catalíticos de  $\alpha$ -1,2-manosidasa fusionados con péptidos de dirección requirió la amplificación de dominios de manosidasa con longitudes variables de delecciones *N*-terminales de varios organismos. Para conseguir este objetivo, las fases de lectura abierta (ORF) de longitud completa de  $\alpha$ -1,2-manosidasas se amplificaron por PCR de ADNc o ADN genómico obtenido de las siguientes fuentes: *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Aspergillus nidulans* y *Penicillium citrinum*. En cada caso, el ADN se incubó en presencia de cebadores oligonucleotídicos específicos para la secuencia deseada de manosidasa además de reactivos requeridos para realizar la reacción de PCR. Por ejemplo, para amplificar la ORF de la  $\alpha$ -1,2-manosidasa IA de *M. musculus*, los cebadores 5'-ATGCCCGTGGGGGGCCTGTTGCCGCTCTT CAGTAGC (SEC ID N°: 52) y el cebador 3'-TCATTTCTCTTTGCCATCAATTCCTTCTTCTGTTACGG (SEC ID N°: 53) se incubaron en presencia de ADN polimerasa Pfu (Stratagene, La Jolla, CA) y se amplificaron en las condiciones recomendadas por Stratagene usando los parámetros de ciclado: 94°C durante 1 min (1 ciclo); 94°C durante 30 s, 68°C durante 30 s, 72°C durante 3 min (30 ciclos). Después de la amplificación, la secuencia de ADN que codifica la ORF se incubó a 72°C durante 5 min con 1 U de ADN polimerasa de Taq (Promega, Madison, WI) antes de la ligación en pCR2.1-TOPO (Invitrogen, Carlsbad, CA) y se introdujo por transformación en *E. coli* TOP10 químicamente competente, como se recomienda por Invitrogen. El producto de PCR clonado se confirmó por secuenciación ABI usando cebadores específicos para la ORF de manosidasa.

Para generar los truncamientos *N*-terminales deseados de cada manosidasa, toda la ORF de cada manosidasa se usó como el molde en una ronda posterior de reacciones de PCR en la que la posición de hibridación del cebador 5' era específica para el extremo 5' del truncamiento deseado y el cebador 3' permaneció específico para el extremo 3'

## ES 2 335 345 T3

original de la ORF. Para facilitar la subclonación del fragmento de manosidasa truncado en el vector de expresión de levadura, se introdujeron sitios de restricción *Ascl* y *PacI* de pJN347 (Figura 2C) en cada producto de truncamiento, respectivamente en los extremos 5' y 3'. El número y la posición de los truncamientos *N*-terminales generados para cada ORF de manosidasa dependía de la posición de la región transmembrana (TM) en relación al dominio catalítico (CD). Por ejemplo, si la región del tallo localizada entre la TM y CD era menor de 150 pb, entonces solamente se generó un truncamiento para esa proteína. Sin embargo, si la región del tallo era mayor de 150 pb, entonces se generaron uno o dos truncamientos dependiendo de la longitud de la región de tallo.

Un ejemplo de cómo se generaron truncamientos para la manosidasa IA de *M. musculus* (Genbank AN 6678787) se describe en este documento, usándose una estrategia similar para las demás manosidasas. La Figura 3 ilustra la ORF de la  $\alpha$ -1,2-manosidasa IA de *M. musculus* con los dominios predichos transmembrana y catalíticos resaltados en negrita. Basándose en esta estructura, se diseñaron tres cebadores 5' (posiciones de hibridación subrayadas en la Figura 3) para generar las delecciones *N*-terminales  $\Delta 65$ ,  $\Delta 105$  y  $\Delta 187$ . Usando la delección  $\Delta 65$  *N*-terminal como un ejemplo, el cebador 5' usado fue 5'-**GGCGCGCCGACTCCTCCAAGCTGCTCAGCGGGTCTGTTCCAC**-3' (SEC ID N°: 54) (con el sitio de restricción *Ascl* resaltado en negrita) junto con el cebador 3' 5'-**CCTTAATTAATCATTTCTCTTTGCCATCAATTTCTTCTGTTCCACGG**-3' (SEC ID N°: 55) (con el sitio de restricción de *PacI* resaltado en negrita). Estos dos cebadores se usaron para amplificar un fragmento de 1561 pb en las condiciones resumidas anteriormente para amplificar la ORF de manosidasa IA de *M. musculus* de longitud completa. Además, al igual que el producto obtenido para la ORF de longitud completa, el producto truncado también se incubó con ADN polimerasa de Taq, se ligó en pCR2.1-TOPO (Invitrogen, Carlsbad, CA), se introdujo por transformación en TOP 10 y se secuenció por ABI. Después de haber amplificado y confirmado la secuencia del fragmento de manosidasa truncado, el plásmido resultante, pCR2.1- $\Delta 65$ mMannIA, se digirió con *Ascl* y *PacI* en tampón N° 4 de New England Biolabs (Beverly, MA) durante 16 h a 37°C. 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 de pJN347 (Figura 2C) como el dominio catalítico truncado se extrajeron en gel y se ligaron usando el Kit de Ligación Quick (New England Biolabs, Beverly, MA), como se recomienda por los fabricantes, y se introdujo por transformación en células DH5 $\alpha$  químicamente competentes (Invitrogen, Carlsbad, CA). Se usó PCR de colonias para confirmar la generación de la construcción de manosidasa IA  $\Delta 65$  de ratón pJN347.

Al haber generado una biblioteca de dominios catalíticos de  $\alpha$ -1,2-manosidasa truncados en el vector de expresión de levadura pJN347 (Figura 2C), la etapa remanente para generar la biblioteca de péptido de dirección/dominio catalítico era clonar en fase las secuencias de péptido de dirección (Figura 2). Tanto las construcciones de manosidasa de pJN347 (Figura 2D) como las construcciones de péptido de dirección de pCR2.1-TOPO (Figura 2B) como tales se incubaron durante una noche a 37°C en tampón N° 4 de New England Biolabs en presencia de las enzimas de restricción *NotI* y *Ascl*. Después de la digestión, tanto la cadena principal de manosidasa de pJN347 como las regiones de péptido de dirección se extrajeron en gel y se ligaron usando el Kit de Ligación Quick (New England Biolabs, Beverly, MA), como se recomienda por los fabricantes, y se introdujeron por transformación en células DH5 $\alpha$  químicamente competentes (Invitrogen, Carlsbad, CA). Posteriormente, las construcciones de péptido de dirección/manosidasa de pJN347 se secuenciaron por ABI para confirmar que las fusiones generadas estaban en fase. El tamaño estimado de la biblioteca final de péptido de dirección/alfa-1,2-manosidasa contiene más de 1300 construcciones generadas por la estrategia que se ha descrito anteriormente. La Figura 2 ilustra la construcción de la genoteca de ADN combinatoria.

### Modificación de una cepa knock-out de *OCH1* de *P. pastoris* con múltiples marcadores auxótrofos

La primera etapa en la construcción de plásmido implicó crear un conjunto de plásmidos universales que contienen regiones de ADN del gen *KEX1* de *P. pastoris* (Boehm y col. Yeast 1999 May; 15(7): 563-72) como separadores para las regiones 5' y 3' de los genes a anular. Los plásmidos también contenían el Ura-blasto de *S. cerevisiae* (Alani y col. (1987) Genetics 116: 541-545) como un separador para los marcadores auxótrofos y un casete de expresión con un sitio de clonación múltiple para la inserción de un gen extraño. Un fragmento de 0,9 kb de la región *KEX1*-5' de *P. pastoris* se amplificó por PCR usando los cebadores **GGCGAGCTCGGCTACCCGGCCAAGGCTGAGATCATTTGTCCAGCTTCA GA** (SEC ID N°: 56) y **GCCCACGTGACCGATCCGTTTAAACATCGATTGGAGAGGCTGACACC GCTACTA** (SEC ID N°: 57) y ADN genómico de *P. pastoris* como un molde y se clonó en los sitios *SacI*, *Sall* de pUC19 (New England Biolabs, Beverly, MA). El plásmido resultante se cortó con *BamHI* y *Sall* y un fragmento de 0,8 kb de la región *KEX1*-3' que se había amplificado usando los cebadores **CGGGATCCACTAGTATTTAAATCATATGTGCGAGTGTAACAACCTTTCCC ACATGG** (SEC ID N°: 58) y **GGACGCGTCCGACGGCCTACCCGGCCGTACGAGGAATTTCTCGG ATGACTCTTTTC** (SEC ID N°: 59) se clonó en los sitios abiertos creando pJN262. Este plásmido se cortó con *BamHI* y el fragmento de 3,8 kb de *BamHI*, *BglII* de pNKY51 (Alani y col. (1987) Genetics 116: 541-545) se insertó en las dos posibles orientaciones 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 de GAPDH de *P. pastoris* se amplificó usando los cebadores **CGGGATCCCTCGAGAGATCTTTTTTGTAGAAATGTCTTGGTGCCCT** (SEC ID N°: 60) y **GGACATGCATGCACTAGTGCGGCCGCCACGTGATAGTTGTTCAATTGATTGAAATAGGGAC AA** (SEC ID N°: 61) 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 de terminador de la transcripción ("TT") CYC1 que se había amplificado usando los cebadores **CCTTGCTAGCTTAATTAACCGCGGCACGTCCGACGGCGGCCCA CGGGTCCCA** (SEC ID N°: 62) y **GGACATGCATGCGGATCCCT**

## ES 2 335 345 T3

TAAGAGCCGGCAGCTTGCAAATTAAGCCTTCGAGCGTCC C (SEC ID N°: 63) y el plásmido pPICZ-A (Invitrogen) como un molde se clonó en los sitios abiertos creando pJN261 (Figura 4B).

Se creó un plásmido knockout para el gen *OCH1* de *P. pastoris* digiriendo pJN263 con *Sall* y *SpeI* y un fragmento de ADN de 2,9 kb de la región *OCH1-5'*, que se había amplificado usando los cebadores GAACCACGTCGAC GGCCATTGCGGCCAAAACCTTTTTTCCTATT CAAACACAAGGCATTGC (SEC ID N°: 64) y CTCCAATAC TAGTCGAAGATTATCTTCTACGGTGCCTGGACTC (SEC ID N°: 65) 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 un fragmento de ADN de 1,0 kb de la región 3' de *OCH* que se había generado usando los cebadores TGGGAAG GTTTAAACAAAGCTAGAGTAAAATAGATATAGCGAG ATTAGAGAATG (SEC ID N°: 66) y AAGAATTC GGCTGGAAGGCCTTGACCTTGATGTAGTTCCCGTT TTCATC (SEC ID N°: 67) se insertó para generar pJN298 (Figura 4C). Para permitir la posibilidad de usar simultáneamente el plásmido para introducir un nuevo gen, el casete de expresión de *BamHI* de pJN261 (Figura 4B) se clonó en el único sitio *BamHI* de pJN298 (Figura 4C) para crear pJN299 (Figura 4E).

El casete de *P. pastoris* Ura3-blaster se construyó usando una estrategia similar a la descrita en Lu y col. (1998) Appl. Microbiol. Biotechnol. 49: 141-146. Un fragmento de 2,0 kb de *PstI*, *SpeI* de URA3 de *P. pastoris* se insertó en los sitios *PstI*, *XbaI* de pUC19 (New England Biolabs, Beverly, MA) para crear pJN306 (Figura 4D). Después, un fragmento de ADN de 0,7 kb de *SacI*, *PvuII* de la fase de lectura abierta de *lacZ* se clonó en los sitios *SacI*, *SmaI* para producir pJN308 (Figura 4D). Después de la digestión de pJN308 (Figura 4D) con *PstI* y tratamiento con ADN polimerasa de T4, el fragmento *SacI-PvuII* de *lacZ* en el que se habían formado extremos romos con ADN polimerasa de T4 se insertó generando pJN315 (Figura 4D). El casete *lacZ/URA3* se liberó por digestión con *SacI* y *SphI*, se crearon extremos romos con ADN polimerasa de T4 y se clonó en la cadena principal de pJN299 que se había digerido con *PmeI* y *AflIII* y se crearon extremos romos con ADN polimerasa de T4. El plásmido resultante se denominó pJN329 (Figura 4E).

Un plásmido de expresión marcado con *HIS4* se creó cortando pJN261 (Figura 4F) con *EcoICRI* (Figura 4F). Un fragmento de 2,7 kb del gen *HIS4* de *Pichia pastoris* que se había amplificado usando los cebadores GCCCAAGCCGGCCTTAAGGGATCTCCTGATGACTGACTCACTGATAATA AAAATACGG (SEC ID N°: 68) y GGGCGCGTATTTAAATACTAGTGGATCTATCGAATCTAAATGTAAGTTA AAATCTCTAA (SEC ID N°: 69) cortado con *NgoMIV* y *SwaI* y en los que después se habían formado extremos romos usando ADN polimerasa de T4, se ligó entonces en el sitio abierto. Este plásmido se denominó pJN337 (Figura 4F). Para construir un plásmido con un sitio de clonación múltiple adecuado para la construcción de biblioteca de fusión, pJN337 se cortó con *NotI* y *PacI* y los dos oligonucleótidos GGCCGCGTGCAGATTAAATGAATTCGGCGCGCCTTAAT (SEC ID N°: 70) y TAAGGCGCGCCGAATTCATTTAAATCTGCAGGGC (SEC ID N°: 71) que se habían hibridado *in vitro* se ligaron en los sitios abiertos, creando pJN347 (Figura 4F).

Para crear una cepa knockout de *och1* que contiene múltiples marcadores auxótrofos, 100  $\mu$ g de pJN329 se digirieron con *SfiI* y se usaron para transformar la cepa de *P. pastoris* JC308 (Cereghino y col. (2001) Gene 263: 159-169) por electroporación. Después de la transformación, las placas de retirada de URA se incubaron a temperatura ambiente durante 10 días. Mil (1000) colonias se seleccionaron y volvieron a sembrar en estrías. Los 1000 clones después se sembraron en estrías en 2 conjuntos de placas de retirada de URA. Un conjunto se incubó a temperatura ambiente, mientras que el segundo conjunto se incubó a 37°C. Los clones que no eran capaces de desarrollarse a 37°C, pero que se desarrollaron a temperatura ambiente, se sometieron a PCR de colonias para ensayar el knockout correcto de *OCH1*. Un clon que mostró la señal de PCR esperada (aproximadamente 4,5 kb) se denominó YJN153.

### Ejemplo 5

#### 50 Caracterización de la Genoteca de ADN combinatoria

Los transformantes positivos explorados por PCR de colonias que confirman la integración de la construcción de manosidasa en el genoma de *P. pastoris* se cultivaron posteriormente a temperatura ambiente en medio complejo de metanol tamponado con BMGY 50 mol constituido por extracto de levadura al 1%, peptona al 2%, tampón fosfato potásico 100 mM, pH 6,0, base de nitrógeno de levadura al 1,34%, 4 X 10<sup>-5</sup>% de biotina y glicerol al 1% como un medio de cultivo) hasta DO<sub>600</sub> nm 2-6, punto en el cual se lavaron con medio BMMY 10 mol (medio complejo de metanol tamponado constituido por extracto de levadura al 1%, peptona al 2%, tampón fosfato potásico 100 mM, pH 6,0, base de nitrógeno de levadura al 1,34%, 4 X 10<sup>-5</sup>% de biotina y metanol al 1,5% como un 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 como se describe en el Ejemplo 3 para caracterizar su estructura de glucano. Usando los péptidos de dirección en la Tabla 6, los dominios catalíticos de manosidasa localizados en el RE o el Golgi mostraron un nivel significativo de corte de un glucano que contiene predominantemente Man<sub>8</sub>GlcNAc<sub>2</sub> hasta un glucano que contiene predominantemente Man<sub>5</sub>GlcNAc<sub>2</sub>. Esto es evidente cuando la estructura de glucano de la glucoproteína indicadora se compara entre la del knockout de *och1* de *P. pastoris* en las Figuras 5C y 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 genoteca de ADN combinatoria que muestran actividad de manosidasa significativa en *P. pastoris*. La expresión de pGC5 (*MNS1*) (m) de *Saccharomyces/manosidasa* IB  $\Delta$ 99 de ratón (Figuras 5D y 6E) produjo una proteína que tiene aproximadamen-

te el 30% de todos los glucanos cortados hasta Man<sub>5</sub>GlcNAc<sub>2</sub>, mientras que la expresión de pFB8 (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) (Figura 6F) produjo aproximadamente el 50% de Man<sub>5</sub>GlcNAc<sub>2</sub> y la expresión de pBC18-5 (*VANI* (s) de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*) (Figura 5E) produjo el 70% de Man<sub>5</sub>GlcNAc<sub>2</sub>.

5

## Ejemplo 6

*Corte in vivo por α-1,2-manosidasa*

10

Para garantizar que las cepas modificadas novedosas del Ejemplo 4 de hecho producían la estructura de Man<sub>5</sub>GlcNAc<sub>2</sub> deseada *in vivo*, los sobrenadantes celulares se ensayaron para actividad de manosidasa (véase las Figuras 7-9). Para cada construcción/cepa huésped descrita más adelante, se realizó HPLC a 30°C con una columna de 4,0 mm x 250 mm de Altech (Avondale, PA, EE.UU.) de resina Econosil-NH<sub>2</sub> (5 μm) a un caudal de 1,0 ml/min durante 40 min. En las Figuras 7 y 8, la degradación del Man<sub>9</sub>GlcNAc<sub>2</sub> convencional [b] se demostró que tenía lugar dando como resultado un pico que se correlaciona con Man<sub>8</sub>GlcNAc<sub>2</sub>. En la Figura 7, el Man<sub>9</sub>GlcNAc<sub>2</sub> [b] convencional eluyó a 24,61 min y Man<sub>5</sub>GlcNAc<sub>2</sub> [a] eluyó a 18,59 min. En la Figura 8, Man<sub>9</sub>GlcNAc<sub>2</sub> eluyó a 21,37 min y Man<sub>5</sub>GlcNAc<sub>2</sub> a 15,67 min. En la Figura 9, el Man<sub>8</sub>GlcNAc<sub>2</sub> convencional [b] se demostró que eluía a 20,88 min.

15

20

Las células de *P. pastoris* que comprenden el plásmido pFB8 (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) se cultivaron a 30°C en BMGY hasta una DO<sub>600</sub> de aproximadamente 10. Las células se recuperaron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación para producir un sobrenadante esencialmente claro. Una alícuota del sobrenadante se retiró para ensayos de manosidasa y el resto se usó para la recuperación de K3 soluble secretada. Una única etapa de purificación usando cromatografía de CM-sepharose 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 una K3 al 95% pura que eluía entre NaCl 300-500 mM. El análisis de *N*-glucano de los glucanos obtenidos de K3 se muestra en la Figura 6F. La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente para la presencia de actividad de manosidasa secretada. Un patrón disponible en el mercado de oligomanosa 9 de tipo ligado a N marcada con 2-aminobenzamida (Man9-2-AB) (Glyko, Novato, CA) se añadió a: BMMY (Figura 7A), el sobrenadante de la anterior alícuota (Figura 7B) y BMMY que contenía 10 ng de 75 mU/ml de α-1,2-manosidasa de *Trichoderma reesei* (obtenido de Contreras y col., documento WO 02/00856 A2) (Figura 7C). Después de la incubación durante 24 horas a temperatura ambiente, las muestras se analizaron por HPLC de amino sílice para determinar el alcance del corte de manosidasa.

25

30

35

Las células de *P. pastoris* que comprendían el plásmido pGC5 (*MNS1* (m) de *Saccharomyces*/manosidasa IB Δ99 de ratón) se cultivaron y ensayaron de forma similar. Las células se cultivaron a temperatura ambiente en BMGY hasta una DO<sub>600</sub> de aproximadamente 10. Las células se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 bajo el control de un promotor *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación para producir un sobrenadante esencialmente claro. Una alícuota del sobrenadante se retiró para ensayos de manosidasa y el resto se usó para la recuperación de K3 soluble secretada. Una única etapa de purificación usando cromatografía de CM-sepharose 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 una K3 al 95% pura que eluía entre NaCl 300-500 mM. El análisis de *N*-glucano de los glucanos obtenidos de K3 se muestra en la Figura 5D. La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente para la presencia de actividad de manosidasa secretada como se muestra en la Figura 8B. Un patrón disponible en el mercado de Man9-2-AB (Glyko, Novato, CA) se añadió a: BMMY (Figura 8A), el sobrenadante de la anterior alícuota (Figura 8B) y BMMY que contenía 10 ng de 75 mU/ml de α-1,2-manosidasa de *Trichoderma reesei* (obtenido de Contreras y col., documento WO 02/00856 A2) (Figura 8C). Después de la incubación durante 24 horas a temperatura ambiente, las muestras se analizaron por HPLC de amino sílice para determinar el alcance del corte de manosidasa.

40

45

50

Man9-2-AB se usó como sustrato y es evidente que después de 24 horas de incubación, la actividad de manosidasa prácticamente estaba ausente en el sobrenadante del producto de digestión de la cepa pFB8 (*SEC12* (m) de *Saccharomyces*/manosidasa IA Δ187 de ratón) (Figura 7B) y el producto de digestión de 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<sub>9</sub>GlcNAc<sub>2</sub> a Man<sub>5</sub>GlcNAc<sub>2</sub> en las mismas condiciones, como se muestra en las Figuras 7C y 8C. Estos son datos concluyentes que muestran el corte de manosidasa *in vivo* en la cepa pGC5 de *P. pastoris*; y la cepa pFB8, que es claramente diferente de lo que se había descrito hasta la fecha (Contreras y col., documento WO 02/00856 A2).

55

60

La Figura 9 confirma adicionalmente la localización y la actividad de la enzima manosidasa. Se cultivó *P. pastoris* que comprende pBC18-5 (*VANI*(s) de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans*) a temperatura ambiente en BMGY hasta una DO<sub>600</sub> de aproximadamente 10. Las células se recuperaron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 bajo el control de un promotor *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación para producir un sobrenadante esencialmente claro. Una alícuota del sobrenadante se retiró para ensayos de manosidasa y el resto se usó para la recuperación de K3 soluble secretada. Una única etapa de purificación usando cromatografía de CM-sepharose 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 una K3 al 95% pura que eluía entre NaCl 300-500 mM. El

65

## ES 2 335 345 T3

análisis de *N*-glucano de los glucanos obtenidos de K3 se muestra en la Figura 5E. La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente para la presencia de actividad de manosidasa secretada como se muestra en la Figura 9B. Un patrón disponible en el mercado de ManB-2-AB (Glyko, Novato, CA) se añadió a: BMMY (Figura 9A), sobrenadante de la anterior alícuota de pBC18-5 (*VAN1*(s) de *Saccharomyces*/manosidasa IB Δ80 de *C. elegans* (Figura 9B) y BMMY que contenía medio 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 por HPLC de amino sílice para determinar el alcance del corte de manosidasa. La Figura 9B demuestra la actividad de 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 7

#### *Ensayo de Óptimo de pH de α-1,2-manosidasa modificada*

Se cultivaron 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 hasta una DO600 de aproximadamente 17. Aproximadamente 80 μl de estas células se inocularon en 600 μl de BMGY y se cultivaron durante una noche. Posteriormente, las células se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor AOX1. Después de 24 horas de inducción, las células se retiraron por centrifugación para producir un sobrenadante esencialmente claro (pH 6,43). El sobrenadante se retiró para ensayos de óptimo de pH de manosidasa. Se añadió Man<sub>8</sub>GlcNAc<sub>2</sub> marcado con fluorescencia (0,5 μg) a 20 μl de sobrenadante ajustado a diversos pH (Figura 11) y se incubó durante 8 horas a temperatura ambiente. Después de la incubación, la muestra se analizó por HPLC usando una columna de sílice unida a amino, de perlas de 5 micrómetros, Econosil NH2 4,6 X 250 mm (Altech, Avondale, PA). El caudal fue 1,0 ml/min durante 40 min y la columna se mantuvo a 30°C. Después de eluir de forma isocrática (68% A:32% B) durante 3 min, se empleó un gradiente de disolvente lineal (68% A:32% B a 40% A:60% B) a lo largo de 27 min para eluir los glucanos (18). Disolvente A (acetonitrilo) y el disolvente B (formato de amonio, 50 mM, pH 4,5). La columna se equilibró con disolvente (68% A:32% B) durante 20 min entre ciclos.

### Ejemplo 8

#### *Modificación de P. pastoris para Producir N-glucanos con la Estructura GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>*

Se requiere actividad de GlcNAc Transferasa I para la maduración de *N*-glucanos complejos e híbridos (Patente de Estados Unidos N° 5.834.251). Man<sub>5</sub>GlcNAc<sub>2</sub> se puede cortar solamente por 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 por GlcNAc Transferasa I (Schachter, 1991 Glycobiology 1(5): 453-461). Por consiguiente, se preparó una genoteca 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 supuestas α-1,2-manosiltransferasas de *S. cerevisiae* y *P. pastoris* basándose en la homología de *S. cerevisiae*: D2, D9 y J3, que son homólogos de *KTR*. La Tabla 10 incluye, pero no limita, secuencias de péptido de dirección tales como *SEC* y *OCH1*, de GnTI de *P. pastoris* y *K. lactis* (Véase la Tabla 6 y la Tabla 10).

TABLA 10

*Una biblioteca combinatoria representativa de secuencias de péptido de dirección/dominio catalítico para UDP-N-Acetilglucosaminil Transferasa I (GnTI)*

| D<br>E |                                     | Péptido de dirección |                 |                 |                 |                 |
|--------|-------------------------------------|----------------------|-----------------|-----------------|-----------------|-----------------|
|        |                                     | <i>OCH1</i> (s)      | <i>OCH1</i> (m) | <i>OCH1</i> (l) | <i>MNN9</i> (s) | <i>MNN9</i> (m) |
| 55     | <b>Ser humano, GnTI, Δ38</b>        | PB105                | PB106           | PB107           | PB104           | N/A             |
|        | <b>Ser humano, GnTI, Δ86</b>        | NB12                 | NB13            | NB14            | NB15            | NB              |
| 60     | <b><i>C. elegans</i>, GnTI, Δ88</b> | OA12                 | OA13            | OA14            | OA15            | OA16            |
|        | <b><i>C. elegans</i>, GnTI, Δ35</b> | PA12                 | PA13            | PA14            | PA15            | PA16            |
|        | <b><i>C. elegans</i>, GnTI, Δ63</b> | PB12                 | PB13            | PB14            | PB15            | PB16            |
| 65     | <b><i>X. laevis</i>, GnTI, Δ33</b>  | QA12                 | QA13            | QA14            | QA15            | QA16            |
|        | <b><i>X. laevis</i>, GnTI, Δ103</b> | QB12                 | QB13            | QB14            | QB15            | QB16            |

## ES 2 335 345 T3

Las secuencias de péptido de dirección se seleccionaron de *OCH1* en *P. pastoris* (largo, medio y corto) (véase Ejemplo 4) y *MNN9* (SwissProt P39107) en *S. cerevisiae* (corto y medio). Se seleccionaron dominios catalíticos 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. laevis* con una deleción N-terminal de 33 y 103 aminoácidos, respectivamente.

Una porción del gen que codifica *N*-acetilglucosaminil Transferasa I humana (MGATI, N° de Acceso NM002406), que carece de los primeros 154 pb, se amplificó por PCR usando los oligonucleótidos 5'-TGGCAGGCGCGCCT CAGTCAGCGCTCTCG-3' (SEC ID N°: 72) y 5'-AGGTAAATTA AGTGCTAATTCCAGCTAGG-3' (SEC ID N°: 73) y el vector pHG4.5 (ATCC# 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* truncada se insertó en el plásmido pJN346 para crear pNA. Después de la digestión de pJN271 con *NotI* y *Asc*, el inserto de 120 pb se ligó en pNa para generar una fusión en fase del dominio transmembrana de *MNN9* con la *GnTI*, creando pNA1 5.

El organismo huésped es una cepa de *P. pastoris* que es deficiente en hipermanosilación (por ejemplo, un mutante de *och1*), proporciona el sustrato UDP-GlcNAc en el Golgi y/o RE (es decir, contiene un transportador funcional de UDP-GlcNAc) y proporciona *N*-glucanos de la estructura Man<sub>5</sub>GlcNAc<sub>2</sub> en el Golgi y/o RE (por ejemplo, pFB8 de *P. pastoris* (SEC12(m) de *Saccharomyces/*manosidasa IA Δ187 de ratón) de antes). En primer lugar, pFB8 de *P. pastoris* se transformó con pPB103 que contenía el gen *MNN2-2* de *Kluyveromyces lactis* (Genbank AN AF106080) (que codifica el transportador de UDP-GlcNAc) clonado en el sitio *BamHI* y *BglIII* del plásmido pBLADE-SX (Cereghino y col. (2001) Gene 263: 159-169). Después, la genoteca de ADN combinatoria que se ha mencionado anteriormente que codifica una combinación de genes de GnTI/localización exógenos o endógenos se transformó y las colonias se seleccionaron y analizaron para la presencia de la construcción de GnTI por PCR de colonias. La eficacia de transformación e integración estaba generalmente por encima del 80% y la exploración de PCR se puede omitir una vez que se han establecido parámetros de transformación robustos.

### Purificación de Proteína

Se purificó K3 del medio por cromatografía de afinidad de Ni utilizando 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. Se puede realizar otro procedimiento de exploración usando un anticuerpo de unión de GlcNAc terminal específico o una lectina tal como la lectinas GSII de *Griffonia simplicifolia*, que se une a la GlcNAc (EY Laboratories, San Mateo, CA). Estas exploraciones se pueden automatizar usando lectinas o anticuerpos que se han modificado con marcadores fluorescentes tales como FITC o analizar por MALDI-TOF.

La K3 secretada se puede purificar por cromatografía de afinidad de Ni, cuantificar y cantidades iguales de proteína se pueden unir a una placa de 96 pocillos de alta unión a proteína. Después del bloqueo con BSA, las placas se pueden sondear con una lectina GSII-FACS y explorar para una respuesta fluorescente máxima. Un procedimiento preferido para detectar las anteriores proteínas glucosiladas implica la exploración por espectrometría de masas MALDI-TOF después de la purificación por afinidad de K3 secretada del sobrenadante de transformantes cultivados en 96 pocillos. Las colonias transformadas se seleccionaron y desarrollaron hasta una DO600 de 10 en una placa de 2 ml, 96 pocillos en BMGY a 30°C. 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 de extracción retrasada MALDI-TOF como se describe en este documento.

En resumen, los procedimientos de la invención produjeron cepas de *P. pastoris* que producen GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> con alto rendimiento, como se muestra en la Figura 10B. Al menos el 60% de los *N*-glucanos son GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>. Hasta la fecha, no existen informes que describen la formación de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> en glucoproteínas solubles secretadas en ninguna levadura. Los resultados presentados en este documento muestran que la adición del transportador de UDP-GlcNAc junto con la actividad de GnTI produce una estructura predominante de GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>, que se confirma por el pico en 1457 (m/z) (Figura 10B).

### Construcción de la cepa **PBP-3**

La cepa de *P. pastoris* que expresa K3, (*Δoch1*, arg-, ade-, his-) se transformó sucesivamente con los siguientes vectores. En primer lugar, **pFB8** (*SEC12* (m) de *Saccharomyces/*manosidasa IA Δ1887 de ratón) se introdujo por transformación en la cepa de *P. pastoris* por electroporación. En segundo lugar, pPB103 que contenía el gen *MNN2-2* de *Kluyveromyces lactis* (Genbank AN AF106080) (que codifica el transportador de UDP-GlcNAc) clonado en el plásmido pBLADE-SX (Cereghino y col. (2001) Gene 263: 159-169) digerido con las enzimas *BamHI* y *BglIII* se introdujo por transformación en la cepa de *P. pastoris*. En tercer lugar, **pPB104** que contenía el gen que codifica *MNN9*(s) de *Saccharomyces/GnTI* Δ38 humana clonado como fragmento de *NotI-PacI* en pJN336 se introdujo por transformación en la cepa de *P. pastoris*.

## Ejemplo 9

*Modificación de Células de K. lactis para Producir N-glucanos con la Estructura Man<sub>5</sub>GlcNAc<sub>2</sub>*5 *Identificación y Alteración del gen OCHI de K. lactis*

El gen *OCHI* de la levadura de gemación *S. cerevisiae* codifica una 1,6-manosiltransferasa que es responsable de la primera adición de manosa localizada en Golgi a la estructura de *N*-glucano Man<sub>8</sub>GlcNAc<sub>2</sub> en proteínas secretadas (Nakanishi-Shindo y col. (1993) J. Biol. Chem.; 268(35): 26338-45). Esta transferencia de manosa se reconoce generalmente como la etapa inicial clave en la polimanosilación específica de hongos de estructuras de *N*-glucano (Nakanishi-Shindo y col. (1993) J. Biol. Chem. 268 (35): 26338-26345; Nakayama y col. (1992) EMBO J. 11 (7):2511-19; Morin-Ganet y col. (2000) Traffic 1 (1): 56-68). La delección 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 crecimiento a temperaturas elevadas (Nakayama y col. (1992) EMBO J. 11 (7): 2511-19).

15 La secuencia de *Och1p* de *S. cerevisiae* se alineó con homólogos conocidos de *Candida albicans* (Nº de acceso de Genbank AAL49987) y *P. pastoris* junto con las proteínas Hoc1 de *S. cerevisiae* (Neiman y col (1997) Genetics 145 (3): 637-45) y *K. lactis* (base de datos PENDANT EST) que son manosiltransferasas relacionadas pero distintas. Las regiones de alta homología que eran comunes entre homólogos de *Och1p* pero distintos de los homólogos de *Hoc1p* se usaron para diseñar pares de cebadores degenerados que se dirigieron contra ADN genómico de la cepa de *K. lactis* MG1/2 (Bianchi y col (1987) Current Genetics 12: 185-192). La amplificación por PCR con cebadores RCD33 (CCAGAAGAATTCAATTTGTCARTGG) (SEC ID Nº: 74) y RCD34 (CAGTGAAAATACCTGGNCCNGTCCA) (SEC ID Nº: 75) dio como resultado un producto de 302 pb que se clonó y secuenció y se demostró que la traducción predicha tenía un alto grado de homología con la proteínas de *Och1* (>55% con *Och1p* de *S. cerevisiae*).

25 El producto de PCR de 302 pb se usó para sondar una transferencia de Southern de ADN genómico de la cepa de *K. lactis* (MG1/2) con alta rigurosidad (Sambrook y col., Molecular Cloning: A Laboratory Manual, 2a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Se observó hibridación en un patrón coherente con un único gen que indica que este segmento de 302 pb se corresponde a una porción del genoma de *K. lactis* y que *K. lactis* (*KIOCHI*) contiene una única copia del gen. Para clonar todo el gen *KIOCHI*, la transferencia de Southern se usó para mapear el locus genómico. Por consiguiente, un fragmento de 5,2 kb de *BamHI/PstI* se clonó digiriendo ADN genómico y ligando esos fragmentos en el intervalo de 5,2 kb en pUC19 (New England Biolabs, Beverly, MA) para crear una biblioteca subgenómica de *K. lactis*. Esta biblioteca subgenómica se introdujo por transformación en *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 predicho de *KIOCHI* se secuenció y una fase de lectura abierta de 1362 pb que codificaba una proteína predicha, es decir, idéntica en el 46,5% con el gen *OCHI* de *S. cerevisiae*. La secuencia de 5,2 kb se usó para preparar cebadores para la construcción de un alelo de delección *och1::KAN<sup>R</sup>* usando un procedimiento de solapamiento de PCR (Davidson y col. (2002) Microbiol. 148(Pt8): 2607-15). Este alelo de delección se introdujo por transformación en dos cepas de *K. lactis* y se seleccionaron las colonias resistentes a G418. Estas colonias se exploraron tanto por PCR como por sensibilidad a temperatura para obtener una cepa a la que se había delecionado la ORF de *OCHI*. Los resultados del experimento muestran cepas que ponen de manifiesto un patrón de PCR mutante, que se caracterizaron por análisis de desarrollo a diversas temperaturas y análisis de hidratos de carbono de *N*-glucano de proteínas secretadas y de pared celular después de la digestión con PNGasa. La mutación de *och1* otorgó una sensibilidad a temperatura que permitió que las cepas se desarrollaran a 30°C pero no a 35°C. La Figura 12A muestra un análisis de MALDI-TOF de una cepa de *K. lactis* de tipo silvestre que produce *N*-glucanos de Man<sub>8</sub>GlcNAc<sub>2</sub> [c] y más.

*Identificación, Clonación y Alteración del gen MNN1 de K. lactis*

50 El *MNN1* de *S. cerevisiae* es el gen estructural de la  $\alpha$ -1,3-manosiltransferasa de Golgi. El producto de *MNN1* es una proteína de membrana de tipo II de 762 aminoácidos (Yip y col. (1994) Proc Natl Acad Sci USA. 91 (7): 2723-7). Tanto los oligosacáridos ligados a N como los ligados a O aislados de mutantes de *mnn1* carecen de enlaces  $\alpha$ -1,3-manosa (Raschke y col. (1973) J Biol Chem. 248(13):4660-66).

55 La secuencia de *Mnn1p* de *S. cerevisiae* se usó para buscar las secuencias genómicas traducidas de *K. lactis* (PENDANT). Se identificó una secuencia de ADN de 405 pb que codifica un supuesto fragmento de proteína de similitud significativa con *Mnn1p*. Un segmento interno de esta secuencia se amplificó posteriormente por PCR con los cebadores KMN1 (TGCCATCTTT-TAGGTCCAGGCCCGTTC) (SEC ID Nº: 76) y KMN2 (GATCCCACGACG CATCGTATTTCTTTC), (SEC ID Nº: 77) y se usó para sondar 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 de 4,2 kb de *BamHI-PstI* generando una biblioteca seleccionada por tamaño como se describe en este documento. Un único clon que contenía el gen *MNN1* de *K. lactis* se identificó por PCR de colonias completas usando los cebadores *KMN1* y *KMN2* y se secuenció. Dentro de este clon se identificó una ORF de 2241 pb que codificaba una proteína predicha que era idéntica en el 34% con el gen *MNN1* de *S. cerevisiae*. Los cebadores se diseñaron para la construcción de un alelo de delección de *mnn1::NAT<sup>R</sup>* usando el procedimiento de solapamiento de PCR (Davidson y col. (2002) Microbiol. 148 (Pt 8): 2607-15).

Este alelo de alteración se introdujo por transformación en una cepa de *K. lactis* por electroporación y se seleccionaron transformantes resistentes a nourseotricina y se amplificaron por PCR para inserción homóloga del alelo de

## ES 2 335 345 T3

alteración. Las cepas que ponen de manifiesto un patrón de PCR mutante se pueden someter a análisis de hidratos de carbono de *N*-glucano de un gen indicador conocido.

La Figura 12B ilustra los *N*-glucanos de la cepa de delección de *och11 mm1* de *K. lactis* observada después de la digestión con PNGasa por MALDI-TOF como se describe en este documento. El pico predominante en 1908 (m/z) indicó que [d] es coherente con la masa de Man<sub>9</sub>GlcNAc<sub>2</sub>.

Se describen procedimientos y reactivos adicionales que se pueden usar en los procedimientos para modificar la glucosilación en la bibliografía, tal como la Patente de Estados Unidos N° 5.955.422, Patente de Estados Unidos N° 4.775.622, Patente de Estados Unidos N° 6.017.743, Patente de Estados Unidos N° 4.925.796, Patente de Estados Unidos N° 5.766.910, Patente de Estados Unidos N° 5.834.251, Patente de Estados Unidos N° 5.910.570, Patente de Estados Unidos N° 5.849.904, Patente de Estados Unidos N° 5.955.347, Patente de Estados Unidos N° 5.962.294, Patente de Estados Unidos N° 5.135.854, Patente de Estados Unidos N° 4.935.349, Patente de Estados Unidos N° 5.707.828 y Patente de Estados Unidos N° 5.047.335. Se pueden obtener sistemas de expresión en levadura apropiados de fuentes tales como la Colección Americana de Cultivos Tipo, Rockville, MD. Los vectores están disponibles en el mercado en una diversidad de fuentes.

### Ejemplo 10

#### Identificación, clonación y delección del gen *ALG3* en *P. pastoris* y *K. lactis*

Se generaron cebadores degenerados basándose en un alineamiento de secuencias de proteínas Alg3 de *S. cerevisiae*, *H. sapiens* y *D. melanogaster* y se usaron para amplificar un producto de 83 pb del ADN genómico de *P. pastoris*: 5'-GGTGTGTTTTGTTTTCTAGATCTTTGCAITAYCARTT-3' (SEC ID N°: 78) y 5'-AGAATTTGGTGGGTAAGAATTCCARCACCAYTCRTG-3' (SEC ID N°: 79). El producto de PCR resultante se clonó en el vector pCR2.1 (Invitrogen, Carlsbad, CA) y el análisis de secuencia puso de manifiesto homología con homólogos conocidos de *ALG3/RHK1/NOT56* (Genbank NC\_001134.2, AF309689, NC\_003424.1). Posteriormente, 1929 pb cadena arriba y 2738 pb cadena abajo del producto de PCR inicial se amplificaron de una genoteca de ADN genómico de *P. pastoris* (Boehm (1999) Yeast 15 (7): 563-72) usando los oligonucleótidos internos 5'-CCTAAGCTGGTATGCGTTCTCTTTGCCATATC-3' (SEC ID N°: 80) y 5'-GCGG CATAACAATAATAGATG CTATAAAG-3' (SEC ID N°: 81) junto con T3 (5'-AATTAACCCTCACTAAAGGG-3') (SEC ID N°: 49) y T7 (5'-GTAATACGACTCACTATAGGGC-3') (SEC ID N°: 48) (Integrated DNA Technologies, Coralville, IA) en la cadena principal del plásmido lambda ZAP II que lleva la biblioteca (Stratagene, La Jolla, CA). Los fragmentos resultantes se clonaron en el vector pCR2.1-TOPO (Invitrogen) y se secuenciaron. A partir de esta secuencia se identificó una ORF de 1395 pb que codifica una proteína con una identidad del 35% y una similitud del 53% con el gen *ALG3* de *S. cerevisiae* (usando programas BLAST). El gen se denominó *PpALG3*.

La secuencia de *PpALG3* se usó para crear un conjunto de cebadores para generar una construcción de delección del gen *PpALG3* por solapamiento de PCR (Davidson y col (2002) Microbiol. 148(Pt 8): 2607-15). Los siguientes cebadores se usaron para amplificar regiones de 1 kb 5' y 3' de la ORF de *PpALG3* y el gen KAN<sup>R</sup>, respectivamente:

RCD142 (5'-CCACATCATCCGTGCTACATATAG-3') (SEC ID N°: 82),

**RCD144 (5'-ACGAGGCAAGCTAAACAGATCTCGAAGTATCGAGGGTTAT CCAG-3')**  
(SEC ID N°: 83),

**RCD145 (5'-CCATCCAGTGTGCGAAAACGAGCCAATGGTTCATGTCTATAAATC-3')**  
(SEC ID N°: 84),

RCD147 (5'-AGCCTCAGCGCCAACAAGCGATGG-3') (SEC ID N°: 85),

**RCD143 (5'-CTGGATAACCCTCGATACTTCGAGATCTGTTTAGCTTGCCCTCGT-3')**  
(SEC ID N°: 86)

y

**RCD146 (5'-GATTTATAGACATGAACCATTGGCTCGTTTTGACACTGGATGG-3')**  
(SEC ID N°: 87).

## ES 2 335 345 T3

Posteriormente, los cebadores RCD 142 y RCD 147 se usaron para solapar los tres productos de PCR resultantes en un único alelo de delección de 3,6 kb *alg3::KAN<sup>R</sup>*.

### Identificación, clonación y delección del gen *ALG3* en *K. lactis*

5 Las secuencias de *ALG3p* de *S. cerevisiae*, *Drosophila melanogaster*, *Homo sapiens* etc., se alinearon con secuencias de *K. lactis* (base de datos PENDANT EST). Las regiones de alta homología que eran homólogos en común pero distintos en la secuencia exacta de los homólogos se usaron para crear pares de cebadores degenerados que se dirigieron contra ADN genómico de la cepa de *K. lactis* MG1/2 (Bianchi y col., 1987). En el caso de *ALG3*, la amplificación por PCR con los cebadores KAL-1 (5'-ATCCTTTACCGATGCTGTAT-3') (SEC ID N°: 88) y KAL-2 (5'-ATAACAG TATGTGTTACACGCGTGAG-3') (SEC ID N°: 89) dio como resultado un producto que se clonó y secuenció y se demostró que la traducción predicha tenía un alto grado de homología con proteínas de Alg3p (>50% con Alg3p de *S. cerevisiae*).

15 El producto de PCR se usó para sondar una transferencia de Southern de ADN genómico de la cepa de *K. lactis* (MG1/2) con alta rigurosidad (Sambrook y col., Molecular Cloning: A Laboratory Manual, 2a ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Se observó hibridación en un patrón coherente con un único gen. Esta transferencia de Southern se usó para mapear los loci genómicos. Los fragmentos genómicos se clonaron digiriendo ADN genómico y ligando esos fragmentos en el intervalo de tamaño apropiado en pUC19 para crear una biblioteca subgenómica de *K. lactis*. Esta biblioteca subgenómica se introdujo por transformación en *E. coli* y se ensayaron varios cientos de clones por PCR de colonias, usando los cebadores KAL-1 y KAL-2. Los clones que contenían los genes predichos *KIALG3* y *KIALG61* se secuenciaron y se identificaron las fases de lectura abiertas.

25 Los cebadores para la construcción de un alelo de delección *alg3::NAT<sup>R</sup>*, usando un procedimiento de solapamiento de PCR (Davidson y col. (2002) Microbiol. 148(Pt8): 2607-15), se diseñaron y el alelo de delección resultante se introdujo por transformación en dos cepas de *K. lactis* y se seleccionaron las colonias resistentes a NAT. Estas colonias se exploraron por PCR y se obtuvieron transformantes en los que la ORF de *ALG3* se sustituyó con el alelo mutante *och1::NAT<sup>R</sup>*.

### 30 Ejemplo 11

#### Generación de una cepa mutante de *alg3 och1* que expresa una $\alpha$ -1,2-manosidasa, *GnTI* y *GnTII* para la producción de una glucoproteína de tipo humano

35 Se generó una construcción de delección de *alg3::KAN<sup>R</sup>* de *P. pastoris* como se ha descrito en el Ejemplo 10. Aproximadamente 5  $\mu$ g del producto de PCR resultante se introdujeron por transformación en la cepa **PBP-3** (véase el Ejemplo 3) y se seleccionaron colonias en medio YPD que contenía 200  $\mu$ g/ml de G418. Se confirmó que una cepa de 20 exploradas por PCR contenía la integración correcta del alelo mutante de *alg3::KAN<sup>R</sup>* y que carecía del alelo de tipo silvestre. Esta cepa se denominó **RDP27** (Figura 36).

45 Después se generó una biblioteca de construcciones de *GnTII*, que comprendía fusiones en fase de la biblioteca líder con los dominios catalíticos de los genes de *GnTII* 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 de *NSP<sup>R</sup>* que otorga resistencia al fármaco nourseotricina. Los plásmidos de biblioteca se linealizaron con *EcoRI*, se introdujeron por transformación en la cepa **RDP27** por electroporación y las cepas resultantes se exploraron por análisis de los glucanos liberados de K3 purificada. Una cepa de *P. pastoris* que expresa la *GnTII* de rata fusionada en fase con la construcción *MNN9* (s) de *S. cerevisiae* se denominó **PBP6-5** (Figura 36).

#### 50 Generación de construcciones de expresión de *GnTI*

La construcción de un vector de expresión de *GnTI* (pNA15) que contenía un gen de *GnTI* humana fusionado con la parte N-terminal del gen *MNN9* de *S. cerevisiae* se describe en Choi y col. (2003) Proc Natl Acad Sci U S A. 100(9): 5022-27). De un modo similar se clonó el gen de *GnTII* de rata. El gen de *GnTII* de rata (número de acceso de GenBank U21662) se amplificó por PCR usando la polimerasa Takara *EX Taq<sup>TM</sup>* (Panvera) de genoteca de ADNc de hígado de rata (Clontech) con los cebadores RAT1 (5'-TTCCTCACTGCAGTCTTCTATAACT-3') (SEC ID N°: 90) y RAT2 (5'-TGGAGACCATGAGGTTCCGCATCTAC-3') (SEC ID N°: 91). El producto de PCR después se clonó en el vector pCR2.1-TOPO (Invitrogen) y se secuenció. Usando este vector como un molde, el fragmento *AscI-PacI* de *GnTII*, que codifica los aminoácidos 88-443, se amplificó con polimerasa Pfu Turbo (Stratagene) y los cebadores,

**RAT44 (5'-TTGGCGCGCCTCCCTAGTGTACCAGTTGAACTTTG-3') (SEC ID N°: 92)**

65 y

**RAT11 (5'-GATTAATTAACTCACTGCAGTCTTCTATAACT-3') (SEC ID N°: 93)**

## ES 2 335 345 T3

respectivamente, los sitios de restricción de *AscI* y *PacI* introducidos están subrayados). Después de la confirmación por secuenciación, el dominio catalítico de GnTII de rata se clonó cadena abajo del promotor *PMAI* como un fragmento de *AscI-PacI* en pBP124. En la etapa final, el fragmento génico que codifica la señal de localización de *Mnn2* de *S. cerevisiae* se clonó de pJN281 como fragmento de *NotI-AscI* para generar una fusión en fase con el dominio catalítico de GnTII, para generar el plásmido pTC53.

### Ejemplo 12

#### 10 *Clonación y Expresión De GnTIII Para Producir GlcNAc de Bisección que Refuerzan la Funcionalidad de Anticuerpo*

La adición de una *N*-acetilglucosamina a la estructura de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  por *N*-acetilglucosaminiltransferasas III produce un denominado *N*-glucano bisechado (véase la Figura 15). Esta estructura se ha implicado en una mayor citotoxicidad celular dependiente de anticuerpos (ADCC) (Umana y col. (1999) Nat. Biotechnol. 17(2): 176-80).

Una célula huésped tal como una cepa de levadura capaz de producir glucoproteínas con *N*-glucanos bisechados se modifica de acuerdo con la invención introduciendo en la célula huésped una actividad de GnTIII. Preferiblemente, la célula huésped se transforma con un ácido nucleico que codifica GnTIII (por ejemplo, una de mamífero tal como la GnTIII murina mostrada en la Figura 24) o un dominio del mismo que tiene actividad enzimática, fusionado opcionalmente con un péptido de dirección señal celular heterólogo (por ejemplo, usando las bibliotecas y procedimientos asociados de la invención).

Las IgG están constituidas por dos cadenas pesadas ( $V_H$ ,  $C_{H1}$ ,  $C_{H2}$  y  $C_{H3}$  en la Figura 22), interconectadas en la región bisagra mediante tres puentes disulfuro, y dos cadenas ligeras ( $V_L$ ,  $C_L$  en la Figura 22). Las cadenas ligeras (dominios  $V_L$  y  $C_L$ ) están unidas por otro puente disulfuro con la porción  $C_{H1}$  de la cadena pesada y junto con el fragmento  $C_{H1}$  y  $V_H$  establecen la denominada región Fab. Los antígenos se unen a la porción terminal de la región Fab. La región Fc de IgG está constituida por la región  $C_{H3}$ , la  $C_{H2}$  y la bisagra y es responsable de ejercer las denominadas funciones efectoras (véase más adelante).

La función principal de anticuerpos es la unión a un antígeno. Sin embargo, a menos que la unión a antígeno inactive directamente el antígeno (tal como en el caso de toxinas bacterianas), la mera unión es insignificativa a menos que se desencadenen las denominadas funciones efectoras. Los anticuerpos de la subclase IgG ejercen dos funciones efectoras principales: la activación del sistema de complemento y la inducción de fagocitosis. El sistema de complemento está constituido por un grupo complejo de proteínas séricas implicadas en el control de acontecimientos inflamatorios, en la activación de fagocitos y en la destrucción lítica de membranas celulares. La activación del complemento comienza con la unión del complejo C1 a la porción Fc de dos IgG en estrecha proximidad. El C1 está constituido por una molécula, C1q y dos moléculas C1e y C1s. La fagocitosis se inicia mediante una interacción entre el fragmento Fc de IgG y los receptores Fc-gamma (FcγRI, II y III en la Figura 22). Los receptores Fc se expresan principalmente sobre la superficie de células efectoras del sistema inmune, en particular, macrófagos, monocitos, células mieloides y células dendríticas.

La porción  $C_{H2}$  aloja un sitio de *N*-glucosilación conservado en asparagina 297 (Asn297). Los *N*-glucanos de Asn297 son altamente heterogéneos y se conoce que afectan a la unión de receptor Fc y la activación del complemento. Solamente una minoría (es decir, aproximadamente el 15-20%) de las IgG llevan un *N*-glucano disialilado y el 3-10% tienen uno monosialilado (revisado en Jefferis (2001) Biopharm. 14:19-26). De forma interesante, la estructura mínima de *N*-glucano que se muestra que es necesaria para anticuerpos completamente funcionales capaces de activación de complemento y unión de receptor Fc es un pentasacárido con restos de *N*-acetilglucosamina terminales ( $\text{GlcNAc}_2\text{Man}_3$ ) (revisado en Jefferis, R., Glycosylation of human IgG Antibodies. BioPharm, 2001). Los anticuerpos con menos de un *N*-glucano  $\text{GlcNAc}_2\text{Man}_3$  o sin *N*-glucosilación en Asn297 todavía pueden ser capaces de unirse a un antígeno pero, muy probablemente, no activarán los acontecimientos cruciales aguas abajo tales como fagocitos y activación de complemento. Además, los anticuerpos con *N*-glucanos de tipo fúngico unidos a Asn297, con toda probabilidad, provocarán una respuesta inmune en un organismo mamífero que convertirá ese anticuerpo en inútil como una glucoproteína terapéutica.

#### Clonación Y Expresión De GnTIII

El fragmento de ADN que codifica parte de la proteína GnTIII de ratón que carece del dominio TM se amplifica por PCR a partir de ADN genómico murino (u otro mamífero) usando los cebadores

**directo (5'-TCCTGGCGCGCCTTCCCGAGAGAACTGGCCTCCCTC-3') (SEC ID N°: 94)**

**e inverso (5'-AATTAATTAACCCTAGCCCTCCGCTGTATCCAACCTTG-3') (SEC ID N°: 95)**

Esos cebadores incluyen sitios de restricción *AscI* y *PacI* que se pueden usar para clonar en el vector adecuado para la fusión con la biblioteca líder.

Las secuencias de ácido nucleico (SEC ID N°: 45) y de aminoácidos (SEC ID N°: 46) de GnTIII murina se muestran en la Figura 24.

#### *Clonación de Secuencias que Codifican Inmunoglobulina*

Se han publicado protocolos para la clonación de las regiones variables de anticuerpos, incluyendo secuencias de cebador, previamente. Las fuentes de anticuerpos y genes codificantes pueden ser, entre otras, células B humanas inmunizadas *in vitro* (véase, por ejemplo, Borreback y col. (1988) Proc. Natl. Acad. Sci. USA 85: 3995-3999), linfocitos de sangre periférica o células B humanas únicas (véase, por ejemplo, Lagerkvist y col. (1995) Biotechniques 18: 862-869; y Temess y col. (1997) Hum. Immunol. 56: 17-27) y ratones transgénicos que contienen loci de inmunoglobulina humana, que permiten la creación de líneas celulares de hibridoma.

Usando técnicas de ADN recombinante convencionales, se pueden clonar secuencias de ácido nucleico que codifican anticuerpo. Las fuentes de información genética que codifican inmunoglobulinas de interés son típicamente preparaciones de ARN total de células de interés, tales como linfocitos sanguíneos o líneas celulares de hibridoma. Por ejemplo, empleando un protocolo basado en PCR con cebadores específicos, se pueden clonar regiones variables mediante transcripción inversa iniciada a partir de un cebador específico de secuencia que hibrida con el sitio del dominio C<sub>H</sub>1 de IgG y un segundo cebador que codifica los aminoácidos 111-118 de la región constante kappa murina. Después, los ADNc que codifican V<sub>H</sub> y V<sub>K</sub> se pueden amplificar como se ha publicado previamente (véase, por ejemplo, Graziano y col. (1995) J Immunol. 155(10): 4996-5002; Welschof y col. (1995) J. Immunol. Methods 179: 203-214; y Orlandi y col. (1988) Proc. Natl. Acad. Sci. USA 86: 3833). También se han publicado procedimientos de clonación para inmunoglobulinas completas (cadenas pesada y ligera) (véase, por ejemplo, Buckel y col. (1987) Gene 51: 13-19; Recinos y col. (1994) Gene 149: 385-386; Recinos y col. (1995) Gene 158: 311-12). Se han descrito protocolos adicionales para la clonación y generación de fragmentos de anticuerpo y construcciones de expresión de anticuerpo en Antibody Engineering, Kontermann y Dübel (2001), Eds., Springer Verlag: Berlin Heidelberg New York.

Se han descrito plásmidos de expresión fúngicos que codifican la cadena pesada y ligera de inmunoglobulinas (véase, por ejemplo, Abdel-Salam y col. (2001) Appl. Microbiol. Biotechnol. 56: 157-164; y Ogunjimi y col. (1999) Biotechnology Letters 21: 561-567). Por tanto, se pueden generar plásmidos de expresión que alojan las regiones constantes de inmunoglobulinas. Para facilitar la clonación de regiones variables en estos vectores de expresión, se pueden poner sitios de restricción adecuados en estrecha proximidad con los extremos de las regiones variables. Las regiones constantes se pueden construir de tal manera que las regiones variables se puedan fusionar de forma sencilla en fase con las mismas por un simple experimento de digestión por restricción/ligación. La Figura 23 muestra una revisión esquemática de tal construcción de expresión, diseñada de un modo muy modular, que permite el intercambio sencillo de promotores, terminadores de la transcripción, dominios de dirección de integración e incluso marcadores de selección.

Como se muestra en la Figura 23, los dominios V<sub>L</sub> así como V<sub>H</sub> de elección se pueden clonar de forma sencilla en fase con las regiones C<sub>L</sub> y C<sub>H</sub>, respectivamente. La integración inicial se dirige al locus de AOX de *P. pastoris* (o un locus homólogo en otra célula fúngica) y el promotor AOX inducible por metanol dirigirá la expresión. Alternativamente se puede usar cualquier otro casete de promotor constitutivo o inducible deseado. Por tanto, si se desea, las regiones AOX 5' y AOX 3' así como los fragmentos de terminador de la transcripción (TT) se pueden sustituir de forma sencilla con diferentes dominios de TT, promotor y de dirección de integración para optimizar la expresión. Inicialmente, la señal de secreción de factor alfa con el sitio de proteasa KEX convencional se emplea para facilitar la secreción de cadenas pesadas y ligeras. Las propiedades del vector de expresión se pueden refinar adicionalmente usando técnicas convencionales.

Un vector de expresión de Ig tal como el que se ha descrito anteriormente se introduce en una célula huésped de la invención que expresa GnTIII, preferiblemente en el aparato de Golgi de la célula huésped. Las moléculas de Ig expresadas en tal célula huésped comprenden *N*-glucanos que tienen GlcNAc de bisección.

#### Ejemplo 13

##### *Generación de la Cepa de Levadura YSH-1 ( $\Delta$ och1, $\alpha$ 1,2-manosidasa, GnTI)*

La cepa de *P. pastoris* que se ha descrito anteriormente BK64 (Choi y col. (2003) Proc Natl Acad Sci U S A. 100(9): 5022-7), un triple auxótrofo (*ADE*, *ARG*, *HIS*) que posee el knock-out de *OCH1* y que expresa el dominio kringle 3 (K3) de plasminógeno humano se usó como la cepa huésped. BK64 se transformó con el plásmido pPB103 linealizado con la enzima de restricción *EcoNI* para introducir el transportador de UDP-*N*-acetilglucosamina de *K. lactis* en la célula huésped, creando de este modo la cepa **PBP-1**. El MnsI de ratón se introdujo en esta cepa por transformación con el plásmido pFB8 linealizado con la enzima de restricción *EcoNI*, generando la cepa **PBP-2**. El análisis de glucano de K3 de proteínas aisladas de la cepa **PBP-2** demostró que la glucoforma principal presente era Man<sub>5</sub>GlcNAc<sub>2</sub>.

## ES 2 335 345 T3

Posteriormente se transformó **PBP-2** con el plásmido de GnTI humana pNA15 linealizado con la enzima de restricción *AatII*, generando la cepa **PBP-3**. El análisis de las glucoformas de K3 producidas en la cepa **PBP-3** demostró que el glucano híbrido  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  era la estructura predominante. Para recuperar el marcador *URA3* de **PBP-3**, esta cepa se cultivó en YPD antes de la selección en medio mínimo que contenía 5-Fluoroorótico (5-FOA, Bio-Vectra) y uracilo (Boeke y col. (1984) Mol. Gen. Genet. 197: 345-346). La cepa Ura-menos recuperada que producía las glucoformas  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  se denominó **YSH-1** (Figura 36). El perfil de *N*-glucano de la cepa **YSH-1** se muestra en la Figura 25 (parte superior) y presenta un pico predominante en 1465 m/z que se corresponde a la masa de  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d].

10 Ejemplo 14

*Generación de la Cepa de Levadura YSH-37 (P. pastoris que expresa manosidasa II)*

15 **YSH-1** (Ejemplo 13) se transformó con el plásmido de manosidasa  $\text{II}\Delta 74$  de *D. melanogaster/MNN2(s)* de *S. cerevisiae* (pKD53) linealizado con la enzima de restricción *ApaI*, generando la cepa **YSH-37** (Figura 36). El análisis de las estructuras de glucano K3 producidas en la cepa **YSH-37** (Figura 25 (parte inferior)) demostró que la glucoforma predominante en 1140 m/z se corresponde a la masa de  $\text{GlcNAcMan}_3\text{GlcNAc}_2$  [b] y otras glucoformas  $\text{GlcNAcMan}_4\text{GlcNAc}_2$  [c] en 1303 m/z y  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  [d] en 1465 m/z.

20 Ejemplo 15

*Generación de la Cepa de Levadura YSH-44*

25 La cepa **YSH-37** (Ejemplo 14) se transformó con un plásmido que codifica un líder de GnTII de rata/MNN2 (s), pTC53, linealizado con la enzima de restricción *EcoRI*. La cepa resultante, **YSH-44** (Figura 36), produjo un *N*-glucano de K3 que tiene una única glucoforma en 1356 m/z, que se corresponde a la masa de  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  [x], por espectrometría de masas MALDI-TOF en modo positivo (Figura 29).

30 Ejemplo 16

*Construcción del Plásmido pJN 348*

35 El plásmido pBLURA-SX (de Jim Cregg) se digirió con *BamHI* y *BglII* para liberar el casete de expresión *AOX*. El fragmento de *BamHI* que contiene el casete de expresión de *GAPDH/CYC1* de pJN261 (Figura 4B) (Ejemplo 4) después se ligó en la cadena principal de pBLURA-SX para crear pJN338. El plásmido pJN338 se cortó con *NotI* y *PacI* y los dos oligonucleótidos

**5'-GGCCGCTGCAGATTTAAATGAATTCGGCGCGCCTTAAT-3' (SEC ID N°: 96)**

y

45 **5'-TAAGGCGCGCC GAATTCATTTAAATCTGCAGGGC-3' (SEC ID N°: 97)**

que se habían hibridado *in vitro* se ligaron en los sitios abiertos, para crear pJN348.

50 Ejemplo 17

*Construcción de un Plásmido de Integración pRCD259*

55 El *PpURA3* que contiene el vector de expresión de *GAPDH* pJN348 se linealizó con *XhoI* y se crearon extremos romos con ADN polimerasa de T4 y se trató con fosfatasa intestinal de ternero (CIP). El marcador de resistencia *HYG* se digirió de pAG32 con *BglII* y *SacI* y se formaron extremos romos, después se ligó en pJN348 para crear pRCD259 que se puede usar como un vector de expresión de *HYG* que se integra en el locus de *PpURA3*.

60 Ejemplo 18

*Generación de Construcciones de Fusión de GnTIII*

65 Se prepararon construcciones de fusión entre GnTIII de mamífero y secuencias de dirección de levadura usando el gen *Mgat3* de ratón (número de acceso de GenBank L39373, Bhaumik y col., 1995). Tres fragmentos de ADN correspondientes a delecciones *N*-terminales  $\Delta 32$ ,  $\Delta 86$  y  $\Delta 212$  del gen de GnTIII de ratón se amplificaron por PCR usando polimerasa *Pfu Turbo* (Stratagene) con los cebadores directos

**MG3-B (5'-TCCTGGCGCGCCTTCCCGAGAGAACTGGCCTCCCTC-3') (SEC ID N°: 98),**

**MG3-C (5'-CCGAGGGCGCGCCACAGAGGAACTGCACCGGGTG-3') (SEC ID N°: 99),**

**MG3-D (5'-ACCGAGGGCGCGCCATCAACGCCATCAACATCAACCAC-3') (SEC ID N°: 100)**

**MG3-D (5'-ACCGAGGGCGCGCCATCAACGCCATCAACATCAACCAC-3') (SEC ID N°: 100)**

e inverso

**MG3-A (5'-AATTAATTAACCCTAGCCCTCCGCTGTATCCAACCTTG-3') (SEC ID N°: 101).**

Después, los productos de PCR se clonaron en el vector pJN 348 como fragmentos de *AscI-PacI* y se secuenciaron. Los vectores resultantes pVA (GnTIII  $\Delta$ 32), pVB (GnTIII  $\Delta$ 86) y pVC (GnTIII  $\Delta$ 212) se digirieron con enzimas *NotI-AscI* y se usaron para la ligación con la biblioteca líder de levadura (líderes 20-67). Estos péptidos de dirección se fusionan con los dominios catalíticos seleccionados de la GnTIII de ratón con delecciones *N*-terminales de 32, 86, 212 aminoácidos. Por ejemplo, el péptido de dirección de *MNN2* de *S. cerevisiae* (largo, medio y corto) y *GNT1* de *K. lactis* (corto y medio) (véase el Ejemplo 11) se muestran en la Tabla 11

TABLA 11

Una biblioteca combinatoria representativa de secuencias de péptido de dirección/dominios catalíticos que muestran actividad de *UDP-N-Acetilglucosamidiltransferasa III (GnTIII) YSH-1 de P. pastoris*

|                    |   | <b>Péptido de dirección</b>            |  |  |                                    |
|--------------------|---|--|--|--|------------------------------------|
|                    |   | <i>MNN2(s)</i> de <i>S. cerevisiae</i> | <i>MNN2(m)</i> de <i>S. cerevisiae</i> | <i>MNN2(l)</i> de <i>S. cerevisiae</i> | <i>GNT1(m)</i> de <i>K. lactis</i> |
| Dominio Catalítico | <b>GnTIII <math>\Delta</math>32 de Ratón</b>  | 50% (pVA53)                            | 30-40% (pVA54)                         | 20-30% (pVA55)                         | 0% (pVA51)                         |
|                    | <b>GnTIII <math>\Delta</math>86 de Ratón</b>  | 20-30% (pVB53)                         | 30-40% (pVB54)                         | 20-30% (pVB55)                         | 0% (pVB51)                         |
|                    | <b>GnTIII <math>\Delta</math>212 de Ratón</b> | 0% (pVC53)                             | 0% (pVC54)                             | 0% (pVC55)                             | 0% (pVC51)                         |

Ejemplo 19

*Modificación de P. pastoris para producir GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub> bisecado*

La cepa de *P. pastoris* que produce GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> (**PBP-3**) (véase el Ejemplo 8) se contraseleccionó en 5-FOA, seleccionando de este modo para pérdida del marcador Ursa3+ y un fenotipo *ura3-*. Esa cepa, denominada **YSH-1** (Figura 36) se transformó con la biblioteca de dominios catalíticos de *N*-acetilglucosaminiltransferasa III (GnTIII) (vectores pVA, pVB y pVC) y líderes. Los transformantes se cultivaron a 30°C en BMGY hasta una DO<sub>600</sub> de aproximadamente 10, se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor *AOX1*. La K3 se purificó del medio por cromatografía de afinidad de Ni utilizando 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 (Ejemplo 3). Los *N*-glucanos se liberaron por digestión con PNGasa (Ejemplo 3). Los *N*-glucanos se analizaron con una EM de MALDI-TOF (Ejemplo 3). Las actividades de GnTIII se muestran en la Tabla 11. El número de (+), como

## ES 2 335 345 T3

se usa en este documento, indica los niveles relativos de producción de *N*-glucano bisecado del % de glucanos neutros. Las secuencias de péptido de dirección se seleccionaron entre el grupo constituido por: GLS1 de *Saccharomyces*, MNS1 de *Saccharomyces*, SEC12 de *Saccharomyces*, SEC de *Pichia*, OCH1 de *Pichia*, MNN9 de *Saccharomyces*, VAN1 de *Saccharomyces*, ANP1 de *Saccharomyces*, HOC1 de *Saccharomyces*, MNN10 de *Saccharomyces*, MNN11 de *Saccharomyces*, MNT1 de *Saccharomyces*, D2 de *Pichia*, D9 de *Pichia*, J3 de *Pichia*, KTR1 de *Saccharomyces*, KTR2 de *Saccharomyces*, GnTI de *Kluyveromyces*, MNN2 de *Saccharomyces*, MNN5 de *Saccharomyces*, YUR1 de *Saccharomyces*, MNN1 de *Saccharomyces* y MNN6 de *Saccharomyces*. Los transformantes de pVA53 que muestran la GlcNAc de bisección (por ejemplo, GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>) se denominaron PBP26 (Figura 36).

10  
Ejemplo 20

### *Modificación de YSH-44 de P. pastoris para producir GlcNAc<sub>3</sub>Ma<sub>3</sub>GlcNAc<sub>2</sub> bisecado*

15 Para la expresión de GnTIII en la cepa **YSH-44** (Figura 36), se transfirieron construcciones de GnTIII de los vectores pVA53, pVB53, pVA54 y pVB54 como fragmentos de *NotI-PacI* en pRCD259 para generar vectores pPB135, pPB137, pPB136 y pPB138. Los vectores contienen un marcador de resistencia a HYG y un gen de *URA3* de *P. pastoris* como una secuencia de dirección para la integración genómica. Los plásmidos se linealizan con *Sall*, se introducen por transformación en la cepa YSH-44 por electroporación, se seleccionan en medio que contiene higromicina y las cepas resultantes se exploran por análisis de los glucanos liberados de K3 purificada. Los transformantes se cultivaron a 24°C en BMGY hasta una DO600 de aproximadamente 10, se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor *AOX1*. Se purificó K3 del medio por cromatografía de afinidad de Ni utilizando un formato de 96 pocillos en un robot de laboratorio Beckman BioMek 2000 (Ejemplo 3). La purificación robótica es una adaptación del protocolo proporcionado por Novagen para su resina HisBind (Ejemplo 3). Los *N*-glucanos se liberaron por digestión con PNGasa. Los *N*-glucanos se analizaron con una EM de MALDI-TOF (Ejemplo 3). Los transformantes de pPB135 que muestran la GlcNAc de bisección (por ejemplo, GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>) se denominaron **YSH-57** (Figura 36). La Tabla 11 ilustra la actividad de la GnTIII de ratón.

30  
Ejemplo 21

### *Modificación de PBP6-5 de P. pastoris para producir GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub> bisecado*

35 El **PBP6-5** de *P. pastoris* (Ejemplo 11) se transformó con el plásmido pPB135 (Tabla 11) que codifica un dominio catalítico de GnTIII de ratón ( $\Delta 32$ ) ligado en fase con un péptido de dirección procedente de *MNN2* de *S. cerevisiae*. Los transformantes se cultivaron a 30°C en BMGY hasta una DO600 de aproximadamente 10, se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor *AOX1*. Se purificó K3 del medio por cromatografía de afinidad de Ni utilizando 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 (Ejemplo 3). Los *N*-glucanos se liberaron por digestión con PNGasa (Ejemplo 3). Los *N*-glucanos se analizaron con una EM de MALDI-TOF (Ejemplo 3). Los transformantes que muestran la GlcNAc de bisección (por ejemplo, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>) se denominaron **PBP-38** (Figura 36). La Tabla 11 ilustra la actividad de la GnTIII de ratón.

45  
Ejemplo 22

### *Ensayo de actividad de GnTIII in vitro usando el sustrato GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> en la cepa de P. pastoris modificada YSH-57*

50 Para ensayar cualquier actividad de GnTIII *ex vivo* potencial en la cepa de *P. pastoris*, los sobrenadantes de cultivo celular de YSH-57 se ensayaron para actividad de GnTIII. Las células de *P. pastoris* YSH-57 se cultivaron a 24°C en BMGY (B) hasta una DO600 de aproximadamente 10. Las células se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor *AOX1*. Después de 24 horas de inducción, las células se retiraron por centrifugación para producir un sobrenadante esencialmente claro. Una alícuota del sobrenadante se retiró para ensayos de GnTIII y el resto se usó para la recuperación de K3 soluble secretada. Se purificó K3 del medio por cromatografía de afinidad de Ni utilizando 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 (Ejemplo 3). Los *N*-glucanos se liberaron por digestión con PNGasa (Ejemplo 3). La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente para presencia de actividad de GnTIII secretada. Se añadió GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub> purificado de K3 expresada en la cepa **PBP-3** a: BMMY (A) UDP-GlcNAc 1 mM (Sigma Chemical Co., St. Louis, MO) en BMMY (B); el sobrenadante de **YSH-44** transformado con pVA53 [**YSH-57**] (C); el sobrenadante de **YSH-57** + UDP-GlcNAc 1 mM (D). Después de la incubación durante 8 horas a temperatura ambiente, las muestras se analizaron por HPLC de amino sílice para determinar el alcance de la actividad de GnTIII.

## ES 2 335 345 T3

### Ejemplo 23

*Ensayo de actividad de GnTIII in vitro usando el sustrato  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  en la cepa de *P. pastoris* modificada YSH-57*

5

Para ensayar cualquier actividad de GnTIII *ex vivo* potencial en la cepa de *P. pastoris* **YSH-57** se ensayaron sobrenadantes de cultivo celular para actividad de GnTIII. Las células de *P. pastoris* **YSH-57** se cultivaron a 24°C en BMGY hasta una DO600 de aproximadamente 10. Las células se recogieron por centrifugación y se transfirieron a BMMY para inducir la producción de K3 (kringle 3 de plasminógeno humano) bajo el control de un promotor *AOXI*. Después de 24 horas de inducción, las células se retiraron por centrifugación para producir un sobrenadante esencialmente claro. Una alícuota del sobrenadante se retiró para ensayos de GnTIII y el resto se usó para la recuperación de K3 soluble secretada. Se purificó K3 del medio por cromatografía de afinidad de Ni utilizando 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 (Ejemplo 3). Los *N*-glucanos se liberaron por digestión con PNGasa (Ejemplo 3). La alícuota retirada anteriormente del sobrenadante se ensayó adicionalmente para la presencia de actividad de GnTIII secretada. Se añadió  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  purificado de K3 expresada en la cepa **YSH-44** a: BMMY (A) UDP-GlcNAc 1 mM (Sigma Chemical Co., St. Louis, MO) en BMMY (B); el sobrenadante de **YSH-44** transformado con **pVA53 [YSH-57]** (C). Después de incubación durante 8 horas a temperatura ambiente, las muestras se analizaron por HPLC de amino sílice para determinar el alcance de actividad de GnTIII.

20

25

30

35

40

45

50

55

60

65

## REIVINDICACIONES

- 5 1. Un procedimiento para fabricar una glucoproteína recombinante en una célula huésped de levadura u hongo unicelular o multicelular filamentoso de eucariota inferior modificada para producir una glucoproteína que tiene una estructura central de  $\text{Man}_3\text{GlcNAc}_2$  o de  $\text{Man}_5\text{GlcNAc}_2$ , que comprende la etapa de introducir en la célula un ácido nucleico que codifica una actividad de *N*-acetilglucosaminiltransferasa III, en el que dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnTIII  $\Delta 32$  de ratón o GnTIII  $\Delta 86$  de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*.
- 10 2. El procedimiento de la reivindicación 1, en el que las glucoproteínas recombinantes producidas por el procedimiento comprenden más del 10% en mol de estructuras de *N*-glucano bisecadas seleccionadas entre el grupo constituido por  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ ,  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  y  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ .
- 15 3. El procedimiento de la reivindicación 1 ó 2, en el que la actividad de *N*-acetilglucosaminiltransferasa III es sustancialmente intracelular.
- 20 4. El procedimiento de una cualquiera de las reivindicaciones 1 a 3, que comprende además la etapa de aislar la glucoproteína de la célula huésped.
- 25 5. El procedimiento de una cualquiera de las reivindicaciones 1 a 4, en el que la célula huésped se selecciona entre el grupo constituido por *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.*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* y *Neurospora crassa*.
- 30 6. El procedimiento de la reivindicación 5, en el que la célula huésped se selecciona entre el grupo constituido por *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica* y *Pichia sp.*
- 35 7. El procedimiento de la reivindicación 6, en el que la célula huésped es *Pichia pastoris*.
- 40 8. El procedimiento de una cualquiera de las reivindicaciones 1 a 7, en el que la glucoproteína es una proteína terapéutica.
- 45 9. El procedimiento de la reivindicación 8, en el que la proteína terapéutica se selecciona entre el grupo constituido por eritropoyetina, citocinas, factores de coagulación, cadena  $\alpha$  de receptor de IgE soluble, IgG, fragmentos de IgG, IgM, interleucinas, urocinasa, quimasa, inhibidor de tripsina de urea, proteína de unión de IGF, factor de crecimiento epidérmico, factor de liberación de hormona de crecimiento, proteína de fusión de anexina V, angiostatina, factor de crecimiento endotelial vascular 2, factor inhibidor de progenitor mielóide 1, osteoprotegerina,  $\alpha$ -1-antitripsina,  $\alpha$ -fetoproteína, ADNasa II, AAt, rhTBP-1, TACI-Ig, FSH, GM-CSF, GLP-1 con y sin agonista de receptor FC de IL-1, sTNFr ATIII, rhTrombina, glucocerebrosidasa y CTLA4-Ig.
- 50 10. Una célula huésped de levadura u hongo unicelular o multicelular filamentoso que comprende una actividad de *N*-acetilglucosaminiltransferasa III, en la que dicha actividad es de una enzima que comprende el dominio catalítico de *N*-acetilglucosaminiltransferasa III GnTIII  $\Delta 32$  de ratón o GnTIII  $\Delta 86$  de ratón fusionado con un péptido de dirección MNN2 de *Saccharomyces cerevisiae*.
- 55 11. La célula huésped de la reivindicación 10, que comprende además una actividad de *N*-acetilglucosaminiltransferasa I y/o de *N*-acetilglucosaminiltransferasa II.
- 60 12. La célula huésped de la reivindicación 10 u 11, en la que la célula es capaz de producir *N*-glucanos en glucoproteínas que comprenden estructuras de  $\text{GlcNAcMan}_3\text{GlcNAc}_2$ ,  $\text{GlcNAcMan}_5\text{GlcNAc}_2$  o  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  que son capaces de reaccionar con una actividad de GnTIII.
- 65 13. La célula huésped de la reivindicación 10, en la que la célula huésped produce glucoproteínas que comprenden estructuras de *N*-glucano bisecados en una estructura central de  $\text{Man}_5\text{GlcNAc}_2$  o  $\text{Man}_3\text{GlcNAc}_2$ .
14. La célula huésped de la reivindicación 10 u 11, en la que la célula huésped produce glucoproteínas que comprenden estructuras de *N*-glucano bisecado seleccionadas entre el grupo constituido por  $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ ,  $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$  y  $\text{GlcNAc}_2\text{Man}_5\text{GlcNAc}_2$ .
15. La célula huésped de una cualquiera de las reivindicaciones 10 a 14, que comprende además una actividad de  $\alpha$ -1,2-manosidasa I y/o una actividad de manosidasa II.

## ES 2 335 345 T3

16. La célula huésped de una cualquiera de las reivindicaciones 10 a 15, que es deficiente en una actividad de *OCHI* manosiltransferasa y/o actividad de Dol-P-Man:Man<sub>5</sub>GlcNAc<sub>2</sub>-PP-Dol manosiltransferasa.

5 17. La célula huésped de una cualquiera de las reivindicaciones 10 a 16, que comprende además una actividad de transportador de UDP-GlcNAc.

18. La célula huésped de una cualquiera de las reivindicaciones 10 a 17, en la que las actividades son sustancialmente intracelulares.

10 19. La célula huésped de una cualquiera de las reivindicaciones 10 a 18, en la que la célula huésped se selecciona entre el grupo constituido por *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.*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*,  
15 *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum* y *Neurospora crassa*.

20. La célula huésped de la reivindicación 19, en la que la célula huésped se selecciona entre el grupo constituido por *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolic* y  
20 *Pichia sp.*

21. La célula huésped de la reivindicación 20, en la que la célula huésped es *Pichia pastoris*.

22. Una composición de glucoproteína, en la que las glucoproteínas en dicha composición carecen de fucosa y más del 10% en mol de los *N*-glucanos en las glucoproteínas en la composición comprenden una GlcNAc de bisección en una estructura central de Man<sub>5</sub>GlcNAc<sub>2</sub>, Man<sub>4</sub>GlcNAc<sub>2</sub> o Man<sub>3</sub>GlcNAc<sub>2</sub>.

23. La composición de glucoproteína de la reivindicación 22, en la que más del 10% en mol de los *N*-glucanos en las glucoproteínas en la composición comprenden una GlcNAc de bisección unida a una estructura central seleccionada entre el grupo constituido por: GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>4</sub>GlcNAc<sub>2</sub> y GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>.

24. La composición de glucoproteína de la reivindicación 22, en la que más del 80% en mol de los *N*-glucanos en las glucoproteínas en la composición comprenden una GlcNAc de bisección unida a una estructura central seleccionada entre el grupo constituido por: Man<sub>5</sub>GlcNAc<sub>2</sub>, Man<sub>4</sub>GlcNAc<sub>2</sub>, Man<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>5</sub>GlcNAc<sub>2</sub>, GlcNAcMan<sub>4</sub>GlcNAc<sub>2</sub> y GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>.

25. La composición de glucoproteína de la reivindicación 22, en la que las glucoproteínas se preparan por el procedimiento de una cualquiera de las reivindicaciones 1 a 12 y en la que más del 10% en mol de los *N*-glucanos en las glucoproteínas en la composición comprenden una estructura de bisección seleccionada entre el grupo constituido por GlcNAc<sub>3</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAc<sub>2</sub>Man<sub>3</sub>GlcNAc<sub>2</sub>, GlcNAc<sub>2</sub>Man<sub>4</sub>GlcNAc<sub>2</sub> y GlcNAc<sub>2</sub>Man<sub>5</sub>GlcNAc<sub>2</sub>.

26. Una composición farmacéutica que comprende la glucoproteína de una cualquiera de las reivindicaciones 22 a 25.

45

50

55

60

65

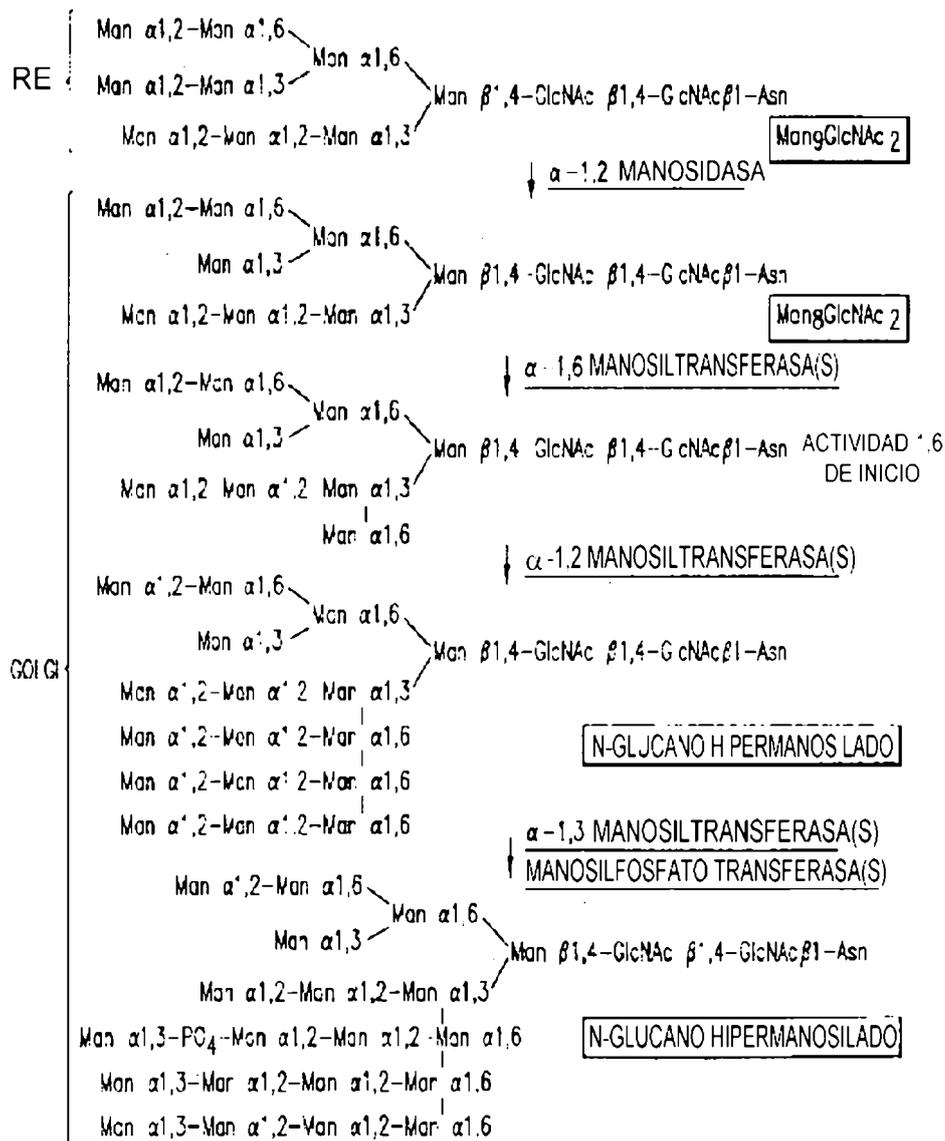


FIG. 1A

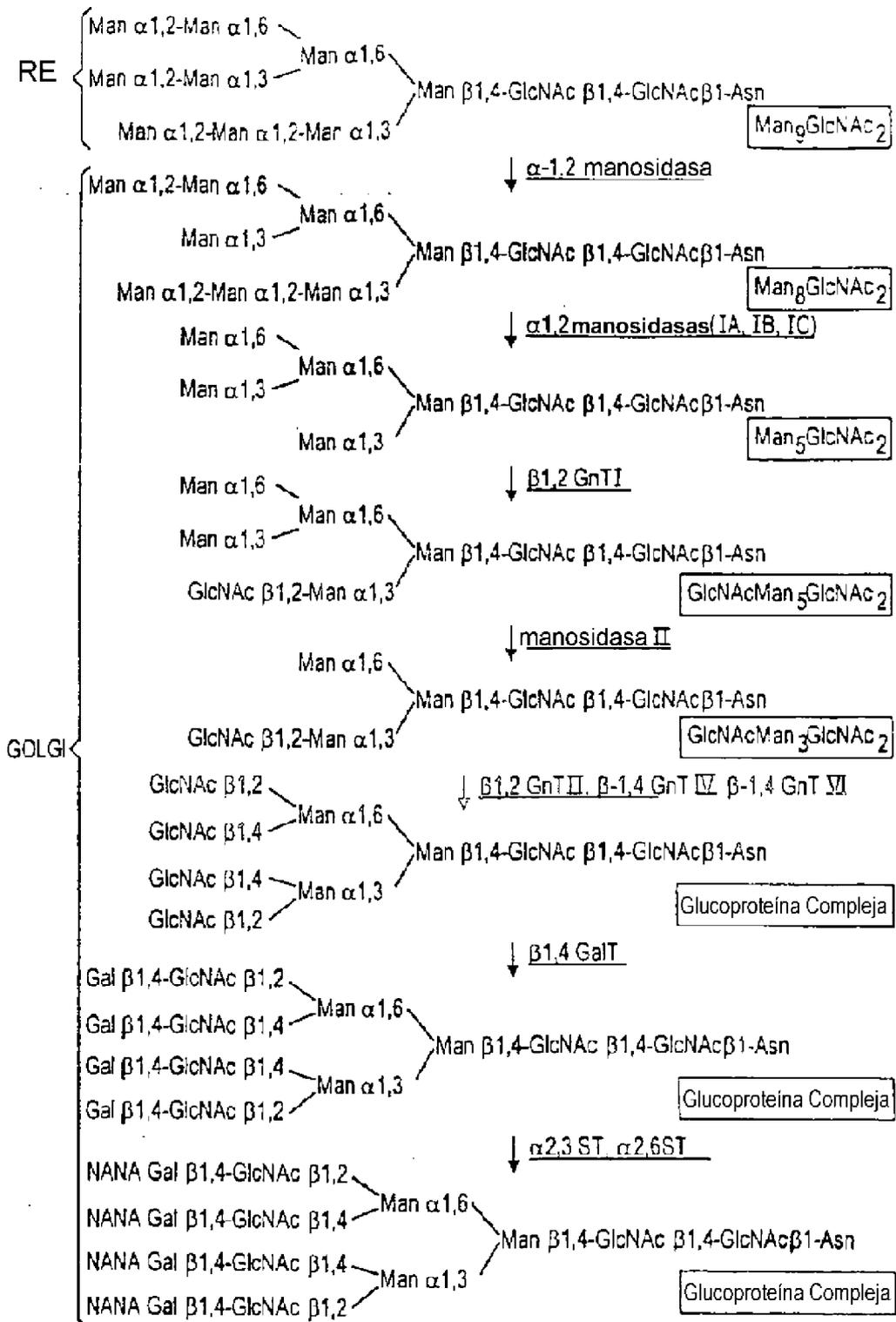


FIG. 1B

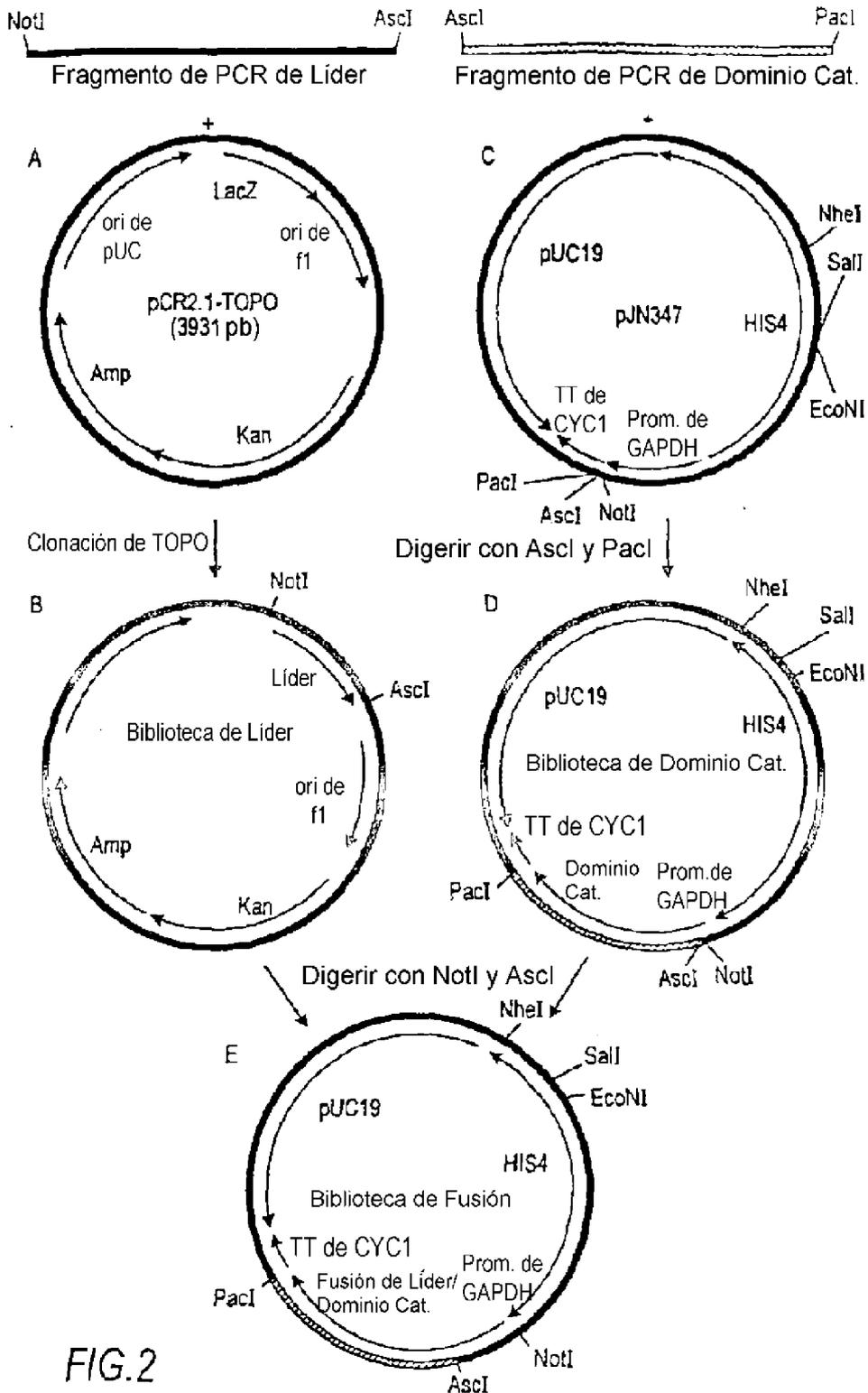


FIG.2





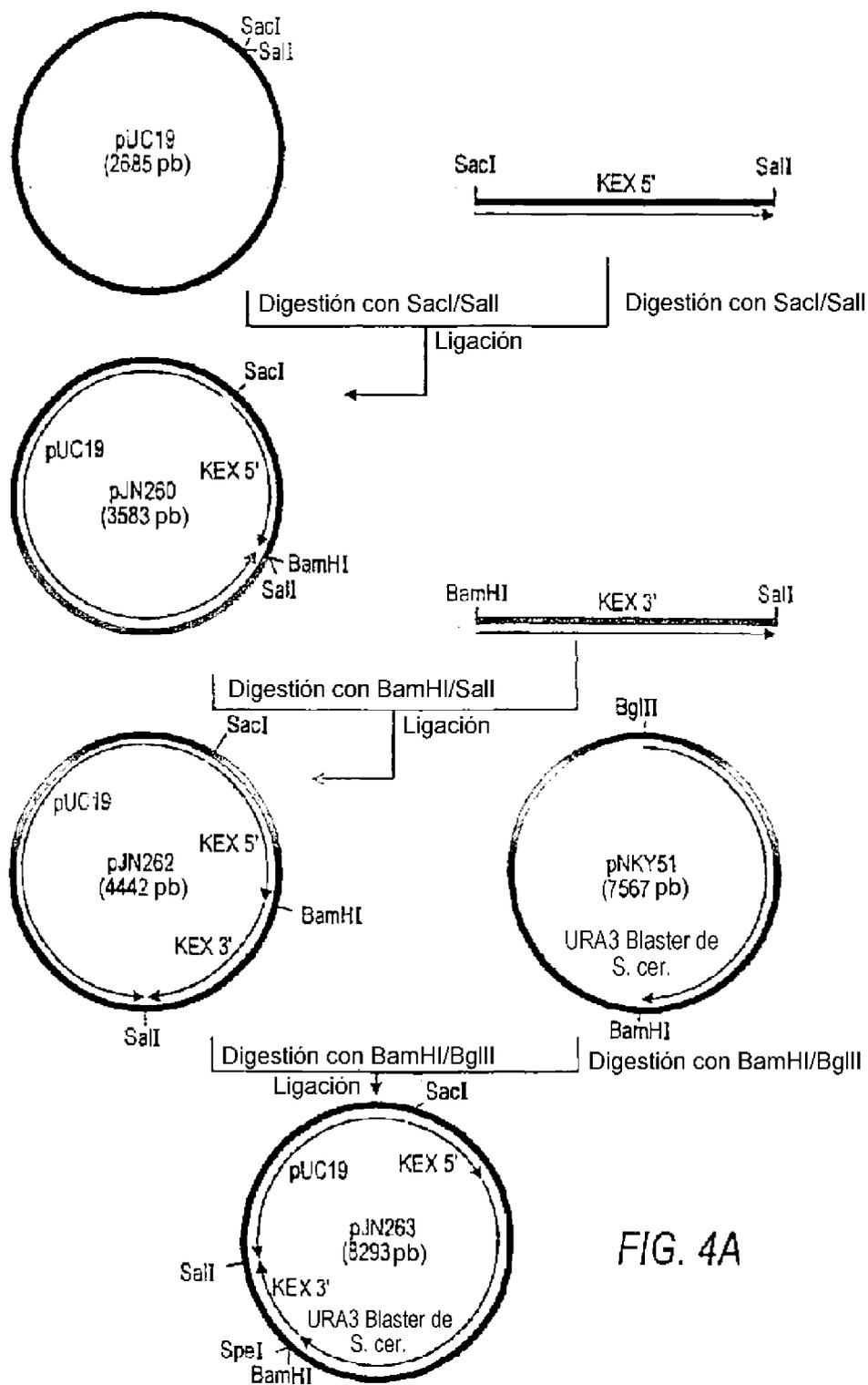


FIG. 4A

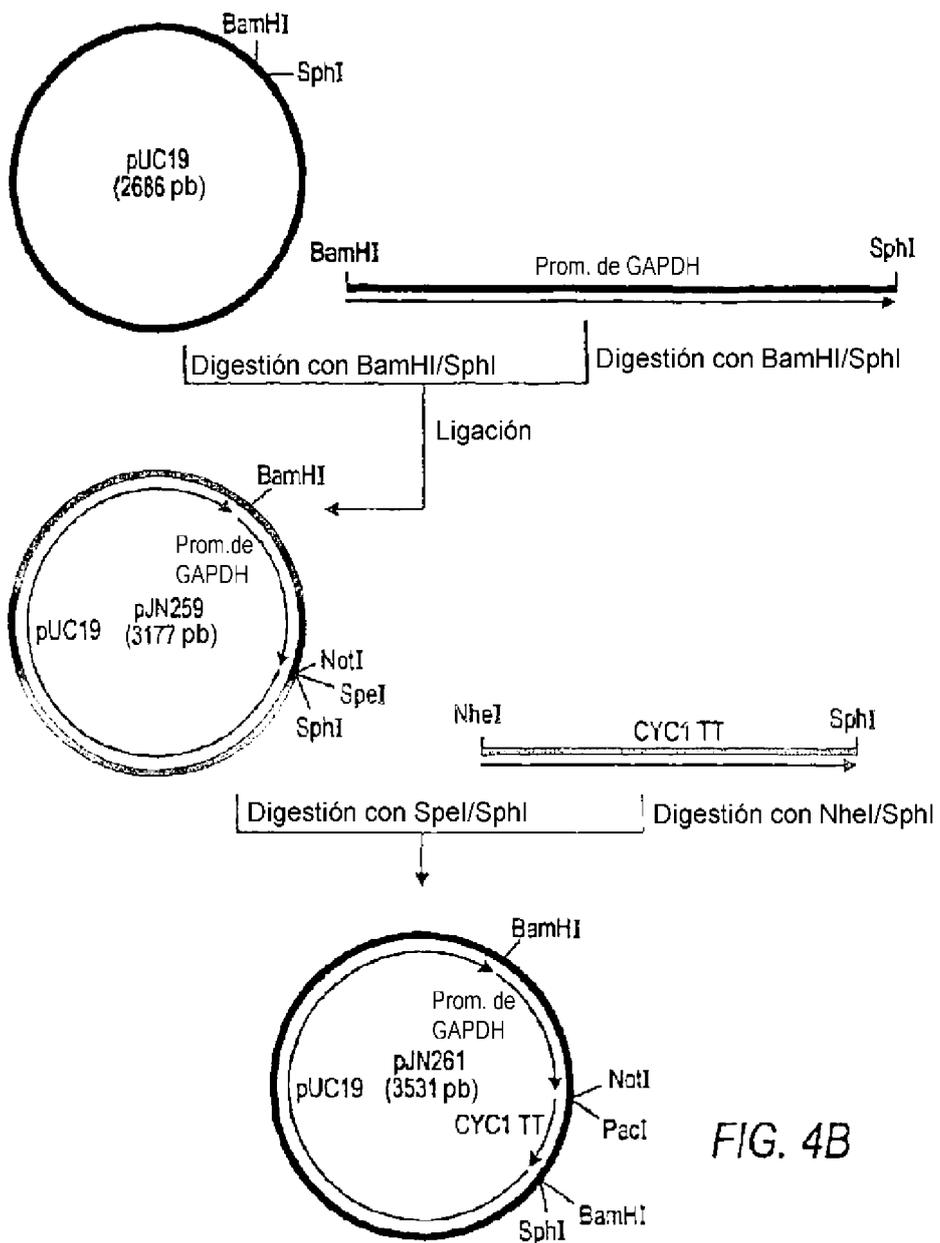


FIG. 4B

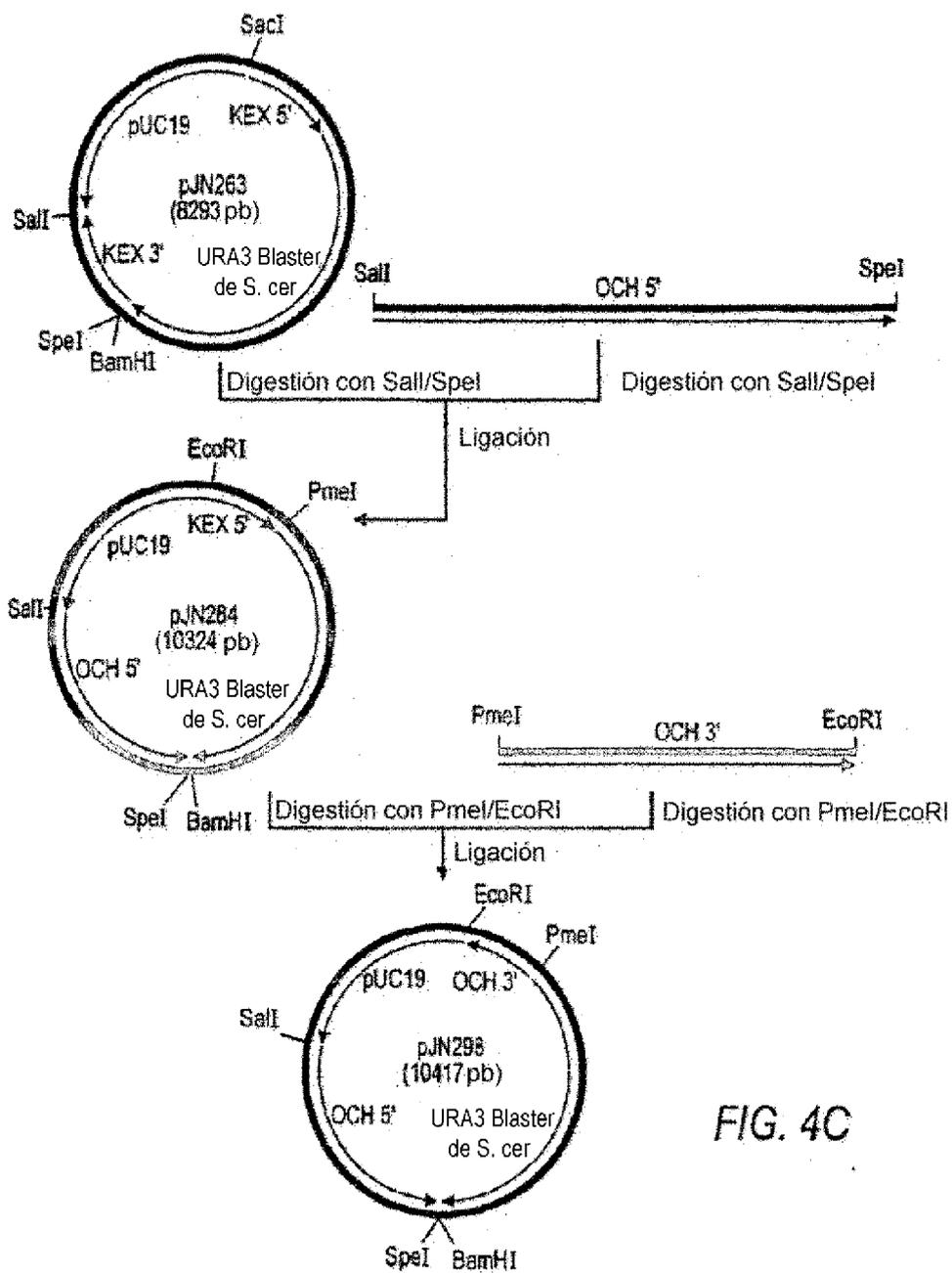


FIG. 4C

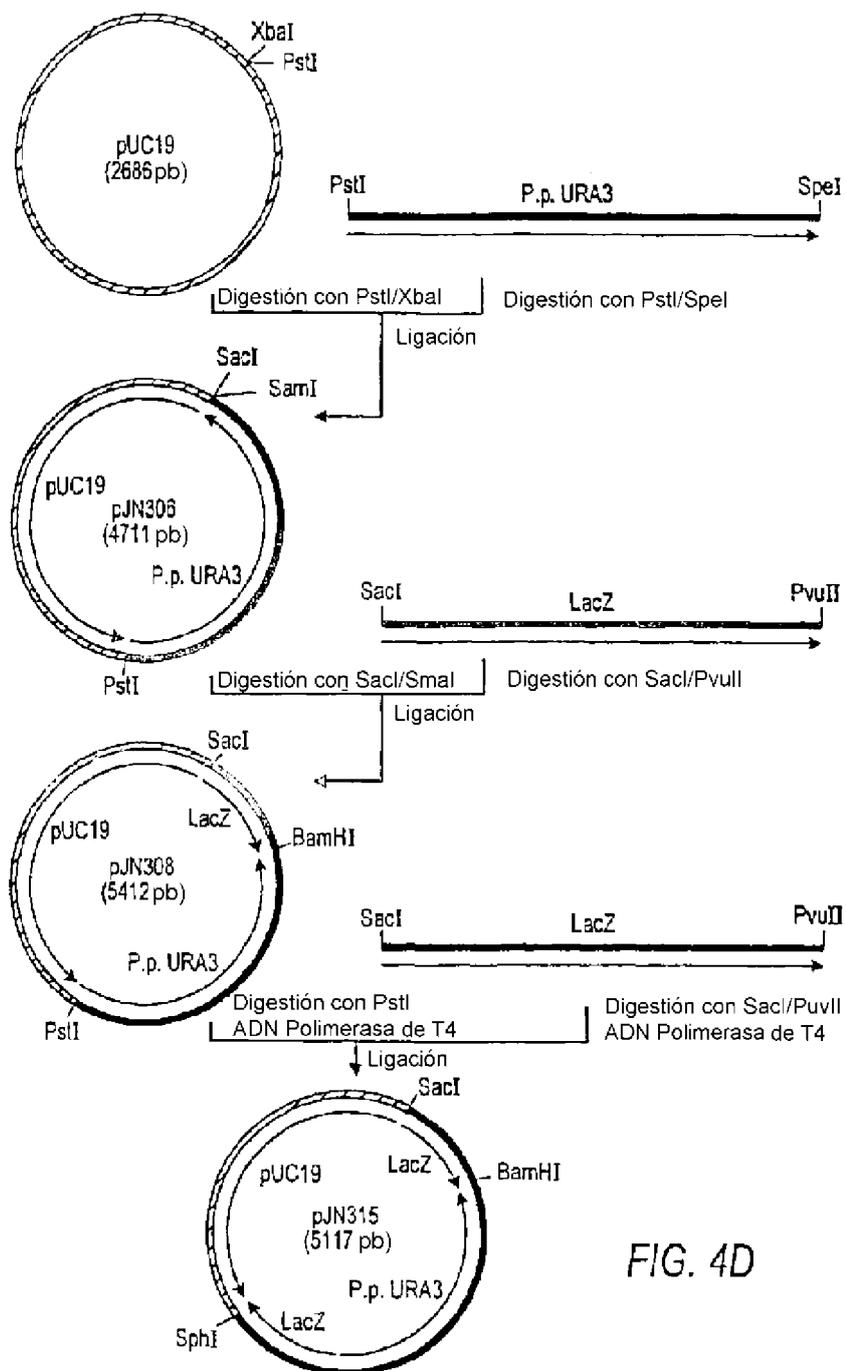


FIG. 4D

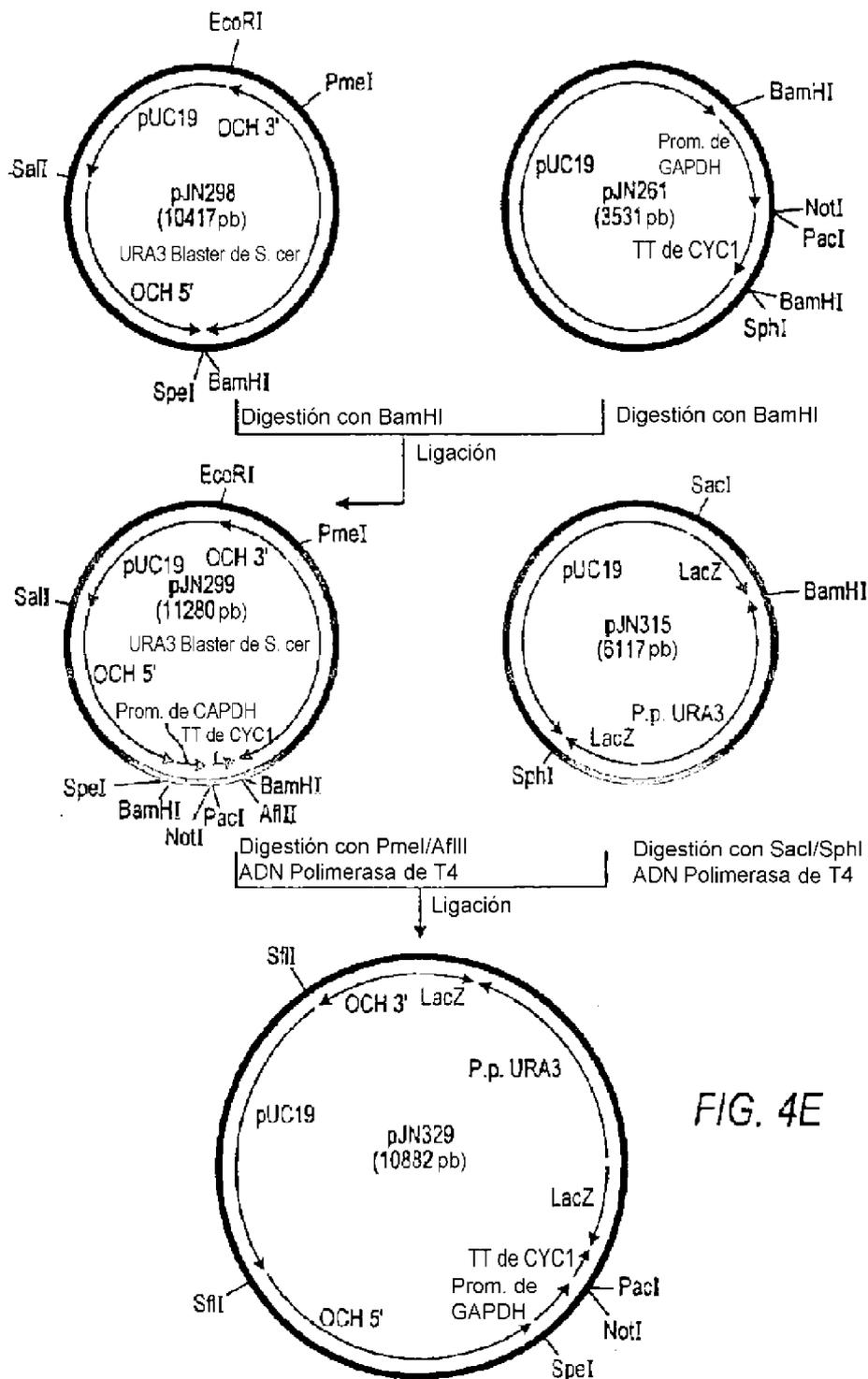


FIG. 4E

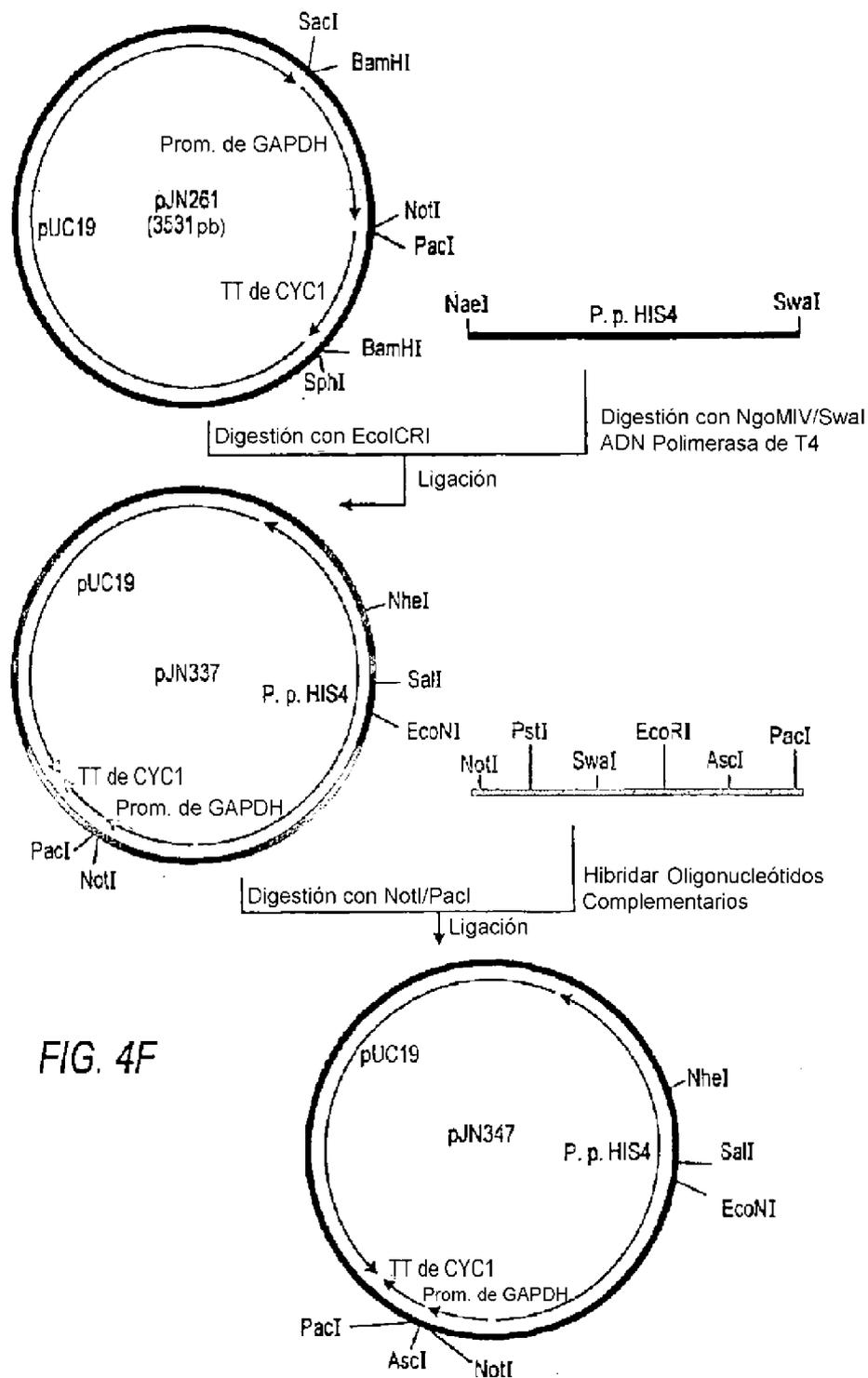


FIG. 4F

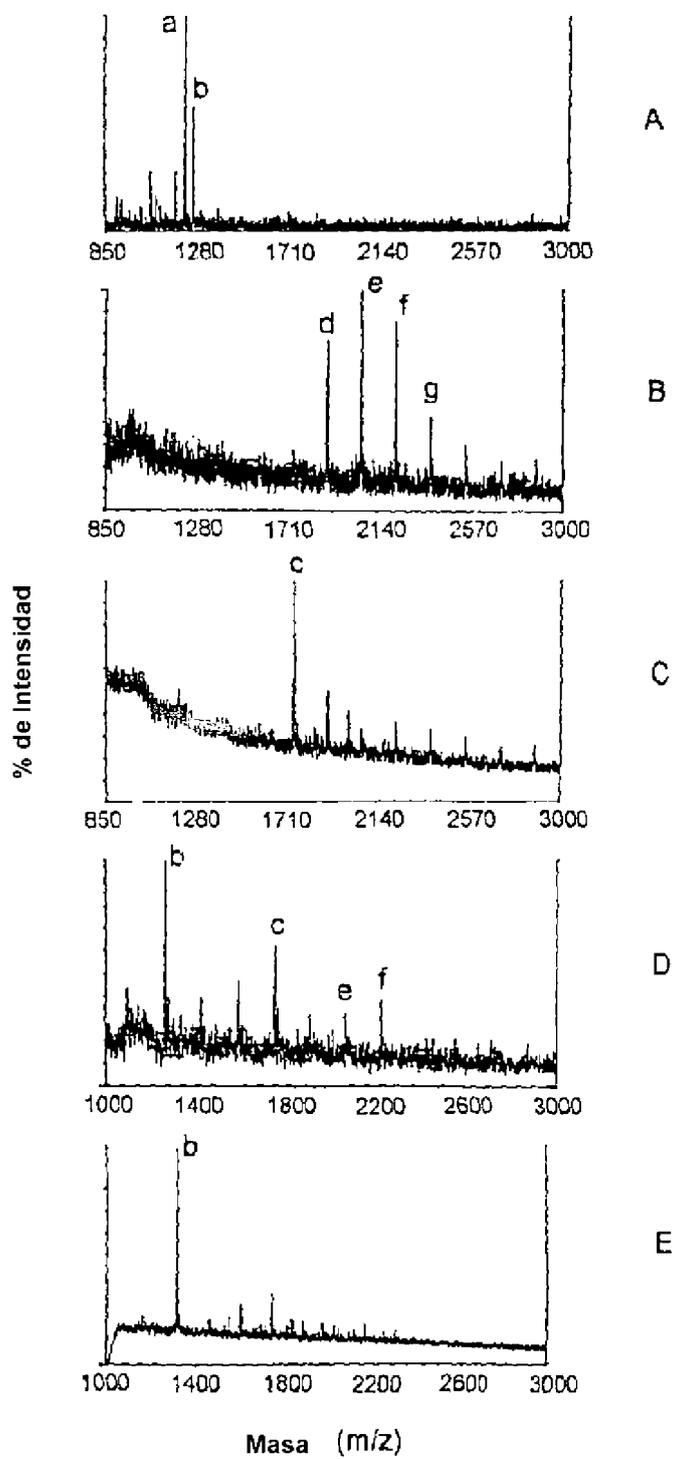
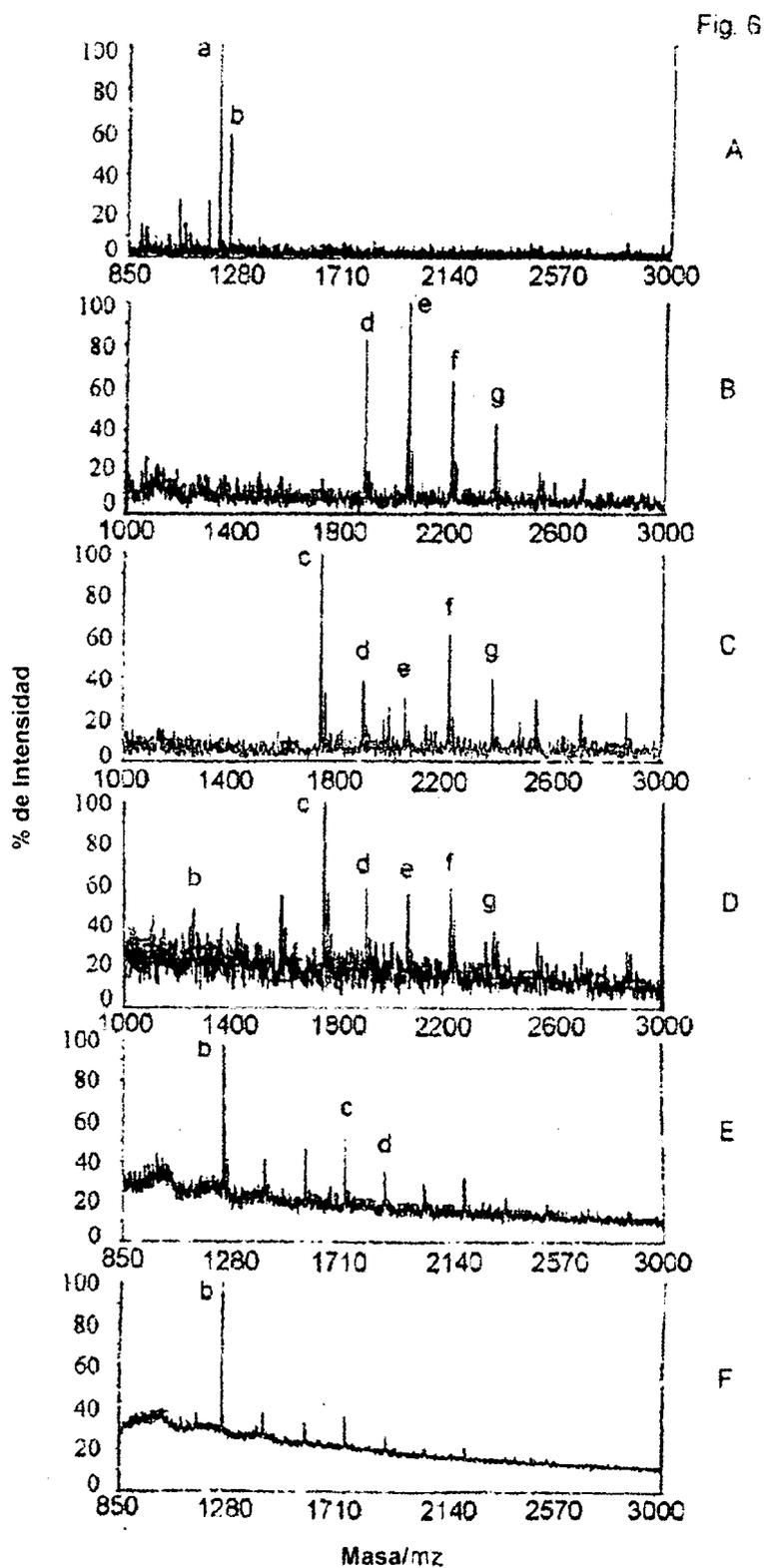
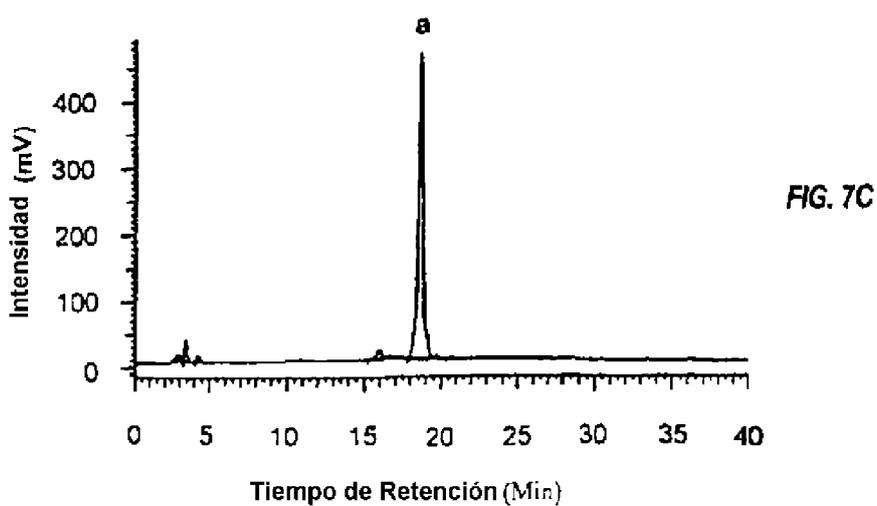
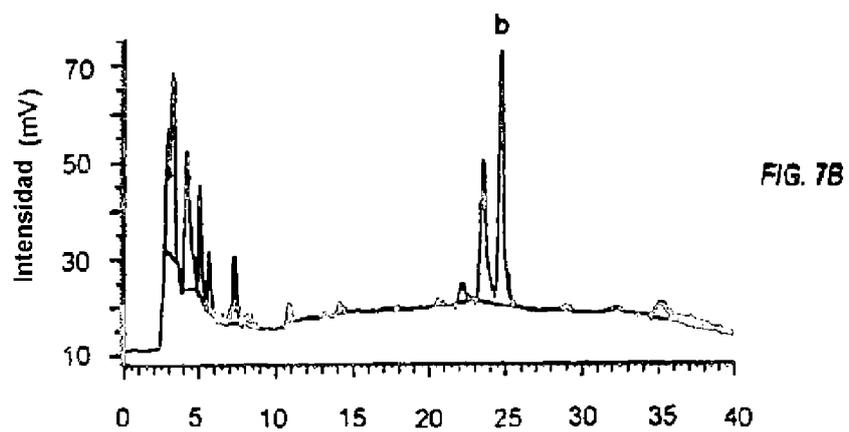
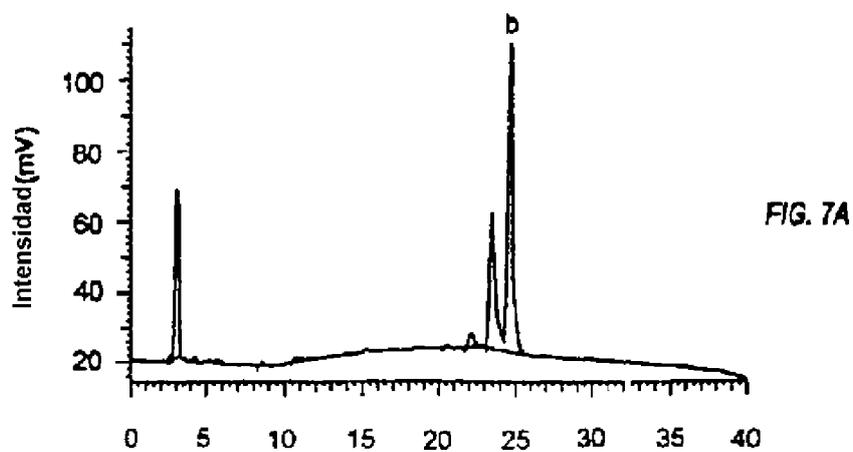


Fig. 5





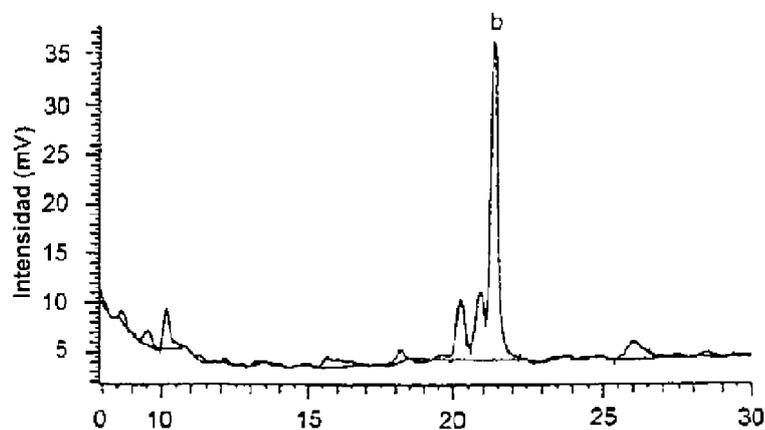


FIG. 8A

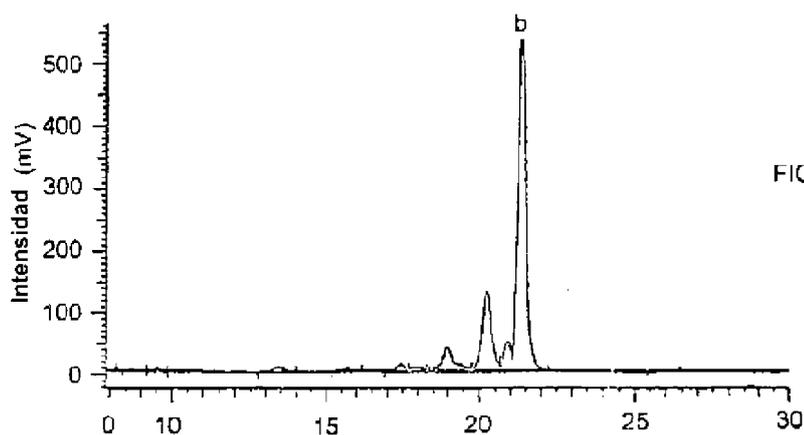


FIG. 8B

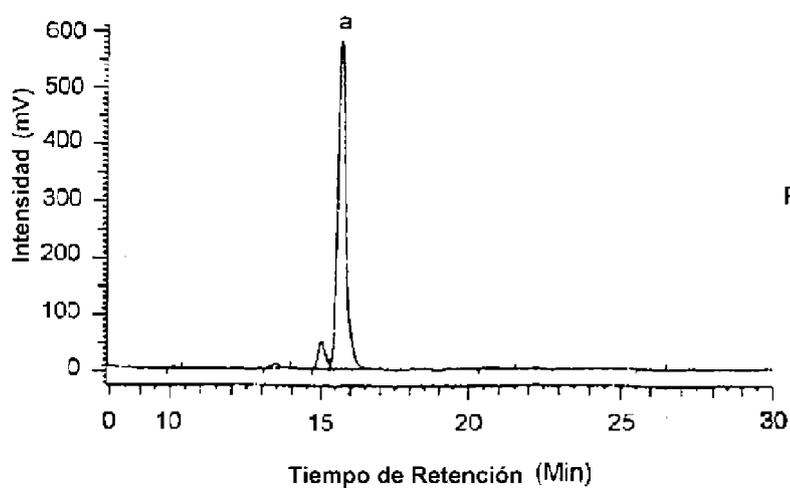
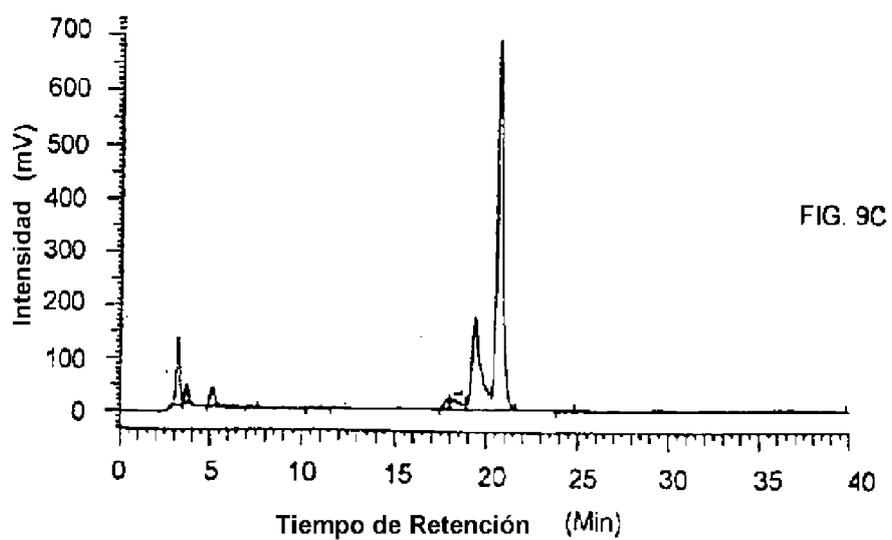
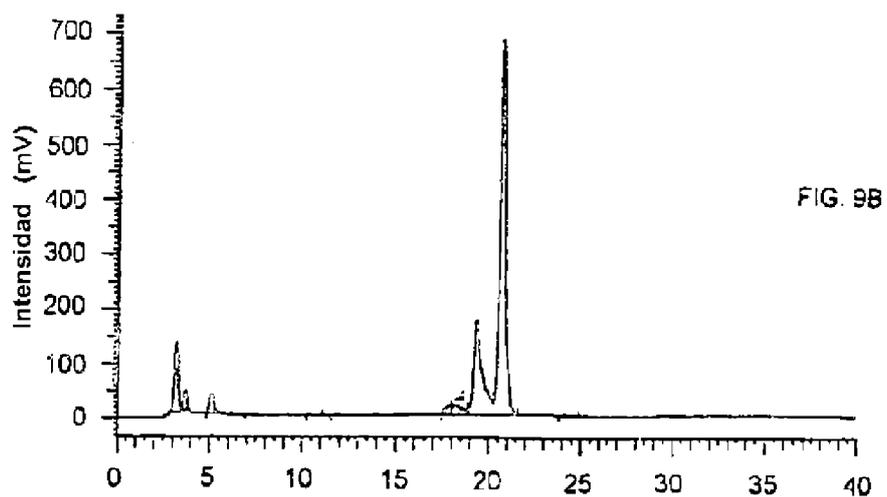
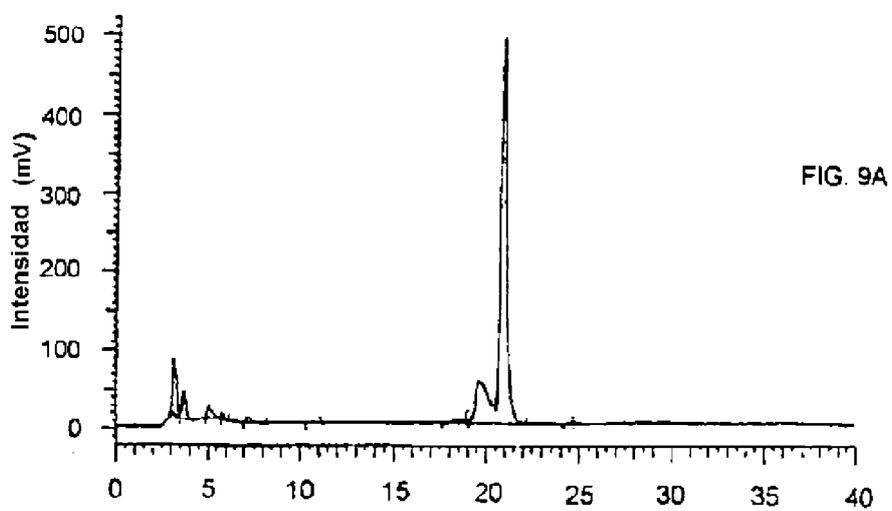
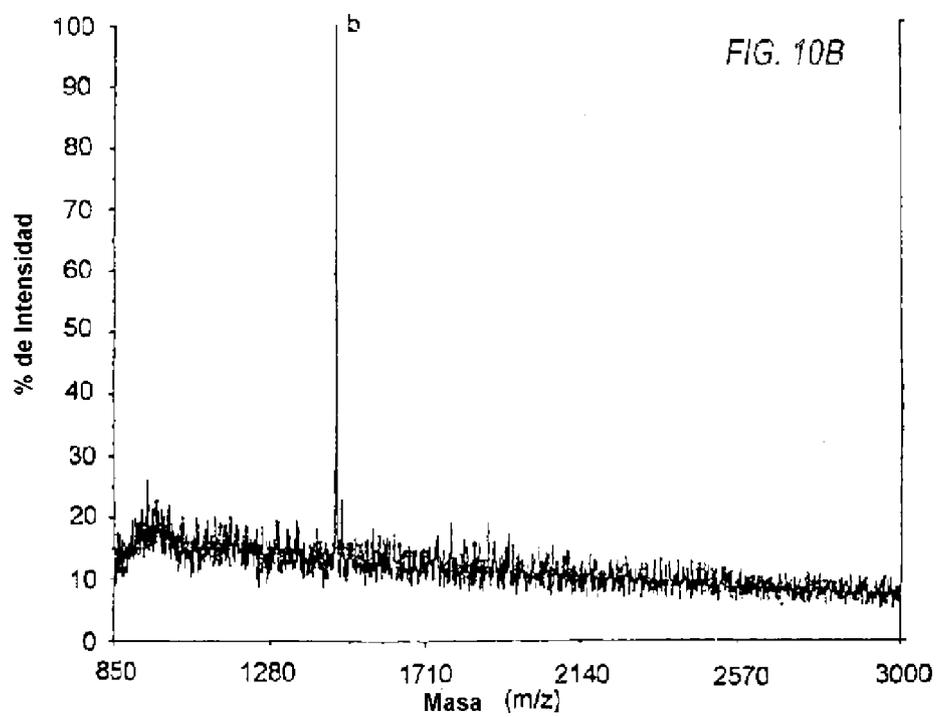
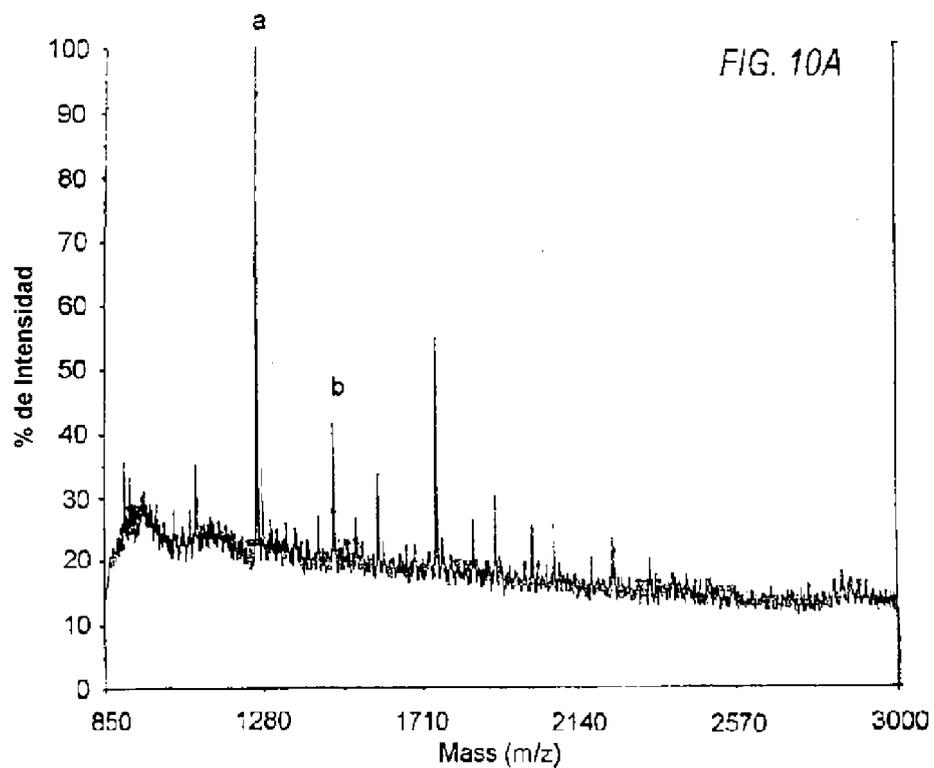


FIG. 8C





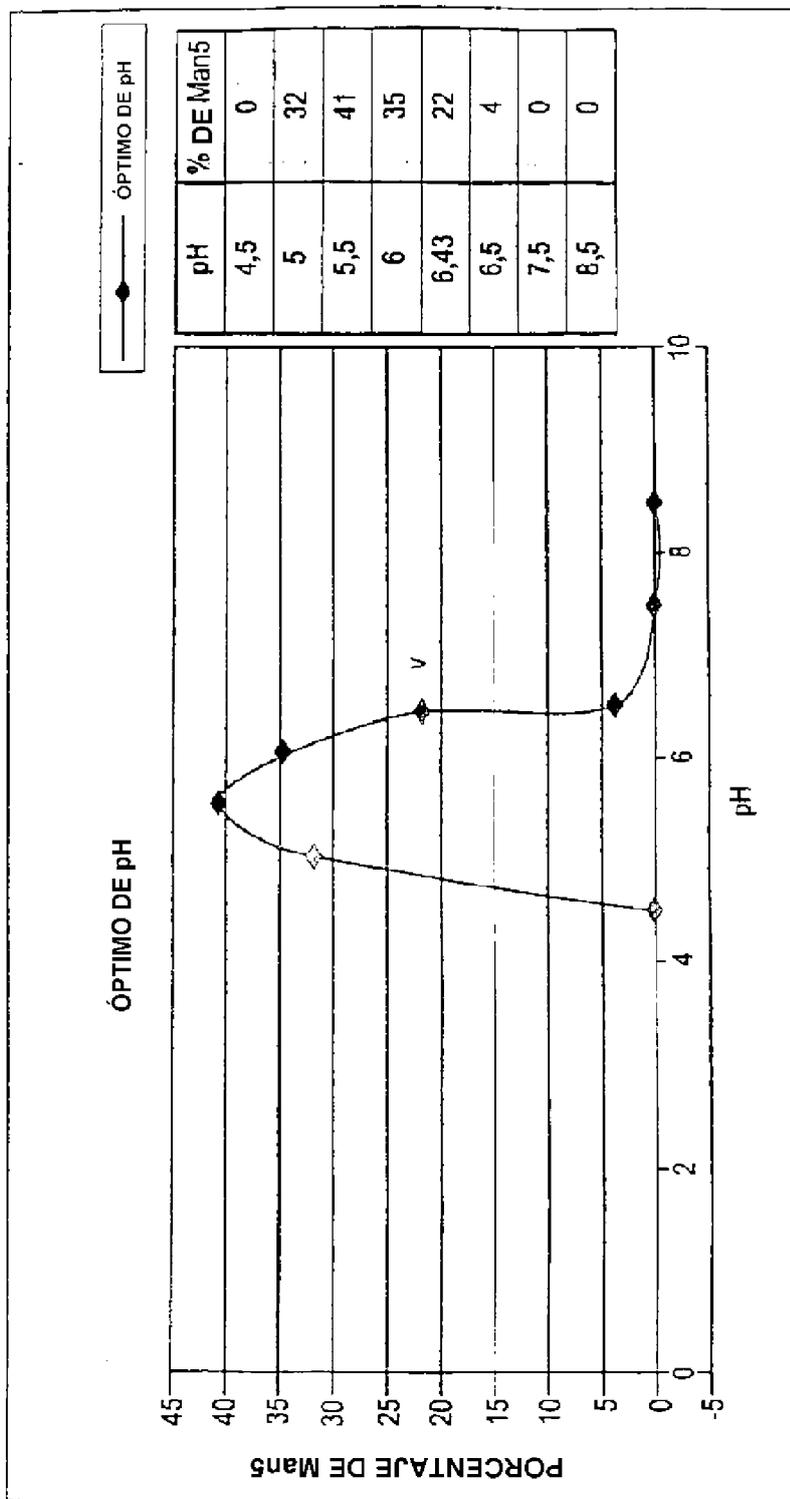
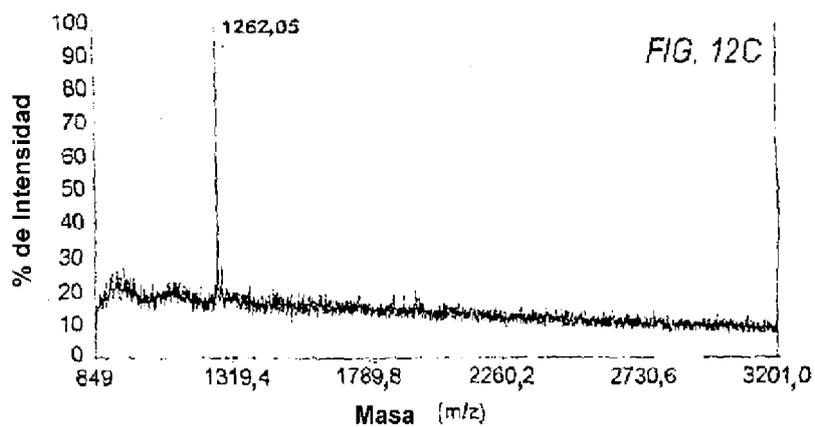
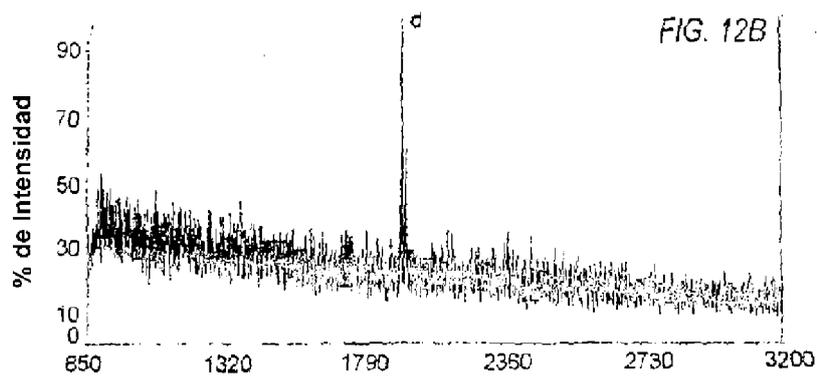
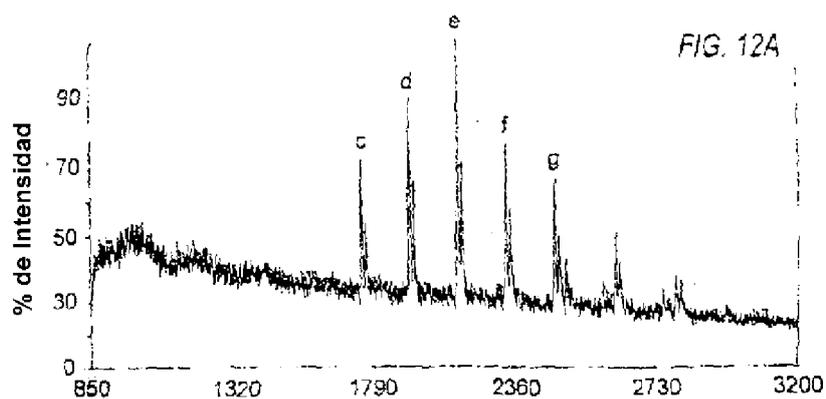


FIG. 11



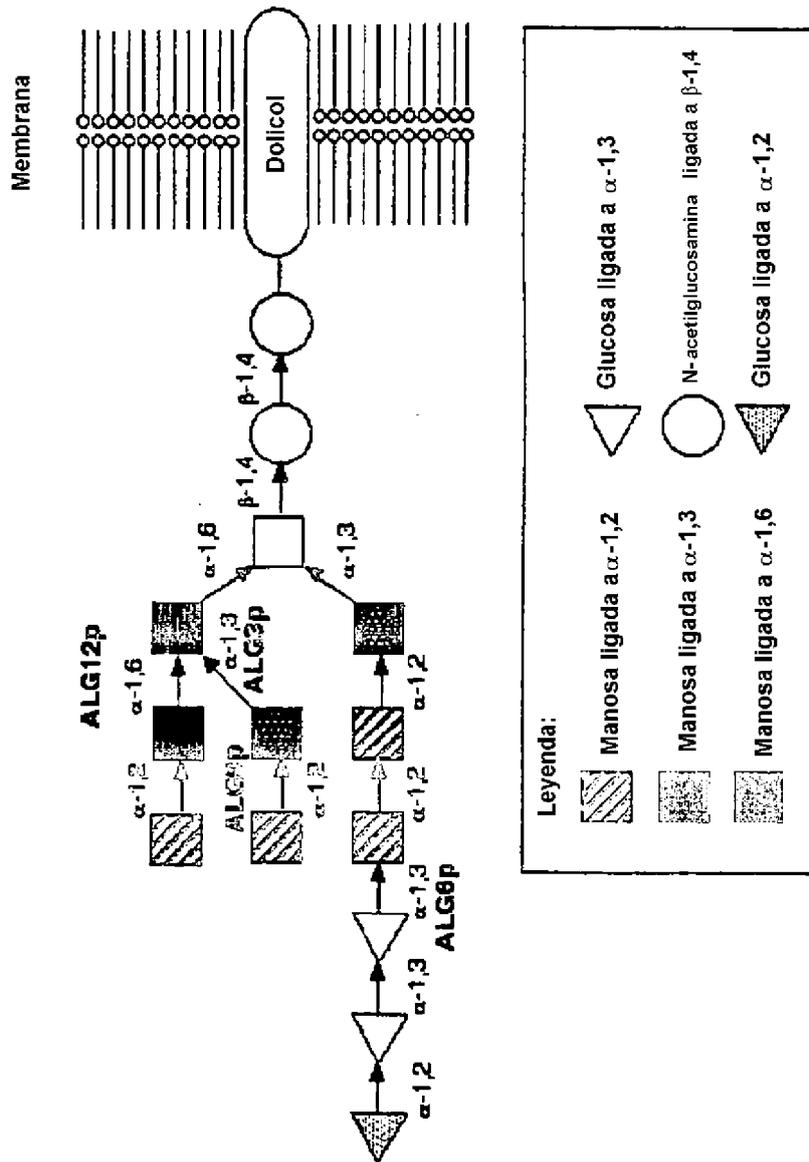


FIG. 13

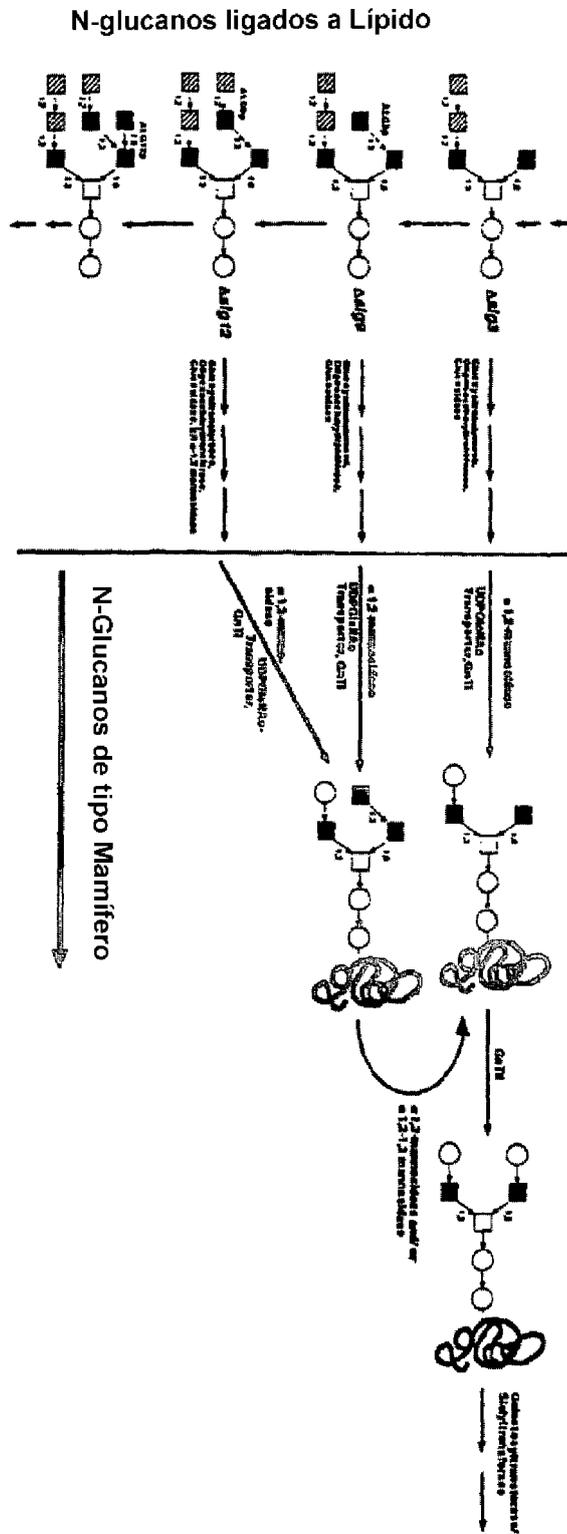


FIG. 14

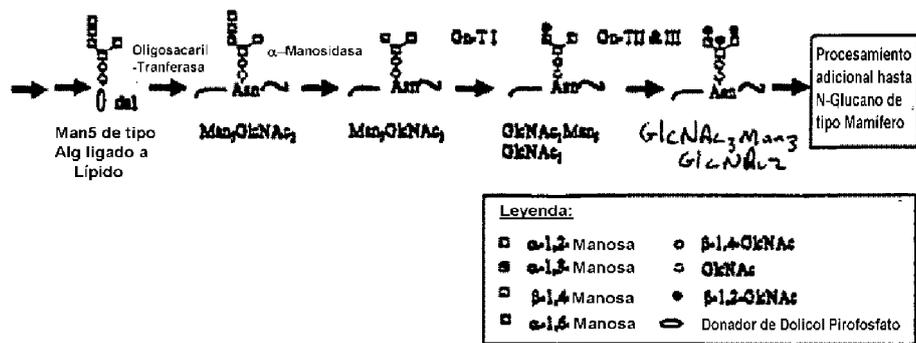


FIG. 15

# ES 2 335 345 T3

Blast de ALG3 05-22-01

|  |                                  |     |       | (bits) | Valor |
|--|----------------------------------|-----|-------|--------|-------|
| <b>Secuencias que producen alineamientos significativos:</b> |                                  |     |       |        |       |
| gi 586444 sp P38179 ALG3_YEAST                               | DOLICHYL-P-MAN:MAN(5)GLCNAC(...) | 797 | 0.0   |        |       |
| gi 3024226 sp Q92685 ALG3_HUMAN                              | DOLICHYL-P-MAN:MAN(5)GLCNAC(...) | 173 | 7e-43 |        |       |
| gi 3024221 sp Q24332 NT56_DROVI                              | LETHAL(2)NEIGHBOUR OF TID P...   | 145 | 3e-34 |        |       |
| gi 3024222 sp Q27333 NT56_DROME                              | LETHAL(2)NEIGHBOUR OF TID P...   | 121 | 3e-27 |        |       |
| gi 10720153 sp P82149 NT53_DROME                             | LETHAL(2)NEIGHBOUR OF TID ...    | 121 | 5e-27 |        |       |
| gi 1707982 sp P40999 GLS2_YEAST                              | 1,3-BETA-GLUCAN SYNTHASE CO...   | 32  | 2.8   |        |       |
| gi 1346146 sp P38631 GLS1_YEAST                              | 1,3-BETA-GLUCAN SYNTHASE CO...   | 31  | 6.6   |        |       |

## Alineamientos

### Levadura

>gi|586444|sp|P38179|ALG3\_YEAST DOLICHYL-P-MAN:MAN(5)GLCNAC(2)-PP-DOLICHYL MANNOSYLTRANSFERASE  
 (DOL-P-MAN DEPENDENT ALPHA(1-3)-MANNOSYLTRANSFERASE)  
 (HM-1 KILLER TOXIN RESISTANCE PROTEIN)  
 Longitud = 458

Puntuación = 797 bits (2059), Esperado = 0,0  
 Identidades = 422/458 (92%), Positivas = 422/458 (92%)

|               |  |     |
|---------------|--|-----|
| Problema: 1   | MEGEQSPQGEKSLQRKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLLIIFESMLCKI  | 60  |
| Sujeto: 1     | MEGEQSPQGEKSLQRKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLLIIFESMLCKI  | 60  |
| Problema: 61  | IIKKVAYTEIDYKAYMEQIEMIQLDGMLDYSQVSGGTGPIVYPAGHVLIYKMMYWLTEGM   | 120 |
| Sujeto: 61    | IIKKVAYTEIDYKAYMEQIEMIQLDGMLDYSQVSGGTGPIVYPAGHVLIYKMMYWLTEGM   | 120 |
| Problema: 121 | DHVERGQVFFRFLYLLTLALQMACYLLHLPPWCVVLAACLKRLHSIYVLRLEFNDCEFTL   | 180 |
| Sujeto: 121   | DHVERGQVFFRFLYLLTLALQMACYLLHLPPWCVVLAACLKRLHSIYVLRLEFNDCEFTL   | 180 |
| Problema: 181 | FMVVTVLGAIVASRCHQRPKLEKSLALVISATYSMAVSIKGNALLYFPAMMISLFIILNDA  | 240 |
| Sujeto: 181   | FMVVTVLGAIVASRCHQRPKLEKSLALVISATYSMAVSIKGNALLYFPAMMISLFIILNDA  | 240 |
| Problema: 241 | NVILTLLDLVAMIAWQVAVAVPFLRSFPQQYLHCAFNGRKKFMYQWSINWQMDDEEAFND   | 300 |
| Sujeto: 241   | NVILTLLDLVAMIAWQVAVAVPFLRSFPQQYLHCAFNGRKKFMYQWSINWQMDDEEAFND   | 300 |
| Problema: 301 | KRFXXXXXXXXXXXXXXXXXVTRYPRILPDLWSSLCHPLRKNVNLNANPAKTIIPFVLIASN | 360 |
| Sujeto: 301   | KRFHLALLISKLIALTTLFVTRYPRILPDLWSSLCHPLRKNVNLNANPAKTIIPFVLIASN  | 360 |
| Problema: 361 | FIGVLFSSRLHYQFLSWYHWTLPILIFWSGMPPFVGPINWVVLHEWCWNSYPPNSQXXXXX  | 420 |
| Sujeto: 361   | FIGVLFSSRLHYQFLSWYHWTLPILIFWSGMPPFVGPINWVVLHEWCWNSYPPNSQ       | 420 |
| Problema: 421 | XXXXXXXXXXXXXXXXXSGSVALAKSHLRRTTSMEEKLN                        | 458 |
| Sujeto: 421   | LALNTVLLLLLALTQLSGSVALAKSHLRRTTSMEEKLN                         | 458 |

FIG. 16-1

## ES 2 335 345 T3

### Ser Humano

>gi|3024226|sp|Q92685|ALG3\_HUMAN DOLICHYL-3-MAN:MAN(5)GLCNAC(2)-PP-DOLICHYL  
 MANNOSYLTRANSFERASE  
 (DOL-P-MAN DEPENDENT ALPHA(1-3)-MANNOSYLTRANSFERASE)  
 (NOT56-LIKE PROTEIN)  
 Longitud = 438

Puntuación = 173 bits (439), Esperado = 7e-43  
 Identidades = 133/396 (33%), Positivos = 195/396 (49%), Huecos = 28/396 (7%)

Problema: 26 WQDLKDGVRVYVIFDCRANLIVMPELLILFESMLCKI I I K K V A Y T E I D Y K A Y M E Q I E M I Q L D 85  
 WO+ R ++ + R L+V L L E + +I +VAYTEID+KAYM ++E + ++  
 Sujeto: 29 WQER----KLLLRFPYRLLVAACLCLAEVGGITFWVIHRVAYTEIDWKAYMAEVEGV-IN 83

Problema: 86 GMLDYSQVSGGTGFLVYPAGHVLIIYKMYWLTGMDHVERGQVFFRYLYLLTLALQMACY 145  
 G DY+Q+ G TGLVYPAG V I+ +Y+ T + Q F LYL TL L Y  
 Sujeto: 84 GTYDYTQLQGDGTGFLVYPAGFVYIFMGLYYATSRGTDIERMAONI FAVLYLATLLLVFLIY 143

Problema: 146 Y-LLELFPWC-VVLACLSKRLHSIYVLRLLFNDCTTLFMVVTVLGAIVASRCHQRPKLLK 203  
 + +PP+ + C S R+HSI+VLRLLFND + + +L + QR  
 Sujeto: 144 HQCKVPPFFVFFMCCASVRVHSIFVLRLLFNDP-----VAMVLLFLSINLLLAQRWNGG- 197

Problema: 204 SLALVISATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVAVPF 263  
 +S+AVS+KMN LL+ P ++ L L L + A + QV + +PF  
 Sujeto: 198 -----CCFFSLAVSVKMNLLFAPGLLFLLLTQGFGRGALPKLGI CAGL--QVVLGLPF 249

Problema: 264 LRSFPQQLHCAFNFRKFMYSINWQWIDEAFNDRFXXXXXXXXXXXXXXXXXVTRY 323  
 L P YL +F+ GR+F++ W++NW+ + E F + F + R+  
 Sujeto: 250 LLENPSGYLSR.SFDLGRQFLFWITVNWRFLEALFLHRAFHLALLTAMLTLLLLFALCRW 309

Problema: 324 PRILPDLWSSLCHPLRKNVLANANPAKTI PFVLIASNFIGVLFSRSLHYQFLSWYHWTLF 383  
 R + S L P ++ I L SNFIG+ FSRSLHYQF WY TLP  
 Sujeto: 310 HRTGESILSLLRDPKRRKVPQPLTPNQIVSTLFTSNFIGICFSRSLHYQFYVWYFHTLP 369

Problema: 384 ILIF----WSGMPFFVGPPIWYVLEWCWNSYPPNS 414  
 L++ W + + + E WY+YE S  
 Sujeto: 370 YLLWAMPARNLTHLLRLLLVGLI--ELSWFTYPTS 403

### Drosophila Vi

>gi|3024221|sp|Q24312|NT56\_DROVI LETHAL(2)NEIGHBOUR OF TID PROTEIN (NOI58)  
 Longitud = 526

Puntuación = 145 bits (366), Esperado = 3e-34  
 Identidades = 133/273 (49%), Positivos = 157/273 (57%), Huecos = 17/273 (6%)

Problema: 33 VRYVIFDCRANLIVMPELLILFESMLCKI I I K K V A Y T E I D Y K A Y M E Q I E M I Q L D G M L D Y S Q 92  
 ++Y+ F+ A IV L++L E+++ ++I++V YTEID+KAYM++ E L+G +YS  
 Sujeto: 34 IKYLAFEPAALPIVSULIVLAEAVINVLVIQRVVPYTEIDWKAYMQECEGF-LNGTTNYSL 92

Problema: 93 VSGGTGFLVYPAGHVLIIYKMYWLTGMDHVERGQVFFRYLYLLTLALQMACYVLLH-LP 151  
 + G TGPLVYPA V IY +Y+LT +V Q F +YLL + L + Y +P  
 Sujeto: 93 LRGDTEPLVYPAFVYIYSGLYLTGQGTNVRLAQYIFACTIYLLQMCLVLRLLYTKSRKVP 152

Problema: 152 PWCVYLAQL-SKRLHSIYVLRLLFNDCTTLFMVVTVLGAIVASRCHQRPKLLKSLALVIS 210  
 P+ +VL+ S R+HSIYVLRLLFND L +L A + QR L S  
 Sujeto: 153 PYVWVLSAFTSYEIHHSIYVLRLLFNDPVAIL-----LLVAALNLFLEDRWTLG-----S 200

Problema: 211 ATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVAVVFLRSFPQ 270  
 YS+AV +KMN + A + LF L + V+ TL+ L Q+ + PFLR+ P +  
 Sujeto: 201 ICYSLAVGVKMN--ILLFAPALLFYLANLGVRLTLVOLTICAVLQLFIGAPFLRTHPME 258

FIG. 16-2

## ES 2 335 345 T3

Problema: 271 YLHCAFNFGRKFMQWSINWQMMDEEAFNDKRF 303  
 YL +F+ GR F ++W++N++ + +E F + F  
 Sujeto: 259 YLRGSFDLGRIFEHKWTVMYRFLSKELFEQREF 291

Puntuación = 53.3 bits (127), Esperado = 1e-06  
 Identities = 31/62 (50%), Positivos = 41/62 (66%), Huecos = 6/62 (9%)

Problema: 352 IPFVLIASNFIVLFSRSLHYQFLSWYHWTLPILIFWSGMPFFVGGPIWYVLH--EWCWNS 409  
 +PF L NFIGV +RSLHYQF WY +LP L+ WS P+ +G + +L E+CWN+  
 Sujeto: 412 LPFFL--CNFIGVACARSLSHYQFYIWFHSLPYLV-WS-TPYSLGVRYLILGIIIEYCWT 467

Problema: 410 YP 411  
 YP  
 Sujeto: 468 YP 468

### Drosophila melanogaster

>gi|3024222|sp|Q27333|NT56\_DROME LETHAL(2)NEIGHBOUR OF TID PROTEIN (NOT56)  
 (NOT45)

Longitud = 510

Puntuación = 121 bits (305), Esperado = 3e-27  
 Identities = 96/272 (35%), Positivos = 154/272 (56%), Huecos = 17/272 (6%)

Problema: 34 RYVIFDCRANLIVMPLLILFESMLCKIIKKVAYTEIDYKAYMEQIEMIQLDGMLEYSQV 93  
 +Y++ + A IV ++L E ++ ++I++V YTEID+ AYM++ E L+G +YS +  
 Sujeto: 36 KYLLLEPAALPIVGLFVLLAEIWINVVVVIQRPVYTEIDWVAYMQECEGF-LNGTTNYSLL 94

Problema: 94 SGGTGPLVYPAGHVLIIYMMYWLTEGMDHVERGQVFFRYLYLLTLALQMACYLLH-LPF 152  
 G TGPLVYPA V IY +Y++T +V Q F +YLL LAL + Y +PP  
 Sujeto: 95 RGDGTGPLVYPAAFVYIYSALYVYVTSHGTVNRLAQYIFAGIYLLQLALVLRYSKSRKVPP 154

Problema: 153 HCVVLACL-SKRLHSIYVLRLEFNDCTILFMVVTVLGAIVASRCHQRPKLKSLALVISA 211  
 + +VL+ S R+HSIYVLRLEND + V +L A + +R L S  
 Sujeto: 155 YVIVLSAFTSYRHSIYVLRLEFNDP-----VAVLLLYAALNLFDRRWTLG-----ST 202

Problema: 212 TYSMAVSIKGNALLYFPAMMISLFIINDANVILTLDDLVAIAWQVAVAVPFLRSFPQOY 271  
 +S+AV +KMN + A + LF L + ++ T+L L Q+ + PFL + P +Y  
 Sujeto: 203 FPSLAVGVKMN--ILLFAPALLLFYLANLGLLRTILQLAVCCVIQLLLGAPFLLTHPVEY 260

Problema: 272 LHCAFNFGRKFMQWSINWQMMDEEAFNDKRF 303  
 L +F+ GR F ++W++N++ + + F ++ F  
 Sujeto: 251 LRGSFDLGRIFEHKWTVMYRFLSRDVFENRTF 292

Puntuación = 49,4 bits (117), Esperado = 2e-05  
 Identities = 27/60 (45%), Positivos = 35/60 (58%), Huecos = 2/60 (3%)

Problema: 352 IPFVLIASNFIVLFSRSLHYQFLSWYHWTLPILIFWSGMPFFVGGPIWYVLHENCWNSYP 411  
 +PF L N +GV SRSLSHYQF WY +LP L + + V + L E+CWN+YP  
 Sujeto: 437 LPFFL--CNLVGVACSRSLHYQFYWYFHSLSPLYLAWSTPYSLGVRCLILGLIEYCWTYP 464

FIG. 16-3

## ES 2 335 345 T3

**Matriz: BLOSUM62**

**Penalizaciones por Huecos: Existencia: 11, Extensión: 1**

**Número de Coincidencias con DB: 28883317**

**Número de Secuencias: 96469**

**Número de extensiones: 1107545**

**Número de extensiones exitosas: 2870**

**Número de secuencias mejores que 10,0: 16**

**Número de HSP mejores que 10,0 sin introducción de huecos: 5**

**Número de HSP con huecos introducidos de forma exitosa en ensayo preliminar: 11**

**Número de HSP que han intentado la introducción de huecos en ensayos preliminar: 2839**

**Número de HSP con huecos (no preliminar): 23**

**longitud de problema: 458**

**longitud de base de datos: 35.174.128**

**longitud de HSP eficaz: 45**

**longitud de problema eficaz: 413**

**longitud de base de datos eficaz: 30.833.023**

**espacio de búsqueda eficaz: 12734038499**

**espacio de búsqueda eficaz usado: 12734038499**

**T: 11**

**A: 40**

**X1: 15 (7,1 bits)**

**X2: 38 (14,6 bits)**

**X3: 64 (24,7 bits)**

**S1: 40 (21,8 bits)**

**S2: 67 (30,4 bits)**

**FIG.16-4**

*ALG3 de S. cerevisiae*

ATGGAAGGTGAACAGTCTCCGCAAGGTGAAAAGTCTCTGCAAAGGAAGC  
 AATTTGTCAGACCTCCGCTGGATCTGTGGCAGGATCTCAAGGACGGTGTG  
 CGCTACGTGATCTTCGATTGTAGGGCCAATCTTATCGTTATGCCCTTTTG  
 ATTTTGTTGAAAGCATGCTGTGCAAGATTATCATTAAAGAAGGTAGCTTAC  
 ACAGAGATCGATTACAAGGCGTACATGGAGCAGATCGAGATGATTCAGCT  
 CGATGGCATGCTGGACTACTCTCAGGTGAGTGGTGGAAACGGGCCCGCTGG  
 TGTATCCAGCAGGCCACGTCTTGATCTACAAGATGATGTACTGGCTAACA  
 GAGGGAATGGACCACGTTGAGCGCGGGCAAGTGTTTTTCAGATACTTGTA  
 TCTCCTTACACTGGCGTTACAAATGGCGTGTTACTACCTTTTACATCTACC  
 ACCGTGGTGTGTGGTCTTGGCGTGCCTCTCTAAAAGATTGCACTCTATTTA  
 CGTGCTACGGTTATTCAATGATTGCTTCACTACTTTGTTTATGGTCGTCACG  
 GTTTTGGGGGCTATCGTGGCCAGCAGGTGCCATCAGCGCCCCAAATAAA  
 GAAGTCCCTTGCGCTGGTGATCTCCGCAACATACAGTATGGCTGTGAGCA  
 TTAAGATGAATGCGCTGTTGTATTTCCCTGCAATGATGATTTCTCTATTCAT  
 CCTTAATGACGCGAACGTAATCCTTACTTTGTTGGATCTCGTTGCGATGAT  
 TGCATGGCAAGTCGCAGTTGCAGTGCCCTTCCTGCGCAGCTTTCGCAACA  
 GTACCTGCATTGCGCTTTTAATTTCCGGCAGGAAGTTTATGTACCAATGGAG  
 TATCAATTGGCAAATGATGGATGAAGAGGCTTTCAATGATAAGAGGTTCC  
 ACTTGGCCCTTTTAATCAGCCACCTGATAGCGCTCACCACACTGTTTCGTCA  
 CAAGATAACCCTCGCATCCTGCCCGATTTATGGTCTTCCCTGTGCCATCCGC  
 TGAGGAAAAATGCAGTGCTCAATGCCAATCCCGCCAAGACTATTCCATTC  
 GTTCTAATCGCATCCAACCTTCATCGGCGTCTTATTTCAAGGTCCCTCCAC  
 TACCAGTTTCTATCCTGGTATCACTGGACTTTGCCTATACTGATCTTTTGGT  
 CGGGAATGCCCTTCTTCGTTGGTCCCATTGGTACGTCTTGCACGAGTGGT  
 GCTGGAATTCCTATCCACCAAACTCACAAGCAAGCACGCTATTGTTGGCA  
 TTGAATACTGTTCTGTTGCTTCTATTGGCCTTGACGCAGCTATCTGGTTCCG  
 TCGCCCTCGCCAAAAGCCATCTTCGTACCACCAGCTCTATGGAAAAAAG  
 CTCAACTGA

*Alg3 de S. cerevisiae*

MEGEQSPQGEKSLQRKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLLIL  
 FESMLCKIIKKVAYTEIDYKAYMEQIEMIQLDGMLDYSQVSGGTGPLVYPAG  
 HVLIYKMMYWLTEGMDHVERGQVFFRYLYLLTLALQMACYLLHLPPWCV  
 VLACLSKRLHSIYVLRLFNDCFITLFMVVTVLGAIVASRCHQRPKLKSLALV  
 ISATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVA  
 VPFLRSFPQQYLHCAFNFGRKFMYQWSINWQMMDEEAFNDKRFHLALLISHL  
 IALTTLVTRYPRILPDLWSSLCHPLRKNVLANPAKTIPFVLLASNFIGVLFS  
 RSLHYQFLSWYHWTLPIILIFWSGMPFFVGPWIWYVLHEWCWNSYPPNSQASTL  
 LLALNTVLLLLLALTQLSGSVALAKSHLRTTSSMEKCLN

FIG. 17

*ALG3 de P. pastoris*

ATGCCTCCGATAGAGCCAGCTGAAAGGCCAAAGCTTACGCTGAAAAATGT  
 TATCGGTGATCTAGTGGCTCTTATTCAAACGTTTTATTAAACCCAGATTTT  
 AGTGTCTTCGTTGCACCTCTTTATGGTTAGCTGATTCCATTGTTATCAAGG  
 TGATCATTGGCACTGTTTCCCTACACAGATATTGATTTTTCTTCATATATGCA  
 ACAAATCTTTAAAATTGACAAGGAGAATTAGATTATAGCAACATATTTG  
 GTGACACCCGGTCCATTGGTTTACCCAGCCGGCCATGTTTCATGCTTACTCAG  
 TACTTTTCGTGGTACAGTGATGGTGGAGAAGACGTCAGTTTCGTTCAACAA  
 GCATTTGGTTGGTTATACCTAGGTTGCTTGTACTATCCATCAGCTCCTACT  
 TTTTCTCTGGCTTAGGGAAAATACCTCCGGTTTATTTTGTTTTGTTGGTAGC  
 GTCCAAGAGACTGCATTCAATATTTGTATTGAGACTCTTCAATGACTGTTT  
 AACAACATTTTTGATGTTGGCAACTATAATCATCCTTCAACAAGCAAGTAG  
 CTGGAGGAAAGATGGCACAACCTATCCATTATCTGTCCCTGATGCTGCAG  
 ATACGTACAGTTTAGCCATCTCTGAAAGATGAATGCGCTGCTATACCTCC  
 CAGCATTCTACTACTCATATATCTCATTGTGACGAAAATTTGATTAAG  
 CCTGGCACCTGTTCTAGTTTTGATATTGGTGCAAGTAGGAGTCGGTTATT  
 CGTTCATTTTACCGTTGCACTATGATGATCAGGCAAATGAAATTCGTTCTG  
 CCTACTTTAGACAGGCTTTTGACTTTAGTCGCCAATTTCTTTATAAGTGA  
 CGGTTAATTGGCGCTTTTTGAGCCAAGAACTTTCAACAATGTCCATTTTC  
 ACCAGCTCCTGTTTGCTCTCCATATTATTACGTTAGTCTTGTTTCATCCTCAA  
 GTTCTCTCTCTAAAACATTGGAAAACCGCTTGGTAGATTTGTGTTGGA  
 CTTTTCAAATTTTGGAAAGCCAACCTTATCTCCAACCAATATTATCAACGA  
 CCCAGAAAGAAGCCCAGATTTTGTACACCGTCATGGCTACTACCAACTT  
 AATAGGGGTGCTTTTTGCAAGATCTTACACTACCAGTTCCTAAGCTGGTA  
 TCGGTTCTCTTGGCATACTCCTTACAAGGCTCGTCTGAACTTTATAGCA  
 TCTATTATTGTTATGCCGCTCACGAGTATTGCTGGTTGGTTTTCCCAGCTA  
 CAGAACAAGTTCCGCGTTGTTGGTATCTATCTTACTACTTATCCTGATTC  
 TCATTTTTACCAACGAACAGTATTTCTTCTCAATCGGTCCCTGCAGAAA  
 AAAAGAATACATAA

*Alg3p de P. pastoris*

MPPIEPAERPKLTLKNVIGDLVALIQNVLFNPDFSVFVAPLLWLADSIVIKVUG  
 TVSYTDIDFSSYMQQIFKIRQGELDYSNIFGDTGPLVYPAGHVHAYSVLSWYS  
 DGGEDVSFVQQAFGWLYLGCLLSISSYFFSGLGKIPPVYFVLLVASKRLHSIF  
 VLRLFNDCLTFLMLATIIILQQASSWRKDGTIPLSVPDAADTYSLAISVKMN  
 ALLYLPAFLLLIYLICDENLIKALAPVLVLLVQVGVGYSFILPLHYDDQANEIR  
 SAYFRQAFDFSRQFLYKWTVNWRFLSQETFNNVHFHQLLFALHIITLVLFILKF  
 LSPKNIGKPLGRFVLDIFKFWKPTLSPTNIINDPERSPDFVYTVMATTNLIGVLF  
 ARSLHYQFLSWYAFSLPYLLYKARLNFIASIVYAAHEYCWLVPATEQSSAL  
 LVSILLILLILIFTNEQLFSPQSVPAAEKNT

FIG. 18

ES 2 335 345 T3

BLAST de ALG3 de *P. pastoris*

(bits) Valor

|   |   |                     |       |
|---|---|---------------------|-------|
| <a href="#">gi 586444 sp P36179 ALG3 YEAST</a>        | <a href="#">Dolichyl-P-Man:Man(5)GlcNAC(...)</a>        | <a href="#">228</a> | 2e-58 |
| <a href="#">gi 12802365 gb AAK07848.1 AF309689.10</a> | <a href="#">putative NOT-56 manno...</a>                | <a href="#">212</a> | 6e-54 |
| <a href="#">gi 984725 gb AAA75352.1</a>               | <a href="#">ORF 1</a>                                   | <a href="#">206</a> | 4e-52 |
| <a href="#">gi 7492702 pir T39084</a>                 | <a href="#">probable mannosyltransferase - fissi...</a> | <a href="#">176</a> | 8e-43 |
| <a href="#">gi 16226531 gb AAL16193.1 AF428424.1</a>  | <a href="#">At2g47760/F17A22.15 [A...</a>               | <a href="#">164</a> | 2e-39 |
| <a href="#">gi 25367230 pir BB4919</a>                | <a href="#">Not56-like protein (imported) - Ara...</a>  | <a href="#">164</a> | 3e-39 |
| <a href="#">gi 25814791 emb CAB70171.2</a>            | <a href="#">Hypothetical protein K09E4.2 [C...</a>      | <a href="#">161</a> | 2e-38 |
| <a href="#">gi 17535001 ref NP_496950.1</a>           | <a href="#">Putative plasma membrane membr...</a>       | <a href="#">160</a> | 3e-38 |
| <a href="#">gi 1654000 emb CAA70220.1</a>             | <a href="#">Not56-like protein (Homo sapiens...</a>     | <a href="#">155</a> | 2e-36 |
| <a href="#">gi 13279206 gb AAH04313.1 AAH04313</a>    | <a href="#">Unknown (protein for IMA...</a>             | <a href="#">154</a> | 2e-36 |
| <a href="#">gi 22122365 ref NP_666051.1</a>           | <a href="#">hypothetical protein MGC36684 ...</a>       | <a href="#">150</a> | 3e-35 |
| <a href="#">gi 21292031 gb EAA04176.1</a>             | <a href="#">agCP3388 [Anopheles gambiae str...</a>      | <a href="#">120</a> | 4e-26 |
| <a href="#">gi 1780792 emb CAA71167.1</a>             | <a href="#">lethal(2)neighbour of tid [Dros...</a>      | <a href="#">114</a> | 3e-24 |

Alineamientos

*S. cerevisiae*

Puntuación = 228 bits (580), Esperado = 2e-58

Identidades = 154/429 (35%), Positivos = 229/429 (53%), Huecos = 37/429 (8%)

|               |  |
|---------------|--|
| Problema: 9   | RPKLTLNKVIQDLVALIQNVLFNEDFSVFPVAPLLWLAQSIIVIKVIIIGTVSYTDIDFSSYM 68 |
|               | RP L L DL ++ V+F+ ++ V PLL L +S++ K+II V+YT+ID+ +YM                |
| Sujeto: 20    | RPPLDLNQ---DLKDGVRVVFDCRANLIUMPLLLIFESMLCKI IKKVAYTEIDYKAYM 76     |
| Problema: 69  | QQIFKIR-CGELDYSHIFGDTGPLVYPAGHVHAYSVLSWYSDGGEDVSVFQQAFGWLVLG 127   |
|               | +QI I+ G LDYS + G TGPLVYPAGHV Y ++ W ++G + V Q F +LYL              |
| Sujeto: 77    | EQIENIQDLDGHLDYSDVSGGTGPLVYPAGHVLIYKIMYWLTEGMDHVERGQVFRYLYLL 136   |
| Problema: 125 | CLLLSISYFFSGLGKIPPVYFVLLVQSKRLHSIFVLRLEFNDCLTTFLLALATI---IILQ 184  |
|               | L L +- Y+ L +PP VL SIQLHSI+VLRLEFNDCL TT M+ T+ I+                  |
| Sujeto: 137   | TLALQMACYY---LLRLPPWCVVLACLKRLHSIYVLRLEFNDCLTTFLLMVTVLGAIVAS 193   |
| Problema: 185 | QASSWRKDGTTIPLSVPDAADTYSLAISVKNXXXXXXXXXXXXXXXXXCDENLIKALAPXX 244  |
|               | + K ++ L + + TYS+A+S+KMN D N+I L                                   |
| Sujeto: 194   | RCHQRPKPKKSLALVI---SATYSMAVSIKKNALLYFPAMMISLFIILNDANVILTLDDL 250   |
| Problema: 245 | XXXXXXXXXXYSFILPLHYDDQANEIRSAYFROAFDFSRQFLYKWTVNWRFLSOETFN 304     |
|               | F+ Y AF+F R+F+Y+W++NW+ + +E FN+                                    |
| Sujeto: 251   | AMIAMQVAVAVPFL-----RSFPQQYLHCAFNFGRKEMYQWSINWQMMDEEAFNDK 301       |
| Problema: 305 | HFHQLLFALHITL-VLFILKFLSPKNIGKPLGRFVLDIFKFWKPTLSPTNIIN-OPERS 362    |
|               | FH L H+I L LF+ ++ R + D++ L ++N +P ++                              |
| Sujeto: 302   | RFHLALLSHLIALTTLFVTRY-----PRILPDLWSSLCHPLRKNVAVLNANPAKT 351        |
| Problema: 363 | PDFVYTVMATINLIGVLFARSLHYQFLSWYAFSLPYLLYKARLNFIASIIVYAAHEYCNL 422   |
|               | F V+ +N IGVLF+RSLHYQFLSWY ++LP L++ + + F I Y HE+CW                 |
| Sujeto: 352   | IFP---VLIASFVIGVLFSPSLHYQFLSWYHWTLELIFWQGMFFVCGPIWYVLEHEWCWN 408   |
| Problema: 423 | VFPATEQSS 431  |
|               | +P O+S   |
| Sujeto: 409   | SYFPNSQAS 417  |

FIG. 19-1

ES 2 335 345 T3

*Neurospora crassa*

Puntuación = 212 bits (540), Esperado = 8e-54  
 Identidades = 140/400 (35%), Positivos = 212/400 (53%), Huecos = 29/400 (7%)

Problema: 35 SVFVAPLLWLADSIIVIKVIIGTVSYTDIDFSSYMQQIFKIROGELDYSNIFGDTGPLVYP 94  
 S + P L+L D+++ +II V YT+ID+++YM+Q+ +I GE DY+ + G TGPLVYP  
 Sujeto: 33 SKLIPPALFLVDALLCGLIINKVYPTEIDWAAYMEQVSCILSGERDYTEKVRGGTGPLVYP 92

Problema: 95 AGHVHAYSVLSWYSDGGEDVSVFVQQAAGWLYLGCILLSISSYFFSGLGKIPFVYFVLLVA 154  
 A HV+ Y+ L +D G ++ QQ F LY+ L + + Y+ K PP F LL  
 Sujeto: 93 AAKVYIYTGLYHLTDEGRHLLAQQLFAGLYMVLAVVMGCYN---QAKAPPYLFLLTL 149

Problema: 155 SKRLHSIFVLRLENDCLITTFMLATIIILQOASSWRKDGTTIPLSVPDAADTYSLAISVK 214  
 SKRLHSIFVLR FNDC + I Q+ +W+ A Y+L + VK  
 Sujeto: 150 SKRLHSIFVLRFCFNDCFAVLFVLAIFTFQR-RNMQA-----GALLYTLGLGVK 197

Problema: 215 MXXXXXXXXXXXXXXXXXCDENLIKALAPXXXXXXXXXXXXXYSFILPLHYDDQANEIRSAY 274  
 M + + L F+ HY + Y  
 Sujeto: 198 MTLLESIPAVGIVLFLGSG-SFVITLQIVATMGLVQILIGVFFL--AHYPT-----Y 247

Problema: 275 FROAFDFSRQFLYKWTVNWRFLSQETFNNVHFHQLLFAHIIITLVLFILKFLSPKNIGX 333  
 +AF+ SRQF +KWTVNWRF+ +E F + F L ALH++ L +FI +++ P K  
 Sujeto: 248 LSRAFELSRQFFKWTVNWRFFVGEIPLSKGFALTLALHVLVLCIFITTRWIKPAR--K 305

Problema: 334 PLGRFVLDIFKFKPPTLS-PTNIINDPERSPDFVYTMATTNLIGVLFARSLHYQFLSWY 392  
 L + + + KP L+ P + + +P ++ T + + N +G+LFARSLHYQF ++  
 Sujeto: 306 SLVQLISFVLLAGKPLTVPEHRAARDVTPRYIMTILSAMAVGLLFAKSLHYQFYAYV 365

Problema: 393 AFSLPYLLYKARLNFIASIIVYAAHEYCWLVFPATEQSSA 432  
 A+S P+LL--A L+ + +++A HE+ W VFP+T SSA  
 Sujeto: 366 AWTSPFLLRAGLHPVLVYLLWAVHEWAVNVFPSTPSSA 405

*Schizosaccharomyces pombe*

Puntuación = 176 bits (445), Esperado = 8e-43  
 Identidades = 132/390 (33%), Positivos = 194/390 (49%), Huecos = 35/390 (8%)

Problema: 42 LNLADSIIVIKVIIGTVSYTDIDFSSYMQQIFKIROGELDYSNIFGDTGPLVYPAGHVHAY 101  
 L L + + II V YT+ID+ +YM+Q+ GE DY ++ G TGPLVYP GHV Y  
 Sujeto: 30 LLLLEIPFVFAIISKVYPTEIDWIAAYMEQVNSFLLGERDYKSLVGCCTGPLVYPGGHVFLY 89

Problema: 102 SVLSWYSDGGEDVSVFVQQAAGWLYLGCILLSISSYFFSGLGKIPFVYFVLLVASKRLHSI 161  
 ++L + +DGG ++ Q F ++Y + +I Y F + + P +VLL+ SKRLHSI  
 Sujeto: 90 TLLYYLTDGGTNIVRAQYIFAFVW--ITTAIVGYLKF-IVRAPFYIYVLLILSKRLHSI 146

Problema: 162 FVLRLENDCLITTFMLATIIILQOASSWRKDGTTIPLSVPDAADTYSLAISVKMNXXXX 221  
 F+LRLFNDC + L + I+ W + A+ S+A SVKM+  
 Sujeto: 147 FILRLFNDCGFNS-LFSSLFILSSCKKXWVR-----ASTLLSVACSVKMSLLYV 194

Problema: 222 XXXXXXXXXXXXCDENLIKALAPXXXXXXXXXXXXXYSFILPLHYDDQANEIRSAYFROAFDF 261  
 L++ L P + + + +Y+ QAFDF  
 Sujeto: 195 PAYLVL-----LLQILGPKKTWGHIFVIIIVQILFSIPF----LAYFWSYWTQAFDF 242

Problema: 282 SRQFLYKWTVNWRFLSQETFNNVHFHQLLFAHIIITLVLFILKFLSPKNIGKPLGRFVLD 341  
 R F YKWTVNWRF+ + F + F + LH+ LV F K + + P  
 Sujeto: 243 GRAFDYKWTVNWRFI PRSIFESTSFSSTILFLHVALLVAFTCKHWNKLSRATP----- 295

Problema: 342 IFKFKPPTLSPTNIINDPERSPDFVYTMATTNLIGVLFARSLHYQFLSWYAFSLPYLLY 401  
 F L+ + +P+F++T +AT+NLIG+L ARSLHYQF +W+A+ PYL Y

FIG. 19-2

ES 2 335 345 T3

Sujeto: 296 -FAMVNSMLTLKPLPKLQLATPNFIFTALATSNLIGILCARSLHYQFYANFANWSPYLCY 354

Problema: 402 KARLNFIASIIVYAAHEYCWLVFPATEQSS 431  
 +A I ++ EY W VFP+T+ SS

Sujeto: 355 QASFPAPIVIGLWMLQEYAWNVPSTKLS 384  
*Arabidopsis thaliana*

Puntuación = 164 bits (415), Esperado = 2e-39  
 Identidades = 131/391 (33%), Positivos = 194/391 (49%), Huecos = 29/391 (7%)

Problema: 42 LWLADSIKVIIGTVSYTDIDFSSYMQQIFKIROGELDYSNIFGDTGPLVYPAGHVHAY 101  
 L LAD+I++ +II V YT ID+ +YM Q+ GE DY N+ GDTGPLVYPAG ++ Y

Sujeto: 39 LILADAILVALIIAYVPYTKIDWDAYMSQVSGFLGGERDYGNLKGDTGPLVYPAGFLYVY 98

Problema: 102 SVLSWYSDGGEDVSFVQAFGWLYLGCLLSISYFFSGLGKIPVYFVLLVASKRLHSI 161  
 S + + G +V Q FG LY+ L + + Y + + +P LL SKR+HSI

Sujeto: 99 SAVQNLTGG--EVYPAQILFGVLYIVNLGIVLIYVKTQV--VPHWALSLLCLSKRIHSI 154

Problema: 162 FVLRLFNDCLTTFMLLATIIILQOASSWRKDGTTIPLSVDAADTYSLAISVKMNXKX 221  
 FVLRLFNDCL L+ A++ + +RK + + +S A+SVKMN

Sujeto: 155 FVLRLFNDCLTTFMLLATIIILQOASSWRKDGTTIPLSVDAADTYSLAISVKMNXKX 202

Problema: 222 XXXXXXXXXXXCDENLIKALAPXXXXXXXXXXXXXSYFILPLHYDDQANEIRSAYFRQAFDF 281  
 N+I ++ F++ +Y AFD

Sujeto: 203 PTLLELLKAM--NIIGVVSALAGAALAIQILVGLPFLITYPV-----SYIANAFDL 251

Problema: 282 SRQFLYKWTNVNRFSLQETFNNVHFHQLLFALHIITLVLPILKFLSPKNIGKPLGRFVLD 341  
 R F++ W+VN++F+ + F + F L H+ LV F + K+ G +G

Sujeto: 252 GRVFIHFWSVNFKVPERVFVSKEFAVCLLIAHLFLLVAFA-NYKWCHEGGIIGFMRSR 310

Problema: 342 IFKFWKP-TLSPTNIINDPERSPDFVYTVMTTNLIGVLFARSLHYQFLSNYAFSLPYLL 400  
 F P +LS +++ + + V T M N IG++FARSLHYQF SWY +SLPYLL

Sujeto: 311 HFFLTLPSSLSFSDVSASRIITKEHVVTAMFVGNFICIVFARSLHYQFYSWYFSLPYLL 370

Problema: 401 YKARLNFIASIIVYAAHEYCWLVFPATEQSS 431  
 ++ +I++ E CW V+P+T SS

Sujeto: 371 WRTPFPTWLRRLIMFLGIELCWNVYPSTPSSS 401

FIG. 19-3

*ALG3 de K. lactis*

TTTGTTTACAAGCTGATACCAACGAACATGAATACACCGGCAGGTTTACT  
GAAGATTGGCAAAGCTAACCTTTTACATCCTTTTACCGATGCTGTATT CAG  
TGCGATGAGAGTAAACGCAGAACAAATTGCATACATTTTACTTGTTACCA  
ATTACATTGGAGTACTATTTGCTCGATCATTACACTACCAATTCCTATCTT  
GGTACCATTGGACGTTACCAGTACTATTGAATTGGGCCAATGTTCCGTATC  
CGCTATGTGTGCTATGGTACCTAACACATGAGTGGTGCTGGAACAGCTAT  
CCGCCAAACGCTACTGCATCCACACTGCTACACGCGTGTAACACATACTG  
TTATTGGCTGTATTCTTAAGAGGACCCGCAAACCTCGAAAAGTGGTGATAA  
CGAAACAACACACGAGAAAGCTGAG

*Alg3p de K. lactis*

FVYKLIPTNMNTPAGLLKIGKANLLHPFTDAVFSAMRVNAEQIAYILLVTNYI  
GVLFARSLHYQFLSWYHWTLPVLLNWANVPYPLCVLWYLTHEWCWNSYPP  
NATASTLLHACNTYCYWLYSZEDPQTRKVVITKQHTRKL

FIG. 20

# ES 2 335 345 T3

## BLAST de ALG3 de *K. lactis*

| Puntuación   | E                                      | (bits) | Valor |
|--|--|--------|-------|
| Secuencias que producen alineamientos significativos |  |        |       |
| <a href="#">gi 586444 sp P38179 ALG3 YEAST</a>       | Dolichyl-P-Man:Man(5)GlcNAc(...)       | 125    | 1e-28 |
| <a href="#">gi 984725 gb AAA75352.1</a>              | ORF 1                                  | 94     | 4e-19 |
| <a href="#">gi 16226531 gb BA116193.1 AF428424.1</a> | At2g47760/F17A22.15 [A...]             | 72     | 1e-12 |
| <a href="#">gi 25367230 pir [B64919</a>              | Not56-like protein [imported] - Ara... | 72     | 1e-12 |
| <a href="#">gi 21292031 gb EAA04176.1</a>            | agCP388 [Anopheles gambiae str....]    | 69     | 2e-11 |
| <a href="#">gi 20892051 ref XP_148657.1</a>          | similar to Lerhal(2)neighbour ...      | 65     | 2e-10 |

### Alineamientos

#### *S. cerevisiae*

Puntuación = 125 bits (314), Esperado = 1e-28  
 Identidades = 60/120 (50%), Positivos = 83/120 (69%), Huecos = 1/120 (0%)  
 Marco = +3

Problema: 66 ANLLHPFT-DAVFSAMRVNAEQIAYILLVNTNYIGVLFARSLHYQFLSWYHWTLPLVLLNWA 242  
 ++L HP +AV +A A+ I ++L+ +N+IGVLF-RSLHYQFLSWYHWTLPL+L+ N+  
 Sujeto: 332 SSLCHPLRKNVLANP--ARTIPFVLIASNFIGVLFARSLHYQFLSWYHWTLPLILFWS 389

Problema: 243 NVFYPLCVLWYLTHEWCWNSYPPNATASTLLHACNTYCYWLYS\*EDPQTRKQVITKQHTR 422  
 +P+ + +WY+ HEWCWNSYPPN+ ASTLL A NT L + + V + K H R  
 Sujeto: 390 GMPFFVGPVWYVLEWCWNSYPPNSQASTLLALALNTVLLLLLA-LTQLSGSVALAKSHLR 448

#### *A. thaliana*

Puntuación = 72,0 bits (175), Esperado = 1e-12  
 Identidades = 42/107 (39%), Positivos = 57/107 (53%), Huecos = 3/107 (2%)  
 Marco = +3

Problema: 84 FTDVFSAMRVNAEQIAYILLVNTNYIGVLFARSLHYQFLSWYHWTLPLVLLNWANVPYPLC 263  
 F+D S + + E + + V N+IG++FARSLHYQF SWY ++LP LL P L  
 Sujeto: 122 FSDVSASRI-ITKEHVVTAMFVGNFIGIVFARSLHYQFYSWYFYSLEPYLLWRTPEPTWLR 380

Problema: 264 VLWYLTHEWCWNSYPPNATASTL--LHACNTYCYWLYS\*EDPQTRK 395  
 ++ +L E CWN YP ++S L LH WL DP K  
 Sujeto: 381 LIMFLGIELCWNVYFSTPSSSGLLLCLHLYIIVGLWLPASVDPYQLK 427

FIG. 21

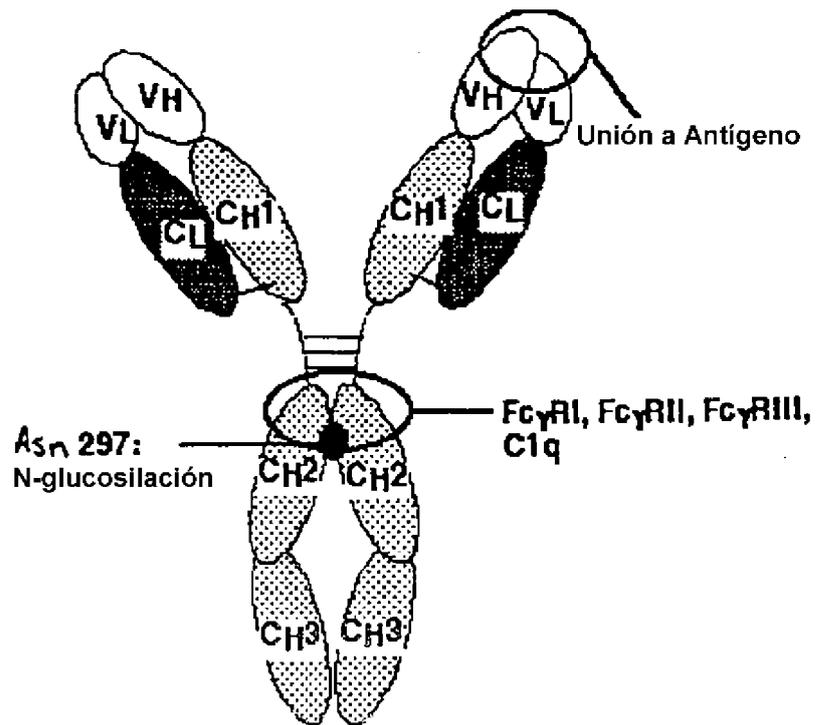


FIG. 22

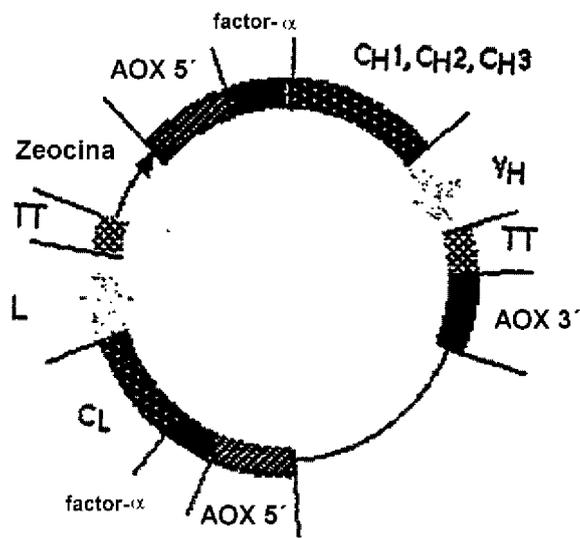


FIG. 23

ES 2 335 345 T3

>gi|6754685|ref|NM\_010795.1| Mus musculus mannoside acetyl  
glucosaminyltransferase 3 (Mgat3), mRNA

ATGAAGATGAGACGCTACAAGCTCTTTCTCATGTTCTGTATGGCTGGCCTGTGCCTCATATCCTTCTCTGC  
ACTTCTTTAAGACCTTATCCTATGTCACCTTCCCGAGAGAAGTGGCCCTCCCTCAGCCCTAACCTCGTATC  
CAGCTTCTTCTGGAAACAATGCCCTGTCACTCCCAGGCCAGTCCGGAGCCGGGTGGCCCGGACCTATTG  
CGGACACCCCTCTACTCCCACTCTCCCTGCTCCAGCCACTGTCCCCGAGCAAGGCCAGAGGAACTGC  
ACCGGGTGGACTTCGTGTTGCCGGAGGACACCACGGAGTATTTTGTGCGCACCAAAGCTGGTGTGTGTG  
CTTCAAACCAGGTACCAGGATGCTGGAGAAACCTTCGCCAGGGCGGACAGAGGAGAAGCCCGAAGTGTCT  
GAGGGCTCCTCAGCCCGGGGACCTGCTCGGAGGCCATGAGGCACGTGTTGAGTACGGGGAGCGCCTGG  
GCAGCCGGGGCACTAGGCGCAAGTGGGTTGAGTGTGTGTGCTGCCAGGCTGGCACGGGCCAGTTGCCG  
GGTGCCCAACGGTGGTGCAGTATCCCAACCTGCCCAACGAAGAAACGCTGGTACCCAGGGAGGTACCGAGG  
CGGGTTATCAACGCCATCAACATCAACCACGAGTTCGACCTGCTGGATGTGGCTTCCATGAGCTGGGAG  
ATGTTGTGGACGCCCTTCGTGGTCTGTGAATCTAATTTACCCGCTACGGGGAGCCTCGGCCGCTCAAGTT  
CCGAGAGATGCTGACCAATGGCACCTTCGAGTACATCCGCCACAGGTGCTCTATGCTTCTCTGGACCAT  
TCCCACCTGGTGGCCGTCAGGACGGCTGGATTGCGGATGACTACCTGCCACCTTCCCTACCCAGGATG  
GGTCTCCCGCCTGCCAACCTGCCGCCCGATGACGTCTTATCATCGACGATGGGAGGAGATCCCTGC  
GGTGTATGGTGTGCTGTTCTCAAACCTTACGATGGCTGGACAGAGCCCTTCGCCCTCCACATGCGGAAG  
TCCCTGTATGTTCTTCTGGAAGCAGCCGGCCACACTGGAGGTGGTGTGAGGCTGCACCATGGACATGC  
TGCAGGCCGTGTATGGCTGGATGGCATCCGCCCTGCCGCCCGCCAGTACTACACCATGCCCAACTTCCG  
GCAGTATGAGAACCACCGGCCACATCCTAGTGCAGTGGTCTCTCGGCAGCCCCCTGCACCTTCGGGGC  
TGGCATGCTCCTGGTGCTTACACCCGAGGGCATCTACTTTAAACTCGTGTGAGCCAGAAATGGCGACT  
TCCCCCGCTGGGGTACTATGAGGACAAGAGGGACCTCAATTACATCCGAGCTTGATCCGCACTGGGGG  
ATGGTTCGACGGAAACGAGCAGGAGTACCCTCCTGCCGACCCAGTGAACACATGTATGCTCCTAAATAC  
CTGCTCAAGAACTATGACCAGTTCGGCTACTTGTGCTGGAATAATCCCTACCCGGAGCCCAAGAGCACTGTAG  
AGGCTGGCCGCCAGAACCCAGGGCTCAGATCGAAGGCCATCTGCTGTGAGGGCAAGTTGGATACAGTGGAA  
GGCTAG

>gi|2117717|pir||JC4362 beta-1,4-mannosyl-glycoprotein 4-beta-N-  
acetylglucosaminyltransferase (EC 2.4.1.144) III - mouse  
MRRYKLFMFCMAGLCLISFLHFKTILSYVTFPEELASLSPNLISSFFHNNAPVTPQASPEPGDPDLLRT  
PLYSHSPILLOPLSPSKATEELHRVDFVLPEDTTEYFVRTRAGGVCFKPGTRMLEKPSFCRTEEKTEVEEG  
SSARGPARPPHREVLSSREPLGSRGTRPKWVECVCLPGWEGPSCGVPTVVOYSNLPKIERLVPREVP  
INAININHEFDLLDVRFHGLGDVVDVAVVCDVSNFTAYGEPRPLKFRMLTNGTTFEYIRHKVLYVFLDHF  
PGRQDQWIADYYLRTFLTQDGVSRLLRNLRPDDVFIIDDADEIPARDGVLFLKLYDQWTEPFAFHMRKSL  
YGFYKQPGTLEVVSGCTMDMLQAVYGLDGIIRLRRRQYYTMPNFRQYENRTGHILVQWLSGSLPHFAGWH  
CSWCFTEPGIYFKLVSAQNGDFPRWGDYEDKRDLYIRSLIRTGGWFDGTQOEYPPADPSEHMYAPKYL  
KNYDQFRYLLENPYREPSTVEGGRQNGSDGRSSAVRGLDTEAG

FIG. 24

**Fig. 25**

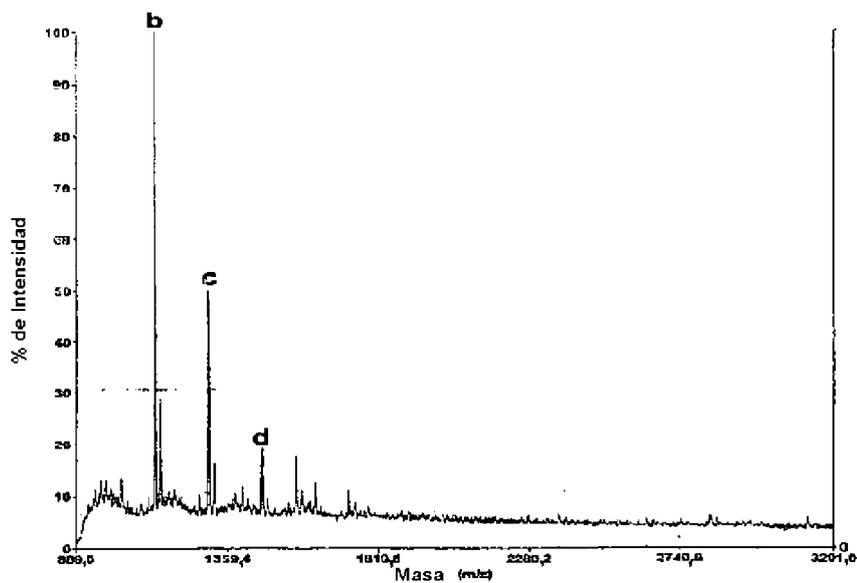
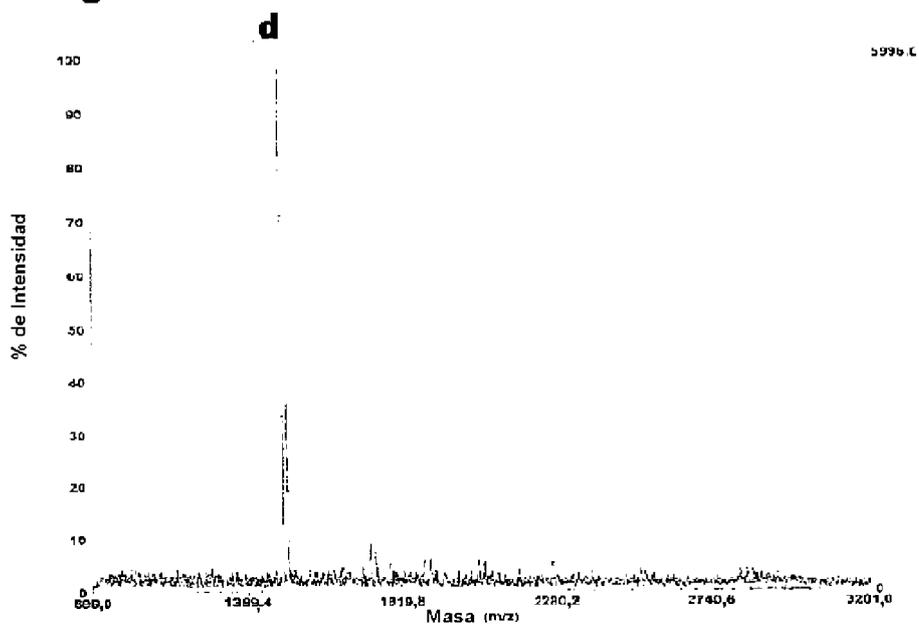


Fig. 26

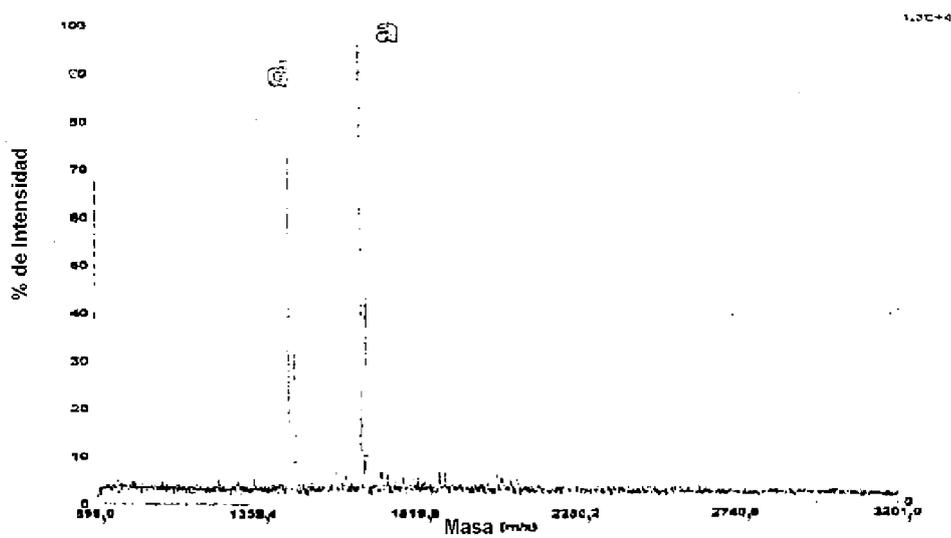
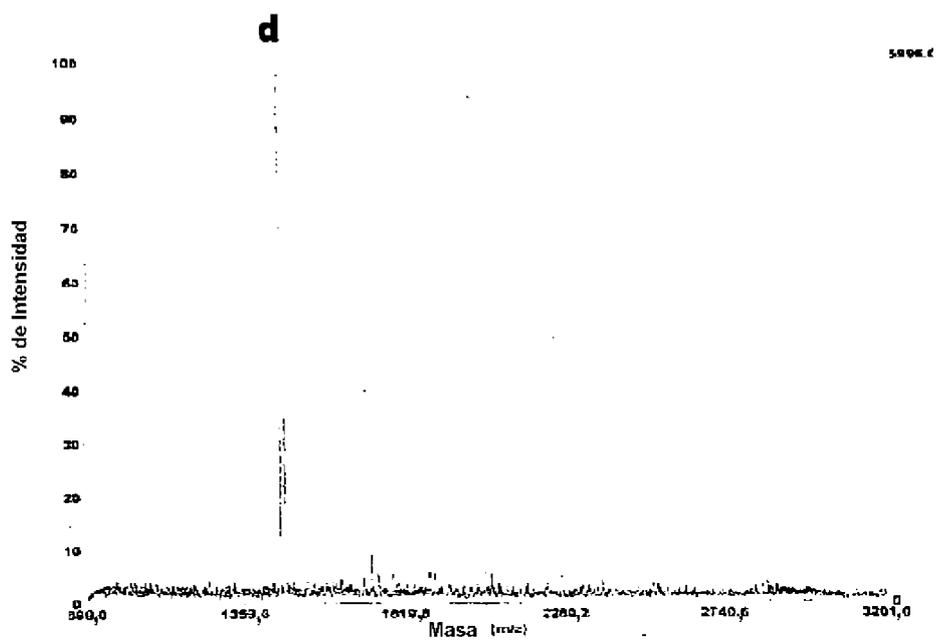
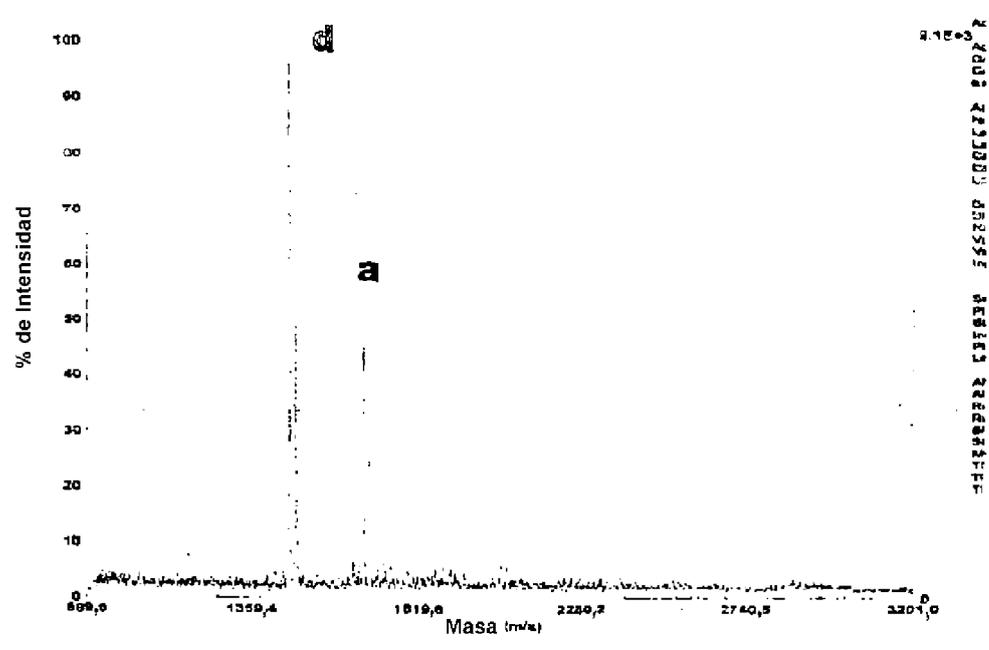
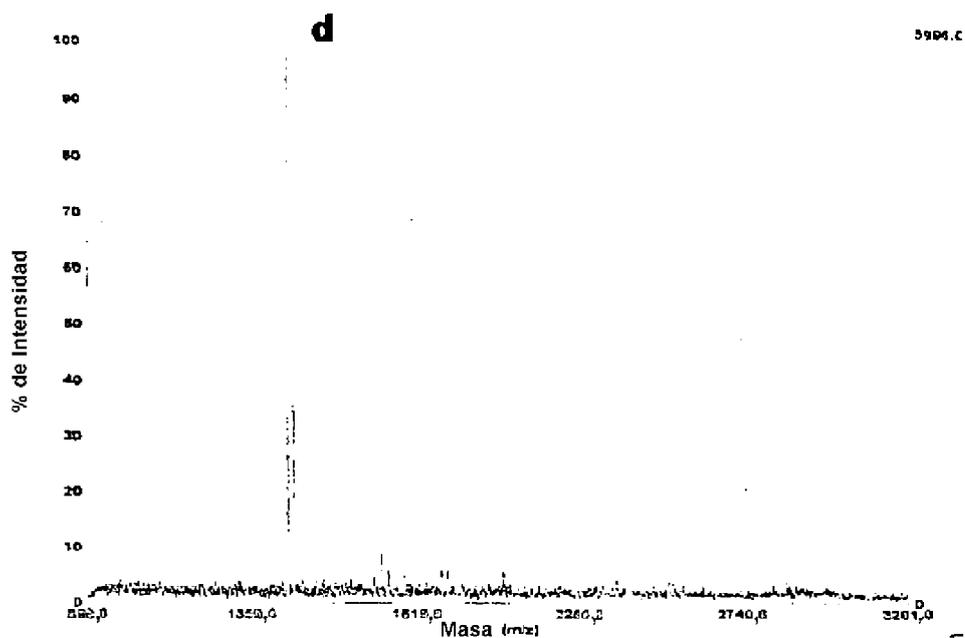
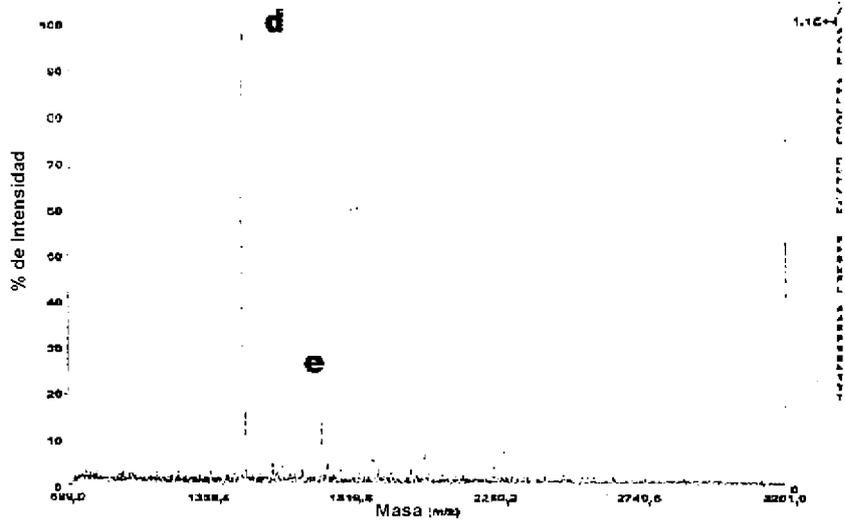
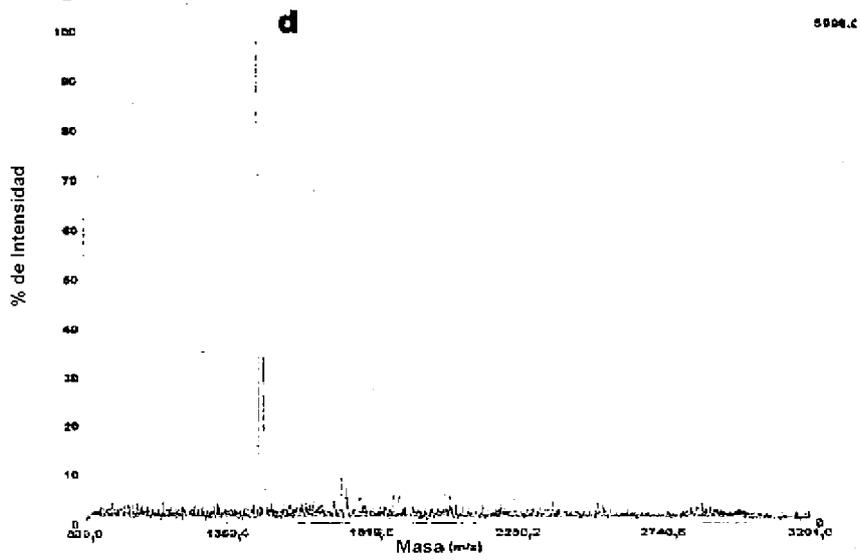


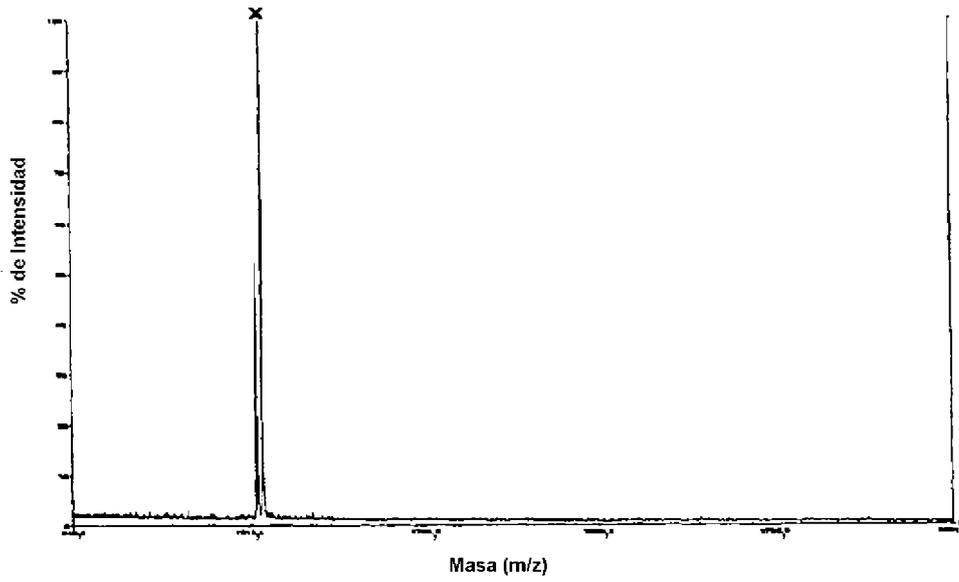
Fig. 27



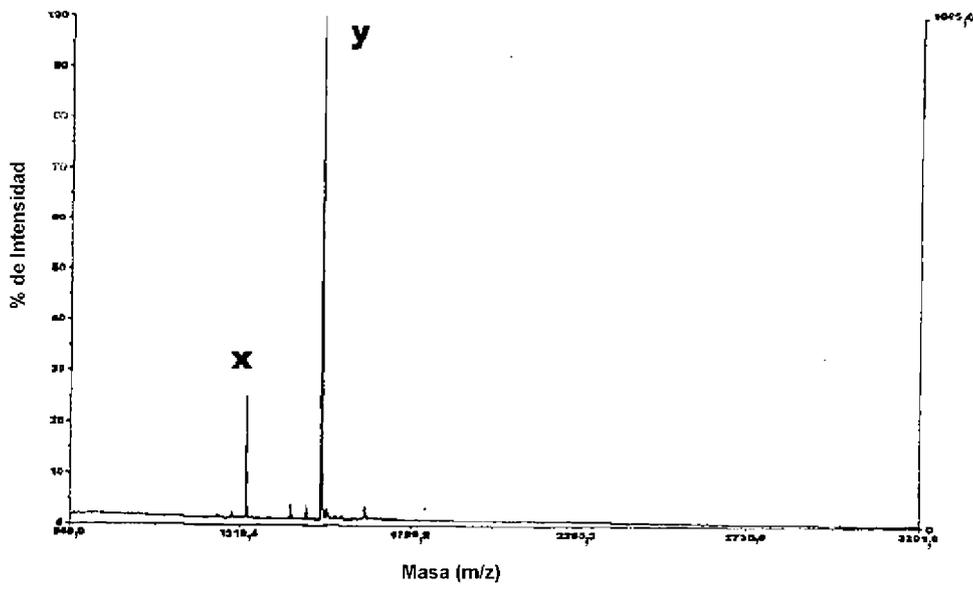
**Fig. 28**



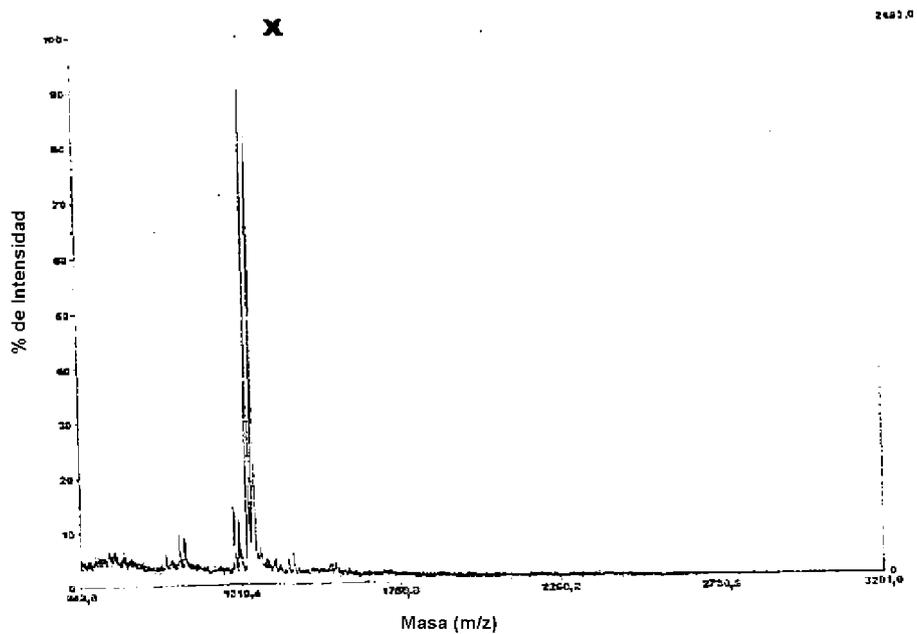
**Fig. 29**



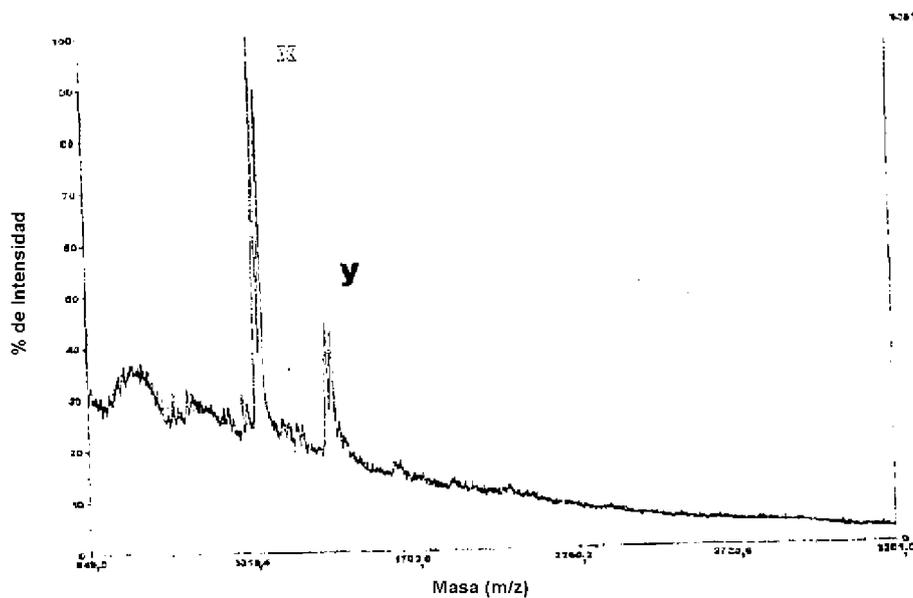
**Fig. 30**

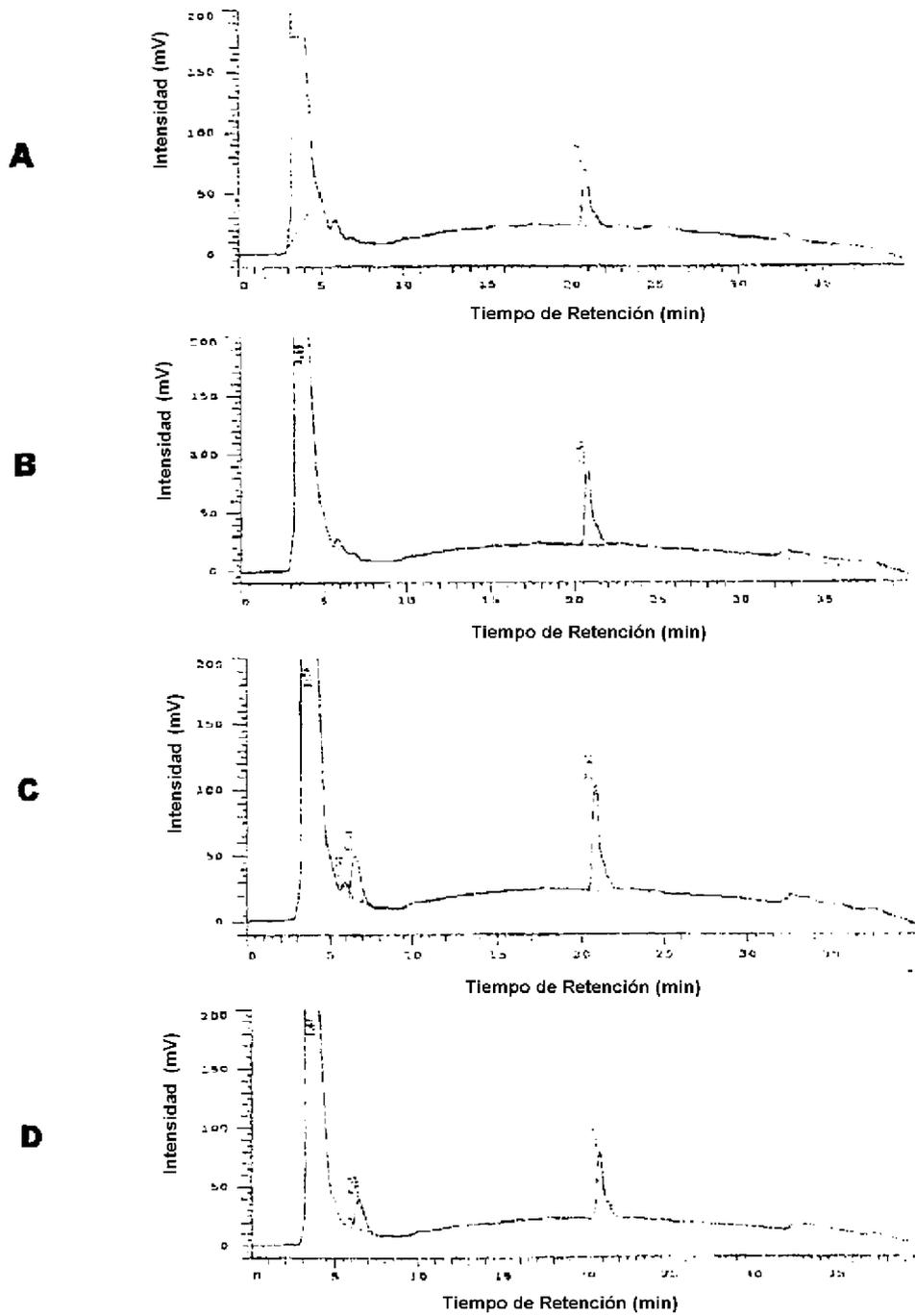


**Fig. 31**

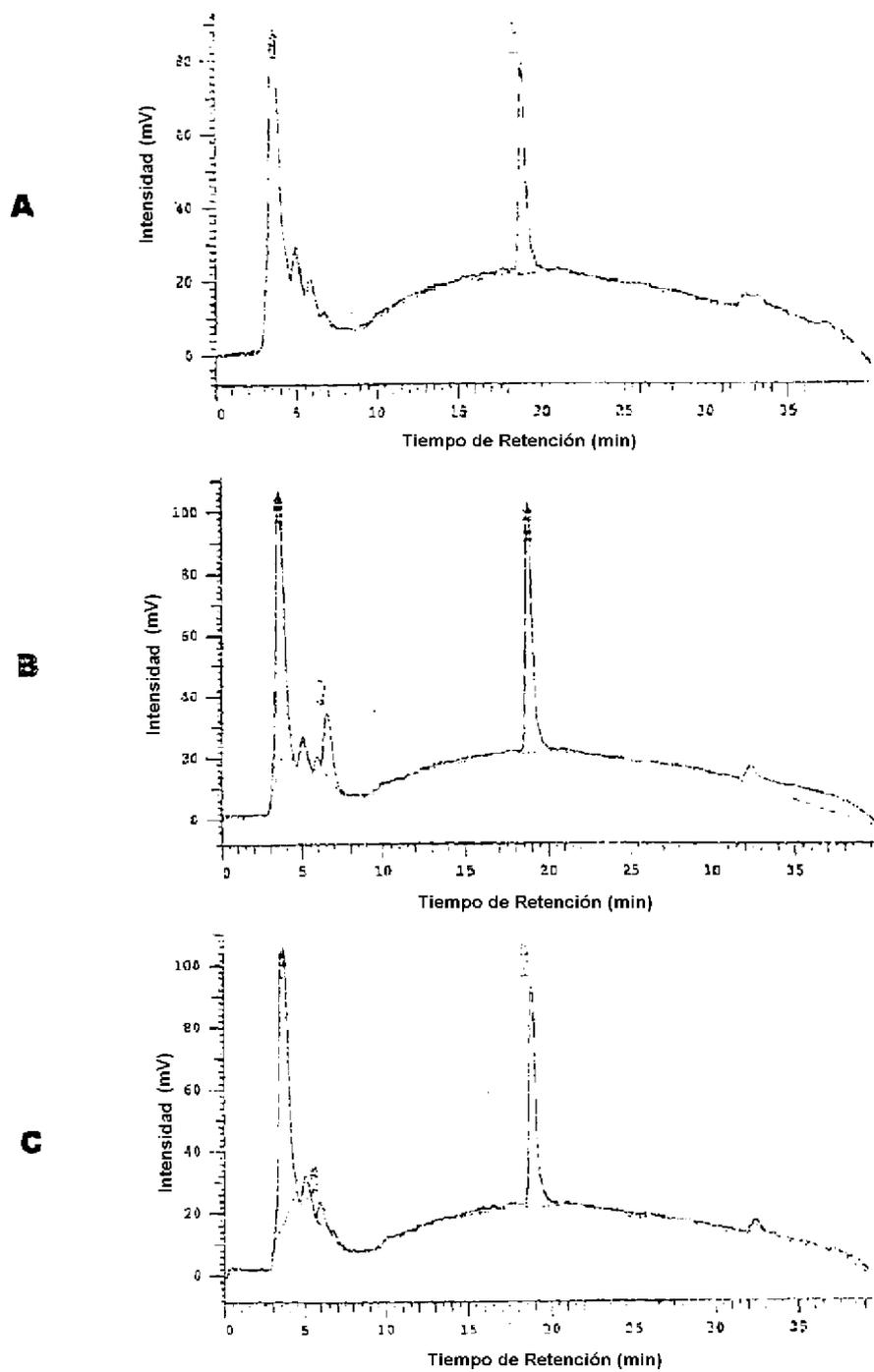


**Fig. 32**

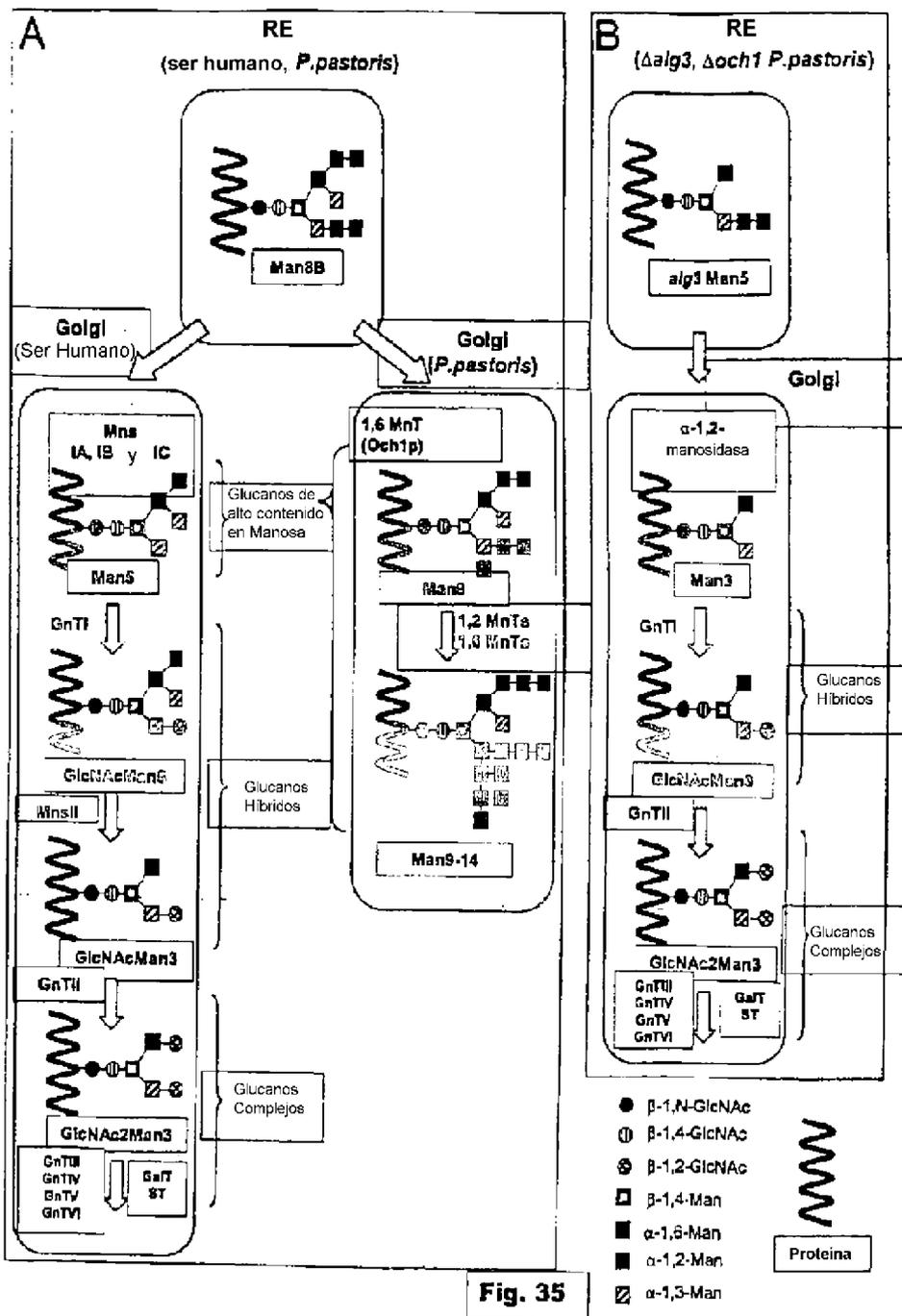




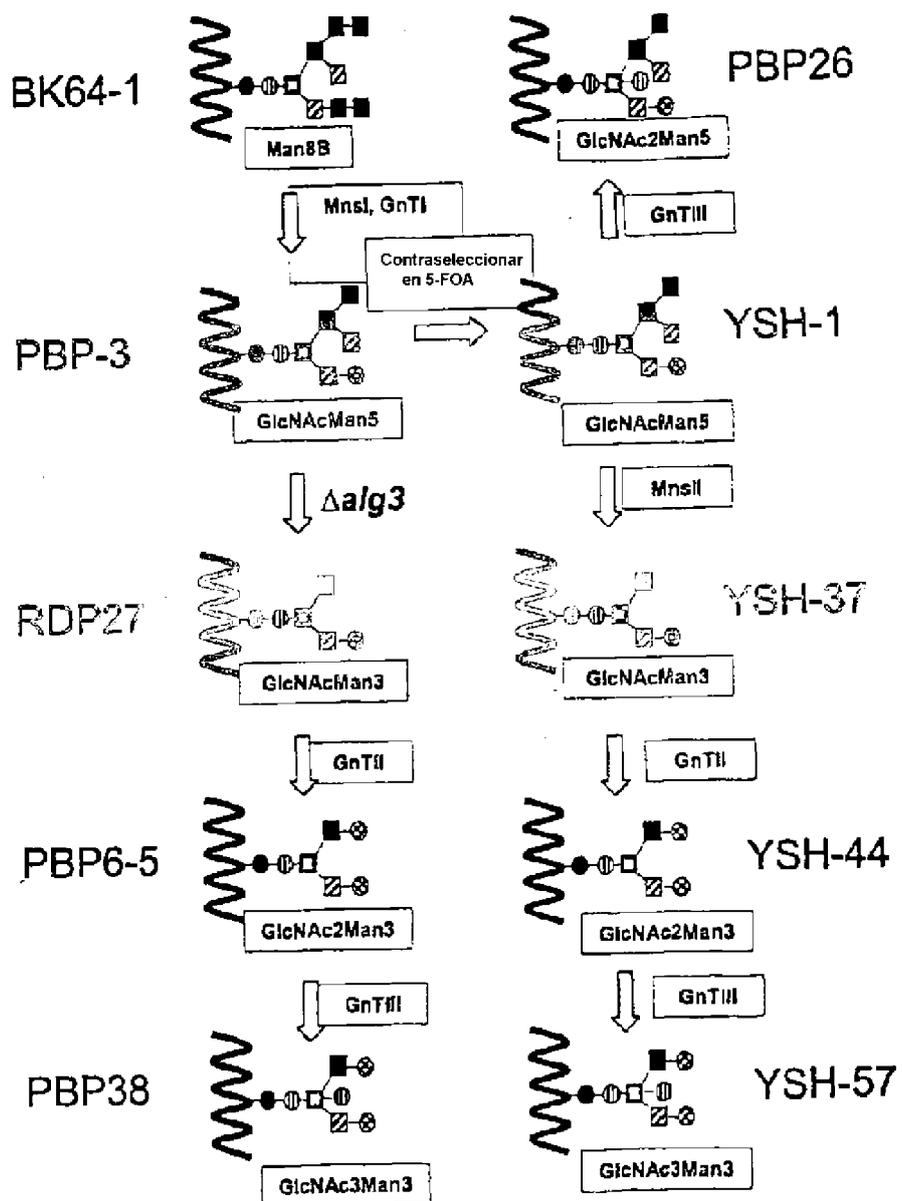
**Fig. 33**



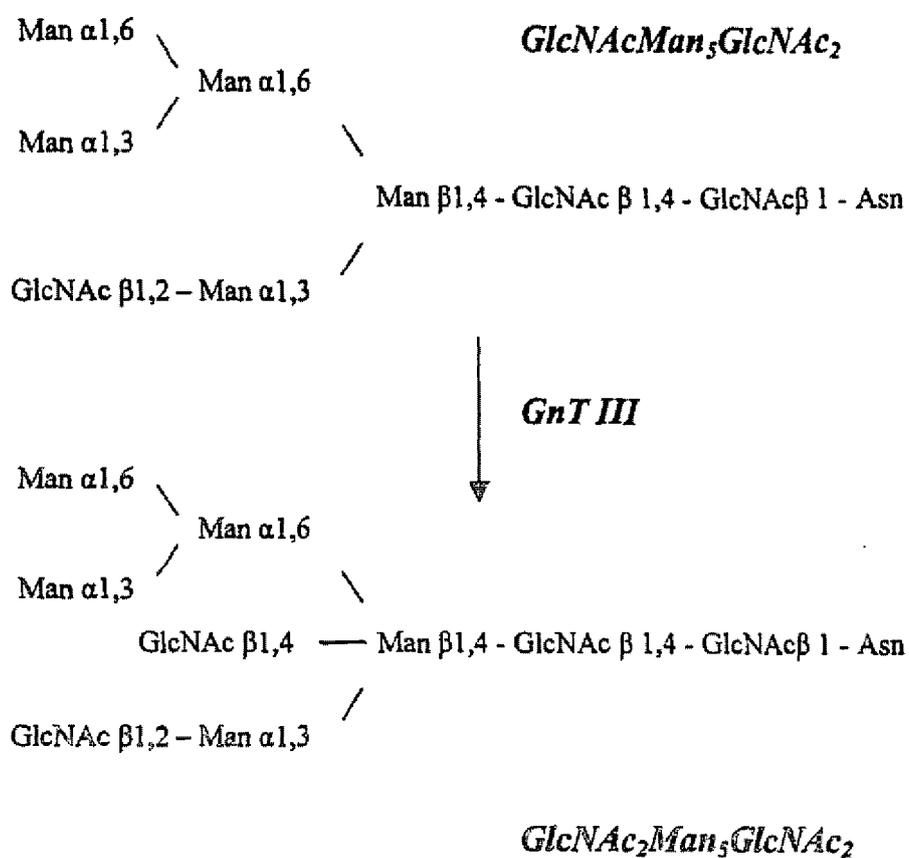
**Fig. 34**



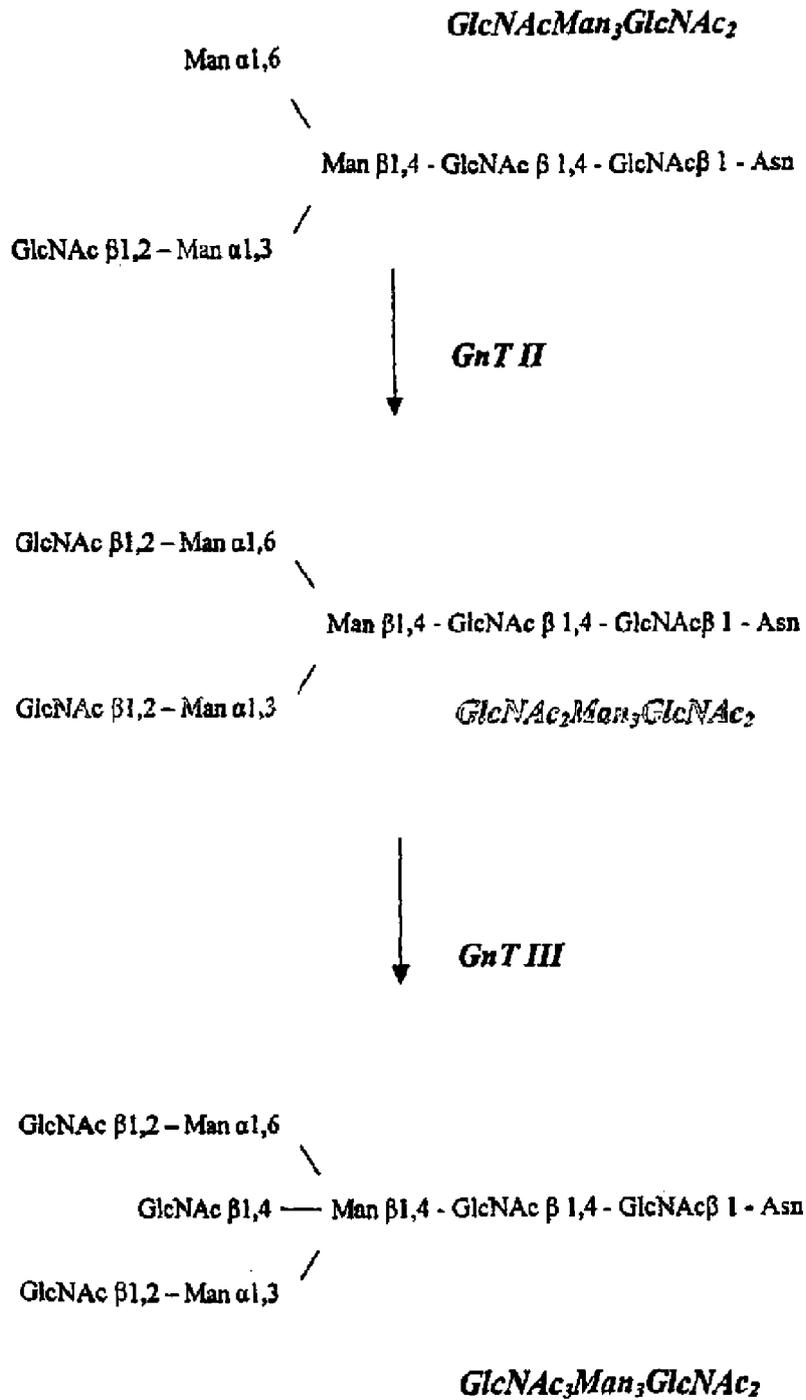
**Fig. 36**



**Fig. 37**



**Fig. 38**



# ES 2 335 345 T3

## LISTA DE SECUENCIAS

- <110> BOBROWICZ, PIOTR
- 5 <120> EXPRESIÓN DE *N*-ACETILGLUCOSAMINILTRANSFERASA III EN EUCARIOTAS INFERIORES
- <130> L2575 EPS3
- 10 <140> PCT/US04/005128
- <141> 20-02-2004
- <150> 10/680.963
- 15 <151> 07-10-2003
- <150> 10/371.877
- 20 <151> 20-02-2003
- <150> 09/892.591
- <151> 27-06-2001
- 25 <150> 60/214.358
- <151> 28-06-2000
- <150> 60/215.638
- 30 <151> 30-06-2000
- <150> 60/279.997
- 35 <151> 30-03-2001
- <150> PCT/US02/41510
- <151> 24-12-2002
- 40 <150> 60/344.169
- <151> 27-12-2001
- 45 <160> 95
- <170> PatentInVer. 3.2
- <210> 1
- 50 <211> 3
- <212> PRT
- <213> Secuencia Artificial
- 55 <220>
- <223> Descripción de Secuencia Artificial: péptido sintético
- <220>
- 60 <221> MOD\_RES
- <222> (2)
- <223> Cualquier aminoácido excepto Pro
- 65

## ES 2 335 345 T3

<400> 1  
Asn Xaa Ser  
1  
5  
<210> 2  
<211> 3  
<212> PRT  
10 <213> Secuencia Artificial  
  
<220>  
15 <223> Descripción de Secuencia Artificial: péptido sintético  
<220>  
<221> MOD\_RES  
  
20 <222> (2)  
<223> Cualquier aminoácido excepto Pro  
  
<400> 2  
25 Asn Xaa Thr  
1  
  
30 <210> 3  
<211> 21  
<212> ADN  
<213> Secuencia Artificial  
  
35 <220>  
<223> Descripción de Secuencia Artificial: cebador sintético  
  
40 <400> 3  
  
atggcgaagg cagatggcag t 21  
  
45 <210> 4  
<211> 21  
<212> ADN  
<213> Secuencia Artificial  
  
50 <220>  
<223> Descripción de Secuencia Artificial: cebador sintético  
  
55 <400> 4  
  
ttagtccttc caacttcctt c 21  
  
60 <210> 5  
<211> 26  
<212> ADN  
<213> Secuencia Artificial  
  
65 <220>  
<223> Descripción de Secuencia Artificial: cebador sintético

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <220>  |    |
|    | <221> base modificada  |    |
|    | <222> (9)  |    |
| 5  | <223> g, a, t o c  |    |
|    | <220>  |    |
|    | <221> base modificada  |    |
| 10 | <222> (12)   |    |
|    | <223> g, a, t o c  |    |
|    | <220>  |    |
| 15 | <221> base modificada  |    |
|    | <222> (18)   |    |
|    | <223> g, a, t o c  |    |
| 20 | <400> 5  |    |
|    | taytgmgng tngarcynga yathaa  | 26 |
| 25 | <210> 6  |    |
|    | <211> 20   |    |
|    | <212> ADN  |    |
| 30 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
| 35 | <200>  |    |
|    | <221> base modificada  |    |
|    | <222> (6)  |    |
| 40 | <223> g, a, t o c  |    |
|    | <220>  |    |
|    | <221> base modificada  |    |
| 45 | <222> (12)   |    |
|    | <223> g, a, t o c  |    |
|    | <400> 6  |    |
| 50 | gertncccc anckytarta   | 20 |
|    | <210> 7  |    |
| 55 | <211> 39   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 60 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
|    | <400> 7  |    |
| 65 | ggccgcctgc agatttaaat gaattggcg cgcctaat                             | 39 |

## ES 2 335 345 T3

<210> 8  
<211> 34  
<212> ADN  
5 <213> Secuencia Artificial  
  
<220>  
<223> Descripción de Secuencia Artificial: oligonucleótido sintético  
10  
<400> 8

15 taaggcgcgc cgaattcatt taaatctgca gggc

34

<210> 9  
<211> 458  
20 <212> PRT  
<213> *Saccharomyces cerevisiae*

<220>  
25 <221> MOD\_RES  
  
<222> (304)..(318)  
<223> Aminoácido variable  
30

<220>  
<221> MOD\_RES  
35 <222> (416)..(436)  
<223> Aminoácido variable

40

45

50

55

60

65

# ES 2 335 345 T3

<400> 9

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5  | Met | Glu | Gly | Glu | Gln | Ser | Pro | Gln | Gly | Glu | Lys | Ser | Leu | Gln | Arg | Lys |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
|    | Gln | Phe | Val | Arg | Pro | Pro | Leu | Asp | Leu | Trp | Gln | Asp | Leu | Lys | Asp | Gly |
|    |     |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| 10 | Val | Arg | Tyr | Val | Ile | Phe | Asp | Cys | Arg | Ala | Asn | Leu | Ile | Val | Met | Pro |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    | Leu | Leu | Ile | Leu | Phe | Glu | Ser | Met | Leu | Cys | Lys | Ile | Ile | Ile | Lys | Lys |
| 15 |     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |
|    | Val | Ala | Tyr | Thr | Glu | Ile | Asp | Tyr | Lys | Ala | Tyr | Met | Glu | Gln | Ile | Glu |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| 20 | Met | Ile | Gln | Leu | Asp | Gly | Met | Leu | Asp | Tyr | Ser | Gln | Val | Ser | Gly | Gly |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Gly | His | Val | Leu | Ile | Tyr | Lys | Met |
| 25 |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
|    | Met | Tyr | Trp | Leu | Thr | Glu | Gly | Met | Asp | His | Val | Glu | Arg | Gly | Gln | Val |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 30 | Phe | Phe | Arg | Tyr | Leu | Tyr | Leu | Leu | Thr | Leu | Ala | Leu | Gln | Met | Ala | Cys |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | Tyr | Tyr | Leu | Leu | His | Leu | Pro | Pro | Trp | Cys | Val | Val | Leu | Ala | Cys | Leu |
| 35 | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Ser | Lys | Arg | Leu | His | Ser | Ile | Tyr | Val | Leu | Arg | Leu | Phe | Asn | Asp | Cys |
|    |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| 40 | Phe | Thr | Thr | Leu | Phe | Met | Val | Val | Thr | Val | Leu | Gly | Ala | Ile | Val | Ala |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
|    | Ser | Arg | Cys | His | Gln | Arg | Pro | Lys | Leu | Lys | Lys | Ser | Leu | Ala | Leu | Val |
| 45 |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Ile | Ser | Ala | Thr | Tyr | Ser | Met | Ala | Val | Ser | Ile | Lys | Met | Asn | Ala | Leu |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 50 | Leu | Tyr | Phe | Pro | Ala | Met | Met | Ile | Ser | Leu | Phe | Ile | Leu | Asn | Asp | Ala |

# ES 2 335 345 T3

|    |   |   |     |     |
|----|---|---|-----|-----|
|    | 225   | 230   | 235 | 240 |
| 5  | Asn Val Ile Leu Thr   | Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln |     |     |
|    |   | 245   | 250 | 255 |
|    | Val Ala Val Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu | 260   | 265 | 270 |
| 10 | His Cys Ala Phe Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile | 275   | 280 | 285 |
| 15 | Asn Trp Gln Met Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe Xaa | 290   | 295 | 300 |
|    | Xaa Phe Val | 305   | 310 | 315 |
| 20 | Thr Arg Tyr Pro Arg Ile Leu Pro Asp Leu Trp Ser Ser Leu Cys His | 325   | 330 | 335 |
| 25 | Pro Leu Arg Lys Asn Ala Val Leu Asn Ala Asn Pro Ala Lys Thr Ile | 340   | 345 | 350 |
|    | Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser Arg | 355   | 360 | 365 |
| 30 | Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Ile | 370   | 375 | 380 |
| 35 | Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp Tyr | 385   | 390 | 395 |
|    | Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ser Gln Xaa | 405   | 410 | 415 |
| 40 | Xaa | 420   | 425 | 430 |
| 45 | Xaa Xaa Xaa Xaa Ser Gly Ser Val Ala Leu Ala Lys Ser His Leu Arg | 435   | 440 | 445 |
|    | Thr Thr Ser Ser Met Glu Lys Lys Leu Asn                         | 450   | 455 |     |
| 50 | <210> 10  |   |     |     |
|    | <211> 458   |   |     |     |
|    | <212> PRT   |   |     |     |
| 55 | <213> <i>Saccharomyces cerevisiae</i>                           |   |     |     |
|    | <400> 10  |   |     |     |
| 60 | Met Glu Gly Glu Gln Ser Pro Gln Gly Glu Lys Ser Leu Gln Arg Lys | 1   | 5   | 10  |
|    | Gln Phe Val Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly | 20  | 25  | 30  |
| 65 | Val Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro | 35  | 40  | 45  |

## ES 2 335 345 T3

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Leu | Leu | Ile | Leu | Phe | Glu | Ser | Met | Leu | Cys | Lys | Ile | Ile | Ile | Lys | Lys |
|    |     | 50  |     |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |
| 5  | Val | Ala | Tyr | Thr | Glu | Ile | Asp | Tyr | Lys | Ala | Tyr | Met | Glu | Gln | Ile | Glu |
|    |     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |
|    | Met | Ile | Gln | Leu | Asp | Gly | Met | Leu | Asp | Tyr | Ser | Gln | Val | Ser | Gly | Gly |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| 10 | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Gly | His | Val | Leu | Ile | Tyr | Lys | Met |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
|    | Met | Tyr | Trp | Leu | Thr | Glu | Gly | Met | Asp | His | Val | Glu | Arg | Gly | Gln | Val |
| 15 |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Phe | Phe | Arg | Tyr | Leu | Tyr | Leu | Leu | Thr | Leu | Ala | Leu | Gln | Met | Ala | Cys |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 20 | Tyr | Tyr | Leu | Leu | His | Leu | Pro | Pro | Trp | Cys | Val | Val | Leu | Ala | Cys | Leu |
|    |     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Ser | Lys | Arg | Leu | His | Ser | Ile | Tyr | Val | Leu | Arg | Leu | Phe | Asn | Asp | Cys |
| 25 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Phe | Thr | Thr | Leu | Phe | Met | Val | Val | Thr | Val | Leu | Gly | Ala | Ile | Val | Ala |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
|    | Ser | Arg | Cys | His | Gln | Arg | Pro | Lys | Leu | Lys | Lys | Ser | Leu | Ala | Leu | Val |
| 30 |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Ile | Ser | Ala | Thr | Tyr | Ser | Met | Ala | Val | Ser | Ile | Lys | Met | Asn | Ala | Leu |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 35 | Leu | Tyr | Phe | Pro | Ala | Met | Met | Ile | Ser | Leu | Phe | Ile | Leu | Asn | Asp | Ala |
|    |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
|    | Asn | Val | Ile | Leu | Thr | Leu | Leu | Asp | Leu | Val | Ala | Met | Ile | Ala | Trp | Gln |
| 40 |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Val | Ala | Val | Ala | Val | Pro | Phe | Leu | Arg | Ser | Phe | Pro | Gln | Gln | Tyr | Leu |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 45 | His | Cys | Ala | Phe | Asn | Phe | Gly | Arg | Lys | Phe | Met | Tyr | Gln | Trp | Ser | Ile |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|    | Asn | Trp | Gln | Met | Met | Asp | Glu | Glu | Ala | Phe | Asn | Asp | Lys | Arg | Phe | His |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| 50 | Leu | Ala | Leu | Leu | Ile | Ser | His | Leu | Ile | Ala | Leu | Thr | Thr | Leu | Phe | Val |
|    |     | 305 |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
|    | Thr | Arg | Tyr | Pro | Arg | Ile | Leu | Pro | Asp | Leu | Trp | Ser | Ser | Leu | Cys | His |
| 55 |     |     |     |     | 325 |     |     |     |     |     | 330 |     |     |     | 335 |     |
|    | Pro | Leu | Arg | Lys | Asn | Ala | Val | Leu | Asn | Ala | Asn | Pro | Ala | Lys | Thr | Ile |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| 60 | Pro | Phe | Val | Leu | Ile | Ala | Ser | Asn | Phe | Ile | Gly | Val | Leu | Phe | Ser | Arg |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |

65

# ES 2 335 345 T3

Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Ile  
 370 375 380

5
   
Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp Tyr  
 385 390 395 400

Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ser Gln Ala  
 405 410 415

10
   
Ser Thr Leu Leu Leu Ala Leu Asn Thr Val Leu Leu Leu Leu Ala  
 420 425 430

15
   
Leu Thr Gln Leu Ser Gly Ser Val Ala Leu Ala Lys Ser His Leu Arg  
 435 440 445

Thr Thr Ser Ser Met Glu Lys Lys Leu Asn  
 450 455

20
   
<210> 11  
 <211> 389  
 <212> PRT

25
   
<213> *Saccharomyces cerevisiae*  
 <220>  
 <221> MOD\_RES  
 <222> (279)..(293)  
 <223> Aminoácido variable

30
   
<400> 11

35
   
Trp Gln Asp Leu Lys Asp Gly Val Arg Tyr Val Ile Phe Asp Cys Arg  
 1 5 10 15

40
   
Ala Asn Leu Ile Val Met Pro Leu Leu Ile Leu Phe Glu Ser Met Leu  
 20 25 30

Cys Lys Ile Ile Ile Lys Lys Val Ala Tyr Thr Glu Ile Asp Tyr Lys  
 35 40 45

45
   
Ala Tyr Met Glu Gln Ile Glu Met Ile Gln Leu Asp Gly Met Leu Asp  
 50 55 60

50
   
Tyr Ser Gln Val Ser Gly Gly Thr Gly Pro Leu Val Tyr Pro Ala Gly  
 65 70 75 80

His Val Leu Ile Tyr Lys Met Met Tyr Trp Leu Thr Glu Gly Met Asp  
 85 90 95

55
   
His Val Glu Arg Gly Gln Val Phe Phe Arg Tyr Leu Tyr Leu Leu Thr  
 100 105 110

60
   
Leu Ala Leu Gln Met Ala Cys Tyr Tyr Leu Leu His Leu Pro Pro Trp  
 115 120 125

Cys Val Val Leu Ala Cys Leu Ser Lys Arg Leu His Ser Ile Tyr Val  
 130 135 140

65
   
Leu Arg Leu Phe Asn Asp Cys Phe Thr Thr Leu Phe Met Val Val Thr  
 145 150 155 160

# ES 2 335 345 T3

Val Leu Gly Ala Ile Val Ala Ser Arg Cys His Gln Arg Pro Lys Leu  
 165 170 175  
 5 Lys Lys Ser Leu Ala Leu Val Ile Ser Ala Thr Tyr Ser Met Ala Val  
 180 185 190  
 10 Ser Ile Lys Met Asn Ala Leu Leu Tyr Phe Pro Ala Met Met Ile Ser  
 195 200 205  
 Leu Phe Ile Leu Asn Asp Ala Asn Val Ile Leu Thr Leu Leu Asp Leu  
 210 215 220  
 15 Val Ala Met Ile Ala Trp Gln Val Ala Val Ala Val Pro Phe Leu Arg  
 225 230 235 240  
 20 Ser Phe Pro Gln Gln Tyr Leu His Cys Ala Phe Asn Phe Gly Arg Lys  
 245 250 255  
 Phe Met Tyr Gln Trp Ser Ile Asn Trp Gln Met Met Asp Glu Glu Ala  
 260 265 270  
 25 Phe Asn Asp Lys Arg Phe Xaa  
 275 280 285  
 Xaa Xaa Xaa Xaa Xaa Phe Val Thr Arg Tyr Pro Arg Ile Leu Pro Asp  
 290 295 300  
 30 Leu Trp Ser Ser Leu Cys His Pro Leu Arg Lys Asn Ala Val Leu Asn  
 305 310 315 320  
 35 Ala Asn Pro Ala Lys Thr Ile Pro Phe Val Leu Ile Ala Ser Asn Phe  
 325 330 335  
 Ile Gly Val Leu Phe Ser Arg Ser Leu His Tyr Gln Phe Leu Ser Trp  
 340 345 350  
 40 Tyr His Trp Thr Leu Pro Ile Leu Ile Phe Trp Ser Gly Met Pro Phe  
 355 360 365  
 45 Phe Val Gly Pro Ile Trp Tyr Val Leu His Glu Trp Cys Trp Asn Ser  
 370 375 380  
 Tyr Pro Pro Asn Ser  
 385  
 50 <210> 12  
 <211> 375  
 <212> PRT  
 55 <213> *Homo sapiens*  
 <400> 12  
 60 Trp Gln Glu Arg Arg Leu Leu Leu Arg Glu Pro Arg Tyr Thr Leu Leu  
 1 5 10 15  
 Val Ala Ala Cys Leu Cys Leu Ala Glu Val Gly Ile Thr Phe Trp Val  
 20 25 30  
 65 Ile His Arg Val Ala Tyr Thr Glu Ile Asp Trp Lys Ala Tyr Met Ala

# ES 2 335 345 T3

|    | 35  |     | 40  |     | 45  |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Glu | Val | Glu | Gly | Val | Ile | Asn | Gly | Thr | Tyr | Asp | Tyr | Thr | Gln | Leu | Gln |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 5  | Gly | Asp | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Gly | Phe | Val | Tyr | Ile | Phe |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
|    | Met | Gly | Leu | Tyr | Tyr | Ala | Thr | Ser | Arg | Gly | Thr | Asp | Ile | Arg | Met | Ala |
| 10 |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Gln | Asn | Ile | Phe | Ala | Val | Leu | Tyr | Leu | Ala | Thr | Leu | Leu | Leu | Val | Phe |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| 15 | Leu | Ile | Tyr | His | Gln | Thr | Cys | Lys | Val | Pro | Pro | Phe | Val | Phe | Phe | Phe |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Met | Cys | Cys | Ala | Ser | Tyr | Arg | Val | His | Ser | Ile | Phe | Val | Leu | Arg | Leu |
| 20 |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | Phe | Asn | Asp | Pro | Val | Ala | Met | Val | Leu | Leu | Phe | Leu | Ser | Ile | Asn | Leu |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Leu | Leu | Ala | Gln | Arg | Trp | Gly | Trp | Gly | Cys | Cys | Phe | Phe | Ser | Leu | Ala |
| 25 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Val | Ser | Val | Lys | Met | Asn | Val | Leu | Leu | Phe | Ala | Pro | Gly | Leu | Leu | Phe |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 30 | Leu | Leu | Leu | Thr | Gln | Phe | Gly | Phe | Arg | Gly | Ala | Leu | Pro | Lys | Leu | Gly |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Ile | Cys | Ala | Gly | Leu | Gln | Val | Val | Leu | Gly | Leu | Pro | Phe | Leu | Leu | Glu |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 35 | Asn | Pro | Ser | Gly | Tyr | Leu | Ser | Arg | Ser | Phe | Asp | Leu | Gly | Arg | Gln | Phe |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
|    | Leu | Phe | His | Trp | Thr | Val | Asn | Trp | Arg | Phe | Leu | Pro | Glu | Ala | Leu | Phe |
| 40 |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Leu | His | Arg | Ala | Phe | His | Leu | Ala | Leu | Leu | Thr | Ala | His | Leu | Thr | Leu |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 45 | Leu | Leu | Leu | Phe | Ala | Leu | Cys | Arg | Trp | His | Arg | Thr | Gly | Glu | Ser | Ile |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|    | Leu | Ser | Leu | Leu | Arg | Asp | Pro | Ser | Lys | Arg | Lys | Val | Pro | Pro | Gln | Pro |
| 50 |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
|    | Leu | Thr | Pro | Asn | Gln | Ile | Val | Ser | Thr | Leu | Phe | Thr | Ser | Asn | Phe | Ile |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
|    | Gly | Ile | Cys | Phe | Ser | Arg | Ser | Leu | His | Tyr | Gln | Phe | Tyr | Val | Trp | Tyr |
| 55 |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
|    | Phe | His | Thr | Leu | Pro | Tyr | Leu | Leu | Trp | Ala | Met | Pro | Ala | Arg | Trp | Leu |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| 60 | Thr | His | Leu | Leu | Arg | Leu | Leu | Val | Leu | Gly | Leu | Ile | Glu | Leu | Ser | Trp |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
|    | Asn | Thr | Tyr | Pro | Ser | Thr | Ser |     |     |     |     |     |     |     |     |     |
| 65 |     | 370 |     |     |     |     | 375 |     |     |     |     |     |     |     |     |     |

<210> 13  
 <211> 271

# ES 2 335 345 T3

<212> PRT

<213> *Saccharomyces cerevisiae*

5 <400> 13

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 10 | Val | Arg | Tyr | Val | Ile | Phe | Asp | Cys | Arg | Ala | Asn | Leu | Ile | Val | Met | Pro |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
|    | Leu | Leu | Ile | Leu | Phe | Glu | Ser | Met | Leu | Cys | Lys | Ile | Ile | Ile | Lys | Lys |
|    |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| 15 | Val | Ala | Tyr | Thr | Glu | Ile | Asp | Tyr | Lys | Ala | Tyr | Met | Glu | Gln | Ile | Glu |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    | Met | Ile | Gln | Leu | Asp | Gly | Met | Leu | Asp | Tyr | Ser | Gln | Val | Ser | Gly | Gly |
| 20 |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
|    | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Gly | His | Val | Leu | Ile | Tyr | Lys | Met |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
|    | Met | Tyr | Trp | Leu | Thr | Glu | Gly | Met | Asp | His | Val | Glu | Arg | Gly | Gln | Val |
| 25 |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Phe | Phe | Arg | Tyr | Leu | Tyr | Leu | Leu | Thr | Leu | Ala | Leu | Gln | Met | Ala | Cys |
| 30 |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
|    | Tyr | Tyr | Leu | Leu | His | Leu | Pro | Pro | Trp | Cys | Val | Val | Leu | Ala | Cys | Leu |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Ser | Lys | Arg | Leu | His | Ser | Ile | Tyr | Val | Leu | Arg | Leu | Phe | Asn | Asp | Cys |
| 35 |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | Phe | Thr | Thr | Leu | Phe | Met | Val | Val | Thr | Val | Leu | Gly | Ala | Ile | Val | Ala |
| 40 |     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
|    | Ser | Arg | Cys | His | Gln | Arg | Pro | Lys | Leu | Lys | Lys | Ser | Leu | Ala | Leu | Val |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Ile | Ser | Ala | Thr | Tyr | Ser | Met | Ala | Val | Ser | Ile | Lys | Met | Asn | Ala | Leu |
| 45 |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
|    | Leu | Tyr | Phe | Pro | Ala | Met | Met | Ile | Ser | Leu | Phe | Ile | Leu | Asn | Asp | Ala |
| 50 |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Asn | Val | Ile | Leu | Thr | Leu | Leu | Asp | Leu | Val | Ala | Met | Ile | Ala | Trp | Gln |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
|    | Val | Ala | Val | Ala | Val | Pro | Phe | Leu | Arg | Ser | Phe | Pro | Gln | Gln | Tyr | Leu |
| 55 |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
|    | His | Cys | Ala | Phe | Asn | Phe | Gly | Arg | Lys | Phe | Met | Tyr | Gln | Trp | Ser | Ile |
| 60 |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Asn | Trp | Gln | Met | Met | Asp | Glu | Glu | Ala | Phe | Asn | Asp | Lys | Arg | Phe |     |
| 65 |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |

ES 2 335 345 T3

<210> 14

<211> 258

<212> PRT

5 <213> *Drosophila virilis*

<400> 14

10

Ile Lys Tyr Leu Ala Phe Glu Pro Ala Ala Leu Pro Ile Val Ser Val  
1 5 10 15

15

Leu Ile Val Leu Ala Glu Ala Val Ile Asn Val Leu Val Ile Gln Arg  
20 25 30

Val Pro Tyr Thr Glu Ile Asp Trp Lys Ala Tyr Met Gln Glu Cys Glu  
35 40 45

20

Gly Phe Leu Asn Gly Thr Thr Asn Tyr Ser Leu Leu Arg Gly Asp Thr  
50 55 60

25

Gly Pro Leu Val Tyr Pro Ala Ala Phe Val Tyr Ile Tyr Ser Gly Leu  
65 70 75 80

Tyr Tyr Leu Thr Gly Gln Gly Thr Asn Val Arg Leu Ala Gln Tyr Ile  
85 90 95

30

Phe Ala Cys Ile Tyr Leu Leu Gln Met Cys Leu Val Leu Arg Leu Tyr  
100 105 110

35

Thr Lys Ser Arg Lys Val Pro Pro Tyr Val Leu Val Leu Ser Ala Phe  
115 120 125

Thr Ser Tyr Arg Ile His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp  
130 135 140

40

Pro Val Ala Ile Leu Leu Leu Tyr Ala Ala Leu Asn Leu Phe Leu Asp  
145 150 155 160

45

Gln Arg Trp Thr Leu Gly Ser Ile Cys Tyr Ser Leu Ala Val Gly Val  
165 170 175

Lys Met Asn Ile Leu Leu Phe Ala Pro Ala Leu Leu Leu Phe Tyr Leu  
180 185 190

50

Ala Asn Leu Gly Val Leu Arg Thr Leu Val Gln Leu Thr Ile Cys Ala  
195 200 205

55

Val Leu Gln Leu Phe Ile Gly Ala Pro Phe Leu Arg Thr His Pro Met  
210 215 220

Glu Tyr Leu Arg Gly Ser Phe Asp Leu Gly Arg Ile Phe Glu His Lys  
225 230 235 240

60

Trp Thr Val Asn Tyr Arg Phe Leu Ser Lys Glu Leu Phe Glu Gln Arg  
245 250 255

Glu Phe

65

ES 2 335 345 T3

<210> 15

<211> 60

<212> PRT

5 <213> *Saccharomyces cerevisiae*

<400> 15

10 Ile Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser  
 1 5 10 15  
 15 Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro  
 20  
 25 Ile Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp  
 35 40 45  
 30 Tyr Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro  
 50 55 60

<210> 16

<211> 58

<212> PRT

<213> *Drosophila virilis*

30

<400> 16

35 Leu Pro Phe Phe Leu Cys Asn Phe Ile Gly Val Ala Cys Ala Arg Ser  
 1 5 10 15  
 40 Leu His Tyr Gln Phe Tyr Ile Trp Tyr Phe His Ser Leu Pro Tyr Leu  
 20 25 30  
 45 Val Trp Ser Thr Pro Tyr Ser Leu Gly Val Arg Tyr Leu Ile Leu Gly  
 35 40 45  
 50 Ile Ile Glu Tyr Cys Trp Asn Thr Tyr Pro  
 55

<210> 17

<211> 270

50 <212> PRT

<213> *Saccharomyces cerevisiae*

55

60

65

# ES 2 335 345 T3

<400> 17

```

5      Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro Leu
      1          5          10          15
      Leu Ile Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys Val
      20          25          30
10     Ala Tyr Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu Met
      35          40          45
      Ile Gln Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly Thr
15     50          55          60
      Gly Pro Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met Met
      65          70          75          80
20
      Tyr Trp Leu Thr Glu Gly Met Asp His Val Glu Arg Gly Gln Val Phe
25     85          90          95
      Phe Arg Tyr Leu Tyr Leu Leu Thr Leu Ala Leu Gln Met Ala Cys Tyr
      100         105         110
30     Tyr Leu Leu His Leu Pro Pro Trp Cys Val Val Leu Ala Cys Leu Ser
      115         120         125
      Lys Arg Leu His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Cys Phe
35     130         135         140
      Thr Thr Leu Phe Met Val Val Thr Val Leu Gly Ala Ile Val Ala Ser
40     145         150         155         160
      Arg Cys His Gln Arg Pro Lys Leu Lys Lys Ser Leu Ala Leu Val Ile
      165         170         175
45     Ser Ala Thr Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu Leu
      180         185         190
      Tyr Phe Pro Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala Asn
50     195         200         205
      Val Ile Leu Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln Val
      210         215         220
55     Ala Val Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu His
      225         230         235         240
      Cys Ala Phe Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile Asn
      245         250         255
60     Trp Gln Met Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe
      260         265         270

```

<210> 18

<211> 257

65 <212> PRT

<213> *Drosophila melanogaster*

ES 2 335 345 T3

<400> 18

5 Lys Tyr Leu Leu Leu Glu Pro Ala Ala Leu Pro Ile Val Gly Leu Phe  
 1 5 10 15  
 Val Leu Leu Ala Glu Leu Val Ile Asn Val Val Val Ile Gln Arg Val  
 20 25 30  
 10 Pro Tyr Thr Glu Ile Asp Trp Val Ala Tyr Met Gln Glu Cys Glu Gly  
 35 40 45  
 15 Phe Leu Asn Gly Thr Thr Asn Tyr Ser Leu Leu Arg Gly Asp Thr Gly  
 50 55 60  
 Pro Leu Val Tyr Pro Ala Ala Phe Val Tyr Ile Tyr Ser Ala Leu Tyr  
 65 70 75 80  
 20 Tyr Val Thr Ser His Gly Thr Asn Val Arg Leu Ala Gln Tyr Ile Phe  
 85 90 95  
 25 Ala Gly Ile Tyr Leu Leu Gln Leu Ala Leu Val Leu Arg Leu Tyr Ser  
 100 105 110  
 Lys Ser Arg Lys Val Pro Pro Tyr Val Leu Val Leu Ser Ala Phe Thr  
 115 120 125  
 30 Ser Tyr Arg Ile His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Pro  
 130 135 140  
 35 Val Ala Val Leu Leu Leu Tyr Ala Ala Leu Asn Leu Phe Leu Asp Arg  
 145 150 155 160  
 Arg Trp Thr Leu Gly Ser Thr Phe Phe Ser Leu Ala Val Gly Val Lys  
 165 170 175  
 40 Met Asn Ile Leu Leu Phe Ala Pro Ala Leu Leu Leu Phe Tyr Leu Ala  
 180 185 190  
 45 Asn Leu Gly Leu Leu Arg Thr Ile Leu Gln Leu Ala Val Cys Gly Val  
 195 200 205  
 Ile Gln Leu Leu Leu Gly Ala Pro Phe Leu Leu Thr His Pro Val Glu  
 210 215 220  
 50 Tyr Leu Arg Gly Ser Phe Asp Leu Gly Arg Ile Phe Glu His Lys Trp  
 225 230 235 240  
 55 Thr Val Asn Tyr Arg Phe Leu Ser Arg Asp Val Phe Glu Asn Arg Thr  
 245 250 255  
 Phe

60 <210> 19

<211> 60

<212> PRT

<213> *Saccharomyces cerevisiae*

65

# ES 2 335 345 T3

<400> 19

5           Ile Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser  
            1                           5                           10                           15  
          Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro  
                          20                           25                           30  
10        Ile Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp  
                  35                           40                           45  
          Tyr Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro  
15                   50                           55                           60

<210> 20

<211> 58

20 <213> *Drosophila melanogaster*

<400> 20

25        Leu Pro Phe Phe Leu Cys Asn Leu Val Gly Val Ala Cys Ser Arg Ser  
            1                           5                           10                           15  
          Leu His Tyr Gln Phe Tyr Val Trp Tyr Phe His Ser Leu Pro Tyr Leu  
30                            20                           25                           30  
          Ala Trp Ser Thr Pro Tyr Ser Leu Gly Val Arg Cys Leu Ile Leu Gly  
                  35                           40                           45  
35        Leu Ile Glu Tyr Cys Trp Asn Thr Tyr Pro  
                  50                           55

<210> 21

40 <211> 1377

<212> ADN

<213> *Saccharomyces cerevisiae*

45

50

55

60

65

# ES 2 335 345 T3

<400> 21

```

5      atggaagggtg aacagttctcc gcaaggtgaa aagtctctgc aaaggaagca atttgtcaga 60
      cctccgctgg atctgtggca ggatctcaag gacgggtgtgc gctacgtgat cttcgattgt 120
      agggccaatc ttatcgttat gccccttttg attttgttcg aaagcatgct gtgcaagatt 180
      atcattaaga aggttagctta cacagagatc gattacaagg cgtacatgga gcagatcgag 240
      atgattcagc togatggcat gctggactac tctcaggtga gtggtggaac gggcccgctg 300
10     gtgtatccag caggccacgt cttgatctac aagatgatgt actggctaac agaggggaatg 360
      gaccacgttg agcgcgggca agtggttttc agatacttgt atctccttac actggcggtta 420
      caaatggcgt gttactacct ttacatctta ccaccgtggg gtgtggtctt ggcgtgcctc 480
      tctaaaagat tgcactctat ttacgtgcta cggttattca atgattgctt cactactttg 540
      tttatggtcg tcacgggtttt gggggctatc gtggccagca ggtgccatca gcgccccaaa 600
15     ttaaagaagt ccocttgcgct ggtgatctcc gcaacataca gtatggctgt gagcattaag 660
      atgaatgcgc tgtttgtattt ccctgcaatg atgatttctc tattcatcct taatgacgcg 720
      aacgtaatcc ttactttggt ggatctcgtt gcgatgattg catggcaagt cgcagttgca 780
      gtgccccttc tgcgcagctt tccgcaacag tacctgcatt gcgcttttaa tttcggcagg 840
20     aagtttatgt accaatggag tatcaattgg caaatgatgg atgaagaggc tttcaatgat 900
      aagaggttcc acttggccct tttaatcagc cacctgatag cgtcaccac actgttcgtc 960
      acaagatacc ctgcatect gcccgattta tggctctccc tgtgccatcc gctgaggaaa 1020
      aatgcagtgc tcaatgcca acccgccaag actattccat tcgttcta atcgatccaac 1080
      ttcacggcgc toctattttc aaggtcctc cactaccagt ttctatcctg gtatcactgg 1140
25     actttgcta tactgatctt ttggtcggga atgcccttct tcgttggctc catttggtag 1200
      gtcttgcaag agtgggtgctg gaattcctat ccacaaact cacaagcaag cacgctattg 1260
      ttggcattga atactgttct gttgcttcta ttggccttga cgcagctatc tggttcggtc 1320
      gcctcgcga aaagccatct tcgtaccacc agctctatgg aaaaaagct caactga 1377

```

30

<210> 22

<211> 458

35 <212> PRT

<213> *Saccharomyces cerevisiae*

<400> 22

40

```

      Met Glu Gly Glu Gln Ser Pro Gln Gly Glu Lys Ser Leu Gln Arg Lys
          1             5             10             15

```

45

```

      Gln Phe Val Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly
          20             25             30

```

50

```

      Val Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro
          35             40             45

```

55

60

65

# ES 2 335 345 T3

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Leu | Leu | Ile | Leu | Phe | Glu | Ser | Met | Leu | Cys | Lys | Ile | Ile | Ile | Lys | Lys |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 5  | Val | Ala | Tyr | Thr | Glu | Ile | Asp | Tyr | Lys | Ala | Tyr | Met | Glu | Gln | Ile | Glu |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
|    | Met | Ile | Gln | Leu | Asp | Gly | Met | Leu | Asp | Tyr | Ser | Gln | Val | Ser | Gly | Gly |
| 10 |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Gly | His | Val | Leu | Ile | Tyr | Lys | Met |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
|    | Met | Tyr | Trp | Leu | Thr | Glu | Gly | Met | Asp | His | Val | Glu | Arg | Gly | Gln | Val |
| 15 |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Phe | Phe | Arg | Tyr | Leu | Tyr | Leu | Leu | Thr | Leu | Ala | Leu | Gln | Met | Ala | Cys |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 20 | Tyr | Tyr | Leu | Leu | His | Leu | Pro | Pro | Trp | Cys | Val | Val | Leu | Ala | Cys | Leu |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Ser | Lys | Arg | Leu | His | Ser | Ile | Tyr | Val | Leu | Arg | Leu | Phe | Asn | Asp | Cys |
| 25 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Phe | Thr | Thr | Leu | Phe | Met | Val | Val | Thr | Val | Leu | Gly | Ala | Ile | Val | Ala |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
|    | Ser | Arg | Cys | His | Gln | Arg | Pro | Lys | Leu | Lys | Lys | Ser | Leu | Ala | Leu | Val |
| 30 |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Ile | Ser | Ala | Thr | Tyr | Ser | Met | Ala | Val | Ser | Ile | Lys | Met | Asn | Ala | Leu |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 35 | Leu | Tyr | Phe | Pro | Ala | Met | Met | Ile | Ser | Leu | Phe | Ile | Leu | Asn | Asp | Ala |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
|    | Asn | Val | Ile | Leu | Thr | Leu | Leu | Asp | Leu | Val | Ala | Met | Ile | Ala | Trp | Gln |
| 40 |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Val | Ala | Val | Ala | Val | Pro | Phe | Leu | Arg | Ser | Phe | Pro | Gln | Gln | Tyr | Leu |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 45 | His | Cys | Ala | Phe | Asn | Phe | Gly | Arg | Lys | Phe | Met | Tyr | Gln | Trp | Ser | Ile |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|    | Asn | Trp | Gln | Met | Met | Asp | Glu | Glu | Ala | Phe | Asn | Asp | Lys | Arg | Phe | His |
| 50 |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
|    | Leu | Ala | Leu | Leu | Ile | Ser | His | Leu | Ile | Ala | Leu | Thr | Thr | Leu | Phe | Val |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
|    | Thr | Arg | Tyr | Pro | Arg | Ile | Leu | Pro | Asp | Leu | Trp | Ser | Ser | Leu | Cys | His |
| 55 |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
|    | Pro | Leu | Arg | Lys | Asn | Ala | Val | Leu | Asn | Ala | Asn | Pro | Ala | Lys | Thr | Ile |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| 60 | Pro | Phe | Val | Leu | Ile | Ala | Ser | Asn | Phe | Ile | Gly | Val | Leu | Phe | Ser | Arg |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |

65

ES 2 335 345 T3

Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Ile  
 370 375 380

5 Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp Tyr  
 385 390 395 400

10 Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ser Gln Ala  
 405 410 415

15 Ser Thr Leu Leu Leu Ala Leu Asn Thr Val Leu Leu Leu Leu Leu Ala  
 420 425 430 435

20 Leu Thr Gln Leu Ser Gly Ser Val Ala Leu Ala Lys Ser His Leu Arg  
 435 440 445

Thr Thr Ser Ser Met Glu Lys Lys Leu Asn  
 450 455

<210> 23

<211> 1395

25 <212> ADN

<213> *Pichia pastoris*

<400> 23

30 atgcctccga tagagccagc tgaaggcca aagcttaagc tgaaaaatgt tatecggat 60  
 ctagtggcgc ttattcaaaa cgttttattt aaccagatt ttagtgtctt cgttgcacct 120  
 cttttatggt tagctgattc cattgttattc aagtgatca ttggcactgt tccctacaca 180  
 35 gatattgatt tttcttcata tatgcaacaa atctttaaaa ttcgacaagg aqaattagat 240  
 tatagcaaca tatttggtga caccggcca ttggtttacc cagccggcca tgttcattgt 300  
 tactcagtac tttcgtggta cagtgatggt ggagaagacg tcagtttcgt tcaacaagca 360  
 tttggttggg tatacctagg ttgcttggtta ctatccatca gctcctactt tttctctggc 420  
 ttagggaaaa tacctcoggt ttattttggt ttggttgtag cgtccaagag actgcattca 480  
 40 atatttgat tgagactctt caatgactgt ttaacaacat ttttgatggt ggcaactata 540  
 atcatccttc aacaagcaag tagctggagg aaagatggca caactattcc attatctgtc 600  
 cctgatgctg cagatacgtc cagtttagcc atctctgtaa agatgaatgc gctgctatac 660  
 ctcccagcat tctactact catatatctc atttgtgacg aaaatttgat taaagccttg 720  
 gcacctgttc tagttttgat attggtgcaa gtaggagtcg gttattcgtt cattttaccg 780  
 45 ttgcactatg atgatcaggc aaatgaaatt cgttctgcct actttagaca ggcttttgac 840  
 tttagtcgcc aattttctta taagtggacg gtttaattggc gctttttgag ccaagaaact 900  
 ttcaacaatg tccattttca ccagctcctg tttgctctcc atattattac gttagtcttg 960  
 ttcactctca agttcctctc tcttaaaaac attggaaaac cgcttggtag atttggtgtg 1020  
 50 gacattttca aattttggaa gccaacctta tctccaacca atattatcaa cgaccaga 1080  
 agaagcccag attttggtta caccgtcatg gctactacca acttaatagg ggtgcttttt 1140  
 gcaagatctt tacactacca gttcctaagc tggatgctg tctctttgcc atatctcctt 1200  
 tacaaggctc gctgaaact tatagcatct attattggtt atgcccctca cgagtattgc 1260  
 55 tggttggttt tcccagctac agaacaaagt tccgcgttgt tggatctat ctactactt 1320  
 atoctgattc tcatttttac caacgaacag ttatttctt ctcaatcggg cctgcagaa 1380  
 aaaaagaata cataa 1395

60 <210> 24

<211> 464

<212> PRT

<213> *Pichia pastoris*

65

# ES 2 335 345 T3

<400> 24

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Met | Pro | Pro | Ile | Glu | Pro | Ala | Glu | Arg | Pro | Lys | Leu | Thr | Leu | Lys | Asn |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| 5  | Val | Ile | Gly | Asp | Leu | Val | Ala | Leu | Ile | Gln | Asn | Val | Leu | Phe | Asn | Pro |
|    |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
|    | Asp | Phe | Ser | Val | Phe | Val | Ala | Pro | Leu | Leu | Trp | Leu | Ala | Asp | Ser | Ile |
| 10 |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    | Val | Ile | Lys | Val | Ile | Ile | Gly | Thr | Val | Ser | Tyr | Thr | Asp | Ile | Asp | Phe |
|    | 50  |     |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 15 | Ser | Ser | Tyr | Met | Gln | Gln | Ile | Phe | Lys | Ile | Arg | Gln | Gly | Glu | Leu | Asp |
|    | 65  |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     |     | 80  |
|    | Tyr | Ser | Asn | Ile | Phe | Gly | Asp | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Gly |
| 20 |     |     |     | 85  |     |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | His | Val | His | Ala | Tyr | Ser | Val | Leu | Ser | Trp | Tyr | Ser | Asp | Gly | Gly | Glu |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| 25 | Asp | Val | Ser | Phe | Val | Gln | Gln | Ala | Phe | Gly | Trp | Leu | Tyr | Leu | Gly | Cys |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
|    | Leu | Leu | Leu | Ser | Ile | Ser | Ser | Tyr | Phe | Phe | Ser | Gly | Leu | Gly | Lys | Ile |
|    | 130 |     |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 30 | Pro | Pro | Val | Tyr | Phe | Val | Leu | Leu | Val | Ala | Ser | Lys | Arg | Leu | His | Ser |
|    | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Ile | Phe | Val | Leu | Arg | Leu | Phe | Asn | Asp | Cys | Leu | Thr | Thr | Phe | Leu | Met |
| 35 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
|    | Leu | Ala | Thr | Ile | Ile | Ile | Leu | Gln | Gln | Ala | Ser | Ser | Trp | Arg | Lys | Asp |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 40 | Gly | Thr | Thr | Ile | Pro | Leu | Ser | Val | Pro | Asp | Ala | Ala | Asp | Thr | Tyr | Ser |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Leu | Ala | Ile | Ser | Val | Lys | Met | Asn | Ala | Leu | Leu | Tyr | Leu | Pro | Ala | Phe |
| 45 |     |     |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
|    | Leu | Leu | Leu | Ile | Tyr | Leu | Ile | Cys | Asp | Glu | Asn | Leu | Ile | Lys | Ala | Leu |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| 50 | Ala | Pro | Val | Leu | Val | Leu | Ile | Leu | Val | Gln | Val | Gly | Val | Gly | Tyr | Ser |
|    |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
|    | Phe | Ile | Leu | Pro | Leu | His | Tyr | Asp | Asp | Gln | Ala | Asn | Glu | Ile | Arg | Ser |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 55 | Ala | Tyr | Phe | Arg | Gln | Ala | Phe | Asp | Phe | Ser | Arg | Gln | Phe | Leu | Tyr | Lys |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
|    | Trp | Thr | Val | Asn | Trp | Arg | Phe | Leu | Ser | Gln | Glu | Thr | Phe | Asn | Asn | Val |
| 60 |     |     | 290 |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
|    | His | Phe | His | Gln | Leu | Leu | Phe | Ala | Leu | His | Ile | Ile | Thr | Leu | Val | Leu |
|    | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| 65 | Phe | Ile | Leu | Lys | Phe | Leu | Ser | Pro | Lys | Asn | Ile | Gly | Lys | Pro | Leu | Gly |
|    |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |

## ES 2 335 345 T3

|    |  |                              |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|--|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    |  | Arg                          | Phe | Val | Leu | Asp | Ile | Phe | Lys | Phe | Trp | Lys | Pro | Thr | Leu | Ser | Pro |
|    |  |                              |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| 5  |  | Thr                          | Asn | Ile | Ile | Asn | Asp | Pro | Glu | Arg | Ser | Pro | Asp | Phe | Val | Tyr | Thr |
|    |  |                              |     |     | 355 |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
|    |  | Val                          | Met | Ala | Thr | Thr | Asn | Leu | Ile | Gly | Val | Leu | Phe | Ala | Arg | Ser | Leu |
|    |  |                              | 370 |     |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |
| 10 |  | His                          | Tyr | Gln | Phe | Leu | Ser | Trp | Tyr | Ala | Phe | Ser | Leu | Pro | Tyr | Leu | Leu |
|    |  | 385                          |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     | 400 |     |
|    |  | Tyr                          | Lys | Ala | Arg | Leu | Asn | Phe | Ile | Ala | Ser | Ile | Ile | Val | Tyr | Ala | Ala |
| 15 |  |                              |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |
|    |  | His                          | Glu | Tyr | Cys | Trp | Leu | Val | Phe | Pro | Ala | Thr | Glu | Gln | Ser | Ser | Ala |
|    |  |                              |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |
| 20 |  | Leu                          | Leu | Val | Ser | Ile | Leu | Leu | Leu | Ile | Leu | Ile | Leu | Ile | Phe | Thr | Asn |
|    |  |                              |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
|    |  | Glu                          | Gln | Leu | Phe | Pro | Ser | Gln | Ser | Val | Pro | Ala | Glu | Lys | Lys | Asn | Thr |
| 25 |  |                              | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
|    |  | <210> 25                     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <211> 423                    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 30 |  | <212> PRT                    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <213> <i>Pichia pastoris</i> |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <220>                        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 35 |  | <221> MOD_RES                |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <222> (209)..(223)           |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <223> Aminoácido variable    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <220>                        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 40 |  | <221> MOD_RES                |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <222> (235)..(246)           |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | <223> Aminoácido variable    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
| 45 |  | <400> 25                     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|    |  | Arg                          | Pro | Lys | Leu | Thr | Leu | Lys | Asn | Val | Ile | Gly | Asp | Leu | Val | Ala | Leu |
| 50 |  | 1                            |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
|    |  | Ile                          | Gln | Asn | Val | Leu | Phe | Asn | Pro | Asp | Phe | Ser | Val | Phe | Val | Ala | Pro |
|    |  |                              |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| 55 |  | Leu                          | Leu | Trp | Leu | Ala | Asp | Ser | Ile | Val | Ile | Lys | Val | Ile | Ile | Gly | Thr |
|    |  |                              |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    |  | Val                          | Ser | Tyr | Thr | Asp | Ile | Asp | Phe | Ser | Ser | Tyr | Met | Gln | Gln | Ile | Phe |
| 60 |  |                              | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
|    |  | Lys                          | Ile | Arg | Gln | Gly | Glu | Leu | Asp | Tyr | Ser | Asn | Ile | Phe | Gly | Asp | Thr |
|    |  | 65                           |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| 65 |  | Gly                          | Pro | Leu | Val | Tyr | Pro | Ala | Gly | His | Val | His | Ala | Tyr | Ser | Val | Leu |

# ES 2 335 345 T3

|    | 85   | 90 | 95 |
|----|--|----|----|
|    | Ser Trp Tyr Ser Asp Gly Gly Glu Asp Val Ser Phe Val Gln Gln Ala<br>100 105 110     |    |    |
| 5  | Phe Gly Trp Leu Tyr Leu Gly Cys Leu Leu Leu Ser Ile Ser Ser Tyr<br>115 120 125     |    |    |
| 10 | Phe Phe Ser Gly Leu Gly Lys Ile Pro Pro Val Tyr Phe Val Leu Leu<br>130 135 140     |    |    |
|    | Val Ala Ser Lys Arg Leu His Ser Ile Phe Val Leu Arg Leu Phe Asn<br>145 150 155 160 |    |    |
| 15 | Asp Cys Leu Thr Thr Phe Leu Met Leu Ala Thr Ile Ile Ile Leu Gln<br>165 170 175     |    |    |
|    | Gln Ala Ser Ser Trp Arg Lys Asp Gly Thr Thr Ile Pro Leu Ser Val<br>180 185 190     |    |    |
| 20 | Pro Asp Ala Ala Asp Thr Tyr Ser Leu Ala Ile Ser Val Lys Met Asn<br>195 200 205     |    |    |
| 25 | Xaa Cys<br>210 215 220         |    |    |
|    | Asp Glu Asn Leu Ile Lys Ala Leu Ala Pro Xaa Xaa Xaa Xaa Xaa Xaa<br>225 230 235 240 |    |    |
| 30 | Xaa Xaa Xaa Xaa Xaa Xaa Tyr Ser Phe Ile Leu Pro Leu His Tyr Asp<br>245 250 255     |    |    |
|    | Asp Gln Ala Asn Glu Ile Arg Ser Ala Tyr Phe Arg Gln Ala Phe Asp<br>260 265 270     |    |    |
| 35 | Phe Ser Arg Gln Phe Leu Tyr Lys Trp Thr Val Asn Trp Arg Phe Leu<br>275 280 285     |    |    |
| 40 | Ser Gln Glu Thr Phe Asn Asn Val His Phe His Gln Leu Leu Phe Ala<br>290 295 300     |    |    |
|    | Leu His Ile Ile Thr Leu Val Leu Phe Ile Leu Lys Phe Leu Ser Pro<br>305 310 315 320 |    |    |
| 45 | Lys Asn Ile Gly Lys Pro Leu Gly Arg Phe Val Leu Asp Ile Phe Lys<br>325 330 335     |    |    |
|    | Phe Trp Lys Pro Thr Leu Ser Pro Thr Asn Ile Ile Asn Asp Pro Glu<br>340 345 350     |    |    |
| 50 | Arg Ser Pro Asp Phe Val Tyr Thr Val Met Ala Thr Thr Asn Leu Ile<br>355 360 365     |    |    |
|    | Gly Val Leu Phe Ala Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr<br>370 375 380     |    |    |
| 55 | Ala Phe Ser Leu Pro Tyr Leu Leu Tyr Lys Ala Arg Leu Asn Phe Ile<br>385 390 395 400 |    |    |
| 60 | Ala Ser Ile Ile Val Tyr Ala Ala His Glu Tyr Cys Trp Leu Val Phe<br>405 410 415     |    |    |
|    | Pro Ala Thr Glu Gln Ser Ser<br>420   |    |    |

65

# ES 2 335 345 T3

<210> 26

<211> 398

<212> PRT

5 <213> *Saccharomyces cerevisiae*

<400> 26

```

10      Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly Val Arg Tyr
        1          5          10          15
15      Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro Leu Leu Ile
        20          25          30
20      Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys Val Ala Tyr
        35          40          45
25      Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu Met Ile Gln
        50          55          60
30      Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly Thr Gly Pro
        65          70          75          80
35      Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met Met Tyr Trp
        85          90          95
40      Leu Thr Glu Gly Met Asp His Val Glu Arg Gly Gln Val Phe Phe Arg
        100          105          110
45      Tyr Leu Tyr Leu Leu Thr Leu Ala Leu Gln Met Ala Cys Tyr Tyr Leu
        115          120          125
50      Leu His Leu Pro Pro Trp Cys Val Val Leu Ala Cys Leu Ser Lys Arg
        130          135          140
55      Leu His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Cys Phe Thr Thr
        145          150          155          160
60      Leu Phe Met Val Val Thr Val Leu Gly Ala Ile Val Ala Ser Arg Cys
        165          170          175
65      His Gln Arg Pro Lys Leu Lys Lys Ser Leu Ala Leu Val Ile Ser Ala
        180          185          190
70      Thr Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu Leu Tyr Phe
        195          200          205
75      Pro Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala Asn Val Ile
        210          215          220
80      Leu Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln Val Ala Val
        225          230          235          240
85      Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu His Cys Ala
        245          250          255

```





# ES 2 335 345 T3

<400> 28

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5  | Ser | Lys | Leu | Ile | Pro | Pro | Ala | Leu | Phe | Leu | Val | Asp | Ala | Leu | Leu | Cys |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
|    | Gly | Leu | Ile | Ile | Trp | Lys | Val | Pro | Tyr | Thr | Glu | Ile | Asp | Trp | Ala | Ala |
|    |     |     | 20  |     |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| 10 | Tyr | Met | Glu | Gln | Val | Ser | Gln | Ile | Leu | Ser | Gly | Glu | Arg | Asp | Tyr | Thr |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    | Lys | Val | Arg | Gly | Gly | Thr | Gly | Pro | Leu | Val | Tyr | Pro | Ala | Ala | His | Val |
| 15 |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
|    | Tyr | Ile | Tyr | Thr | Gly | Leu | Tyr | His | Leu | Thr | Asp | Glu | Gly | Arg | Asn | Ile |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     | 80  |     |
| 20 | Leu | Leu | Ala | Gln | Gln | Leu | Phe | Ala | Gly | Leu | Tyr | Met | Val | Thr | Leu | Ala |
|    |     |     |     | 85  |     |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Val | Val | Met | Gly | Cys | Tyr | Trp | Gln | Ala | Lys | Ala | Pro | Pro | Tyr | Leu | Phe |
| 25 |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
|    | Pro | Leu | Leu | Thr | Leu | Ser | Lys | Arg | Leu | His | Ser | Ile | Phe | Val | Leu | Arg |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 30 | Cys | Phe | Asn | Asp | Cys | Phe | Ala | Val | Leu | Phe | Leu | Trp | Leu | Ala | Ile | Phe |
|    | 130 |     |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | Phe | Phe | Gln | Arg | Arg | Asn | Trp | Gln | Ala | Gly | Ala | Leu | Leu | Tyr | Thr | Leu |
| 35 |     |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Gly | Leu | Gly | Val | Lys | Met | Thr | Leu | Leu | Leu | Ser | Leu | Pro | Ala | Val | Gly |
|    |     |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     | 175 |     |
| 40 | Ile | Val | Leu | Phe | Leu | Gly | Ser | Gly | Ser | Phe | Val | Thr | Thr | Leu | Gln | Leu |
|    |     |     |     | 180 |     |     |     |     |     | 185 |     |     |     | 190 |     |     |
|    | Val | Ala | Thr | Met | Gly | Leu | Val | Gln | Ile | Leu | Ile | Gly | Val | Pro | Phe | Leu |
| 45 |     |     |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
|    | Ala | His | Tyr | Pro | Thr | Glu | Tyr | Leu | Ser | Arg | Ala | Phe | Glu | Leu | Ser | Arg |
|    |     | 210 |     |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |
| 50 | Gln | Phe | Phe | Phe | Lys | Trp | Thr | Val | Asn | Trp | Arg | Phe | Val | Gly | Glu | Glu |
|    | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
|    | Ile | Phe | Leu | Ser | Lys | Gly | Phe | Ala | Leu | Thr | Leu | Leu | Ala | Leu | His | Val |
| 55 |     |     |     |     | 245 |     |     |     |     |     | 250 |     |     |     | 255 |     |
|    | Leu | Val | Leu | Gly | Ile | Phe | Ile | Thr | Thr | Arg | Trp | Ile | Lys | Pro | Ala | Arg |
|    |     |     |     | 260 |     |     |     |     |     | 265 |     |     |     | 270 |     |     |
| 60 | Lys | Ser | Leu | Val | Gln | Leu | Ile | Ser | Pro | Val | Leu | Leu | Ala | Gly | Lys | Pro |
|    |     |     | 275 |     |     |     |     |     | 280 |     |     |     |     | 285 |     |     |

65

## ES 2 335 345 T3

```

Pro Leu Thr Val Pro Glu His Arg Ala Ala Ala Arg Asp Val Thr Pro
  290                               295                               300

Arg Tyr Ile Met Thr Thr Ile Leu Ser Ala Asn Ala Val Gly Leu Leu
5  305                               310                               315                               320

Phe Ala Arg Ser Leu His Tyr Gln Phe Tyr Ala Tyr Val Ala Trp Ser
                               325                               330                               335

10 Thr Pro Phe Leu Leu Trp Arg Ala Gly Leu His Pro Val Leu Val Tyr
                               340                               345                               350

Leu Leu Trp Ala Val His Glu Trp Ala Trp Asn Val Phe Pro Ser Thr
15                               355                               360                               365

Pro Ala Ser Ser Ala
  370

<210> 29
20 <211> 390
   <212> PRT
   <213> Pichia pastoris

25 <220>
   <221> MOD_RES
   <222> (176)..(190)
30 <223> Aminoácido variable

   <220>
   <221> MOD_RES
35 <222> (202)..(213)
   <223> Aminoácido variable

40 <400> 29
   Leu Trp Leu Ala Asp Ser Ile Val Ile Lys Val Ile Ile Gly Thr Val
     1                               5                               10                               15

   Ser Tyr Thr Asp Ile Asp Phe Ser Ser Tyr Met Gln Gln Ile Phe Lys
45                               20                               25                               30

   Ile Arg Gln Gly Glu Leu Asp Tyr Ser Asn Ile Phe Gly Asp Thr Gly
                               35                               40                               45

50 Pro Leu Val Tyr Pro Ala Gly His Val His Ala Tyr Ser Val Leu Ser
     50                               55                               60

   Trp Tyr Ser Asp Gly Gly Glu Asp Val Ser Phe Val Gln Gln Ala Phe
55   65                               70                               75                               80

   Gly Trp Leu Tyr Leu Gly Cys Leu Leu Leu Ser Ile Ser Ser Tyr Phe
                               85                               90                               95

60 Phe Ser Gly Leu Gly Lys Ile Pro Pro Val Tyr Phe Val Leu Leu Val
                               100                               105                               110

   Ala Ser Lys Arg Leu His Ser Ile Phe Val Leu Arg Leu Phe Asn Asp
                               115                               120                               125

65 Cys Leu Thr Thr Phe Leu Met Leu Ala Thr Ile Ile Ile Leu Gln Gln

```



# ES 2 335 345 T3

<400> 30

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     | Leu | Leu | Leu | Glu | Ile | Pro | Phe | Val | Phe | Ala | Ile | Ile | Ser | Lys | Val |     |
|     | 1   |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |     |
| 5   | Pro | Tyr | Thr | Glu | Ile | Asp | Trp | Ile | Ala | Tyr | Met | Glu | Gln | Val | Asn | Ser |
|     |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| 10  | Phe | Leu | Leu | Gly | Glu | Arg | Asp | Tyr | Lys | Ser | Leu | Val | Gly | Cys | Thr | Gly |
|     |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| 15  | Pro | Leu | Val | Tyr | Pro | Gly | Gly | His | Val | Phe | Leu | Tyr | Thr | Leu | Leu | Tyr |
|     |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 20  | Tyr | Leu | Thr | Asp | Gly | Gly | Thr | Asn | Ile | Val | Arg | Ala | Gln | Tyr | Ile | Phe |
|     | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| 25  | Ala | Phe | Val | Tyr | Trp | Ile | Thr | Thr | Ala | Ile | Val | Gly | Tyr | Leu | Phe | Lys |
|     |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| 30  | Ile | Val | Arg | Ala | Pro | Phe | Tyr | Ile | Tyr | Val | Leu | Leu | Ile | Leu | Ser | Lys |
|     |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| 35  | Arg | Leu | His | Ser | Ile | Phe | Ile | Leu | Arg | Leu | Phe | Asn | Asp | Gly | Phe | Asn |
|     |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 40  | Ser | Leu | Phe | Ser | Ser | Leu | Phe | Ile | Leu | Ser | Ser | Cys | Lys | Lys | Lys | Trp |
|     |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 45  | Val | Arg | Ala | Ser | Ile | Leu | Leu | Ser | Val | Ala | Cys | Ser | Val | Lys | Met | Ser |
|     | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
| 50  | Ser | Leu | Leu | Tyr | Val | Pro | Ala | Tyr | Leu | Val | Leu | Leu | Leu | Gln | Ile | Leu |
|     |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
| 55  | Gly | Pro | Lys | Lys | Thr | Trp | Met | His | Ile | Phe | Val | Ile | Ile | Ile | Val | Gln |
|     |     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 60  | Ile | Leu | Phe | Ser | Ile | Pro | Phe | Leu | Ala | Tyr | Phe | Trp | Ser | Tyr | Trp | Thr |
|     |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| 65  | Gln | Ala | Phe | Asp | Phe | Gly | Arg | Ala | Phe | Asp | Tyr | Lys | Trp | Thr | Val | Asn |
|     |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 70  | Trp | Arg | Phe | Ile | Pro | Arg | Ser | Ile | Phe | Glu | Ser | Thr | Ser | Phe | Ser | Thr |
|     | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| 75  | Ser | Ile | Leu | Phe | Leu | His | Val | Ala | Leu | Leu | Val | Ala | Phe | Thr | Cys | Lys |
|     |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| 80  | His | Trp | Asn | Lys | Leu | Ser | Arg | Ala | Thr | Pro | Phe | Ala | Met | Val | Asn | Ser |
|     |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 85  | Met | Leu | Thr | Leu | Lys | Pro | Leu | Pro | Lys | Leu | Gln | Leu | Ala | Thr | Pro | Asn |
|     |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| 90  | Phe | Ile | Phe | Thr | Ala | Leu | Ala | Thr | Ser | Asn | Leu | Ile | Gly | Ile | Leu | Cys |
|     |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| 95  | Ala | Arg | Ser | Leu | His | Tyr | Gln | Phe | Tyr | Ala | Trp | Phe | Ala | Trp | Tyr | Ser |
|     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| 100 | Pro | Tyr | Leu | Cys | Tyr | Gln | Ala | Ser | Phe | Pro | Ala | Pro | Ile | Val | Ile | Gly |
|     |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |

## ES 2 335 345 T3

Leu Trp Met Leu Gln Glu Tyr Ala Trp Asn Val Phe Pro Ser Thr Lys  
 340 345 350

Leu Ser Ser  
 355

5

<210> 31

<211> 390

<212> PRT

10

<213> *Pichia pastoris*

<220>

<221> MOD\_RES

15

<222> (176)..(190)

<223> Aminoácido variable

<220>

20

<221> MOD\_RES

<222> (202)..(213)

<223> Aminoácido variable

25

<400> 31

Leu Trp Leu Ala Asp Ser Ile Val Ile Lys Val Ile Ile Gly Thr Val  
 1 5 10 15

30

Ser Tyr Thr Asp Ile Asp Phe Ser Ser Tyr Met Gln Gln Ile Phe Lys  
 20 25 30

Ile Arg Gln Gly Glu Leu Asp Tyr Ser Asn Ile Phe Gly Asp Thr Gly  
 35 40 45

35

Pro Leu Val Tyr Pro Ala Gly His Val His Ala Tyr Ser Val Leu Ser  
 50 55 60

40

Trp Tyr Ser Asp Gly Gly Glu Asp Val Ser Phe Val Gln Gln Ala Phe  
 65 70 75 80

Gly Trp Leu Tyr Leu Gly Cys Leu Leu Leu Ser Ile Ser Ser Tyr Phe  
 85 90 95

45

Phe Ser Gly Leu Gly Lys Ile Pro Pro Val Tyr Phe Val Leu Leu Val  
 100 105 110

Ala Ser Lys Arg Leu His Ser Ile Phe Val Leu Arg Leu Phe Asn Asp  
 115 120 125

50

Cys Leu Thr Thr Phe Leu Met Leu Ala Thr Ile Ile Ile Leu Gln Gln  
 130 135 140

55

Ala Ser Ser Trp Arg Lys Asp Gly Thr Thr Ile Pro Leu Ser Val Pro  
 145 150 155 160

Asp Ala Ala Asp Thr Tyr Ser Leu Ala Ile Ser Val Lys Met Asn Xaa  
 165 170 175

60

Xaa Cys Asp  
 180 185 190

Glu Asn Leu Ile Lys Ala Leu Ala Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa

65

# ES 2 335 345 T3

|    | 195  | 200 | 205 |
|----|--|-----|-----|
| 5  | Xaa Xaa Xaa Xaa Xaa Tyr Ser Phe Ile Leu Pro Leu His Tyr Asp Asp<br>210 215 220     |     |     |
|    | Gln Ala Asn Glu Ile Arg Ser Ala Tyr Phe Arg Gln Ala Phe Asp Phe<br>225 230 235 240 |     |     |
| 10 | Ser Arg Gln Phe Leu Tyr Lys Trp Thr Val Asn Trp Arg Phe Leu Ser<br>245 250 255     |     |     |
| 15 | Gln Glu Thr Phe Asn Asn Val His Phe His Gln Leu Leu Phe Ala Leu<br>260 265 270     |     |     |
|    | His Ile Ile Thr Leu Val Leu Phe Ile Leu Lys Phe Leu Ser Pro Lys<br>275 280 285     |     |     |
| 20 | Asn Ile Gly Lys Pro Leu Gly Arg Phe Val Leu Asp Ile Phe Lys Phe<br>290 295 300     |     |     |
| 25 | Trp Lys Pro Thr Leu Ser Pro Thr Asn Ile Ile Asn Asp Pro Glu Arg<br>305 310 315 320 |     |     |
|    | Ser Pro Asp Phe Val Tyr Thr Val Met Ala Thr Thr Asn Leu Ile Gly<br>325 330 335     |     |     |
| 30 | Val Leu Phe Ala Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr Ala<br>340 345 350     |     |     |
| 35 | Phe Ser Leu Pro Tyr Leu Leu Tyr Lys Ala Arg Leu Asn Phe Ile Ala<br>355 360 365     |     |     |
|    | Ser Ile Ile Val Tyr Ala Ala His Glu Tyr Cys Trp Leu Val Phe Pro<br>370 375 380     |     |     |
| 40 | Ala Thr Glu Gln Ser Ser<br>385 390   |     |     |
| 45 | <210> 32<br><211> 363<br><212> PRT<br><213> <i>Arabidopsis thaliana</i>            |     |     |
| 50 | <400> 32   |     |     |
|    | Leu Ile Leu Ala Asp Ala Ile Leu Val Ala Leu Ile Ile Ala Tyr Val<br>1 5 10 15       |     |     |
| 55 | Pro Tyr Thr Lys Ile Asp Trp Asp Ala Tyr Met Ser Gln Val Ser Gly<br>20 25 30        |     |     |
| 60 | Phe Leu Gly Gly Glu Arg Asp Tyr Gly Asn Leu Lys Gly Asp Thr Gly<br>35 40 45        |     |     |
|    | Pro Leu Val Tyr Pro Ala Gly Phe Leu Tyr Val Tyr Ser Ala Val Gln<br>50 55 60        |     |     |
| 65 | Asn Leu Thr Gly Gly Glu Val Tyr Pro Ala Gln Ile Leu Phe Gly Val<br>65 70 75 80     |     |     |

## ES 2 335 345 T3

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Leu | Tyr | Ile | Val | Asn | Leu | Gly | Ile | Val | Leu | Ile | Ile | Tyr | Val | Lys | Thr |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| 5  | Asp | Val | Val | Pro | Trp | Trp | Ala | Leu | Ser | Leu | Leu | Cys | Leu | Ser | Lys | Arg |
|    |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| 10 | Ile | His | Ser | Ile | Phe | Val | Leu | Arg | Leu | Phe | Asn | Asp | Cys | Phe | Ala | Met |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 15 | Thr | Leu | Leu | His | Ala | Ser | Met | Ala | Leu | Phe | Leu | Tyr | Arg | Lys | Trp | His |
|    |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| 20 | Leu | Gly | Met | Leu | Val | Phe | Ser | Gly | Ala | Val | Ser | Val | Lys | Met | Asn | Val |
|    |     | 145 |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
| 25 | Leu | Leu | Tyr | Ala | Pro | Thr | Leu | Leu | Leu | Leu | Leu | Lys | Ala | Met | Asn |     |
|    |     |     |     |     | 165 |     |     |     | 170 |     |     |     |     | 175 |     |     |
| 30 | Ile | Ile | Gly | Val | Val | Ser | Ala | Leu | Ala | Gly | Ala | Ala | Leu | Ala | Gln | Ile |
|    |     |     | 180 |     |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| 35 | Leu | Val | Gly | Leu | Pro | Phe | Leu | Ile | Thr | Tyr | Pro | Val | Ser | Tyr | Ile | Ala |
|    |     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| 40 | Asn | Ala | Phe | Asp | Leu | Gly | Arg | Val | Phe | Ile | His | Phe | Trp | Ser | Val | Asn |
|    |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| 45 | Phe | Lys | Phe | Val | Pro | Glu | Arg | Val | Phe | Val | Ser | Lys | Glu | Phe | Ala | Val |
|    |     | 225 |     |     |     | 230 |     |     |     |     | 235 |     |     |     | 240 |     |
| 50 | Cys | Leu | Leu | Ile | Ala | His | Leu | Phe | Leu | Leu | Val | Ala | Phe | Ala | Asn | Tyr |
|    |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |
| 55 | Lys | Trp | Cys | Lys | His | Glu | Gly | Gly | Ile | Ile | Gly | Phe | Met | Arg | Ser | Arg |
|    |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| 60 | His | Phe | Phe | Leu | Thr | Leu | Pro | Ser | Ser | Leu | Ser | Phe | Ser | Asp | Val | Ser |
|    |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| 65 | Ala | Ser | Arg | Ile | Ile | Thr | Lys | Glu | His | Val | Val | Thr | Ala | Met | Phe | Val |
|    |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| 70 | Gly | Asn | Phe | Ile | Gly | Ile | Val | Phe | Ala | Arg | Ser | Leu | His | Tyr | Gln | Phe |
|    |     | 305 |     |     |     | 310 |     |     |     |     | 315 |     |     |     | 320 |     |
| 75 | Tyr | Ser | Trp | Tyr | Phe | Tyr | Ser | Leu | Pro | Tyr | Leu | Leu | Trp | Arg | Thr | Pro |
|    |     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
| 80 | Phe | Pro | Thr | Trp | Leu | Arg | Leu | Ile | Met | Phe | Leu | Gly | Ile | Glu | Leu | Cys |
|    |     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| 85 | Trp | Asn | Val | Tyr | Pro | Ser | Thr | Pro | Ser | Ser | Ser |     |     |     |     |     |
|    |     |     | 355 |     |     |     |     | 360 |     |     |     |     |     |     |     |     |

<210> 33

<211> 428

<212> ADN

<213> *Kluyveromyces lactis*

# ES 2 335 345 T3

<400> 33

```

5      tttgtttaca agctgatacc aacgaacatg aatacaccgg caggtttact gaagattggc 60
      aaagctaacc ttttacatcc ttttaccgat gctgtattca gtgcgatgag agtaaacgca 120
      gaacaaattg catacatttt acttgttacc aattacattg gagtactatt tgctcgatca 180
      ttacactacc aattcctatc ttggtaccat tggacggtac cagtactatt gaattgggcc 240
      aatgttccgt atccgctatg tgtgctatgg tacctaacac atgagtggtg ctggaacagc 300
      tatccgccaa acgctactgc atccacactg ctacacgggt gtaacacata ctgttattgg 360
10     ctgtattcct aagaggaccc gcaaaactcg aaagtgggtg taacgaaaca acacacgaga 420
      aagctgag                                     428
  
```

<210> 34

15 <211> 142

<212> PRT

<213> *Kluyveromyces lactis*

20 <400> 34

```

25     Phe Val Tyr Lys Leu Ile Pro Thr Asn Met Asn Thr Pro Ala Gly Leu
      1           5           10           15
      Leu Lys Ile Gly Lys Ala Asn Leu Leu His Pro Phe Thr Asp Ala Val
      20           25           30
30     Phe Ser Ala Met Arg Val Asn Ala Glu Gln Ile Ala Tyr Ile Leu Leu
      35           40           45
      Val Thr Asn Tyr Ile Gly Val Leu Phe Ala Arg Ser Leu His Tyr Gln
      50           55           60
      Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Val Leu Leu Asn Trp Ala
      65           70           75           80
40     Asn Val Pro Tyr Pro Leu Cys Val Leu Trp Tyr Leu Thr His Glu Trp
      85           90           95
      Cys Trp Asn Ser Tyr Pro Pro Asn Ala Thr Ala Ser Thr Leu Leu His
      100          105          110
45     Ala Cys Asn Thr Tyr Cys Tyr Trp Leu Tyr Ser Glx Glu Asp Pro Gln
      115          120          125
50     Thr Arg Lys Val Val Ile Thr Lys Gln His Thr Arg Lys Leu
      130          135          140
  
```

<210> 35

55 <211> 118

<212> PRT

<213> *Kluyveromyces lactis*

60

65

# ES 2 335 345 T3

<400> 35

```

5      Ala Asn Leu Leu His Pro Phe Thr Asp Ala Val Phe Ser Ala Met Arg
       1          5          10          15
      Val Asn Ala Glu Gln Ile Ala Tyr Ile Leu Leu Val Thr Asn Tyr Ile
          20          25          30
10     Gly Val Leu Phe Ala Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr
          35          40          45
15
      His Trp Thr Leu Pro Val Leu Leu Asn Trp Ala Asn Val Pro Tyr Pro
          50          55          60
20     Leu Cys Val Leu Trp Tyr Leu Thr His Glu Trp Cys Trp Asn Ser Tyr
       65          70          75          80
      Pro Pro Asn Ala Thr Ala Ser Thr Leu Leu His Ala Cys Asn Thr Tyr
          85          90          95
25     Cys Tyr Trp Leu Tyr Ser Glu Asp Pro Gln Thr Arg Lys Val Val Ile
          100         105         110
30     Thr Lys Gln His Thr Arg
          115

```

<210> 36

<211> 117

<212> PRT

<213> *Saccharomyces cerevisiae*

40 <400> 36

```

45     Ser Ser Leu Cys His Pro Leu Arg Lys Asn Ala Val Leu Asn Ala Asn
       1          5          10          15
      Pro Ala Lys Thr Ile Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly
          20          25          30
50     Val Leu Phe Ser Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His
          35          40          45
      Trp Thr Leu Pro Ile Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val
          50          55          60
55     Gly Pro Ile Trp Tyr Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro
       65          70          75          80
      Pro Asn Ser Gln Ala Ser Thr Leu Leu Leu Ala Leu Asn Thr Val Leu
          85          90          95
60     Leu Leu Leu Leu Ala Leu Thr Gln Leu Ser Gly Ser Val Ala Leu Ala
          100         105         110
65     Lys Ser His Leu Arg
          115

```

## ES 2 335 345 T3

<210> 37

<211> 103

<212> PRT

5 <213> *Kluyveromyces lactis*

<400> 37

```

10      Phe Thr Asp Ala Val Phe Ser Ala Met Arg Val Asn Ala Glu Gln Ile
        1          5          10          15
      Ala Tyr Ile Leu Leu Val Thr Asn Tyr Ile Gly Val Leu Phe Ala Arg
        20          25          30
15      Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Val

20          35          40          45
      Leu Leu Asn Trp Ala Asn Val Pro Tyr Pro Leu Cys Val Leu Trp Tyr
        50          55          60
25      Leu Thr His Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ala Thr Ala
        65          70          75          80
30      Ser Thr Leu Leu His Ala Cys Asn Thr Tyr Cys Tyr Trp Leu Tyr Ser
        85          90          95
      Glu Asp Pro Gln Thr Arg Lys
        100

```

35

<210> 38

<211> 106

<212> PRT

40 <213> *Arabidopsis thaliana*

<400> 38

```

45      Phe Ser Asp Val Ser Ala Ser Arg Ile Ile Thr Lys Glu His Val Val
        1          5          10          15
      Thr Ala Met Phe Val Gly Asn Phe Ile Gly Ile Val Phe Ala Arg Ser
        20          25          30
50      Leu His Tyr Gln Phe Tyr Ser Trp Tyr Phe Tyr Ser Leu Pro Tyr Leu
        35          40          45
55      Leu Trp Arg Thr Pro Phe Pro Thr Trp Leu Arg Leu Ile Met Phe Leu
        50          55          60
60      Gly Ile Glu Leu Cys Trp Asn Val Tyr Pro Ser Thr Pro Ser Ser Ser
        65          70          75          80
      Gly Leu Leu Leu Cys Leu His Leu Ile Ile Leu Val Gly Leu Trp Leu
        85          90          95
65      Ala Pro Ser Val Asp Pro Tyr Gln Leu Lys
        100          105

```





ES 2 335 345 T3

|    |   |      |
|----|---|------|
|    | 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   |      |
| 5  | 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   |      |
| 10 | 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   |      |
| 15 | 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   |      |
| 20 | 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   |      |
| 25 | 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   |      |
| 30 | 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   |      |
| 35 | 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   |      |
| 40 | gga cag tgg ggt caa cat cat gtg tgg 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   |      |
| 45 | 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   |      |
| 50 | 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   |      |
| 55 | 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   |      |
| 60 | 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   |      |
| 65 | 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   |      |
| 70 | 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   |      |

# ES 2 335 345 T3

```

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

5
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

10
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

15
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

20
cta gaa agt cac tgc aga gtg aac gga gcc tac tca gcc tta cgg gat 1776
Leu Glu Ser His Cys Arg Val Asn Gly Gly Tyr Ser Gly Leu Arg Asp
      580                585                590

25
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

30
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

35
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

40
cct ata ctc cgt gaa cag aag aag gaa att gat gcc aaa gag aaa tga 1968
Pro Ile Leu Arg Glu Gln Lys Lys Glu Ile Asp Gly Lys Glu Lys
      645                650                655

<210> 40
<211> 655
<212> PRT
45 <213> Mus musculus

<400> 40

50 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

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

60 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

65 Ala Leu Gln Pro Pro Ala Glu His Lys Pro Gly Leu Gly Ala Arg Ala

```

# ES 2 335 345 T3

|    | 85   | 90                                     | 95                                     |
|----|--|--|--|
| 5  | Glu Asp Ala Ala<br>100                     | Glu Gly Arg Val<br>105                 | Arg His Arg Glu Glu Gly Ala Pro<br>110 |
|    | Gly Asp Pro Gly Ala Gly Leu<br>115         | Glu Asp Asn Leu Ala<br>120             | Arg Ile Arg Glu<br>125                 |
| 10 | Asn His Glu Arg Ala Leu<br>130             | Arg Glu Ala Lys Glu<br>135             | Thr Leu Gln Lys Leu<br>140             |
|    | Pro Glu Glu Ile Gln Arg Asp Ile<br>145     | Leu Leu Glu Lys Glu Lys Val Ala<br>155 | 160                                    |
| 15 | Gln Asp Gln Leu Arg Asp Lys Asp<br>165     | Leu Phe Arg Gly Leu Pro Lys Val<br>170 | 175                                    |
|    | Asp Phe Leu Pro Pro Val Gly Val<br>180     | Glu Asn Arg Glu Pro Ala Asp Ala<br>185 | 190                                    |
| 20 | Thr Ile Arg Glu Lys Arg Ala Lys<br>195     | Ile Lys Glu Met Met Thr His Ala<br>200 | 205                                    |
|    | Trp Asn Asn Tyr Lys Arg Tyr Ala Trp<br>210 | Gly Leu Asn Glu Leu Lys Pro<br>215     | 220                                    |
|    | Ile Ser Lys Glu Gly His Ser Ser Ser<br>225 | Leu Phe Gly Asn Ile Lys Gly<br>230     | 235                                    |
| 30 | Ala Thr Ile Val Asp Ala Leu Asp Thr<br>245 | Leu Phe Ile Met Gly Met Lys<br>250     | 255                                    |
|    | Thr Glu Phe Gln Glu Ala Lys Ser Trp<br>260 | Ile Lys Lys Tyr Leu Asp Phe<br>265     | 270                                    |
| 35 | Asn Val Asn Ala Glu Val Ser Val<br>275     | Phe Glu Val Asn Ile Arg Phe Val<br>280 | 285                                    |
|    | Gly Gly Leu Leu Ser Ala Tyr Tyr Leu<br>290 | Ser Gly Glu Glu Ile Phe Arg<br>295     | 300                                    |
|    | Lys Lys Ala Val Glu Leu Gly Val Lys<br>305 | Leu Leu Pro Ala Phe His Thr<br>310     | 315                                    |
| 45 | Pro Ser Gly Ile Pro Trp Ala Leu Leu<br>325 | Asn Met Lys Ser Gly Ile Gly<br>330     | 335                                    |
|    | Arg Asn Trp Pro Trp Ala Ser Gly Gly<br>340 | Ser Ser Ile Leu Ala Glu Phe<br>345     | 350                                    |
| 50 | Gly Thr Leu His Leu Glu Phe Met<br>355     | His Leu Ser His Leu Ser Gly Asp<br>360 | 365                                    |
|    | Pro Val Phe Ala Glu Lys Val Met Lys<br>370 | Ile Arg Thr Val Leu Asn Lys<br>375     | 380                                    |
| 55 | Leu Asp Lys Pro Glu Gly Leu Tyr Pro<br>385 | Asn Tyr Leu Asn Pro Ser Ser<br>390     | 395                                    |
| 60 | Gly Gln Trp Gly Gln His His Val Ser<br>395 | Val Gly Gly Leu Gly Asp Ser<br>400     |  |

65

# ES 2 335 345 T3

|    | 405  | 410 | 415 |
|----|--|-----|-----|
| 5  | Phe Tyr Glu Tyr Leu Leu Lys Ala Trp Leu Met Ser Asp Lys Thr Asp<br>420 | 425 | 430 |
|    | Leu Glu Ala Lys Lys Met Tyr Phe Asp Ala Val Gln Ala Ile Glu Thr<br>435 | 440 | 445 |
| 10 | His Leu Ile Arg Lys Ser Ser Gly Gly Leu Thr Tyr Ile Ala Glu Trp<br>450 | 455 | 460 |
| 15 | Lys Gly Gly Leu Leu Glu His Lys Met Gly His Leu Thr Cys Phe Ala<br>465 | 470 | 475 |
|    | Gly Gly Met Phe Ala Leu Gly Ala Asp Gly Ala Pro Glu Ala Arg Ala<br>485 | 490 | 495 |
| 20 | Gln His Tyr Leu Glu Leu Gly Ala Glu Ile Ala Arg Thr Cys His Glu<br>500 | 505 | 510 |
| 25 | Ser Tyr Asn Arg Thr Tyr Val Lys Leu Gly Pro Glu Ala Phe Arg Phe<br>515 | 520 | 525 |
|    | Asp Gly Gly Val Glu Ala Ile Ala Thr Arg Gln Asn Glu Lys Tyr Tyr<br>530 | 535 | 540 |
| 30 | Ile Leu Arg Pro Glu Val Ile Glu Thr Tyr Met Tyr Met Trp Arg Leu<br>545 | 550 | 555 |
| 35 | Thr His Asp Pro Lys Tyr Arg Thr Trp Ala Trp Glu Ala Val Glu Ala<br>565 | 570 | 575 |
|    | Leu Glu Ser His Cys Arg Val Asn Gly Gly Tyr Ser Gly Leu Arg Asp<br>580 | 585 | 590 |
| 40 | Val Tyr Ile Ala Arg Glu Ser Tyr Asp Asp Val Gln Gln Ser Phe Phe<br>595 | 600 | 605 |
| 45 | Leu Ala Glu Thr Leu Lys Tyr Leu Tyr Leu Ile Phe Ser Asp Asp Asp<br>610 | 615 | 620 |
|    | Leu Leu Pro Leu Glu His Trp Ile Phe Asn Thr Glu Ala His Pro Phe<br>625 | 630 | 635 |
| 50 | Pro Ile Leu Arg Glu Gln Lys Lys Glu Ile Asp Gly Lys Glu Lys<br>645     | 650 | 655 |

<210> 41  
 <211> 4  
 <212> PRT  
 <213> Secuencia Artificial  
 <220>  
 <223> Descripción de Secuencia Artificial: péptido sintético  
 <400> 41  
 65 His Asp Glu Leu  
 1

## ES 2 335 345 T3

<210> 42  
<211> 4  
<212> PRT  
5 <213> Secuencia Artificial

<220>  
<223> Descripción de Secuencia Artificial: péptido sintético  
10  
<400> 42  
Lys Asp Glu Leu  
15 1

<210> 43  
<211> 36  
20 <212> ADN  
<213> Secuencia Artificial

<220>  
25 <223> Descripción de Secuencia Artificial: cebador sintético  
<400> 43  
30 tcttgcgcg cctcccgag agaactggcc tcctc 36

<210> 44  
<211> 33  
35 <212> ADN  
<213> Secuencia Artificial

<220>  
40 <223> Descripción de Secuencia Artificial: cebador sintético  
<400> 44  
45 ccgagcgcg ccacagagga actgcaccgg gtg 33

<210> 45  
<211> 1617  
50 <212> ADN  
<213> *Mus musculus*

55  
60  
65

## ES 2 335 345 T3

<400> 45

```

5      atgaagatga gacgctacaa gctctttctc atgttctgta tggctggcct gtgcctcata 60
      tcttctctgc acttctttaa gaccttatcc tatgtcacct tcccgagaga actggcctcc 120
      ctccagcccta acctcgtatc cagcttcttc tggacaatg cccctgteac tcccagggcc 180
      agtccgggagc cgggtggccc cgacctattg cggacacccc tctactecca ctctcccctg 240
      ctccagccac tgtccccgag caaggccaca gaggaactgc accgggtgga cttcgtgttg 300
10     ccggaggaca ccacggagta ttttgtgcgc accaaagctg gtggtgtgtg cttcaaacca 360
      ggtaccagga tgctggagaa accttcgcca gggcggacag aggagaagcc cgaagtgtct 420
      gagggtcctc cagcccgggg acctgctcgg aggccatga ggcacgtgtt gactacgagg 480
      gagcgcctgg gcagccgggg cactagggcg aagtgggttg agtgtgtgtg cctgccaggc 540
      tggcacgggc ccagttggg ggtgcccacg gtggtgcagt attccaacct gccaccaag 600
15     gaacgcctgg taccagggg ggtaccgagg cgggttatca acgccatcaa catcaaccac 660

```

```

20     gagttcgacc tgctggatgt gcgcttccat gagctgggag atgttgtgga cgccttcgtg 720
      gtctgtgaat ctaatttcac cgcctacggg gagcctcggc cgcctcaagt cccgagagatg 780
      ctgaccaatg gcacctcga gtacatccgc cacaaggtgc tctatgtctt cctggaccat 840
      tccccacctg gtggccgtca ggacggctgg attgcgatg actacctgcg caccttctc 900
      acccaggatg gcgtctccc cctgcgcaac ctggggccc atgacgtctt tatcatcgac 960
25     gatgoggagc agatccctgc gcgtgatggt gtgctgttcc tcaaactcta cgatggctgg 1020
      acagagccct tcgcttcca catgcggaag tccctgtatg gtttcttctg gaagcagccg 1080
      ggcacactgg aggtggtgtc aggtgcacc atggacatgc tgcaggccgt gtatgggctg 1140
      gatggcatcc gcctgcgcc cgcacagtac tacaccatgc ccaacttccg gcagtatgag 1200
      aacgcaccg gccacatcct agtgcagtgg tctctcggca gcccctgca cttcgcgggc 1260
30     tggcattget cctggtgctt cacacccgag ggcatctact ttaaactcgt gtcagcccag 1320
      aatggcgact tccccgctg ggtgactat gaggacaaga gggacctcaa ttacatccgc 1380
      agcttgatcc gcactggggg atggttcgac ggaacgcagc aggagtacc tctgcggac 1440
      cccagtgagc acatgtatgc tctaaatac ctgctcaaga actatgacca gttccgctac 1500
35     ttgctggaaa atccctaccg ggagcccaag agcactgtag aggggtgggc ccagaaccag 1560
      ggctcagatg gaaggccatc tgctgtcagg ggcaagtgg atacagtgga gggctag 1617

```

<210> 46

40 <211> 536

<212> PRT

<213> *Mus musculus*

45

50

55

60

65

# ES 2 335 345 T3

<400> 46

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 5  | Met | Arg | Arg | Tyr | Lys | Leu | Phe | Leu | Met | Phe | Cys | Met | Ala | Gly | Leu | Cys |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
|    | Leu | Ile | Ser | Phe | Leu | His | Phe | Phe | Lys | Thr | Leu | Ser | Tyr | Val | Thr | Phe |
|    |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| 10 | Pro | Arg | Glu | Leu | Ala | Ser | Leu | Ser | Pro | Asn | Leu | Ile | Ser | Ser | Phe | Phe |
|    |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| 15 | Trp | Asn | Asn | Ala | Pro | Val | Thr | Pro | Gln | Ala | Ser | Pro | Glu | Pro | Gly | Asp |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
|    | Pro | Asp | Leu | Leu | Arg | Thr | Pro | Leu | Tyr | Ser | His | Ser | Pro | Leu | Leu | Gln |
|    | 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| 20 | Pro | Leu | Ser | Pro | Ser | Lys | Ala | Thr | Glu | Glu | Leu | His | Arg | Val | Asp | Phe |
|    |     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
|    | Val | Leu | Pro | Glu | Asp | Thr | Thr | Glu | Tyr | Phe | Val | Arg | Thr | Lys | Ala | Gly |
| 25 |     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
|    | Gly | Val | Cys | Phe | Lys | Pro | Gly | Thr | Arg | Met | Leu | Glu | Lys | Pro | Ser | Pro |
|    |     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| 30 | Gly | Arg | Thr | Glu | Glu | Lys | Thr | Glu | Val | Ser | Glu | Gly | Ser | Ser | Ala | Arg |
|    | 130 |     |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
|    | Gly | Pro | Ala | Arg | Arg | Pro | Met | Arg | His | Val | Leu | Ser | Ser | Arg | Glu | Arg |
| 35 | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |
|    | Leu | Gly | Ser | Arg | Gly | Thr | Arg | Arg | Lys | Trp | Val | Glu | Cys | Val | Cys | Leu |
|    |     |     |     | 165 |     |     |     |     |     | 170 |     |     |     |     | 175 |     |
| 40 | Pro | Gly | Trp | His | Gly | Pro | Ser | Cys | Gly | Val | Pro | Thr | Val | Val | Gln | Tyr |
|    |     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |

45

50

55

60

65

# ES 2 335 345 T3

Ser Asn Leu Pro Thr Lys Glu Arg Leu Val Pro Arg Glu Val Pro Arg  
 195 200 205  
 Arg Val Ile Asn Ala Ile Asn Ile Asn His Glu Phe Asp Leu Leu Asp  
 210 215 220  
 Val Arg Phe His Glu Leu Gly Asp Val Val Asp Ala Phe Val Val Cys  
 225 230 235 240  
 Asp Ser Asn Phe Thr Ala Tyr Gly Glu Pro Arg Pro Leu Lys Phe Arg  
 245 250 255  
 Glu Met Leu Thr Asn Gly Thr Phe Glu Tyr Ile Arg His Lys Val Leu  
 260 265 270  
 Tyr Val Phe Leu Asp His Phe Pro Pro Gly Gly Arg Gln Asp Gly Trp  
 275 280 285  
 Ile Ala Asp Asp Tyr Leu Arg Thr Phe Leu Thr Gln Asp Gly Val Ser  
 290 295 300  
 Arg Leu Arg Asn Leu Arg Pro Asp Asp Val Phe Ile Ile Asp Asp Ala  
 305 310 315 320  
 Asp Glu Ile Pro Ala Arg Asp Gly Val Leu Phe Leu Lys Leu Tyr Asp  
 325 330 335  
 Gly Trp Thr Glu Pro Phe Ala Phe His Met Arg Lys Ser Leu Tyr Gly  
 340 345 350  
 Phe Phe Trp Lys Gln Pro Gly Thr Leu Glu Val Val Ser Gly Cys Thr  
 355 360 365  
 Met Asp Met Leu Gln Ala Val Tyr Gly Leu Asp Gly Ile Arg Leu Arg  
 370 375 380  
 Arg Arg Gln Tyr Tyr Thr Met Pro Asn Phe Arg Gln Tyr Glu Asn Arg  
 385 390 395 400  
 Thr Gly His Ile Leu Val Gln Trp Ser Leu Gly Ser Pro Leu His Phe  
 405 410 415  
 Ala Gly Trp His Cys Ser Trp Cys Phe Thr Pro Glu Gly Ile Tyr Phe  
 420 425 430  
 Lys Leu Val Ser Ala Gln Asn Gly Asp Phe Pro Arg Trp Gly Asp Tyr  
 435 440 445  
 Glu Asp Lys Arg Asp Leu Asn Tyr Ile Arg Ser Leu Ile Arg Thr Gly  
 450 455 460  
 Gly Trp Phe Asp Gly Thr Gln Gln Glu Tyr Pro Pro Ala Asp Pro Ser  
 465 470 475 480  
 Glu His Met Tyr Ala Pro Lys Tyr Leu Leu Lys Asn Tyr Asp Gln Phe  
 485 490 495  
 Arg Tyr Leu Leu Glu Asn Pro Tyr Arg Glu Pro Lys Ser Thr Val Glu  
 500 505 510  
 Gly Gly Arg Gln Asn Gln Gly Ser Asp Gly Arg Ser Ser Ala Val Arg  
 515 520 525  
 Gly Lys Leu Asp Thr Ala Glu Gly  
 530 535

65 <210> 47  
 <211> 21

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 5  | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
|    | <400> 47   |    |
| 10 | actgcatct gccttcgcca t   | 21 |
|    | <210> 48   |    |
| 15 | <211> 22   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 20 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
|    | <400> 48   |    |
| 25 | gtaatacgac tcactatagg gc   | 22 |
|    | <210> 49   |    |
| 30 | <211> 20   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 35 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
|    | <400> 49   |    |
| 40 | aattaaccct cactaaaggg  | 20 |
|    | <210> 50   |    |
| 45 | <211> 37   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 50 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
|    | <400> 50   |    |
| 55 | accgaggcgc gccatcaacg ccatcaacat caaccac                             | 37 |
|    | <210> 51   |    |
| 60 | <211> 37   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 65 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |

## ES 2 335 345 T3

|  |    |
|--|----|
| <400> 51   |    |
| aattaattaa ccctagccct ccgctgtatc caacttg                     | 37 |
| 5  |    |
| <210> 52   |    |
| <211> 36   |    |
| <212> ADN  |    |
| 10 <213> Secuencia Artificial                                |    |
| <220>  |    |
| <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 15   |    |
| <400> 52   |    |
| atgccctgg gggcctgtt gccgcttcc agtagc                         | 36 |
| 20   |    |
| <210> 53   |    |
| <211> 40   |    |
| <212> ADN  |    |
| 25 <213> Secuencia Artificial                                |    |
| <220>  |    |
| <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 30   |    |
| <400> 53   |    |
| tcatttctct ttgccatcaa ttccttctt ctgttcacgg                   | 40 |
| 35   |    |
| <210> 54   |    |
| <211> 44   |    |
| <212> ADN  |    |
| 40 <213> Secuencia Artificial                                |    |
| <220>  |    |
| <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 45   |    |
| <400> 54   |    |
| ggcgcgccga ctctccaag ctgctcagcg ggtcctgtt ccac               | 44 |
| 50   |    |
| <210> 55   |    |
| <211> 50   |    |
| <212> ADN  |    |
| 55 <213> Secuencia Artificial                                |    |
| <220>  |    |
| <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 60   |    |
| <400> 55   |    |
| ccctaattaa tcatttctct ttgccatcaa ttccttctt ctgttcacgg        | 50 |
| 65   |    |
| <210> 56   |    |
| <211> 51   |    |

## ES 2 335 345 T3

|    |   |    |
|----|---|----|
|    | <212> ADN   |    |
|    | <213> Secuencia Artificial                                    |    |
| 5  | <220>   |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético  |    |
|    | <400> 56  |    |
| 10 | ggcgagctcg gcctaccgg ccaaggctga gatcattgt ccagctcag a         | 51 |
|    | <210> 57  |    |
| 15 | <211> 55  |    |
|    | <212> ADN   |    |
|    | <213> Secuencia Artificial                                    |    |
|    | <220>   |    |
| 20 | <223> Descripción de Secuencia Artificial: cebador sintético  |    |
|    | <400> 57  |    |
| 25 | gccacgctcg acggatccgt taaacatcg atggagagg ctgacaccgc tacta    | 55 |
|    | <210> 58  |    |
|    | <211> 55  |    |
| 30 | <212> ADN   |    |
|    | <213> Secuencia Artificial                                    |    |
|    | <220>   |    |
| 35 | <223> Descripción de Secuencia Artificial: cebador sintético  |    |
|    | <400> 58  |    |
| 40 | cgggatccac tagtatttaa atcatatgtg cgagtgtaca actctccca catgg   | 55 |
|    | <210> 59  |    |
|    | <211> 55  |    |
| 45 | <212> ADN   |    |
|    | <213> Secuencia Artificial                                    |    |
|    | <220>   |    |
| 50 | <223> Descripción de Secuencia Artificial: cebador sintético  |    |
|    | <400> 59  |    |
| 55 | ggacgcgctcg acggcctacc cggccgtacg aggaatttct cggatgactc ttttc | 55 |
|    | <210> 60  |    |
| 60 | <211> 45  |    |
|    | <212> ADN   |    |
|    | <213> Secuencia Artificial                                    |    |
|    | <220>   |    |
| 65 | <223> Descripción de Secuencia Artificial: cebador sintético  |    |

## ES 2 335 345 T3

<400> 60  
cgggatccct cgagagatct tttttaga aatgtcttgg tgcct 45

5  
<210> 61  
<211> 63  
<212> ADN  
10 <213> Secuencia Artificial

<220>  
<223> Descripción de Secuencia Artificial: cebador sintético

15  
<400> 61  
ggacatgcat gcactagtgc ggccgccacg tgatagttgt tcaattgatt gaaataggga 60  
caa 63

20  
<210> 62  
<211> 52  
<212> ADN  
25 <213> Secuencia Artificial

<220>  
<223> Descripción de Secuencia Artificial: cebador sintético

30  
<400> 62  
ccttgctage ttaattaacc ggggcacgtc cgacggcggc ccacgggtcc ca 52

35  
<210> 63  
<211> 61  
<212> ADN  
40 <213> Secuencia Artificial

<220>  
<223> Descripción de Secuencia Artificial: cebador sintético

45  
<400> 63  
ggacatgcat ggggatccct taagagccgg cagcttgcaa attaaagcct tcgagcgtcc 60  
c 61

50  
<210> 64  
<211> 61  
<212> ADN  
55 <213> Secuencia Artificial

<220>  
<223> Descripción de Secuencia Artificial: cebador sintético

60  
<400> 64  
gaaccacgtc gagggccatt ggggccaaaa ctttttttcc tattcaaaca caaggcattg 60  
c 61

65  
<210> 65  
<211> 43

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                     |    |
| 5  | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético   |    |
|    | <400> 65   |    |
| 10 | ctccaatact agtcgaagat tatcttctac ggtgcctgga etc                | 43 |
|    | <210> 66   |    |
| 15 | <211> 53   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                     |    |
| 20 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético   |    |
|    | <400> 66   |    |
| 25 | tggaaggttt aaacaaagct agagtaaaat agatatagcg agattagaga atg     | 53 |
|    | <210> 67   |    |
| 30 | <211> 50   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                     |    |
| 35 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético   |    |
|    | <400> 67   |    |
| 40 | aagaattcgg ctggaaggcc ttgtacctg atgtagttcc cgttttcac           | 50 |
|    | <210> 68   |    |
| 45 | <211> 58   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                     |    |
| 50 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético   |    |
|    | <400> 68   |    |
| 55 | gcccaagccg gccttaagg atctcctgat gactgactca ctgataataa aaatacgg | 58 |
|    | <210> 69   |    |
| 60 | <211> 59   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                     |    |
| 65 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético   |    |

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <400> 69   |    |
|    | gggcgcgtat ttaaacta gtagctat cgaatctaaa tgtaagtaa aatctctaa          | 59 |
| 5  | <210> 70   |    |
|    | <211> 39   |    |
|    | <212> ADN  |    |
| 10 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
| 15 | <400> 70   |    |
|    | ggccgctgc agattaaat gaattcggcg cgcctaat                              | 39 |
| 20 | <210> 71   |    |
|    | <211> 34   |    |
|    | <212> ADN  |    |
| 25 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
| 30 | <400> 71   |    |
|    | taaggcgcgc cgaatlcatt taaatctgca gggc                                | 34 |
| 35 | <210> 72   |    |
|    | <211> 29   |    |
|    | <212> ADN  |    |
| 40 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
| 45 | <400> 72   |    |
|    | tggcagcgc gcctcagta gcgctctcg  | 29 |
| 50 | <210> 73   |    |
|    | <211> 29   |    |
|    | <212> ADN  |    |
| 55 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
| 60 | <400> 73   |    |
|    | aggtaatta agtgtaatt ccagctagg  | 29 |
| 65 | <210> 74   |    |
|    | <211> 26   |    |

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 5  | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
|    | <400> 74   |    |
| 10 | ccagaagaat tcaattytgy cartgg   | 26 |
|    | <210> 75   |    |
| 15 | <211> 25   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 20 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
|    | <400> 75   |    |
| 25 | cagtgaaaat acctggncn gtcca   | 25 |
|    | <210> 76   |    |
| 30 | <211> 27   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 35 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
|    | <400> 76   |    |
| 40 | tgccatcttt tagtccagg cccgttc   | 27 |
|    | <210> 77   |    |
| 45 | <211> 27   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 50 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
|    | <400> 77   |    |
| 55 | gatcccacga cgcacgtat ttcttc  | 27 |
|    | <210> 78   |    |
| 60 | <211> 35   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial   |    |
| 65 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <400> 78   |    |
|    | gggttttgt ttctagatc ttgcaytay cartt                                  | 35 |
| 5  | <210> 79   |    |
|    | <211> 36   |    |
|    | <212> ADN  |    |
| 10 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
| 15 | <400> 79   |    |
|    | agaatttggg gggtagaagt tccarcacca ytcttg                              | 36 |
| 20 | <210> 80   |    |
|    | <211> 32   |    |
|    | <212> ADN  |    |
| 25 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
| 30 | <400> 80   |    |
|    | cctaagctgg tatgcgttct cttgccata tc                                   | 32 |
| 35 | <210> 81   |    |
|    | <211> 30   |    |
|    | <212> ADN  |    |
| 40 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: oligonucleótido sintético |    |
| 45 | <400> 81   |    |
|    | gcggcataaa caataataga tgctataaag                                     | 30 |
| 50 | <210> 82   |    |
|    | <211> 24   |    |
|    | <212> ADN  |    |
| 55 | <213> Secuencia Artificial   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético         |    |
| 60 | <400> 82   |    |
|    | ccacatcatc cgtgctacat atag   | 24 |
| 65 | <210> 83   |    |
|    | <211> 44   |    |

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 5  | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 83   |    |
| 10 | acgaggcaag ctaaacagat ctggaagtat cgagggttat ccag             | 44 |
|    | <210> 84   |    |
| 15 | <211> 44   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 20 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 84   |    |
| 25 | ccatccagtg tcgaaaacga gccaatggtt catgtctata aatc             | 44 |
|    | <210> 85   |    |
| 30 | <211> 24   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 35 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 85   |    |
| 40 | agcctcagcg ccaacaagcg atgg                                   | 24 |
|    | <210> 86   |    |
| 45 | <211> 44   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 50 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 86   |    |
| 55 | ctggataacc ctcgatactt cgagatctgt ttagcttgcc tcgt             | 44 |
|    | <210> 87   |    |
| 60 | <211> 44   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 65 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <400> 87   |    |
|    | gatttataga catgaacat tggctcgttt tcgacctgg atgg               | 44 |
| 5  | <210> 88   |    |
|    | <211> 20   |    |
|    | <212> ADN  |    |
| 10 | <213> Secuencia Artificial                                   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 15 | <400> 88   |    |
|    | atcctttacc gatgctgtat  | 20 |
| 20 | <210> 89   |    |
|    | <211> 27   |    |
|    | <212> ADN  |    |
| 25 | <213> Secuencia Artificial                                   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 30 | <400> 89   |    |
|    | ataacagtat gtgttacacg cgtgtag                                | 27 |
| 35 | <210> 90   |    |
|    | <211> 25   |    |
|    | <212> ADN  |    |
| 40 | <213> Secuencia Artificial                                   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 45 | <400> 90   |    |
|    | ttctcactg cagtcttcta taact                                   | 25 |
| 50 | <210> 91   |    |
|    | <211> 26   |    |
|    | <212> ADN  |    |
| 55 | <213> Secuencia Artificial                                   |    |
|    | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
| 60 | <400> 91   |    |
|    | tggagacat gaggtccgc atctac                                   | 26 |
| 65 | <210> 92   |    |
|    | <211> 35   |    |

## ES 2 335 345 T3

|    |  |    |
|----|--|----|
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 5  | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 92   |    |
| 10 | ttggcgcgcc tcctagtgt accagttgaa ctttg                        | 35 |
|    | <210> 93   |    |
| 15 | <211> 32   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 20 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 93   |    |
| 25 | gattaattaa ctcaactgcag tcttctataa ct                         | 32 |
|    | <210> 94   |    |
| 30 | <211> 36   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 35 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 94   |    |
| 40 | tctggcgcgc cctcccagag agaactggcc tcctc                       | 36 |
|    | <210> 95   |    |
| 45 | <211> 37   |    |
|    | <212> ADN  |    |
|    | <213> Secuencia Artificial                                   |    |
| 50 | <220>  |    |
|    | <223> Descripción de Secuencia Artificial: cebador sintético |    |
|    | <400> 95   |    |
| 55 | aattaattaa ccctagccct cgctgtatc caacttg                      | 37 |
| 60 |  |    |
| 65 |  |    |