(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 654 664**

(51) Int. Cl.:

C12N 1/19	(2006.01)
C12N 15/09	(2006.01)
C12R 1/84	(2006.01)
C12N 9/10	(2006.01)
C12N 9/24	(2006.01)
C12N 15/81	(2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(86) Fecha de presentación y número de la solicitud internacional: **23.10.2013** **PCT/US2013/066335**

(87) Fecha y número de publicación internacional: **01.05.2014** **WO14066479**

(96) Fecha de presentación y número de la solicitud europea: **23.10.2013** **E 13849948 (8)**

(97) Fecha y número de publicación de la concesión europea: **13.12.2017** **EP 2912162**

(54) Título: **Cepas de Pichia pastoris para la producción de una estructura de glicano predominantemente homogénea**

(30) Prioridad:

23.10.2012 US 201261717423 P

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
14.02.2018

(73) Titular/es:

RESEARCH CORPORATION TECHNOLOGIES, INC. (100.0%)
6440 N. Swan Road, Suite 200
Tucson, AZ 85718, US

(72) Inventor/es:

GEHLSEN, KURT, R. y
CHAPPELL, THOMAS, G.

(74) Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 654 664 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

Cepas de *Pichia pastoris* para la producción de una estructura de glicano predominantemente homogénea

Referencia cruzada con una solicitud relacionada

La presente solicitud reivindica el beneficio de prioridad de la Solicitud Provisional de EE. UU. Nº 61/717.423, presentada el 23 de octubre de 2012.

Antecedentes de la invención

Pichia pastoris es un sistema altamente satisfactorio para la producción de una amplia variedad de proteínas recombinantes. Varios factores han contribuido a su aceptación rápida, incluyendo: (1) un promotor derivado del gen de la alcohol oxidasa I (AOX1) de *P. pastoris* que es adecuado únicamente para la expresión controlada de genes ajenos; (2) la similitud de técnicas necesarias para la modificación genética de P. pastores con las del Saccharomyces cerevisiae; y (3) la fuerte preferencia de *P. pastoris* por el crecimiento respiratorio, un rasgo fisiológico que facilita su cultivo a alta densidad celular con respecto a las levaduras fermentadoras.

Como levadura, la *P. pastoris* en un microorganismo unicelular que es fácil de modificar y cultivar. Sin embargo, también es una eucariota capaz de muchas de las modificaciones post-traduccionales que se llevan a cabo por las células eucariotas superiores tales como el procesamiento proteolítico, plegamiento, formación de enlaces y glicosilación. Por lo tanto, muchas proteínas que terminarían como cuerpos de inclusión en los sistemas bacterianos se producen como moléculas biológicas activas en *P. pastoris*. El sistema de *P. pastoris* también se considera en general como más rápido, fácil, y menos caro de uso que los sistemas de expresión derivados de las eucariotas superiores tales como los sistemas celulares de cultivo tisular de mamíferos e insectos y habitualmente produce mayores niveles de expresión.

La P. pastores tiene el potencial de llevar a cabo muchas de las modificaciones post-traduccionales normalmente asociadas con eucariotas superiores. Estas incluyen el procesamiento de secuencias de señal (tanto de tipo pre y prepro), plegamiento, formación de puentes disulfuro, y glicosilación unida tanto a O como a N. La glicosilación de proteínas ajenas secretadas de eucariotas (superiores) por *P. pastoris* y otros hongos puede ser problemática. En los mamíferos, los oligosacáridos unidos a O están compuestos por una variedad de azúcares que se incluyen la N-acetilgalactosamina, galactosa y ácido siálico. Por el contrario, las eucariotas inferiores, incluyendo la *P. pastoris*, pueden añadir oligosacáridos compuestos solamente de restos de manosa (Man).

La N-glicosilación en *P. pastoris* también es diferente que en las eucariotas superiores. En todas alas eucariotas, comienza en el RE con la transferencia de una unidad de oligosacárido unido a un lípido, Glc3Man9GlcNAc2 (Glc = glucosa; GlcNAc = N-acetilglucosamina), a asparagina en la secuencia de reconocimiento Asn-X-Ser/Thr. Esta unidad nuclear de oligosacárido se recorta posteriormente a Man8GlcNac2. Es en este punto en el que los patrones de glicosilación de eucariotas superiores e inferiores comienzan a diferenciarse. El aparato de Golgi de mamíferos lleva a cabo una serie de recortes y reacciones de adición que generan oligosacáridos compuestos de Man5-6GlcNac2 (tipo manosa superior), una mezcla de varios azúcares diferentes (tipo complejos) o una combinación de ambas (tipo híbrido). Se han observado dos patrones distintos de N-glicosilación sobre las proteínas ajenas secretadas por *P. pastoris*. Algunas proteínas se secretan con estructuras de carbohidrato similares en tamaño y estructura a la unidad nuclear (Man8-11GlcNac2). Otras proteínas ajenas secretadas de *P. pastoris* reciben muchos más carbohidratos y parecen estar hiperglicosiladas.

Los oligosacáridos altos en manosa unidos a N añadidos a proteínas por las levaduras representan un problema en el uso de proteína ajenas secretadas por la industria farmacéutica. Por ejemplo, pueden ser excesivamente antigénicos cuando se introducen por vía intravenosa en mamíferos y adicionalmente pueden producir un rápido aclaramiento de la proteína del cuerpo mediante el hígado.

En un intento para modificar la ruta de N-glicosilación de *Pichia pastoris*, se creó una cepa (a la que se hace referencia de aquí en adelante como “M5-Blast”), como se describe en Jacobs et al., 2009, Nature Protocols 4:58-70. M5-Blast es una modificación de la cepa GS115 de *P. pastoris* en la que el gen endógeno de la manosiltranferasa OCH1 está alterado por la introducción de un casete que comprende un gen de la α-1,2 manosidasa. Sin embargo, la cepa M5-Blast se somete a reordenamientos genómicos que regeneran el gen OCH1 endógeno y en paralelo retiran el gen de la α-1,2 manosidasa después de rondas de congelación descongelación, cultivándola en distintas temperaturas y condiciones, y por transformaciones posteriores con otros plásmidos para introducir genes exógenos.

Sumario de la divulgación

Se desvelan en el presente documento nuevas cepas de *Pichia pastoris* para la expresión de proteínas exógenas con N-glicanos sustancialmente homogéneos. Más específicamente, las cepas se modifican genéticamente para incluir un alelo OCH1 mutante que se transcribe en un ARNm que codifica un producto genético OCH1 mutante (es decir, α-1,6 manosiltransferasa, o “proteína OCH1”). La proteína OCH1 mutante contiene un dominio catalítico al

menos un 95 % idéntico con los aminoácidos 45-404 de la SEQ ID NO: 2 y mantiene sustancialmente la actividad catalítica de la proteína OCH1 tipo silvestre que comprende la SEQ ID NO: 2, pero carece de la secuencia del extremo N para dirigir la proteína OCH1 mutante al aparato de Golgi y carece de un dominio de anclaje a la membrana en la región del extremo N. Las cepas no incluyen ningún otro alelo OCH1 que produce un ARNm que codifica una proteína funcional OCH1. Dichas cepas son robustas, estables, y transformables, y el alelo OCH1 mutante y el fenotipo asociado (es decir la capacidad para producir N-glicanos sustancialmente homogéneos) se mantienen durante generaciones, después de rondas de congelación y descongelación, y después de transformaciones posteriores.

Esta divulgación también caracteriza métodos de construcción de cepas, así como métodos de expresión de proteína mediante las cepas.

Cualquier característica o combinación de características descritas en el presente documento se incluyen en el alcance de la presente divulgación a condición de que las características incluidas en cualquier dicha combinación no sean inconsistentes mutuamente como será evidente por el contexto, la presente memoria descriptiva, y el conocimiento de un experto habituado en la técnica. Las ventajas adicionales y aspectos de la presente invención son evidentes en la siguiente descripción detallada y las reivindicaciones.

Breve descripción de los dibujos

Figura 1. Diagrama de la estrategia de eliminación para retirar el fragmento de homología OCH1 del extremo N del genoma de Pichia M5-Blast. Los brazos de homología que flanquean el fragmento OCH1 del extremo N se utilizan para crear una construcción de intercambio doble que contiene los sitios de recombinación *lox71-lox66*. Las secuencias que intervienen se pueden retirar por recombinación mediada por *cre*.

Figura 2. Los cebadores de la PCR para la amplificación de los brazos que flanquean la construcción de intercambio doble del ADN genómico de M5-Blast.

Figura 3. Reacciones de PCR para la adición de los sitios *lox* a los extremos de los brazos de homología.

Figura 4. Adición de extensiones de ADN genómico de M5-Blast en el casete existente *tox71-MazF-Nat^R-lox66*.

Figura 5. Ensamblaje solapado y amplificación de la secuencia final para generar el fragmento de intercambio doble para la transformación de Pichia M5-Blast. Esta construcción final tiene > 500 pb de brazos de homología que flanquean el casete de selección/contra-selección.

Figura 6. Pares de cebadores de PCR utilizados para generar el fragmento de ADN para el evento de recombinación de doble intercambio.

Figura 7. Disposición teórica de la región LEU5-manosidasaHDEL del genoma de M5-Blast después del evento de recombinación de doble intercambio.

Figura 8. Disposición teórica del ADN genómico entre *LEU5* y manosidasaHDEL después de la recombinación con *cre*.

Figura 9. Cebadores de la PCR 81487118-80765854 que se utilizan para verificar la secuencia de ADN de la región que se podía haber derivado de los productos de la PCR transformados en la cepa M5-Blast.

Figura 10. Análisis de N-glicano de una proteína recombinante expresadas en distintas cepas de *P. pastoris*.

Figura 11. Análisis de N-glicano del trastuzumab obtenidos en el Estudio 1 descrito en el Ejemplo 6.

Figura 12. Comparación de afinidad de unión a Her2 de trastuzumab tipo Man5 (Estudio 1) y Herceptina comercial por ELISA.

Figura 13. Análisis DSA-FACE del agrupamiento total de N-glicano en proteínas medias de cepas ‘Trans’. A. resultado de una referencia malto-dextrosa. Los Paneles B a F muestran los resultados de N-glicanos, de la siguiente manera: B. cepa GS Trans en microescala; C. cepa M5 Trans en microescala: D. cepa GS Trans en biorreactor; E. cepa M5 Trans en biorreactor; F. referencia de N-glicanos de RNasa B bovina.

Figura 14. Análisis DSA-FACE del agrupamiento total de N-glicano de proteínas media en cepas ‘CalB’. A. resultado de una referencia malto-dextrosa. Los Paneles B a F muestran los resultados de N-glicanos, de la siguiente manera: B. cepa GS ‘CalB’ en microescala; C. cepa M5 ‘CalB’ en microescala: D. cepa GS ‘CalB’ en biorreactor; E. cepa M5 ‘CalB’ en biorreactor; F. referencia de N-glicanos de RNasa B bovina.

Figura 15. Análisis DSA-FACE del agrupamiento total de N-glicano de proteínas media en cepas 'CalA'. A. resultado de una referencia malto-dextrosa. Los Paneles B a F muestran los resultados de N-glicanos, de la siguiente manera: B. cepa GS 'CalA' en microescala; C. cepa M5 'CalA' en microescala: D. cepa GS 'CalA' en biorreactor; E. cepa M5 'CalA' en biorreactor; F. referencia de N-glicanos de RNasa B bovina.

Las **Tablas 1-6** se exponen en las páginas 36-49. Las **Tablas 7-8** se encuentran en la página 32 (Ejemplo 7).

La **Tabla 1** enumera la secuencia de ADN (SEQ ID NO: 1) del locus *OCH1* en una cepa SuperM5 descrita en el Ejemplo 1.

La **Tabla 2** enumera la secuencia de aminoácidos de OCH1 tipo silvestre (SEQ ID NO: 2) en *Pichia pastoris*.

La **Tabla 3** enumera los nucleótidos que se pueden eliminar del segmento OCH1 corriente arriba.

La **Tabla 4** enumera la secuencia de ADN del locus *OCH1* (+/- 2 kb) para la cepa M5-Blast de *Pichia pastoris*.

La **Tabla 5** enumera la secuencia de aminoácidos y secuencia de nucleótidos para el segmento *OCH1* corriente arriba.

La **Tabla 6** enumera la secuencia de aminoácidos y secuencia de nucleótidos para el segmento OCH1 corriente abajo.

Tabla 7, análisis de N-glicano de trastuzumab obtenido del Estudio 2 (Ejemplo 6).

Tabla 8, parámetros cinéticos de trastuzumab analizado en BIAcore (Ejemplo 6).

Descripción detallada

Cepas de *Pichia pastoris* modificadas genéticamente

Esta divulgación caracteriza nuevas cepas de *Pichia pastoris* modificadas genéticamente que son robustas, estables, y transformables, y que producen proteínas con estructuras de N-glicanos sustancialmente homogéneas.

Un primer aspecto de la presente invención proporciona una cepa de *Pichia pastoris* modificada estable, que comprende:

un alelo OCH1 mutante que se transcribe en un ARNm que codifica una proteína OCH1 mutante, donde la proteína OCH1 mutante comprende un dominio catalítico a menos un 95 % idéntico con los aminoácidos 45-404 de la SEQ ID NO: 2 y sustancialmente mantiene la actividad catalítica de la proteína OCH1 de tipo silvestre que comprende la SEQ ID NO: 2, donde la proteína OCH1 mutante carece de secuencia del extremo N para dirigir la proteína OCH1 mutante al aparato de Golgi y carece de un dominio de anclaje a la membrana en la región del extremo N, y donde dicha cepa produce N-glicanos sustancialmente homogéneos.

Se definen aspectos adicionales de la invención en las reivindicaciones adjuntas.

Como se describe adicionalmente en el presente documento, las cepas se modifican genéticamente para que incluyan un alelo OCH1 mutante que se transcribe en un ARNm que codifica un producto genético OCH1 mutante (es decir, α-1,6 manosiltransferasa, o "proteína OCH1"). Las cepas no incluyen ningún otro alelo OCH1 que produzca un ARNm que codifica una proteína OCH1 funcional.

Las cepas se pueden modificar genéticamente adicionalmente para contener un ácido nucleico codificante para y que expresen una 1,2-manosidasa que convierte el N-glicano M8, Man8GlcNac2, en el M5 N-glicano, Man5GlcNac2.

Como resultado de las modificaciones genéticas, las cepas desveladas en el presente documento producen N-glicanos sustancialmente homogéneos.

Por N-glicanos "sustancialmente homogéneos" significa que dada una preparación que contiene una población de una glicoproteína particular de interés, al menos un 50 %, 60 %, 75 %, 80 %, 85 %, 90 % o incluso el 95 % de los N-glicanos en las moléculas de proteínas de la población son los mismos.

Por "estructura de N-glicano predominante" o "glicoforma predominante" significa una estructura de N-glicano específico o glicoforma de (por ejemplo, anclada a) una proteína que constituye el mayor porcentaje de todas las estructuras de N-glicano o glicoformas de la proteína. En ciertas realizaciones específicas, una glicoforma predominante representa al menos un 40 %, 50 %, 60 %, 70 %, 80 %, 90 % o 95 % o más de la población de todas las glicoformas de la proteína. Ejemplos de estructuras de N-glicano deseables incluyen, por ejemplo, Man8GlcNAc2 (o "M8") o Man5GlcNAc2 ("M5"). Las estructuras de N-glicano deseables adicionales incluyen, GnM5

($GlcNAcMan_5GlcNAc_2$), GalGnM5 ($GalGlcNAcMan_5GlcNAc_2$), GalGnM3 ($GalGlcNAcMan_3GlcNAc_2$), GnM3 ($GlcNAcMan_3GlcNAc_2$), Gn2M3 ($GlcNAc_2Man_3GlcNAc_2$), y Gal2Gn2M3 ($Gal_2GlcNAc_2Man_3GlcNAc_2$). Las estructuras de estos N-glicanos se han descrito, por ejemplo, en Jacobs et al., 2009, Nature Protocols 4:58-70.

En una realización específica, las cepas de la presente invención incluyen tanto un alelo OCH1 mutante y un ácido nucleico que codifica y expresa una α-1,2 manosidasa, de manera que las cepas producen N-glicanos homogéneos siendo la M5 la glicoforma predominante. También se hace referencia a estas cepas en el presente documento como cepas SuperM5 o SuperMan5. Un ejemplo de una cepa SuperM5 se describe en la sección de los Ejemplos posteriormente.

Las cepas de la presente invención son “robustas”, lo que significa que las cepas (a menos de que se señale de otra manera como una cepa auxótrofa o deficiente, por ejemplo, deficiente en proteasas o mutante AOX1, etc.) tienen aproximadamente la misma tasa de crecimiento y las mismas condiciones de cultivo que las cepas de *Pichia pastoris* no modificadas tales como la cepa GS115. Por ejemplo, las cepas de la presente invención pueden cultivarse a temperaturas elevadas (por ejemplo, 30 ºC, 37 ºC o incluso 42 ºC) y no son sensibles a la temperatura. Por ejemplo, las cepas SuperM5 que se desvelan en el presente documento son robustas y no son sensibles a la temperatura.

Las cepas de la presente invención también son estables, lo que significa que las modificaciones genéticas y el fenotipo que resulta de las modificaciones genéticas (es decir, la producción de N-glicanos homogéneos) se mantienen a lo largo de generaciones, por ejemplo, al menos 10, 20, 30, 40 o 50 generaciones (divisiones celulares), después de rondas de congelación y descongelación, y después de transformaciones posteriores. Por ejemplo, las cepas SuperM5 desveladas en el presente documento mantienen el alelo OCH1 mutante a lo largo de generaciones y son capaces de continuar produciendo M8 sustancialmente homogéneo (u otros N-glicanos corriente abajo), sin inversión.

Modificación genética – alelo OCH1 mutante

Las cepas de la presente invención están modificadas genéticamente para incluir un alelo OCH1 mutante que se transcribe en un ARNm que codifica un producto genético OCH1 mutante (es decir, α-1,6 manosiltransferasa, o “proteína OCH1”). La proteína OCH1 mutante contiene un dominio catalítico al menos un 95 % idéntico a los aminoácidos 45-404 de SEQ ID NO: 2 y mantiene sustancialmente la actividad catalítica de la proteína OCH1 tipo silvestre que comprende la SEQ ID NO: 2 pero carece de una secuencia del extremo N para la dirección de la proteína OCH1 mutante al aparato de Golgi y carece de un dominio de anclaje a la membrana en la región del extremo N.

El gen OCH1 del tipo silvestre de *Pichia pastoris* tiene una fase de lectura abierta que codifica una proteína de 404 aminoácidos (SEQ ID NO: 2). Como otras glicosiltransferasas del aparato de Golgi fúngicas, la proteína OCH1 de *Pichia pastoris* es una proteína de membrana de tipo II, como cola citoplasmática corta (Met1 a Tyr21 (SEQ ID NO: 25) o Ala2 a Tyr21), un dominio de anclaje a la membrana (Phe22 a Ser44, es decir FYMAIFAVSVICVLYGPSQQLSS (SEQ ID NO: 89)), una región de tallo, y una gran región del extremo C que contiene el dominio catalítico. Véase, por ejemplo, Kim et al., J. Biol. Chem. 281:6261-6272 (2006); Nakayama et al., EMBO 11(7): 2511-2519 (1992); y Tu et al., Cell. Mol. Life Sci. 67:29-41 (2010).

La proteína OCH1 de tipo silvestre se localiza generalmente en cis-Golgi. Se cree que la localización en el aparato de Golgi de la proteína OCH1 de tipo silvestre está determinada por la región del extremo N que consiste en la cola citoplasmática, el dominio de anclaje a la membrana, y la región de tallo. En particular, el dominio de anclaje a la membrana, incluyendo sus constituyentes de aminoácidos, y la longitud, tiene un papel importante en la dirección al aparato de Golgi de la proteína. Véase, por ejemplo, Tu et al. (*Supra*).

La proteína OCH1 mutante de la presente divulgación tiene una secuencia del extremo N que altera la localización en el aparato de Golgi de la proteína OCH1 mutante, en comparación con la proteína OCH1 de tipo silvestre. Como resultado de esta secuencia alterada del extremo N, la proteína OCH1 mutante no se dirige o retiene apropiadamente en el aparato de Golgi, o no se dirige o retiene apropiadamente en el compartimento correcto en el aparato de Golgi. El término “direccionamiento” significa los mecanismos biológicos por el que se transportan las proteínas a los destinos apropiados en la célula o fuera de la célula. En realizaciones especificas la proteína OCH1 mutante de la presente divulgación carece de una secuencia del extremo N que permite el direccionamiento hacia el aparato de Golgi de la proteína OCH1 mutante, de manera que la proteína OCH1 mutante no se dirige al aparato de Golgi y se transporta a otra localización celular o se secreta fuera de la célula.

En algunas realizaciones, la alteración de la secuencia del extremo N es el resultado de una mutación, es decir, la adición, eliminación o sustitución, de uno o más aminoácidos del dominio de anclaje a la membrana de la proteína OCH1. En realizaciones específicas, uno o más aminoácidos del dominio de anclaje a la membrana se han eliminado. En realizaciones particulares, al menos 2, 3, 4, 5, 6, 7 o más aminoácidos, contiguos o de otra manera, del dominio de anclaje a la membrana se han eliminado. Por ejemplo, se han eliminado algunos o los 5 aminoácidos (FYMAI, SEQ ID NO: 90) del dominio de anclaje a la membrana.

En otras realizaciones, la alteración de la secuencia del extremo N es el resultado de una mutación, es decir, la adición, eliminación o sustitución, de uno o más aminoácidos de la cola citoplasmática de la proteína OCH1. En realizaciones específicas, uno o más aminoácidos en la cola citoplasmática se han eliminado; por ejemplo, al menos 2, 3, 4, 5, 6, 7 o más aminoácidos, contiguos o de otra manera, de la cola citoplasmática se han eliminado. Ejemplos de eliminaciones en la cola citoplasmáticas se encuentran en la Tabla 3. En otras realizaciones, la eliminación de uno o más aminoácidos se combinan con la adición de uno o más aminoácidos de la cola citoplasmática.

En otras realizaciones más, la alteración de la secuencia del extremo N es el resultado de una mutación de uno o más aminoácidos en la región tallo de la proteína OCH1; por ejemplo una eliminación de uno o más aminoácidos en los primeros 10, 20, 30, 40, 50 o 60 aminoácidos inmediatamente a continuación del dominio de anclaje a la membrana.

En ciertas realizaciones, la alteración en la secuencia del extremo N es un resultado de una combinación de mutaciones en la cola citoplasmática, el dominio de anclaje a la membrana, y/o la región del tallo de la proteína OCH1.

En realizaciones específicas, la alteración en la secuencia del extremo N es el resultado de una combinación de mutaciones en la cola citoplasmática y el dominio de anclaje a la membrana. Por ejemplo, se han eliminado uno o más aminoácidos en la cola citoplasmática y uno o más aminoácidos en el dominio de anclaje a la membrana. Ejemplos de eliminaciones en la región del extremo N de la proteína OCH1 se enumeran en la Tabla 3.

En otras realizaciones, además de las eliminaciones en uno o más dominios, se añaden uno o más aminoácidos al extremo N de la proteína, a condición de que la secuencia mutante resultante en el extremo N siga fallando en dirigirse o localizar apropiadamente la diana de la proteína OCH1 en el aparato de Golgi. Por ejemplo, que la secuencia mutante resultante en el extremo N siga careciendo de un dominio de anclaje a la membrana funcional. Se puede determinar fácilmente si una secuencia mutante incluye un dominio de anclaje a la membrana basándose en la composición de aminoácidos y la longitud. El dominio de anclaje a la membrana de la glicotransferasa del aparato de Golgi normalmente consiste en 16-20 aminoácidos, que son hidrófobos y a menudo contiene aminoácidos aromáticos, y tiene aminoácidos hidrófobos cargados positivamente inmediatamente fuera de ambos extremos del tramo de membrana. Véase, por ejemplo, Nakayama et al. (1992), *supra*. Un ejemplo de una proteína OCH1 mutante se expone en SEQ ID NO: 3, que tiene sus 10 primeros aminoácidos en lugar de los 26 primeros aminoácidos de la proteína OCH1 de tipo silvestre.

La proteína OCH1 mutante desvelada en el presente documento contiene un dominio catalítico que es sustancialmente idéntico a la de la proteína OCH1 de tipo silvestre.

El dominio catalítico de la proteína OCH1 de tipo silvestre se localiza en el fragmento del extremo C de 360 aminoácidos (es decir, en los aminoácidos 45 a 404 de la SEQ ID NO: 2). En algunas realizaciones la proteína OCH1 mutante comprende un fragmento terminal C que es sustancialmente idéntico a los aminoácidos 45-404, 55-404, 65-404, 75-404, 85-404, 95-404, o 105-404 de SEQ ID NO: 2. “Sustancialmente idéntico” significa que las secuencias, cuando se alinean en sus longitudes completas, son al menos un 90 %, 95 %, 98 %, 99 % idénticos o más. En la mayoría de las realizaciones, el dominio catalítico de la proteína OCH1 mutante no se diferencia del dominio de tipo silvestre en más de 10 aminoácidos, 8 aminoácidos, 5 aminoácidos, 3 aminoácidos, o 2 aminoácidos. En realizaciones específicas, el dominio catalítico de la proteína OCH1 mutante es idéntico que el de la proteína OCH1 de tipo silvestre. Cuando uno o más aminoácidos son diferentes, es preferible que las diferencias representen sustituciones conservadoras de aminoácidos. Ejemplos de sustituciones conservadoras incluyen la sustitución de un resto no polar (hidrófobo) tales como I, V, L o M por toro; la sustitución de un resto polar (hidrófilo) por otro resto polar, tal como R por K, Q por N, G por S, o viceversa; y la sustitución de un resto básico tal como K, R o H por otro o la sustitución de un resto ácido tal como D o E por otro.

La proteína OCH1 mutante también mantiene sustancialmente la actividad catalítica de la proteína OCH1 de tipo silvestre, es decir, al menos un 75 %, 80 %, 85 %, 90 %, 95 % o más, de la actividad α-1,6 manosiltransferasa de la proteína OCH1 de tipo silvestre. La actividad de una proteína OCH1 mutante también se puede determinar fácilmente utilizando ensayos *in vitro* e *in vivo* conocidos en la técnica. Véase, por ejemplo, Nakayama (1992), *supra*.

Como se ha descrito anteriormente, las cepas de la presente invención incluyen un alelo OCH1 mutante que se transcribe en un ARNm que codifica una proteína OCH1 mutante, y no incluye otro alelo que produzca un ARNm que codifique una proteína OCH1 funcional. Dichas cepas se pueden modificar por una variedad de medios.

En algunas realizaciones, el alelo OCH1 de tipo silvestre en el locus OCH1 del cromosoma de una cepa de *Pichia pastoris* se ha modificado o mutado para proporcionar un alelo OCH1 mutante (como se ilustra en los ejemplos posteriores del presente documento), o se ha sustituido por un alelo OCH1 mutante (por ejemplo, mediante recombinación homóloga). Las modificaciones deberían ser de tal manera que la cepa resultante sea estable con respecto al alelo OCH1 mutante. Es decir, el alelo mutante se mantiene en la cepa a lo largo de generaciones (por ejemplo, al menos 10, 20, 30, 40, 50 o más divisiones celulares) adecuado tanto para el cultivo en matraces de poco volumen como en cultivos en biorreactor a tamaño industrial, sin la inversión a un alelo OCH1 que codifique una

proteína OCH1 funcional.

En otras realizaciones, un alelo OCH1 mutante se introduce mediante un vector de expresión en una cepa de *Pichia pastoris* cuyo alelo OCH1 de tipo silvestre ("alelo" OCH1 de tipo silvestre si es haploide, o "alelos" OCH1 de tipo silvestre si es diploide) ya se ha alterado de forma que se produce una proteína OCH1 no funcional del alelo OCH1 nativo o locus OCH1 nativo. El vector de expresión puede ser un vector integrador diseñado para integrar el alelo OCH1 mutante en el genoma del huésped; o un vector replicador (por ejemplo, un plásmido) que se replica en la cepa independientemente de los cromosomas.

Tanto si el alelo OCH1 mutante se crea en el locus OCH1 nativo mutando o sustituyendo el alelo OCH1 de tipo silvestre, como si se proporciona mediante un vector de expresión en una cepa cuyo alelo(s) OCH1 de tipo silvestre ("alelo" OCH1 de tipo silvestre si es haploide, o "alelos" OCH1 de tipo silvestre si es diploide) ya se ha alterado, es importante que la cepa mutante resultante no produzca una proteína OCH1 funcional a lo largo de generaciones (por ejemplo, al menos 10, 20, 30, 40, 50 o más divisiones celulares). Por "proteína OCH1 funcional" se significa la proteína OCH1 de tipo silvestre o un equivalente funcional de la proteína OCH1 de tipo silvestre, es decir, una proteína que se dirige al aparato de Golgi y que mantiene sustancialmente la actividad catalítica de la proteína de tipo silvestre (es decir, al menos aproximadamente un 80 %, 85 %, 90 %, 95 % o más, de la actividad α-1,6 manosiltransferasa de la proteína OCH1 de tipo silvestre). Para evitar la inversión, se deberían que eliminar las secuencias homólogas en la cepa para evitar la recombinación homóloga que genere un alelo OCH1 de tipo silvestre.

El alelo OCH1 mutante, si está presente en el cromosoma del huésped o en un vector extra-cromosómico, se transcribe en un ARNm. En otras palabras, la cepa se modifica de manera que la secuencia codificante del alelo OCH1 mutante esté unido operativamente a un promotor para efectuar la transcripción. El promotor puede ser un promotor endógeno tal como el promotor OCH1 endógeno, un promotor heterólogo al alelo OCH1 (por ejemplo, un promotor AOX1, un promotor GAP), y similares; puede ser un promotor exógeno funcional en *Pichia pastoris*. El nivel de transcripción puede ser el mismo, mayor o menor que el nivel de expresión del alelo OCH1 de tipo silvestre en una cepa de *Pichia pastoris* no modificada (tal como la GS115).

Las cepas de *Pichia pastoris* que tienen las modificaciones genéticas del alelo(s) OCH1 que se describen anteriormente incluyen cepas haploides y cepas diploides. Para las cepas diploides que tienen un alelo OCH1 mutante integrado en un cromosoma del huésped, las cepas pueden ser homocigotas o heterocigotas para el alelo OCH1 mutante.

Las cepas de *Pichia pastoris* que tienen modificaciones genéticas en el alelo(s) OCH1 que se han descrito anteriormente son robustas y estables, y producen proteínas con estructuras de N-glicano sustancialmente homogéneas siendo Man8GlcNAc2 el N-glicano predominante.

Modificación genética – Un ácido nucleico que codifica y expresa una α-1,2 manosidasa

Además de las modificaciones genéticas del alelo(s) OCH1 que se han descrito anteriormente, las cepas se pueden modificar para que incluyan una molécula de ácido nucleico que codifica y es capaz de expresar una α-1,2 manosidasa o un fragmento funcional de la misma que convierte el Man_8GlcNA_{c2} en Man_5GlcNA_{c2}, proporcionando de esta manera el Man_5GlcNA_{c2} como la forma de N-glicano predominante.

La α-1,2 manosidasa (MS-I) es una familia de enzimas bien caracterizadas. Se sabe que la mayoría de las enzimas MS-I se localizan en al aparato de Golgi o el retículo endoplásmico, aunque pocas se secretan y tienen actividad extracelular. Véase, Gonzalez et al., Mol Biol Evolution 17:292-300 (2000). La topología de las enzimas que se localizan en el RE y aparato de Golgi incluye en general un dominio catalítico luminal y una región transmembrana del extremo N. Véase, Herscovics, Biochimie 8: 757-62 (2001). La región del extremo N está compuesto por una región de tallo (cercana al dominio catalítico luminal), un dominio transmembrana, y una cola citoplasmática. En las enzimas MS-I secretadas, la región transmembrana extra-catalítica también se conoce como secuencia líder, que sirve como una señal para la secreción de la enzima. Las caracterizaciones detalladas de distintas α-1,2 manosidasas se pueden encontrar en Becker et al. (European J. Cell Biol 79: 986-992 (2000)) que estudia las enzimas MS-I del ratón y *S. cerevisiae* y sus dominios catalíticos; Schneikert y Herscovics (Glycobiology 4: 445-450 (1994)) que caracteriza la actividad de MS-I murina y su dominio catalítico; Gonzalez et al. (J. Biol Chem 274: 21375-86 (1999)) que examina las actividades y dominios de varias enzimas MS-I, incluyendo dos enzimas de *C. elegans*, una MS-I humana y la MS-I de *S. cerevisiae* (del RE); y Maras et al. (J. Biotechnology 77:255-263 (2000)), que caracteriza la α-1,2 manosidasa de *T. reesei* que pertenece a la categoría de MS-I secretoras, que están compuestas de un dominio catalítico y una secuencia líder del extremo N.

La molécula de ácido nucleico que codifica una α-1,2 manosidasa o u fragmento funcional de la misma puede derivarse de cualquier especie para su uso en la presente invención, incluyendo pero sin limitarse a los genes de mamíferos que codifican, por ejemplo, la α-1,2 manosidasa murina (Herscovics et al. J. Biol. Chem. 269: 9864-9871, 1994), α-1,2 manosidasa de conejo (Lal et al. J. Biol. Chem. 269: 9872-9881, 1994), o α-1,2 manosidasa humana (Tremblay et al. Glycobiology 8: 585-595, 1998), genes fúngicos que codifican, por ejemplo, α-1,2 manosidasa de

Aspergillus (gen msdS), α-1,6 manosiltransferasa de *Trichoderma reesei* (Maras et al., J. Biotechnol. 77: 255-263, 2000), o α-1,2 manosidasa de *Saccharomyces cerevisiae*, así como otros genes tales como los de *C. elegans* (Nº de registro del GenBank CAA98114 y CAB01415) y *Drosophila melanogaster* (Nº de registro del GenBank AAF46570) (véase, por ejemplo, Nett et al., Yeast 28:237-252, 2011).

Por "parte funcional" o "fragmento enzimáticamente activo" de una α-1,2 manosidasa, se significa un fragmento de polipéptido de una α-1,2 manosidasa de origen natural o tipo silvestre que mantiene sustancialmente la actividad enzimática de la proteína de longitud completa. Por "sustancialmente" en este contexto se significa al menos aproximadamente un 75 %, 80 %, 85 %, 90 %, 95 % o más, de la actividad enzimática de la proteína de longitud completa se mantiene. Por ejemplo, el dominio catalítico de una α-1,2 manosidasa, con ausencia de cualquier secuencia transmembrana del extremo N o de señal, constituye un "fragmento funcional" de la α-1,2 manosidasa. Los expertos en la técnica pueden identificar fácilmente y fabricar fragmentos funcionales de una α-1,2 manosidasa basándose en la información disponible en la técnica y una combinación de técnicas conocidas en la técnica. La actividad de un fragmento polipeptídico particular también se puede verificar utilizando ensayos *in vitro* o *in vivo* conocidos en la técnica.

En algunas realizaciones, la secuencia de nucleótidos que codifica una α-1,2 manosidasa o un fragmento funcional se deriva de la secuencia codificante de la α-1,2 manosidasa de *Trichoderma reesei*. En realizaciones específicas, la secuencia de nucleótidos codifica la α-1,2 manosidasa de *Trichoderma reesei* descrita por Maras et al. J. Biotechnol. 77: 255-63 (2000), o un fragmento funcional de la misma (tal como el dominio catalítico del extremo C de la proteína de longitud completa).

En algunas realizaciones, las cepas se modifican de manera que la α-1,2 manosidasa o un fragmento funcional se dirigen al RE. En realizaciones específicas, el direccionamiento al RE se consigue incluyendo una secuencia de direccionamiento al RE en la α-1,2 manosidasa o un fragmento funcional. Ejemplos de secuencias de direccionamiento al RE incluyen un fragmento del extremo N de *S. cerevisiae* SEC12, una secuencia del extremo N de α-glucosidasa I de *S. cerevisiae* codificada por *GLS1*, y un fragmento del extremo N de α-1,2 manosidasa de S. cerevisiae codificada por *MNS1*. Véase también, Nett et al. (2011), *supra*. En una realización específica, la α-1,2 manosidasa o un fragmento funcional se dirige al RE incluyendo una señal de retención de RE, HDEL (SEQ ID NO: 91), y el extremo C de la α-1,2 manosidasa o su fragmento funcional.

El ácido nucleico que codifica una α-1,2 manosidasa o un fragmento funcional se puede introducir mediante un vector de expresión en una cepa de *Pichia pastoris*. El vector de expresión puede ser un vector integrador diseñado proa integrar la secuencia codificante de α-1,2 manosidasa en el genoma del huésped; o un vector replicador (por ejemplo, un plásmido) que se replica en la cepa independientemente de los cromosomas. En los casos de un vector integrador, el vector se puede diseñar para conseguir la integración en el ácido nucleico del alelo OCH1 de tipo silvestre (por ejemplo, mediante un intercambio simple o doble por recombinación homóloga) y la alteración simultánea del alelo OCH1 de tipo silvestre.

Cepas SuperM5

La presente divulgación proporciona cepas de *Pichia pastoris* que son robustas, estables, y transformables, y producen proteínas con N-glicanos Man5GlcNAc2 sustancialmente homogéneos. También se hacen referencia a estas cepas como cepas SuperM5 o SuperMan5.

Las cepas SuperM5 se modifican genéticamente para que incluyan un alelo OCH1 mutante que se transcribe en un ARNm que codifica una proteína OCH1 mutante que contiene un dominio catalítico sustancialmente idéntico al de la proteína OCH1 de tipo silvestre, pero que carece de la secuencia del extremo N necesaria para dirigir la proteína OCH1 al aparato de Golgi. Las cepas no incluyen ningún otro alelo OCH1 que produzca un ARNm que codifica una proteína OCH1 funcional. Las cepas se modifican genéticamente adicionalmente para que contengan un ácido nucleico que codifique y exprese una α-1,2 manosidasa o un fragmento funcional de la misma, que se dirige al RE y convierte el Man8GlcNAc2 en Man5GlcNAc2.

Un ejemplo de una cepa SuperM5 se describe en el Ejemplo 1. La secuencia de nucleótidos del locus OCH1 de esta cepa se expone en la Tabla 1 y la SEQ ID NO: 1. Construida utilizando la cepa M5-Blast descrita en Jacobs et al. (2009), la cepa SuperM5 es superior a la M5-Blast en términos de crecimiento robusto, estabilidad y homogeneidad de los glicanos M5 que se producen.

Modificación genética – introducción de enzimas adicionales

Estas cepas se pueden modificar adicionalmente para expresar otras enzimas corriente abajo (o fragmentos funcionales de las mismas) en la ruta de glicosilación hacia la producción de N-glicanos de tipo híbrido y complejos. Dichas enzimas adicionales incluyen, por ejemplo, una o más de GlcNAc transferasa I (GnT-I), β-1,4-galactosiltransferasa 1 (GalT), manosidasa II (Man-II), y GnT-II, entre otras. Véase, Jacobs et al. (2009); Patente de EE. UU. 7.029.872 de Gerngross.

La GnT-I cataliza la adición de un resto de GlcNAc unido en β-1,2 a la α-1,3 manosa del núcleo trimanosil de Man5GlcNAc2. La introducción de la actividad de la GnT-I se puede conseguir transformando con una vector que comprende una secuencia de ácido nucleico que codifica una GlcNAc transferasa I (GnT-I) para su uso en la presente invención. Dicha secuencia de ácido nucleico se puede derivar de cualquier especie, por ejemplo, conejo, rata, ser humano, plantas, insectos, nematodos, y protozoos tales como *Leishmania tarentolae*. En realizaciones específicas, la secuencia de nucleótidos codifica una GnT-I humana. La GnT-I o una parte funcional de la misma se dirige el aparato de Golgi, que se puede alcanzar incluyendo una señal de localización el aparato de Golgi de levadura en la proteína GnT-I o una parte funcional de la misma. En ciertas realizaciones el dominio catalítico de la GnT-I humana se fusiona al dominio del extremo N de *S. cerevisiae* Kre2p, una glicosiltransferasa con una localización cis/media en el aparato de Golgi conocida.

GalT cataliza la adición de un resto de galactosa en la unión β-1,4 de la β-1,2-GlcNAc, utilizando UDP-Gal como sustrato donante. La introducción de la actividad GalT se puede conseguir transformando con un sector que comprende una secuencia de ácido nucleico que codifica una GalT o un fragmento funcional de la misma, que se puede derivar de seres humanos, plantas (por ejemplo, *Arabidopsis thaliana*), insectos (por ejemplo, *Drosophila melanogaster*). La GalT o una parte funcional es modifica genéticamente para que contenga una señal de retención en el aparato de Golgi y se dirija al aparato de Golgi. Un ejemplo de señal de retención en el aparato de Golgi ejemplar está compuesta por los 100 primeros aminoácidos de la proteína Kre2 de *Saccharomyces cerevisiae*.

Man-II actúa para eliminar tanto las manosas α-1,3 como α-1,6 de los glicanos $GlcNAcMan_5GlcNA_{c2}$. La presencia de un resto GlcNAc unido a β-1,2 del brazo α-1,3 es esencial para su actividad. La introducción de la actividad de Man-II se puede conseguir transformando una cepa con un vector de ácido nucleico que codifica una proteína Man.II o un fragmento funcional de la misma, modificado para que contenga una señal de localización en el aparato de Golgi .Como ejemplo, un ácido nucleico adecuado puede codificar el dominio catalítico de la Man-II de *Drosophila melanogaster*, fusionado en fase con el dominio de localización en el aparato de Golgi Mnn2p de *S. cerevisiae*.

GnT-II cataliza la adición de un segundo resto de GlcNAc unido a β-1,2 a la α-1,6 manosa libre del núcleo trimanosil. La introducción de actividad de GnT-II se puede conseguir transformando con un vector que contiene una secuencia de nucleótido que codifica una proteína GnT-II o un fragmento funcional de la misma. Se han clonado enes de GnT-II de varias especies que incluyen especies de mamíferos y se pueden utilizar en la presente invención. Como ejemplo, una secuencia de nucleótidos adecuada codifica el dominio catalítico de GnT-II de rata fusionado a la parte del extremo N de Mnn2p de *S. cerevisiae*.

Otras manipulaciones de las cepas

Las cepas desveladas en el presente documento pueden incluir características adicionales, que se consiguen por distintas modificaciones adecuadas (tales como la modificación cruzada o recombinante), que incluyen, por ejemplo, las que tienen un gen auxótrofo mutante (por ejemplo, his) para facilitar la clonación y selección, que tienen una deficiencia de proteasa para limitar la degradación del producto (por ejemplo, *pep4*- prb1-, y/o *sub2*-), que tienen un fenotipo de utilización lenta del metanol (por ejemplo, *mutS*).

En realizaciones específicas, esta divulgación proporciona las siguientes cepas:

SuperMan5, *P. pastoris*, och1-, resistente a blasticidina, Manosidasa I de *T. reesei* (=His+);
SuperMan5 (his-), *P. pastoris*, och1-, his4-, resistente a blasticidina, Manosidasa I de T. reesei;
SuperMan5 (mutS), *P. pastoris*, och1-, resistente a blasticidina, Manosidasa I de T. reesei (utilización lenta del metanol);
SuperMan5 (pep4-), *P. pastoris*, och1-, resistente a blasticidina, Manosidasa I de *T. reesei* (deficiente en proteasas);
SuperMan5 (prb1-), *P. pastoris*, och1-, resistente a blasticidina, Manosidasa I de *T. reesei* (deficiente en proteasas);
SuperMan5 (pep4-, sub2-), *P. pastoris*, och1-, resistente a blasticidina, Manosidasa I de *T. reesei* (deficiente en proteasas);
SuperMan5 (pep4-, prb1-), *P. pastoris*, och1-, resistente a blasticidina, Manosidasa I de *T. reesei* (deficiente en proteasas).

<u>Uso de las cepas</u>

Se puede expresar una proteína heteróloga con uno o más sitios de N-glicosilación en las cepas de la presente invención transformando una cepa de la presente invención con un vector de expresión que codifica la proteína heteróloga, para obtener una preparación de la proteína heteróloga sustancialmente homogénea en su estructura de N-glicano.

Ejemplo 1 – Generación de una cepa SuperM5

Este Ejemplo describe la creación de una cepa SuperM5 a partir de la cepa M5-Blast descrita en Jacobs et al. (2009), Nature Protocols 4:58-70.

La cepa M5-Blast es una modificación de la cepa G115 de *P. pastoris* en la que el gen OCH1 de la manosiltransferasa se altera por la inserción de un vector que comprende un gen de α-1,2 manosidasa (vector pGlycoSwitchM5-Blast) mediante una recombinación homóloga de intercambio simple. Como resultado de la recombinación homóloga de intercambio simple, el casete de expresión de manosidasa integrado está flanqueado por aproximadamente 450 pb de secuencias homólogas de la ORF de OCH1. La secuencia del locus genómico OCH1 de esta cepa M5-Blast se expone en la SEQ ID NO: 53. La secuenciación revelaba la pérdida de 10 pb en la unión entre la secuencia del vector pGlycoSwitchM5-Blast y el fragmento 3' de la ORF de OCH1, dando como resultado la pérdida de uno de los tres codones de parada deseados del vector pGlycoSwitchM5-Blast corriente arriba del fragmento del extremo C de OCH1, y el cambio de fase del segundo y tercer codones de parada para una fase de lectura diferente a la del fragmento. Como resultado la ORF actual se extendió 28 pb corriente arriba hasta un codón ATG en fase en la matriz del vector. La Phe27 de la proteína de tipo silvestre se vuelve la Phel11 de la nueva ORF, y la nueva secuencia de señal prevista consiste parcialmente en la señal de anclaje anterior y la nueva secuencia fusionada de la matriz del vector. La secuencia de aminoácidos de esta nueva ORF se expone en la SEQ ID NO: 3 (siendo los primeros 25 aminoácidos la nueva secuencia de señal prevista).

La región del extremo N del locus genómico OCH1 después del evento de recombinación homóloga de intercambio simple se representa como un diagrama en la Figura 1, junto con la construcción utilizada para retirar esta región del extremo N por recombinación homóloga de doble de intercambio doble. La construcción contenía marcadores de selección y de contra-selección flanqueados por un para *lox71-lox66*, permitiendo la retirada posterior del casete de selección/contra-selección por recombinación mediada por *cre*. La secuencia del casete de doble intercambio del casete de selección/contra-selección con los brazos de homología se expone en la SEQ ID NO: 58, y su creación se describe posteriormente en el presente Ejemplo.

Con el fin de confirmar la secuencia de la región dirigida antes de crear la construcción de intercambio, los cebadores de PCR se diseñaron para amplificar ~ 1650 pb de ADN que engloba la región corriente arriba de la ORF de manosidasa. Utilizando la Phusion polimerasa (NEB), los cebadores 80670918 y 80670919 amplificaba un fragmento de tamaño apropiado del ADN genómico de M5-Blast. El producto de la PCR se clonó en TOPO y se verificó su secuencia. La secuencia de ADN demostraba que el vector de expresión de manosidasa se había integrado en el genoma de GS115 correctamente al final de la inserción.

Se diseñaron cebadores de los flancos para amplificar las regiones de homología que se muestran en la Figura 1 del ADN genómico de M5-Blast. El alineamiento de estos cebadores de PCR se muestra en la Figura 2. El uso de Phusion polimerasa daba como resultado reacciones PCR satisfactorias a partir del ADN genómico de M5-Blast.

Los productos de PCR de los siguientes pares de cebadores se aislaron en gel y se utilizaron como matrices para la adición de los sitios de recombinación *lox71* y *lox66*:

80765855-80765856 (642 pb, Figura 2A)
80765857-80765858 (658 pb, Figura 2A)
80765852-80765854 (910 pb, Figura 2B)
80765853-80765854 (956 pb, Figura 2B)

Los cebadores de la PCR no coincidentes se diseñaron para añadir los sitios *lox* en los extremos apropiados de los brazos de homología. Estos cebadores no coincidentes se representan en un diagrama en la Figura 3. Las reacciones de PCR con Phusion polimerasa eran satisfactorias en la generación de los productos de ADN de tamaño correcto de cada una de las 3 reacciones:

80765855-80984794 (670 pb, Figura 3A)
80765857-80984794 (681 pb, Figura 3A)
80984795-80765854 (850 pb, Figura 3B)

Además de añadir los sitios *lox* a los brazos, se diseñaron cebadores de PCR para añadir las extensiones de ADN genómico de M5-Blast de *Pichia pastoris* apropiadas en un casete *lox71-MazF-NatR-lox66*. De nuevo, se utilizó Phusion polimerasa para generar el producto de PCR correcto, como se muestra en la Figura 4. El par de cebadores utilizado:

80984793-80984796 (2941 pb, Figura 4)

El producto de PCR del casete de selección/contra-selección se purificó en gel y se llevó a cabo una PCR de tres piezas solapadas para unir los brazos de homología al casete. En resumen, se ciclaron 20 x las tres piezas en ausencia de cebadores para hibridar y extender el solapamiento a los extremos de los fragmentos. La mezcla

ciclada se diluyó entonces o ciclar 35 x en presencia de los cebadores que se representan en diagramas en las Figuras 5.

La reacción PCR se llevó a cabo con una Phusion polimerasa, utilizando un tiempo de extensión de 3 min. Los cebadores se detallan posteriormente:

80765855-80765854 (4311 pb, Figura 5)

Este producto de PCR se aisló y se clonó en TOPO. La selección de la clonación TOPO se llevó a cabo en placas LB-Nat para asegurar la inclusión del casete de selección. La secuenciación de ADN se llevó a cabo en múltiples aislados para determinar las secuencias de los brazos de homología. El aislado final contenía un casete de expresión Nat^R, los sitios de recombinación *lox71* e *lox66* y los brazos de homología correctos.

Los cebadores de la PCR internos del fragmento clonado detallados en la Figura 5 se utilizaron para generar un ADN lineal para la transformación de *Pichia pastoris*. Se diseñaron dos conjuntos independientes de cebadores:

81364233-81364234 (4063 pb, Figura 6)

81364235-81364236 (4060 pb, Figura 6)

Las reacciones de PCR se llevaron a cabo utilizando Phusion polimerasa con un tiempo de extensión de 100 s.

Los productos PCR se purificaron por electroforesis en gel de agarosa y se eluyeron de la matriz de unión con gua. La cepa M5-Blast de *Pichia pastoris* se hizo competente por electroporación utilizando un tratamiento DTT/sorbitol convencional. La electroporación se llevó a cabo utilizando cubetas de 1 mm que contenían 20 ml de células competentes 1-2 ml de ADN lineal purificado. Las mezclas de transformación se colocaron en placas sobre YPD-Nat agar.

Después de la electroporación, se cultivaron las células a 30 ºC durante 3 días. Los transformantes individuales se parchearon en YPD-Nat para el almacenamiento y el análisis. La Figura 7 muestra la disposición teórica del locus OCH1 después de la integración apropiada del intercambio doble del producto(s) de PCR en el genoma del M5-Blast. Los pares de cebadores PCR se diseñaron para comprobar los aislados resistentes a nourseotricina eran el resultado de recombinación homóloga, más que por integración aleatoria de los productos de la PCR en el genoma. Estos pares de cebadores de PCR se representan en un diagrama en la Figura 7.

81487116-81487117 (895 pb, Figura 7)
81487118-81487119 (937 pb, Figura 7)
81487120-81487121 (656 pb, Figura 7)
81487122-81487123 (756 pb, Figura 7)

Un total de 24 aislados independientes se exploraron por PCR y 2 aislados que aparecían correctamente se caracterizaron adicionalmente por secuenciación de ADN de los productos de la PCR. Los dos aislados se prendieron en colonias sencillas en medio YPD y se dejaron en YPD-Nat. Se hicieron preparaciones de ADN genómico a pequeña escala utilizando perlas de lisis de cristal de fenol/cloroformo. Basándose en los resultados de secuenciación los pares de cebadores 81487116-81487117, 81487118-81487119, 81487120-81487121 y 81487122-81487123 de estos genómicos extractos, ambos aislados contenían un casete de selección/contra-selección *lox71-lox66* en la localización apropiada del genoma de M5-Blast. No había mutaciones introducidas en la reacción inicial de PCR para generar el fragmento de transformación, las uniones de recombinación en ambos extremos eran idénticos a la secuencia de ADN de M5-Blast “de tipo silvestre”, y ambos sitios *lox71 y lox66* estaban intactos. La secuencia de ADN del locus OCH1 después de la recombinación de doble intercambio se expone en la SEQ ID NO: 59.

Los dos aislados (A1-2 y A4-3) se transformaron con un plásmido que expresa constitutivamente la *cre* recombinasa. En resumen, ambas cepas se hicieron electrocompetentes utilizando un procedimiento DTT/sorbitol, se electroporaron con un plásmido circular y se colocaron en placas en YPD-G418. Los transformantes se cultivaron a 30 ºC durante varios días y se recogieron las colonias. Las colonias se transfirieron directamente en placas de metanol para inducir la contra –selección *MazF* o se parchearon a YPD para permitir la pérdida del plásmido *cre*-ARS antes de la inducción *MazF*. La inducción por metanol se llevó a cabo en BMMY (un 1 % de metanol) y CSM (medio sintético completo, 0,5 % de metanol). Las placas se suplementaron con metanol diariamente añadiendo 100 ml de metanol en el borde invertido de la placa. Se llevó la incubación a 30 ºC. Había una formación de colonias significativa en todas las condiciones ensayadas; el cultivo en metanol parecía independiente de si el transformante venía directamente de YPD-G418 o se había sometido a un parcheo intermedio en YPD sin G418.

La combinación de cre debería retirar las secuencias de ADN entre los sitios *lox71 y lox66*, dejando solo una cicatriz del sitio *lox* defectuoso en el genoma. El resultado teórico de este evento de recombinación se muestra en la Figura 8. Los cebadores de la PCR se diseñaron para amplificar la región que contiene la cicatriz del *lox* defectuoso. Veinte

colonias que se cultivaron en metanol se exploraron por PCR para determinar la pérdida del casete de selección/contra-selección. Los cebadores de la PCR que se utilizaron eran:

80670916-80670917 (680 pb, Figura 8)
80670918-80670919 (782 pb, Figura 8)

Diecisiete de veinte asilados generaban el producto de PCR apropiado con el primer par de cebadores. La mayoría, pero no todos, los 17 también demostraban un producto apropiado con el segundo par de cebadores. Cada uno de los 17 aislados se parcheó en YPD, YPD-Blast, YPD-Nat e YPD-G418 para ensayar la presencia o ausencia de los marcadores de selección farmacológica. Si el plásmido *cre* tenía retirado apropiadamente el casete de selección/contra-selección y se había perdido posteriormente, la cepa resultante debería ser resistente a blasticidina y sensible a G418 y nourseotricina. Todos los aislados eran resistentes a blasticidina y sensibles a nourseotricina. Unos pocos mantenían la resistencia a G418 (aún contenían el plásmido *cre*, quizás integrado) y se desecharon. Del resto, 4 se recogieron para la secuenciación de ADN de la región intergénica *LEU5*-manosidasaHDEL.

Los cebadores de la PCR existentes se utilizaron para amplificar la región genómica que abarca *LEU5* y la ORF de manosidasaHDEL.

81487118-80765854 (1602 pb, Figura 9)

La amplificación PCR se llevó a cabo utilizando la Phusion polimerasa en el ADN genómico que se preparó por extracción de perlas de cristal de fenol/cloroformo. Se utilizaron cebadores de secuenciación múltiple interna para verificar la secuencia completa del producto de PCR de 1602 pb. Los cuatro productos PCR secuenciados eran correctos, y contenían un sitio *lox* defectuoso en la localización apropiada entre el gen *LEU5* y la ORF de la manosidasaHDEL. Tanto el promotor LEU5 y el promotor GAP que dirigen la expresión de manosidasaHDEL estaban intactos y eran idénticos a los promotores presenten en la cepa M5-Blast de partida. La secuencia de ADN del locus OCH1 después de una recombinación de intercambio doble y la recombinación *cre* se expone en la SEQ ID NO: 1.

Se prepararon las reservas de glicerol de cada uno de los 4 aislados (y las dos cepas parentales antes de la recombinación *cre*).

bG yeast-100015	A1-2 (pre-recombinación)
bG yeast-100016	A4-3 (pre-recombinación)
bG yeast-100017	aislado 1 (post-recombinación)
bG yeast-100018	aislado 2 (post-recombinación)
bG yeast-100019	aislado 3 (post-recombinación)
bG yeast-100020	aislado 4 (post-recombinación)

Cada reserva de glicerol estaba manchada y re-ensayada en cuanto al os marcadores apropiados:

bG yeast-100015	his^-, $blasticidina^R$, $nourseotricina^R$
bG yeast-100016	his^- , $blasticidina^R$, $nourseotricina^R$
bG yeast-100017	his^-, $blasticidina^R$, $nourseotricina^S$
bG yeast-100018	his^-, $blasticidina^R$, $nourseotricina^S$
bG yeast-100019	his^-, $blasticidina^R$, $nourseotricina^S$
bG yeast-100020	his^-, $blasticidina^R$, $nourseotricina^S$

Todas las reservas de glicerol se ensayaron como se esperaba.

Se generaron las punzadas de YPD de los 6 aislados y se sometieron a glicoanálisis. Se utilizó la reserva bG yeast-100017 para generar una gran preparación de ADN genómico para la secuenciación del genoma. Además, se prepararon las muestras de GS115 de tipo silvestre y la cepa M5-Blast. En resumen, los aglomerados celular es de 100 ml de cultivos celulares (cultivo en YPD, 30 ºC) se re-suspendieron en 1 M/100 mM de citrato (pH 6,0) y se trataron con Zimolasa (Zymo Research) que contenía RNasa durante 3 h a 37 ºC. Se añadió SDS al 0,5 % para lisar los esferoplastos. Entonces se añadió proteinasa K y la mezcla se incubó a 50 ºC durante una noche. Se añadió un volumen igual de fenol/cloroformo y la mezcla se agitó suavemente durante 30 min. Después de la centrifugación, la parte acuosa superior se retiró y se precipitó el ADN con isopropanol. El ADN hilado se desenrolló de la solución y se suspendió en TE. El ADN se re-precipitó con etanol y entonces se lavó con un 70 % de etanol, se secó al aire y se re-suspendió un tiempo final en TE.

El ADN se distribuyó en múltiples tubos:

bG DNA-100215	ADN genómico GS115
bG DNA-100216	ADN genómico GS115
bG DNA-100217	ADN genómico GS115

bG DNA-100221	ADN genómico bG yeast-100017
bG DNA-100222	ADN genómico bG yeast-100017
bG DNA-100223	ADN genómico M5-Blast
bG DNA-100224	ADN genómico M5-Blast

Con el fin de ensayar los aislados de ADN genómico y verificar que las manipulaciones que se llevan a cabo en la creación de la cepa bG yeast-100017 no tenía alterada la forma mutante de ORF OCH1, la región del extremo N de la ORF OCH1 se asiló de la bG DNA-100221 (nueva cepa) y bG DNA-100223 (cepa M5-Blast) por PCR y se re-secuenció. Ambas preparaciones de ADN eran idénticas en el locus ORF OCH1, y contenían la eliminación de 10 pb como se ha descrito anteriormente.

Los cebadores utilizados en este Ejemplo se enumeran en la lista siguiente:

SEQ ID		
60	**80670916**	**CAAGTTGCGCCCCCTGGCA**
61	**80670917**	**TGGAGCAGCTAATGCGGAGGA**
62	**80670918**	**AGTTCCGCCGAGACTTCCCCA**
63	**80670919**	**TTCAGCCGGAATTTGTGCCGT**
64	**80765852**	**ATCCAGGGTGACGGTGCCGA**
65	**80765853**	**GCAAGAGGCCCGGCAGTACC**
66	**80765854**	**CCGCCCTCGTAGGGTTGGGAG**
67	**80765855**	**TTCGCGGTCGGGTCACACA**
68	**80765856**	**AACTGCCATCTGCCTTCGCC**
69	**80765857**	**CAAATCGCGGGTTCGCGGTC**
70	**80765858**	**GAGCAAACTGCCATCTGCCTTCG**
71	**80984793**	**GTGTTCGTAGCAAATATCATCAGCCTACCGTTCGTATAGCATACATTA TACGAAGTTATGGATCTAACATCCAAA**
72	**80984794**	**TTTGGATGTTAGATCCATAACTTCGTATAATGTATGCTATACGAACGG TAGGCTGATGATATTTGCTACGAACAC**
73	**80984795**	**GCCGCCATCCAGTGTCATAACTTCGTATAGCATACATTATACGAACGG TACTTTTTTGTAGAAATGTCTTGGTGT**
74	**80984796**	**ACACCAAGACATTTCTACAAAAAAGTACCGTTCGTATAATGTATGCTA TACGAAGTTATGACACTGGATGGCGGC**
75	**81364231**	**GTGTTCGTAGCAAATATCATCAGCCTACCG**
76	**81364232**	**ACACCAAGAAC ACACCAAGAACATTTCTACAACAAAAAAGTAC CGT**
77	**81364233**	**TTCGCGGTCGGGTCACACAC**
78	**81364234**	**GGAGCAGCTAATGCGGAGGATGC**
79	**81364235**	**CGGTCGGGTCACACACGGAG**
80	**81364236**	**TGGAGCAGCTAATGCGGAGGA**
81	**81487116**	**TGAGTCCTGGTGCTCCTGACG**
82	**81487117**	**CCCCTCCTGTTGCGTTTGGC**
83	**81487118**	**AGCGTTCTGAGTCCTGGTGCT**
84	**81487119**	**GGTCCTGCGTTTGCAACGGT**
85	**81487120**	**ACTAACGCCGCCATCCAGTGTC**
86	**81487121**	**GCTTCAGCCGGAATTTGTGCCG**
87	**81487122**	**CGCCTCGACATCATCTGCCC**
88	**81487123**	**TCAGCCGGAATTTGTGCCGT**

Ejemplo 2 – Almacenamiento y manejo

La SuperM5 se almacenó en diferentes condiciones a -80 ºC, -4 ºC, 20 ºC y a temperatura ambiente. Las cepas se almacenaron como reservas de glicerol congeladas y como cultivos de punciones. Se almacenaron los diferentes cultivos y se descongelaron para los diferentes experimentos y para el envío a los colaboradores para el ensayo. En todos los casos las cepas se recuperaron, se colocaron en placas y se cultivaron de manera similar al a cepa *Pichia pastoris* GS115 parental y se cultivó en un medio definido y complejo similar al de las cepas parentales. Las cepas SuperM5 transformadas de manera similar que la cepa parental y las proteínas se expresaron con la manosa-5 glicosilación como la glicoforma predominante, o la única glicoforma. Las cepas se almacenaron y re-cultivaron repetidamente para establecer la robustez de las cepas SuperM5.

Ejemplo 3 – Análisis de proteínas de ensayo en cepas de *P. pastoris*

Los genes de lipasas A y B de *Candida antartica*, transferrina humana, y el dominio CH2 humano de IgG se integraron en el genoma de SuperM5 utilizando métodos de transformación convencionales. En todos los casos, se producían cantidades significativas de proteína y se secretaban en el medio. Las cepas transformadas y la proteína contenida en el medio se ensayaron por análisis de glicanos utilizando previamente los métodos publicados. En todos los casos, los perfiles de glicano de las proteínas de ensayo y para las glicoproteínas de la cepa demostraban

una estructura de manosa-5 glicano con ninguna otra estructura de manosa superior detectada por los métodos utilizados.

Ejemplo 4 – Análisis de las manoproteínas de la pared celular en cepas de *P. pastoris*

Se iniciaron doce cepas de *Pichia pastoris* y la cepa Man5-Blast en placas de 24 pocillos que contenían 2 ml de YPD y se cultivaron durante una noche a 28 ºC mientras se agitaba (250 rpm). Después del cultivo, las células se recolectaron por centrifugación (3000 g durante 5 min a temperatura ambiente) y se extrajeron las manoproteínas de la pared celular de acuerdo con el protocolo de Jacobs et al. (Véase Jacobs et al., 2009, Nature Protocols 4(1):58-70). Las manoproteínas extraídas (en 100 ml ddH20) se diluyeron a 300 ml con tampón RCM (8 M de urea, 3,2 mM de EDTA, 360 mM de Tris-HCl, pH 8,6). Los N-glicanos se prepararon a partir de estas muestras después de un procedimiento de desglicosilación en membrana de 96 pocillos como publicó Laroy et al. (Laroy et al., 2006, Nature Protocols, 1: 397-405).

Después de marcar los N-glicanos secados con ácido 8-aminopireno-1,3,6-trisulfónico 2, se retiró el exceso de marcador utilizando cromatografía de exclusión (resina Sephadex G-10 2). Las muestras se reconstituyeron finalmente en 10 ml de agua ultrapura y se diluyeron 10x antes de su inyección (80'' a 1,2 kV) en los capilares (e.1. 36 cm; i.d. 50 mm) de un secuenciador DNA 3130 ABI. Se aplicaron los siguientes ajustes: temperatura del horno: 60 ºC; tensión de ejecución: 15 kV; tensión pre-ejecución: 180'' tiempo de ejecución 1000 '', tiempo de pre-ejecución: 15 kV. Se utilizó el Genemapper v3.7 para analizar los datos obtenidos y estructuras se asignaron a los picos (véase la Figura 10).

Ejemplo 5 – Materiales y Métodos

A continuación se describe ejemplos no limitantes de materiales y métodos de la presente invención.

Plásmidos y cepas: Se adquirieron el vector de expresión pPICZαA de *Pichia pastoris* en Invitrogen Corporation; el plásmido pUC19/GM-CSF (que contiene la secuencia genética GM-CSF) se sintetizó en Shanghai Qing-Lan Biotech Co., Ltd.; el vector de expresión pYES2 de Saccharomyces cerevisiae, X-33 de *Pichia pastoris* (tipo silvestre), JM109 de *E. coli* eran del laboratorio de los inventores.

Reactivos e instrumentos: La Taq ADN polimerasa, Pfu ADN polimerasa, enzimas de restricción, T4 ligasa, ácido 5-fluoroorótico (5-FOA) se adquirieron en Shanghai Biological Engineering Technology Services Co., Ltd.; la Zymoliasa se adquirió en Sigma Company (USA); la N-glicosidasa F (PNGasa F) se adquirió en New England Biolabs, Inc. (USA); la peptona, extracto de levadura, base nitrogenada de levadura sin aminoácidos (YNB) se adquirieron en BIO BASIC INC (Canadá). La máquina de PCR (PTC100) era de MJ Research, Inc. (USA); los sistemas de electroforesis, sistema de creación de imágenes por gel eran de Bio-Rad (USA); el sistema de purificación AKTA se adquirió de GE (USA).

Cebadores: basados en la secuencia genética informada Pichia URA3 (orotidina-5'-fosfato descarboxilasa) (Gen-Bank: AF321098), se diseñaron dos pares de extensión de cebadores de amplificación basados en el fragmento de homología: URA5F, URA5R y URA3F, URA3R; basándose en la secuencia del vector pYES2 de Saccharomyces cerevisiae se diseñaron los cebadores pYES2F y pYES2R; basándose en la secuencia genética del GenBank (E12456) expuesta de OCH1 de Pichia, se diseñaron dos pares de cebadores de amplificación basada en la secuencia homóloga: OCH5F, OCH5R y OCH3F, OCH3R. Los cebadores de identificación interna (in) 5F, (in) 3R también se basaron en la misma secuencia; las secuencias de cebadores universales 5' AOX1, 3' AOX1 se basaron en referencias. Los cebadores se sintetizaron en Shanghai Biological Engineering Technology Services Co., Ltd.

El cultivo celular de levaduras, extracción genómica y condiciones de PCR se llevaron a cabo basándose en protocolos conocidos.

La construcción de la secuencia de ADN de reemplazamiento homólogo URA3: utilizando el genoma X-33 de la cepa como matriz y los pares de cebadores URA5F, URA5R y URA3F, URA3R, de ambos lados de los genes URA3, URA5' y URA3', de 700 pb y 600 pb, respectivamente, se amplificaron por PCR. Entonces, utilizando URA5' y URA3' como matrices y URA5F y URA3R como par de cebadores, se amplificaron por PCR el URA5-3, el fragmento de ADN de remplazo homólogo diana para el gen URA3, que tenía un tamaño aproximadamente de 1300 pb.

La construcción del plásmido pYXZ: utilizando el plásmido pYES2 y el par de cebadores pYES2F y pYES2R, la secuencia que contiene el gen URA3 se amplificó por PCR. El producto de PCR se purificó un digirió con Sal I y continuó con una reacción de unión. El plásmidos pYXZ auto-ligado se transformó en *E. coli* JM109, y se colocó en placas sobre placas LB que contenían ampicilina para seleccionar los clones positivos.

El brazo homólogo de OCH1: utilizando el genoma de la cepa X-33 como matriz y los pares de cebadores OCH5F, OCH5R y OCH3F, OCH3R, para amplificar por PCR los extremos 5' y 3' de los brazos homólogos del gen OCH1, OCH5' y OCH3' y su fragmento de fusión OCH3-5. El método utilizado era similar a los que se han descrito anteriormente. Los tamaños de fragmento eran 1000 pb, 700 pb y 1700 pb, respectivamente.

La construcción del plásmido inactivado pYXZ-OCH1: Los inventores digirieron el gen OCH1 el fragmento de fusión homólogo 5' y 3' OCH3-5 con Nhe I y Sal I y se clonó el fragmento en el plásmido pYXZ digerido con Sal I y Nhe I para producir el plásmido de inactivación pYXZ-OCH1.

Inactivación del gen URA3 de *Pichia pastoris* X-33 para construir un marcador genético auxotrófico: Las células X-33 competentes se trasformaron por choque utilizando el brazo del fragmento de fusión URA5-3 que tiene la secuencia homóloga con ambos extremos del gen URA 3; las células transformadas se diseminaron en medio MD que contenían 5-FOA y uracilo (YNB 1. 34 %, glucosa 2 %, agar 1,5 %) y medio MDU (YNB 1. 34 %, glucosa 2 %, agar 1,5 %, uracilo 100 mg/ml), y se incubaron a 30 grados Celsius durante 3 – 5 días. Después, se seleccionaron las cepas que crecían bien en el medio MDU pero no podían crecer en el medio MD. El proceso de selección se repitió durante 3 rondas para obtener rasgos estables y las cepas finales se confirmaron por una reacción de PCR utilizando URA5F, URA3R como cebadores y el ADN como matriz.

Gen OCH1 anulado de *Pichia pastoris* X-33: el plásmido de inactivación pYXZ-OCH1 se alinearon en el sitio Mlu I que se localiza entre los dos brazos homólogos y se transformaron por choque en célula competentes X-33 (ura3), y se diseminaron en medio MD, y se incubaron a 25 grados Celsius durante una semana. Se recogieron las colonias únicas con un palillo de dientes y se sembraron hasta la misma coordinación en dos placas con medio YPD (peptona 2 %, extracto de levadura 1 %, glucosa 2 %, agar 1,5 %), y se incubaron a 25 grados Celsius y 37 grados Celsius, respectivamente durante unos días. Las colonias que crecían bien a 25 grados Celsius pero no podían crecer a 37 grados Celsius se extrajeron para obtener el ADN genómico. Los cebadores externos de OCH1, OCH5F, OCH3R y los cebadores internos (in) 5F, (in) 3R se utilizaron para la identificación de PCR.

Construcción del vector de expresión: el plásmido pUC19/GM-CSF del laboratorio del laboratorio se digirió doblemente con EcoRI y Not I. El fragmento del gen GM-CSF se extrajo (se introdujo una secuencia marcadora 6xHis), y se clonó en el vector de expresión pPICZaA de *Pichia pastoris* digerido con las mismas enzimas de restricción para fabricar el vector de expresión pPICZaA/GM-CSF. Los clones positivos se seleccionaron y confirmaron con la digestión por enzimas de restricción y secuenciación.

La expresión y análisis de GM-CSF en *Pichia pastoris* X-33 y X-33 (och1-): se alineó el vector de expresión pPICZaA/GM-CSF con SaI I y se transformó eléctricamente por choque en células competentes X-33 y X-33 (och1-). La mezcla del choque se diseminó en tela de cultivo revestida con medio YPDZ (que contenían 100 mg/ml, 300 mg/ml, o 500 mg/ml de Zeocina), los transformantes X-33 se cultivaron a 30 grados Celsius durante 3-5 días, y las transformantes X-33 (cho1-) se cultivaron a 25 grados Celsius durante aproximadamente 1 semana. Las colonias únicas que crecían bien se recogieron para extraer el ADN genómico y se identificaron con la reacción PCR utilizando los cebadores 5'AOX, 3'AOX1 para seleccionar los transformantes positivos. Las células X-33/PICZaA/GM-CSF se inocularon en 2 ml de medio YPD (2 % de peptona, un 1 % de extracto de levadura, un 2 % de glucosa), se incubaron a 30 grados Celsius durante 24 h. El cultivo se utilizó para inocular (relación de inoculación 5 %) en 10 ml de medio BMGY (un 2 % de peptona, un 1 % de extracto de levadura, YNB 1. 34 %, glicerol 2 %, 100 mmol/l de tampón fosfato, pH 6,0). Después de la incubación a 30 grados Celsius durante 36 h, el cultivo se centrifugó para retirar el sobrenadante y el aglomerado se re-suspendió hasta 3 ml de medio BMMY (extracto de levadura 1 %, YNB 1. 34 %, peptona 2 %, 100 mmol/l de tampón de fosfato, pH 6,0), se añadió un 2 % de metanol para inducir la expresión: se cultivaron las células X-33 (och1-)/pPICZαA/GM-CSF positivas en el medio YPD a 25 grados Celsius. La condición de inducción de la expresión era la misma que se utilizó en las células X-33, se añadió metanol cada 24 h y la inducción era durante 72 h. Una vez que se terminaron, los cultivos celulares se centrifugaron y se recolectó el sobrenadante para el análisis proteico.

Ejemplo 6 – Análisis del transcriptoma de cepas M5-Blast y SuperM5

Cultivo de la cepa para el aislamiento de ARN. Las cepas BG10, GS115, M5-Blast y SuperM5 (descritas en el Ejemplo 1) se mantuvieron en placas de agar YPD como parches. Para el análisis del transcriptoma, un cultivo de 50 ml de cada cepa se inocularon de un parche y se cultivaron en BMGY a 30 ºC, 200 rpm durante aproximadamente 16 horas. El cultivo estacionario se diluyó 100 veces en medio BMGY recién preparado y se cultivó a 30 ºC, 200 rpm durante 6 horas. Este momento se consideró de crecimiento exponencial con glicerol como fuente de carbono. Se centrifugaron alícuotas en tubos de 15 ml, se desecharon los sobrenadantes y se congelaron rápidamente los aglomerados en nitrógeno líquido. Los aglomerados celulares se almacenaron a -80 ºC para el aislamiento posterior del ARN total.

Aislamiento del ARN total. Se utilizaron los kits FastRNA SPIN (MP Bio) para aislar el ARN total. Se llevó a cabo la lisis celular utilizando el BioSpec Mini-Beadbeater 96. Se eluyó el ARN total de la columna de centrifugación en 15 ml de agua libre de RNasa/DNasa, se congeló en nitrógeno líquido y se almacenó a -80 ºC. Las muestras de ARN se transportaron en hielo seco para el análisis de ARN-Seq en una máquina HiSeq Illumina. Las muestras de ARN se analizaron utilizando un Agilent BioAnalyzer, y todos mostraron picos de ARN ribosómicos de levadura intactos.

Generación y secuenciación de una biblioteca de ARN. Se prepararon las bibliotecas de ARNm utilizando reactivos de Illumina. Se utilizó un kit de Preparación de Muestras de ARN TruSeq para generar selectivamente ADNc con código de barras a partir de poliARNA. Después de la codificación por barras y la amplificación, se

agruparon un total de 12 muestras (4 muestras para este estudio) por análisis. Se llevaron a cabo lecturas finales únicas con base cincuenta. Los datos se remitieron a BioGrammatics en formato FASTQ convencional. Las lecturas se recortaron basándose en bases ambiguas y calificación de calidad y después se filtraron para eliminar todas las lecturas recortadas que eran menores de 40 bases de longitud. Aproximadamente 0,3 % de las lecturas se eliminaron de cada conjunto de datos.

El algoritmo ARN-Seq de CLC Genomics Workbench se utilizó para mapear las lecturas de cada conjunto de datos a la secuencia anotada de BG10 tipo silvestre. Nótese que el genoma de BG10 no contiene casetes de expresión para el gen de resistencia de manosidasa y blasticidina presente en las cepas Man5 y SuperM5.

Perfil de expresión genética. Los perfiles de expresión de cada una de las 4 cepas se representaron y agruparon. Los gráficos de dispersión (con valores R) se evaluaron por comparaciones cepa por cepa de los perfiles de expresión totales. Las cepas BG10 y GS115 presentaban la correlación más estrecha (valor R = 0,98), seguido por las cepas Man5 y SuperM5 (valor R = 0,95). Una ligera regulación positiva general se observó en las cepas OCH1 mutantes vs GS115 (valores de R de 0,92 y 0,84 para Man5 y SuperM5, respetivamente). En total, los patrones de expresión genética eran similares entre las 3 cepas (GS115, M5 y SuperM5) cuando se cultivaban en glicerol.

De cada uno de los conjuntos de datos ARN-Seq se extrajo el mapeo de la cepa BG10, el mapeo de CHO1. En las cepas BG10 y GS115, la escala de cobertura era de 0 a aproximadamente 75. Los niveles de expresión de OCH1 eran aproximadamente igual. Las lecturas de secuenciación se distribuyeron igualmente a lo largo de la fase de lectura abierta. Los niveles de expresión de OCH1 en estas dos cepas eran aproximadamente de un 0,2 % de los genes expresados más altamente.

Para la cepa SuperM5, la escala de cobertura era de 0 – 47. El nivel de expresión caía a aproximadamente la mitad de las de las cepas BG10 y GS115. También, no había cobertura para el extremo N de la fase de lectura abierta. Esta falta de cobertura era el resultado de la eliminación completa de estas secuencias de ADN de la cepa SuperM5.

Para la cepa Man5, la escala de cobertura era de 0 – 502. Había significativamente más cobertura de la parte del extremo N de la fase de lectura abierta de la parte del extremo C. Esta cobertura desequilibrada era el resultado de la duplicación de la mayoría de la parte del extremo N de la fase de lectura abierta en la cepa Man5. La parte N-terminal de la ORF se expresaba a partir del ADN corriente arriba de la ORF de manosidasa y la forma mutante de la parte del extremo C de la ORF se expresaba corriente abajo de las ORF de manosidasa y resistencia a blasticidina.

El mapeo de los datos de Man5 y SuperM5 respecto a la forma mutante de la ORF OCH1 muestra una cobertura completa de la ORF OCH1 mutante en ambas cepas, indicando la expresión genética. La cepa Man_5 muestra una cobertura extra en la parte del extremo N de la ORF, por las mismas razones descritas anteriormente para el mapeo de la ORF OCH1 de tipo silvestre.

El mapeo de los datos de Man5 y SuperM5 respecto a la ORF de manosidasa muestra niveles de expresión similares en las dos cepas.

Conclusión. Se llevó a cabo el análisis de transcriptoma en las cepas Man y SuperM5. Las cepas mostraban patrones de expresión genética total similares. En las cepas Man y SuperM5, se expresa una forma mutante de la ORF OCH1 (está presente el ARNm poliadenilado). La ORF manosidasa HDEL se expresa en ambas cepas a aproximadamente el mismo nivel.

Ejemplo 7 – Expresión de trastuzumab en cepas tipo M5-Blast y SuperM5

En el Estudio 1, la cepa SuperM5 descrita en el Ejemplo 1 se transformó con un vector de expresión que codifica trastuzumab por electroporación. Las colonias resistentes a Zeocina se exploraron por PCR del genoma utilizando cebadores AOX1 y cebadores específicos de Herceptina. Los clones positivos de la PCR del genoma se cultivaron y se evaluó la expresión de trastuzumab tipo Man5 por SDS-PAGE.

En el Estudio 2, las cepas de Pichia transformadas con un vector de expresión que codifica el trastuzumab se exploraron para seleccionar una cepa que exprese altos niveles de trastuzumab. La cepa seleccionada se transformó con el plásmido GlycoSwitch® (pGlycoSwitch-M5/2 (GAP, BSD), proporcionado por Gent University) por electroporación. Las colonias resistentes a Blasticidina S se exploraron por PCR del genoma para detectar la inserción pGlycoSwitch-M5 en el locus OCH1 y la presencia del gen MDS1. Los clones positivos por PCR genómica se cultivaron y se evaluó la expresión de trastuzumab tipo Man5 de los sobrenadantes del cultivo por SDS-PAGE.

En el Estudio 3, los clones positivos obtenidos en el Estudio 1 (clon 46) y el Estudio 2 (clon 11) se cultivaron en un matraz aireador. La expresión de trastuzumab se indujo por la sustitución con medio que contenía metanol. 72 horas después de la inducción con metanol, se purificó el trastuzumab utilizando resina de afinidad de proteína A de los sobrenadantes del cultivo. La productividad del trastuzumab del clon 46 y clon 11 era de 3 mg/l y 1,3 mg/l, respectivamente.

En el Estudio 4, se analizaron las estructuras de N-glicano de trastuzumab producidas en el clon 46 (Estudio 1) y clon 11 (Estudio 2). Se evaluó la homogeneidad de las estructuras de N-glicano en los análisis primarios, y se identificaron las estructuras de N-glicano en el análisis secundario de acuerdo con la búsqueda en las bases de datos de N-glicano y la inyección de HPLC junto con la muestra de referencia. De estos análisis, los N-glicanos de trastuzumab obtenidos del clon 46 (Estudio 1) eran virtualmente homogéneos y el N-glicano predominante (o esencialmente el único) se estimó como el Man5GlcNac2 por análisis de masas MALDI-TOF (Figura 11). Se descubrió que las estructuras de N-glicano del trastuzumab obtenido del clon 11 (Estudio 2) eran una mezcla de Man5GlcNac2 y Man8GlcNac2.

Tabla 7. Análisis de N-glicano del trastuzumab obtenido en el Estudio 2

N-glicano	ODS (GU)	Amida (GU)	PM (Da)	Composición (%)	Valor Cuantitativo (pmol/mg)	Estructura de N-glicano estimada
N1-1	4,7	9,7	1962	7,4	161	$(Hexosa)_9(HexNAc)_2$
N1-2		10,7	2124	4,6	101	$(Hexosa)_{10}$ $(HexNAc)_2$
N2-1	5,0	8,8	1800	22,3	487	$Man_8GlcNAc_2$
N2-2		10,1	2124	7,1	154	$(Hexosa)_{10}$ $(HexNAc)_2$
N3	5,2	7,9	1638	7,4	161	$Man_7GlcNAc_2$
N4-1	6,1	7,0	1475	16,9	370	$Man_8GlcNAc_2$
N4-2		7,9	1638	11,1	241	$(Hexosa)_7(HexNAc)_2$
N5	7,3	6,0	1313	22,1	481	$Man_5GlcNAc_2$
Otros				1,1		
Total				100		

Se descubrió que la afinidad de unión por Her2 del trastuzumab tipo Man5 obtenido del clon 46 se analizó en paralelo con la Herceptina comercial por ELISA y análisis BIAcore, era similar a la actividad de unión a HER2 de la Herceptina comercial. Véase la Figura 12 y la Tabla 8.

Tabla 8. Parámetros cinéticos del trastuzumab analizado por BIAcore

mAb	k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	K_A (M^{-1})	K_D (nM)
Man5-trastuzumab	$2,29 \times 10^5$	$2,43 \times 10^{-5}$	$1,20 \times 10^{10}$	0,083
mAb	k_a ($M^{-1}s^{-1}$)	k_d (s^{-1})	K_A (M^{-1})	K_D (nM)
Herceptina de CHO	$4,25 \times 10^5$	$5,21 \times 10^{-5}$	$8,17 \times 10^9$	0,12
trastuzumab de *Pichia*	$4,65 \times 10^5$	$8,72 \times 10^{-5}$	$5,33 \times 10^9$	0,19

Ejemplo 8 – Análisis de Proteínas glucosiladas adicionales expresadas en M5-Blast y SuperM5

Se integraron genes de lipasas A y B (CalA, 2 motivos de N-glicosilación y CalB, 1 motivo de N-glicosilación) de Candida antárctica, así como transferrina sérica humana (2 motivos de N-glicosilación), dirigidos por un promotor AOX1, en el genoma de la cepa M5-Blast así como la cepa SuperM5, ambas descritas en el Ejemplo 1, mediante recombinación homóloga en el locus AOX1 (selección por Zeocina). Se co-transformó con un plásmido que albergaba un casete de complementación para la auxotrofia por histidina próximo a un gen sintético que codifica la PDI nativa de Pichia que está dirigido por un promotor AOX1. La selección se hizo sobre medio sólido mínimo con Zeocina.

Se cultivaron 47 transformantes de cada combinación descrita anteriormente y se exploraron en cuanto a la abundancia y calidad de la proteína con respecto a cambios obvios en el comportamiento de migración de las proteínas secretadas en microCE (electroforesis de capilaridad, GXII, CaliperLS). Se aplicó como control negativo sobrenadante de la cepa simulada (GS115).

Las 3 proteínas secretadas de la cepa SuperM5 mostraban niveles de expresión comparables a la de la cepa M5-Blast. Además, las señales proteicas diana de los sobrenadantes de SuperM5 en microCE presentaban un tiempo menor de migración que las de los sobrenadantes de M5-Blast, cambiando a pesos moleculares aparentes menores. Se cree que la N-glicosilación alterada de las proteínas secretadas de SuperM5 daba como resultado un peso molecular menor en microescala.

Las muestras de los sobrenadantes de los cultivos a microescala y los de los cultivos en un biorreactor se analizaron en cuanto a sus composiciones en N-glicanos. De las muestras obtenidas del cultivo a microescala, se diluyeron 0,5 ml del medio con dos veces el volumen de tampón RCM (8 M de urea, 3,2 mM de EDTA, 360 mM de Tris-HCl, pH 8,6). De las muestras del biorreactor se utilizaron 0,2 ml. Se prepararon los N-glicanos de estas muestras a continuación del procedimiento de desglicosilación en membrana según publicó Laroy et al. (*Supra*). Después de marcar los N-glicanos secados con ácido 8-aminopireno-1,3,6-trisulfónico, se retiró el exceso de marcador utilizando una cromatografía de exclusión por tamaño (resina Sephadex G-10). Las muestras se reconstituyeron finalmente con 10 ml de gua ultrapura y se diluyeron 10x antes de su inyección (80'' a 1,2 kV) en los capilares (e.1. 36 cm; i.d. 50 mm) de un secuenciador de ADN ABI 3730. Se aplicaron los siguientes ajustes:

Temperatura del horno: 60 ºC	Tensión de ejecución: 15 kV
Tensión pre-ejecución: 180"	Tiempo de ejecución: 1000"
Tiempo pre-ejecución: 15 kV	

Se utilizó el Genemapper v3.7 para analizar los datos obtenidos y las estructuras se asignaron a los picos.

Como se muestra en las Figuras 13-15, los N-glicanos de las proteínas producidas por la cepa SuperM5 eran sustancialmente homogéneos, siendo el Man5GlcNAc2 el N-glicano principal. Por el contrario, los N-glicanos de las proteínas producidas por la M5-Blast eran bastante heterogéneos, especialmente en los cultivos en biorreactor.

Ejemplo 9

En este Ejemplo, se crea una cepa diploide por emparejamiento de la cepa SuperM5 del Ejemplo 1 con una cepa de *Pichia pastoris* de tipo silvestre de un fondo genético diferente. La combinación de los dos fondos genéticos permite una determinación de si los segundos sitios represores o amplificadores del fenotipo de alteración de OCH1 existen en cualquier cepa. El diploide se "mantiene junto" utilizando dos marcadores genéticos dominantes en localizaciones genómicas idénticas de cada cepa haploide. En el locus OCH1 diploide esta cepa transcribe dos ARNm diferentes; uno que codifica el Och1p de tipo silvestre (de la copia genómica haploide de tipo silvestre) y la otra que codifica el Och1p mutante (de la copia genómica haploide de SuperMan5).

Se construyó un vector de intercambio doble que contenía un marcador genético de Higromicina B que remplaza una región altamente conservada de Och1p con un marcador de epítopo V5. Este vector se diseñó de manera que la integración en el genoma diploide, remplazara, a una distribución aproximadamente del 50/50, el dominio altamente conservado de la forma tanto de tipo silvestre como mutante de Och1p, creando ambas inserciones de epítopo en el mismo fondo genético de partida. En el caso en el que el vector se integra en la copia genómica de OCH1 de SuperM5, el marcador genético del fármaco en el vector estará estrechamente unido al marcador de blasticidina adyacente al OCH1. La PCR genómica y secuenciación de ADN se pueden utilizar para verificar la construcción de las dos cepas diploides, una con el tipo silvestre y la otra con la forma mutante del epítopo Och1p marcado.

Las diploides se esporularon y las esporas aleatorias se cultivaron y analizaron. Después del cultivo en medio no selectivo, las colonias haploides resultantes se valoraron en cuanto al a resistencia a Higromicina B. Las características de distribución y crecimiento de los haploides resistentes a Higromicina B pueden determinar la letalidad o deficiencia de crecimiento de la inactivación de Och1p por la inserción del epítopo.

Métodos – Se emparejó una cepa SuperMan5 existente con marcador de resistencia a Zeocina en el locus *prb1Δ* con una cepa haploide BG10 con un marcador de resistencia a nourseotricina en el locus *prb1Δ* para crear la cepa diploide de partida. La cepa BG10 se crea a partir de una construcción de ADN con inactivación del *prb1Δ*.

Un vector de Higromicina B se construyó para sustituir 14 aminoácidos en la secuencia de Ochp1 (LFARGGLYADMDTML, SEQ ID NO: 92) con los 15 aminoácidos del epítopo V5 (GKPIPNPLLGLDST, SEQ ID NO: 93). Esto mantiene la región codificante de longitud completa de las formas de tipo silvestre y mutante de Och1p cuando se integran en el genoma.

El vector de Higromicina B se integra por recombinación homóloga en el genoma diploide. La exploración por PCR del ADN genómico se puede utilizar para verificar el cromosoma (de SuperMan5 o BG10) y la localización en la que se produce la recombinación homóloga. Los productos PCR de las cepas positivas se secuenciaron para verificar la sustitución de los 14 aminoácidos del dominio de Och1p con el marcador V5, estando seguros de que las longitudes respectivas de las ORF se mantenían en cada una de las dos copias.

Las dos cepas se cultivaron y esporularon, y los haploides resultantes se verificaron por sensibilidad a uno o el otro marcador farmacológico en el locus *prb1Δ*. Los haploides se podían explorar visualmente en cuanto al fenotipo de crecimiento a nivel de la placa y, si se observaba una distribución de crecimiento marcada, se valoraba la presencia de la construcción marcada con V5 en cualquiera o ambos de los loci *och1* de SuperMan5 o BG10. Si los haploides crecían igualmente bien, se podían valorar por la presencia del marcador de Higromicina B. La pérdida dl marcador de Higromicina B con la esporulación y la germinación posterior indicarán que la inserción del epítopo V5 en la proteína Och1p es letal en ambos casos de tipo silvestre y mutantes.

La detección de la proteína V5 marcada mediante transferencia de Western en los sobrenadantes y extractos de ambas cepas diploides, y, si fueran viables, los haploides resultantes. Como experimentación adicional, se puede llevar a cabo la localización subcelular de las formas de Ochp1 de tipo silvestre y mutantes marcadas con V5 por inmunofluorescencia de las células diploides.

Como se utiliza en el presente documento, el término “aproximadamente” se refiere a más o menos un 10 % del número referenciado.

Tabla 1. SEQ ID NO: 1

1	AACGTCAAAG	ACAGCAATGG	AGTCAATATT	GATAACACCA	CTGGCAGAGC	GGTTCGTACG
61	TCGTTTTGGA	GCCGATATGA	GGCTCAGCGT	GCTAACAGCA	CGATTGACAA	GAAGACTCTC
121	GAGTGACAGT	AGGTTGAGTA	AAGTATTCGC	TTAGATTCCC	AACCTTCGTT	TTATTCTTTC
181	GTAGACAAAG	AAGCTGCATG	CGAACATAGG	GACAACTTTT	ATAAATCCAA	TTGTCAAACC
241	AACGTAAAAC	CCTCTGGCAC	CATTTTCAAC	ATATATTTGT	GAAGCAGTAC	GCAATATCGA
301	TAAATACTCA	CCGTTGTTTG	TAACAGCCCC	AACTTGCATA	CGCCTTCTAA	TGACCTCAAA
361	TGGATAAGCC	GCAGCTTGTG	CTAACATACC	AGCAGCACCG	CCCGCGGTCA	GCTGCGCCCA
421	CACATATAAA	GGCAATCTAC	GATCATGGGA	GGAATTAGTT	TTGACCGTCA	GGTCTTCAAG
481	AGTTTTGAAC	TCTTCTTCTT	GAACTGTGTA	ACCTTTTAAA	TGACGGGATC	TAAATACGTC
541	ATGGATGAGA	TCATGTGTGT	AAAAACTGAC	TCCAGCATAT	GGAATCATTC	CAAAGATTGT
601	AGGAGCGAAC	CCACGATAAA	AGTTTCCCAA	CCTTGCCAAA	GTGTCTAATG	CTGTGACTTG
661	AAATCTGGGT	TCCTCGTTGA	AGACCCTGCG	TACTATGCCC	AAAAACTTTC	CTCCACGAGC
721	CCTATTAACT	TCTCTATGAG	TTTCAAATGC	CAAACGGACA	CGGATTAGGT	CCAATGGGTA
781	AGTGAAAAAC	ACAGAGCAAA	CCCCAGCTAA	TGAGCCGGCC	AGTAACCGTC	TTGGAGCTGT
841	TTCATAAGAG	TCATTAGGGA	TCAATAACGT	TCTAATCTGT	TCATAACATA	CAAATTTTAT
901	GGCTGCATAG	GGAAAAATTC	TCAACAGGGT	AGCCGAATGA	CCCTGATATA	GACCTGCGAC
961	ACCATCATAC	CCATAGATCT	GCCTGACAGC	CTTAAAGAGC	CCGCTAAAAG	ACCCGGAAAA
1021	CCGAGAGAAC	TCTGGATTAG	CAGTCTGAAA	AAGAATCTTC	ACTCTGTCTA	GTGGAGCAAT
1081	TAATGTCTTA	GCGGCACTTC	CTGCTACTCC	GCCAGCTACT	CCTGAATAGA	TCACATACTG
1141	CAAAGACTGC	TTGTCGATGA	CCTTGGGGTT	ATTTAGCTTC	AAGGGCAATT	TTTGGGACAT
1201	TTTGGACACA	GGAGACTCAG	AAACAGACAC	AGAGCGTTCT	GAGTCCTGGT	GCTCCTGACG
1261	TAGGCCTAGA	ACAGGAATTA	TTGGCTTTAT	TTGTTTGTCC	ATTTCATAGG	CTTGGGGTAA
1321	TAGATAGATG	ACAGAGAAAT	AGAGAAGACC	TAATATTTTT	TGTTCATGGC	AAATCGCGGG
1381	TTCGCGGTCG	GGTCACACAC	GGAGAAGTAA	TGAGAAGAGC	TGGTAATCTG	GGGTAAAAGG
1441	GTTCAAAAGA	AGGTCGCCTG	GTAGGGATGC	AATACAAGGT	TGTCTTGGAG	TTTACATTGA
1501	CCAGATGATT	TGGCTTTTTC	TCTGTTCAAT	TCACATTTTT	CAGCGAGAAT	CGGATTGACG
1561	GAGAAATGGC	GGGGTGTGGG	GTGGATAGAT	GGCAGAAATG	CTCGCAATCA	CCGCGAAAGA
1621	AAGACTTTAT	GGAATAGAAC	TACTGGGTGG	TGTAAGGATT	ACATAGCTAG	TCCAATGGAG
1681	TCCGTTGGAA	AGGTAAGAAG	AAGCTAAAAC	CGGCTAAGTA	ACTAGGGAAG	AATGATCAGA
1741	CTTTGATTTG	ATGAGGTCTG	AAAATACTCT	GCTGCTTTTT	CAGTTGCTTT	TTCCCTGCAA
1801	CCTATCATTT	TCCTTTTCAT	AAGCCTGCCT	TTTCTGTTTT	CACTTATATG	AGTTCCGCCG
1861	AGACTTCCCC	AAATTCTCTC	CTGGAACATT	CTCTATCGCT	CTCCTTCCAA	GTTGCGCCCC
1921	CTGGCACTGC	CTAGTAATAT	TACCACGCGA	CTTATATTCA	GTTCCACAAT	TTCCAGTGTT
1981	CGTAGCAAAT	ATCATCAGCC	TACCGTTCGT	ATAGCATACA	TTATACGAAC	GGTACTTTTT
2041	TGTAGAAATG	TCTTGGTGTC	CTCGTCCAAT	CAGGTAGCCA	TCTCTGAAAT	ATCTGGCTCC
2101	GTTGCAACTC	CGAACGACCT	GCTGGCAACG	TAAAATTCTC	CGGGGTAAAA	CTTAAATGTG
2161	GAGTAATGGA	ACCAGAAACG	TCTCTTCCCT	TCTCTCTCCT	TCCACCGCCC	GTTACCGTCC
2221	CTAGGAAATT	TTACTCTGCT	GGAGAGCTTC	TTCTACGGCC	CCCTTGCAGC	AATGCTCTTC
2281	CCAGCATTAC	GTTGCGGGTA	AAACGGAGGT	CGTGTACCCG	ACCTAGCAGC	CCAGGGATGG
2341	AAAAGTCCCG	GCCGTCGCTG	GCAATAATAG	CGGGCGGACG	CATGTCATGA	GATTATTGGA
2401	AACCACCAGA	ATCGAATATA	AAAGGCGAAC	ACCTTTCCCA	ATTTTGGTTT	CTCCTGACCC
2461	AAAGACTTTA	AATTTAATTT	ATTTGTCCCT	ATTTCAATCA	ATTGAACAAC	TATTTCGCGA
2521	AACGATGAGA	TTTCCTTCAA	TTTTTACTGC	TGTTTTATTC	GCAGCATCCT	CCGCATTAGC
2581	TGCTCCAGTC	AACACTACAA	CAGAAGATGA	AACGGCACAA	ATTCCGGCTG	AAGCTGTCAT
2641	CGGTTACTCA	GATTTAGAAG	GGGATTTCGA	TGTTGCTGTT	TTGCCATTTT	CCAACAGCAC
2701	AAATAACGGG	TTATTGTTTA	TAAATACTAC	TATTGCCAGC	ATTGCTGCTA	AAGAAGAAGG
2761	GGTATCTCTC	GAGAAAAGAG	AGGCTGAAGC	TGAATTCGCC	ACAAAACGTG	GATCTCCCAA

2821 CCCTACGAGG GCGGCAGCAG TCAAGGCCGC ATTCCAGACG TCGTGGAACG CTTACCACCA
2881 TTTTGCCTTT CCCCATGACG ACCTCCACCC GGTCAGCAAC AGCTTTGATG ATGAGAGAAA
2941 CGGCTGGGGC TCGTCGGCAA TCGATGGCTT GGACACGGCT ATCCTCATGG GGGATGCCGA
3001 CATTGTGAAC ACGATCCTTC AGTATGTACC GCAGATCAAC TTCACCACGA CTGCGGTTGC
3061 CAACCAAGGC ATCTCCGTGT TCGAGACCAA CATTCGGTAC CTCGGTGGCC TGCTTTCTGC
3121 CTATGACCTG TTGCGAGGTC CTTTCAGCTC CTTGGCGACA AACCAGACCC TGGTAAACAG
3181 CCTTCTGAGG CAGGCTCAAA CACTGGCCAA CGGCCTCAAG GTTGCGTTCA CCACTCCCAG
3241 CGGTGTCCCG GACCCTACCG TCTTCTTCAA CCCTACTGTC CGGAGAAGTG GTGCATCTAG
3301 CAACAACGTC GCTGAAATTG GAAGCCTGGT GCTCGAGTGG ACACGGTTGA GCGACCTGAC
3361 GGGAAACCCG CAGTATGCCC AGCTTGCGCA GAAGGGCGAG TCGTATCTCC TGAATCCAAA
3421 GGGAAGCCCG GAGGCATGGC CTGGCCTGAT TGGAACGTTT GTCAGCACGA GCAACGGTAC
3481 CTTTCAGGAT AGCAGCGGCA GCTGGTCCGG CCTCATGGAC AGCTTCTACG AGTACCTGAT
3541 CAAGATGTAC CTGTACGACC CGGTTGCGTT TGCACACTAC AAGGATCGCT GGGTCCTTGC
3601 TGCCGACTCG ACCATTGCGC ATCTCGCCTC TCACCCGTCG ACGCGCAAGG ACTTGACCTT
3661 TTTGTCTTCG TACAACGGAC AGTCTACGTC GCCAAACTCA GGACATTTGG CCAGTTTTGC
3721 CGGTGGCAAC TTCATCTTGG GAGGCATTCT CCTGAACGAG CAAAAGTACA TTGACTTTGG
3781 AATCAAGCTT GCCAGCTCGT ACTTTGCCAC GTACAACCAG ACGGCTTCTG GAATCGGCCC
3841 CGAAGGCTTC GCGTGGGTGG ACAGCGTGAC GGGCGCCGGC GGCTCGCCGC CCTCGTCCCA
3901 GTCCGGGTTC TACTCGTCGG CAGGATTCTG GGTGACGGCA CCGTATTACA TCCTGCGGCC
3961 GGAGACGCTG GAGAGCTTGT ACTACGCATA CCGCGTCACG GGCGACTCCA AGTGGCAGGA
4021 CCTGGCGTGG GAAGCGTTCA GTGCCATTGA GGACGCATGC CGCGCCGGCA GCGCGTACTC
4081 GTCCATCAAC GACGTGACGC AGGCCAACGG CGGGGGTGCC TCTGACGATA TGGAGAGCTT
4141 CTGGTTTGCC GAGGCGCTCA AGTATGCGTA CCTGATCTTT GCGGAGGAGT CGGATGTGCA
4201 GGTGCAGGCC AACGGCGGGA ACAAATTTGT CTTTAACACG GAGGCGCACC CCTTTAGCAT
4261 CCGTTCATCA TCACGACGGG GCGGCCACCT TGCTCACGAC GAGTTGTAAT CTAGGGCGGC
4321 CGCCAGCTTG GGCCCGAACA AAAACTCATC TCAGAAGAGG ATCTGAATAG CGCCGTCGAC
4381 CATCATCATC ATCATCATTG AGTTTTAGCC TTAGACATGA CTGTTCCTCA GTTCAAGTTG
4441 GGCACTTACG AGAAGACCGG TCTTGCTAGA TTCTAATCAA GAGGATGTCA GAATGCCATT
4501 TGCCTGAGAG ATGCAGGCTT CATTTTTGAT ACTTTTTTAT TTGTAACCTA TATAGTATAG
4561 GATTTTTTTT GTCATTTTGT TTCTTCTCGT ACGAGCTTGC TCCTGATCAG CCTATCTCGC
4621 AGCTGATGAA TATCTTGTGG TAGGGGTTTG GGAAAATCAT TCGAGTTTGA TGTTTTTCTT
4681 GGTATTTCCC ACTCCTCTTC AGAGTACAGA AGATTAAGTG AGACCTTCGT TTGTGCGGAT
4741 CCCCCACACA CCATAGCTTC AAAATGTTTC TACTCCTTTT TTACTCTTCC AGATTTTCTC
4801 GGACTCCGCG CATCGCCGTA CCACTTCAAA ACACCCAAGC ACAGCATACT AAATTTCCCC
4861 TCTTTCTTCC TCTAGGGTGT CGTTAATTAC CCGTACTAAA GGTTTGGAAA AGAAAAAAGA
4921 GACCGCCTCG TTTCTTTTTC TTCGTCGAAA AAGGCAATAA AAATTTTTAT CACGTTTCTT
4981 TTTCTTGAAA ATTTTTTTTT TTGATTTTTT TCTCTTTCGA TGACCTCCCA TTGATATTTA
5041 AGTTAATAAA CGGTCTTCAA TTTCTCAAGT TTCAGTTTCA TTTTTCTTGT TCTATTACAA
5101 CTTTTTTTAC TTCTTGCTCA TTAGAAAGAA AGCATAGCAA TCTAATCTAA GGGCGGTGTT
5161 GACAATTAAT CATCGGCATA GTATATCGGC ATAGTATAAT ACGACAAGGT GAGGAACTAA
5221 ACCATGGCCA AGCCTTTGTC TCAAGAAGAA TCCACCCTCA TTGAAAGAGC AACGGCTACA
5281 ATCAACAGCA TCCCCATCTC TGAAGACTAC AGCGTCGCCA GCGCAGCTCT CTCTAGCGAC
5341 GGCCGCATCT TCACTGGTGT CAATGTATAT CATTTTACTG GGGGACCTTG TGCAGAACTC
5401 GTGGTGCTGG GCACTGCTGC TGCTGCGGCA GCTGGCAACC TGACTTGTAT CGTCGCGATC
5461 GGAAATGAGA ACAGGGGCAT CTTGAGCCCC TGCGGACGGT GCCGACAGGT GCTTCTCGAT
5521 CTGCATCCTG GGATCAAAGC CATAGTGAAG GACAGTGATG GACAGCCGAC GGCAGTTGGG
5581 ATTCGTGAAT TGCTGCCCTC TGGTTATGTG TGGGAGGGCT AAGCACTTCG TGGCCGAGGA
5641 GCAGGACTGA CACGTCCGAC GCGGCCCGAC GGGTCCGAGG CCTCGGAGAT CCGTCCCCCT
5701 TTTCCTTTGT CGATATCATG TAATTAGTTA TGTCACGCTT ACATTCACGC CCTCCCCCCA
5761 CATCCGCTCT AACCGAAAAG GAAGGAGTTA GACAACCTGA AGTCTAGGTC CCTATTTATT
5821 TTTTTATAGT TATGTTAGTA TTAAGAACGT TATTTATATT TCAAATTTTT CTTTTTTTTC
5881 TGTACAGACG CGTGTACGCA TGTAACATTA TACTGAAAAC CTTGCTTGAG AAGGTTTTGG
5941 GACGCTCGAA GGCTTTAATT TGCAAGCTGG AGACCAACAT GTGAGCAAAA GGCCAGCAAA
6001 AGGCCAGGAA CCGTAAAAAG GCCGCGTTGC TGGCGTTTTT CCATAGGCTC CGCCCCCCTG
6061 ACGAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA
6121 GATACCAGGC GTTTCCCCCT GGAAGCTCCC TCGTGCGCTC TCCTGTTCCG ACCCTGCCGC

6181 TTACCGGATA CCTGTCCGCC TTTCTCCCTT CGGGAAGCGT GGCGCTTTCT CATAGCTCAC
6241 GCTGTAGGTA TCTCAGTTCG GTGTAGGTCG TTCGCTCCAA GCTGGGCTGT GTGCACGAAC
6301 CCCCCGTTCA GCCCGACCGC TGCGCCTTAT CCGGTAACTA TCGTCTTGAG TCCAACCCGG
6361 TAAGACACGA CTTATCGCCA CTGGCAGCAG CCACTGGTAA CAGGATTAGC AGAGCGAGGT
6421 ATGTAGGCGG TGCTACAGAG TTCTTGAAGT GGTGGCCTAA CTACGGCTAC ACTAGAAGAA
6481 CAGTATTTGG TATCTGCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTTGGTAGCT
6541 CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTTT TTTTGTTTGC AAGCAGCAGA
6601 TTACGCGCAG AAAAAAAGGA TCTCAAGAAG ATCCTTTGAT CTTTTCTACG GGGTCTGACG
6661 CTCAGTGGAA CGAAAACTCA CGTTAAGGGA TTTTGGTCAT GAGATCAGAT CTAACATCCA
6721 TAATCGTATT CGCCGTTTCT GTCATTTGCG TTTTGTACGG ACCCTCACAA CAATTATCAT
6781 CTCCAAAAAT AGACTATGAT CCATTGACGC TCCGATCACT TGATTTGAAG ACTTTGGAAG
6841 CTCCTTCACA GTTGAGTCCA GGCACCGTAG AAGATAATCT TCGAAGACAA TTGGAGTTTC
6901 ATTTTCCTTA CCGCAGTTAC GAACCTTTTC CCCAACATAT TTGGCAAACG TGGAAAGTTT
6961 CTCCCTCTGA TAGTTCCTTT CCGAAAAACT TCAAAGACTT AGGTGAAAGT TGGCTGCAAA
7021 GGTCCCCAAA TTATGATCAT TTTGTGATAC CCGATGATGC AGCATGGGAA CTTATTCACC
7081 ATGAATACGA ACGTGTACCA GAAGTCTTGG AAGCTTTCCA CCTGCTACCA GAGCCCATTC
7141 TAAAGGCCGA TTTTTTCAGG TATTTGATTC TTTTTGCCCG TGGAGGACTG TATGCTGACA
7201 TGGACACTAT GTTATTAAAA CCAATAGAAT CGTGGCTGAC TTTCAATGAA ACTATTGGTG
7261 GAGTAAAAAA CAATGCTGGG TTGGTCATTG GTATTGAGGC TGATCCTGAT AGACCTGATT
7321 GGCACGACTG GTATGCTAGA AGGATACAAT TTTGCCAATG GGCAATTCAG TCCAAACGAG
7381 GACACCCAGC ACTGCGTGAA CTGATTGTAA GAGTTGTCAG CACGACTTTA CGGAAAGAGA
7441 AAAGCGGTTA CTTGAACATG GTGGAAGGAA AGGATCGTGG AAGTGATGTG ATGGACTGGA
7501 CGGGTCCAGG AATATTTACA GACACTCTAT TTGATTATAT GACTAATGTC AATACAACAG
7561 GCCACTCAGG CCAAGGAATT GGAGCTGGCT CAGCGTATTA CAATGCCTTA TCGTTGGAAG
7621 AACGTGATGC CCTCTCTGCC CGCCCGAACG GAGAGATGTT AAAAGAGAAA GTCCCAGGTA
7681 AATATGCACA GCAGGTTGTT TTATGGGAAC AATTTACCAA CCTGCGCTCC CCCAAATTAA
7741 TCGACGATAT TCTTATTCTT CCGATCACCA GCTTCAGTCC AGGGATTGGC CACAGTGGAG
7801 CTGGAGATTT GAACCATCAC CTTGCATATA TTAGGCATAC ATTTGAAGGA AGTTGGAAGG
7861 ACTAAAGAAA GCTAGAGTAA AATAGATATA GCGAGATTAG AGAATGAATA CCTTCTTCTA
7921 AGCGATCGTC CGTCATCATA GAATATCATG GACTGTATAG TTTTTTTTTT GTACATATAA
7981 TGATTAAACG GTCATCCAAC ATCTCGTTGA CAGATCTCTC AGTACGCGAA ATCCCTGACT
8041 ATCAAAGCAA GAACCGATGA AGAAAAAAAC AACAGTAACC CAAACACCAC AACAAACACT
8101 TTATCTTCTC CCCCCCAACA CCAATCATCA AAGAGATGTC GGAACCAAAC ACCAAGAAGC
8161 AAAAACTAAC CCCATATAAA AACATCCTGG TAGATAATGC TGGTAACCCG CTCTCCTTCC
8221 ATATTCTGGG CTACTTCACG AAGTCTGACC GGTCTCAGTT GATCAACATG ATCCTCGAAA
8281 TGGGTGGCAA GATCGTTCCA GACCTGCCTC CTCTGGTAGA TGGAGTGTTG TTTTTGACAG
8341 GGGATTACAA GTCTATTGAT GAAGATACCC TAAAGCAACT GGGGGACGTT CCAATATACA
8401 GAGACTCCTT CATCTACCAG TGTTTTGTGC ACAAGACATC TCTTCCCATT GACACTTTCC
8461 GAATTGACAA GAACGTCGAC TTGGCTCAAG ATTTGATCAA TAGGGCCCTT CAAGAGTCTG
8521 TGGATCATGT CACTTCTGCC AGCACAGCTG CAGCTGCTGC TGTTGTTGTC GCTACCAACG
8581 GCCTGTCTTC TAAACCAGAC GCTCGTACTA GCAAAATACA GTTCACTCCC GAAGAAGATC
8641 GTTTTATTCT TGACTTTGTT AGGAGAAATC CTAAACGAAG AAACACACAT CAACTGTACA
8701 CTGAGCTCGC TCAGCACATG AAAAACCATA CGAATCATTC TATCCGCCAC AGATTTCGTC
8761 GTAATCTTTC CGCTCAACTT GATTGGGTTT ATGATATCGA TCCATTGACC AACCAACCTC
8821 GAAAAGATGA AAACGGGAAC TACATCAAGG TACAAGATCT TCCACAAGGA ATTCGTGGTC
8881 ATTATTCTGC CCAAGATGAT TACAATTTGT GTTTATCGGT TCAACCTTTC ATTGAATCTG
8941 TAGATGAGAC AACAGGCCAA GAATTTTTCA AACCTCTGAA AGGTGTATTT GATGACTTGG
9001 AATCTCGCTT TCCTCACCAT ACAAAGACTT CCTGGAGAGA CAGATTCAGA AAGTTTGCCT
9061 CTAAATACGG TGTTCGTCAG TACATCGCGT ATTATGAAAA GACTGTTGAA CTCAATGGTG
9121 TTCCTAATCC GATGACGAAC TTTACCTCAA AGGCTTCCAT TGAAAAATTT AGAGAAAGAC
9181 GCGGGACTTC ACGTAACAGT GGCCTTCCAG GCCCGGTTGG TGTAGAAGCT GTAAGCTCTT
9241 TGGACCACAT ATCCCCATTG GTCACATCTA ATTCCAATTC TGCAGCTGCT GCAGCTGCTG
9301 CCGCAGCAGT TGCAGCCTCT GCCTCTGCTT CTTCAGCTCC TAATACTTCA ACTACCAATT
9361 TCTTTGAACA GGAGAATATT GCCCAAGTTC TCTCTGCACA TAACAACGAG CAGTCTATTG
9421 CAGAAGTTAT TGAGTCCGCA CAGAATGTCA ACACCCATGA AAGTGAACCT ATAGCTGATC
9481 ATGTTCGAAA AAATCTTACA GACGATGAAT TGCTTGACAA AATGGATGAT ATTTTAAGCT

```
9541 CCAGAAGTCT  AGGCGGACTA  GATGACTTGA  TAAAGATCCT  CTACACTGAG  CTGGGATTTG
9601 CTCATCGTTA  TACCGAATTT  CTTTTTACCT  CATGTTCTGG  TGATGTGATT  TTCTTCCGAC
9661 CATTAGTGGA  ACATTTCCTT  CTTACTGGTG  AGTGGGAGCT  GGAGAATACT  CGTGGCATCT
9721 GGACCGGTCG  TCAAGACGAA  ATGCTACGTG  CTAGCAATCT  AGATGACCTG  CACAAGTTAA
9781 TTGACCTGCA  TGGGAAAGAA  CGTGTTGAGA  CCAGAAGAAA  AGCCATCAAG  GGAGAATGAT
9841 CATAAGAAAT  GAAAAACGTA  TAAGT
```

Tabla 2. SEQ ID NO: 2

(M)AKADGSLLY	YNPHNPPRRY	YFYMAIFAVS	VICVLYGPSQ
QLSSPKIDYD	PLTLRSLDLK	TLEAPSQLSP	GTVEDNLRRQ
LEFHFPYRSY	EPFPQHIWQT	WKVSPSDSSF	PKNFKDLGES
WLQRSPNYDH	FVIPDDAAWE	LIHHEYERVP	EVLEAFHLLP
EPILKADFFR	YLILFARGGL	YADMDTMLLK	PIESWLTFNE
TIGGVKNNAG	LVIGIEADPD	RPDWHDWYAR	RIQFCQWAIQ
SKRGHPALRE	LIVRVVSTTL	RKEKSGYLNM	VEGKDRGSDV
MDWTGPGIFT	DTLFDYMTNV	NTTGHSGQGI	GAGSAYYNAL
SLEERDALSA	RPNGEMLKEK	VPGKYAQQVV	LWEQFTNLRS
PKLIDDILIL	PITSFSPGIG	HSGAGDLNHH	LAYIRHTFEG
SWKD			

Tabla 3

Nucleótidos eliminados desde corriente arriba de OCH1	Aminoácidos correspondientes a los nucleótidos eliminados	Descripción
GCG AAG GCA GAT GGC (SEQ ID NO: 29)	AKADG (SEQ ID NO: 4)	5 AAs eliminados desde corriente arriba de la porción OCH1
GCG AAG GCA GAT GGC AGT (SEQ ID NO: 30)	AKADGS (SEQ ID NO: 5)	6 AAs eliminados desde corriente arriba de la porción OCH1
GCG AAG GCA GAT GGC AGT TTG (SEQ ID NO: 31)	AKADGSL (SEQ ID NO: 6)	7 AAs eliminados desde corriente arriba de la porción OCH1
GCG AAG GCA GAT GGC AGT TTG CTC (SEQ ID NO: 32)	AKADGSLL (SEQ ID NO: 7)	8 AAs eliminados desde corriente arriba de la porción OCH1
GCG AAG GCA GAT GGC AGT TTG CTC TAC (SEQ ID NO: 33)	AKADGSLLY (SEQ ID NO: 8)	9 AAs eliminados desde corriente arriba de la porción OCH1
GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT (SEQ ID NO: 34)	AKADGSLLYY (SEQ ID NO: 9)	10 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC (SEQ ID NO: 35)	MAKADG (SEQ ID NO: 10)	6 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT (SEQ ID NO: 36)	MAKADGS (SEQ ID NO: 11)	7 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG (SEQ ID NO: 37)	MAKADGSL (SEQ ID NO: 12)	8 AAs eliminados desde corriente arriba de la porción OCH1

Nucleótidos eliminados desde corriente arriba de OCH1	Aminoácidos correspondientes a los nucleótidos eliminados	Descripción
ATG GCG AAG GCA GAT GGC AGT TTG CTC (SEQ ID NO: 38)	MAKADGSLL (SEQ ID NO: 13)	9 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC (SEQIDNO:39)	MAKADGSLLY (SEQ ID NO: 14)	10 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT (SEQ ID NO: 40)	MAKADGSLLYY (SEQ ID NO: 15)	11 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT (SEQ ID NO: 41)	MAKADGSLLYYN (SEQ ID NO: 16) 41)	12 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT (SEQ ID NO: 42)	MAKADGSLLYYNP (SEQ ID NO: 17)	13 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC (SEQ ID NO: 43)	MAKADGSLLYYNPH (SEQ ID NO: 18)	14 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT (SEQ ID NO: 44)	MAKADGSLLYYNPHN (SEQ ID NO: 19)	15 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT CCA (SEQ ID NO: 45)	MAKADGSLLYYNPHNP (SEQ ID NO: 20)	16 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT CCA CCC (SEQ ID NO: 46)	MAKADGSLLYYNPHNP P (SEQ ID NO: 21)	17 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT CCA CCC AGA (SEQ ID NO: 47)	MAKADGSLLYYNPHNP PR (SEQ ID NO: 22)	18 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT CCA CCC AGA AGG (SEQ ID NO: 48)	MAKADGSLLYYNPHNP PRR (SEQ ID NO: 23)	19 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT CCA CCC AGA AGG TAT (SEQ ID NO:49)	MAKADGSLLYYNPHNP PRRY (SEQ ID NO: 24)	20 AAs eliminados desde corriente arriba de la porción OCH1

Nucleótidos eliminados desde corriente arriba de OCH1	Aminoácidos correspondientes a los nucleótidos eliminados	Descripción
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CAC AAT CCA CCC AGA AGG TAT TAC (SEQ ID NO: 50)	MAKADGSLLYYNPHNP PRRYY (SEQ ID NO: 25)	2 1 AAs eliminados desde corriente arriba de la porción OCH1
GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CACA ATC CAC CCA GAA GGT ATT ACT TCT ACA TGG CTA SEQ ID NO: 51)	AKADGSLLYYNPHNPP RRYYFYMA (SEQ ID NO: 26)	24 AAs eliminados desde corriente arriba de la porción OCH1
ATG GCG AAG GCA GAT GGC AGT TTG CTC TAC TAT AAT CCT CACA ATC CAC CCA GAA GGT ATT ACT TCT ACA TGG CTA (SEQ ID NO: 52)	MAKADGSLLYYNPHNP PRRYYFYMA (SEQ ID NO: 27)	25 AAs eliminados desde corriente arriba de la porción OCH1

Tabla 4. SEQ ID NO: 53.

```
   1 AACGTCAAAG  ACAGCAATGG  AGTCAATATT  GATAACACCA CTGGCAGAGC GGTTCGTACG
  61 TCGTTTTGGA  GCCGATATGA  GGCTCAGCGT  GCTAACAGCA CGATTGACAA GAAGACTCTC
 121 GAGTGACAGT  AGGTTGAGTA  AAGTATTCGC  TTAGATTCCG AACCTTCGTT TTATTCTTTC
 181 GTAGACAAAG  AAGCTGCATG  CGAACATAGG  GACAACTTTT ATAAATCCAA TTGTCAAACC
 241 AACGTAAAAC  CCTCTGGCAC  CATTTTCAAC  ATATATTTGT GAAGCAGTAC GCAATATCGA
 301 TAAATACTCA  CCGTT6TTTG  TAACAGCCCC  AACTTGCATA CGCCTTCTAA TGACCTCAAA
 361 TGGATAAGCC  GCAGCTTGT6  CTAAGATACC  AGCAGCACCG CCCGCGGTCA GCTGCGCCCA
 421 CACATATAAA  GGCAATCTAC  GATGATGGGA  GGAATTAGTT TTGACCGTCA GGTCTTCAAG
 481 AGTTTTGAAG  TCTTCTTCTT  GAAGTGTGTA  AGCTTTTAAA T6AGGGGATC TAAATACGTC
 541 ATGGATGAGA  TCATGTGTGT  AAAAACTGAC  TCCAGCATAT GGAATCATTC CAAAGATTGT
 601 AGGAGCGAAC  CCACGATAAA  AGTTTCCCAA  CCTTGCCAAA GTGTCTAATG CTGTGACTTG
 661 AAATCTGGGT  TCCTCGTTGA  AGACCCTGCG  TAGTATGCCC AAAAACTTTC CTCGACGAGC
 721 CCTATTAACT  TCTCTATGAG  TTTCAAATGC  GAAACGGACA CGGATTAGGT CCAATGGGTA
 781 AGTGAAAASC  ACAGAGCAAA  CCCCAGCTAA  TGAGCCGGCC AGTAACCGTC TTGGAGCTGT
 841 TTCATAAGAG  TCATTAGGGA  TCAATAACGT  TCTAATGTGT TCATAACATA CAAATTTTAT
 901 GGCTGCATAG  GGAAAAATTC  TCAACAGGGT  AGGCGAATGA CCCTGATATA GACCTGCGAC
 961 ACCATCATAC  CCATAGATCT  GCCTGACAGC  GTTAAAGAGG CGGCTAAAAG ACCCGGAAAA
1021 CC6AGAGAAC  TCTGGATTAG  CAGTCTGAAA  MGAATCTTC  ACTCTGTCTA GTGGA6CAAT
1081 TAATGTCTTA  GCGGCACTTC  CTGCTACTCC  GCCAGCTACT CGTGAATAGA TCACATACT6
1141 CAAAGACTGC  TT6TC6ATGA  CGTTGGGGTT  ATTTAGCTTC AAGG6CAATT TTTGGGACAT
1201 TTTGGACACA  GGAGACTCAG  AAACAGACAC  AGAGCGTTCT GAGTCCTG6T GCTCCTGAC6
1261 TAGGCCTAGA  ACAGGAATTA  TTGGCTTTAT  TTGTTTGTCC ATTTCATAGG CTTGGGGTAA
1321 TAGATAGATG  ACAGAGAAAT  AGAGAAGACC  TAATATTTTT TGTTCATGGC AAATCGCGGG
1381 TTCGCGGTCG  GGTCACACAC  GGAGAAGTAA  TGAGAAGAGC TGGTAATCTG GGGTAAAAGG
1441 GTTCAAAAGA  AGGTCGCCTG  GTAGGGATGC  AATACAAGGT TGTCTTGGAG TTTACATTGA
1501 CCAGATGATT  TGGCTTTTTC  TCTGTTCAAT  TCACATTTTT CAGCGAGAAT CGGATTGACG
1561 GAGAAATGGC  GGGGTGTGGG  GTGGATA6AT  GGCAGAAATG CTCGCAATCA CCGCGAAAGA
1621 AAGACTTTAT  GGAATAGAA.C TACTGGGTGG  TGTAAGGATT ACATAGCTAG TCCAATGGAG
1681 TCCGTTGGAA  AGGTAAGAAG  AAGCTAAAAC  GGGGTAAGTA ACTAG6GAAG AATGATCAGA
1741 CTTT6ATTTG  ATGAGGTCTG  AAAATACTCT  GCTGCTTTTT CAGTTGCTTT TTCCCTGCAA
1801 CCTATCATTT  TCCTTTTCAT  AAGCCTGCCT  TTTGTGTTTT CACTTATATG AGTTCCGCCG
1861 AGACTTCCCC  AAATTCTCTC  CTGGAACATT  CTCTATCGCT CTCCTTCCAA GTTGCGCCCC
1921 CTGGCACTGC  CTAGTAATAT  TACCACGCGA  CFTATATTCA GTTCCACAAT TTCCAGTGTT
1981 CGTAGCAAAT  ATCATCAGCC  ATGGCGAAGG  CAGATGGCAG TTTGCTCTAC TATAATCCTC
2041 ACAATCCACC  CAGAAGGTAT  TACTTCTAGA  TGGCTATATT CGCCGTTTCT GTCATTTGCG
2101 TTTTGTACGG  ACCCTCACAA  CAATTATCAT  CTCGAAAAAT AGACTATGAT CCATTGACGC
2161 TCCGATCACT  TGATTTGAA6  ACTTTGGAAG  CTCCTTCACA GTTGAGTCCA GGCACCGTAG
2221 AAGATAATCT  TCGaagacaa  ttggagtttc  attttcctta ccgcagttac gaoccttttc
```

2281 cccaacatat ttgataaacg tggaaagttt cttcctctga tagttccttt ccgoaaoact
2341 tcaaagactt aggtgaaagt tggctgcaaa ggttectaaa ttatgatcat tttgtgatac
2401 ccgatgatgc agcatgggaa cttattcacc atgaatacga acgtgtacca goagtcttgg
2461 aagctctaaa tgctcaccgc aatgctgtta aggttegtat ggagaaactg ggacttattt
2521 aattatttag agattttaae ttacatttag attcgataga tccacaggac gggtgtggtc
2581 gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggaggegg
2641 ccaaagcggt cggacagtgi tccgagaacg ggtgcgcata gaaattgcat caacgcatat
2701 agcgctagca acacgccata gtgactggcg atgctgtcgg aatggacgat atcccgcaag
2761 aggcccggca gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg
2821 aggataacga tgagcgcatt gttagatttc atacacggtg cctgactgcg ttagcaattt
2881 aactgtgata aactaccgca ttaaagctga tctttttgt agaaatgtct tggtgtcctc
2941 gtccaatcag atacccatct ctgaaatatc tggctccgtt gcaactccga acgacctgct
3001 ggcaacgtaa aattctccgg ggtaaaactt aaatgtggag taatggaacc agaaacgtct
3061 cttcccttct ctctccttcc accgcccgtt accgtcccta ggaaatttta ttctgctgga
3121 gagcttcttc tacggccccc ttgcagcaat gctcttccca gcattacgtt gcgggtaaaa
3181 cggaggtcgt ctacccgacc tagcagccca gggatggaaa agtcccggcc gtcgctggca
3241 ataatagcgg gcggacgcat gtcatgagat tattygaaac caccagaatc gaatataaaa
3301 ggcgaacacc tttcccaatt ttggtttctc ctgacccaaa gactttaaat ttaatttatt
3361 tgtccctatt tcaatcaatt gaacaactat ttcgcgaaat gatgagattt ccttcaattt
3421 ttactgctgt tttattcgca gcatcctccg cattagctgc tccagtcaac actacaacag
3481 aagatgaaac ggcacaaatt ccggctgaag ctgtcatcgg ttactcagat ttagaagggg
3541 atttcgatgt tgctgttttg ccattttcca acagcacaaa taacgggtta ttgtttataa
3601 atactactat tgccagcatt gctgctaaag aagaaggggt atctctcgag aaaagagagg
3661 ctgaagctga attcgccaca aaacgtggat ctcccaaccc tacgagggcg gcagcagtta
3721 aggazQcatt ccagacgtcg tggaacgctt accaccattt tgcctttccc catgacgacc
3781 tccacccggt cagcaacagc tttgatgatg agagaaacgg ctggggctcg tcggcaattg
3841 atggcttgga cacggctatc ctcatggggg atgccgacat tgtgaacacg atccttcagt
3901 Qtgtaccgca gatcaacttc atcacgactg cggttgccaa ccaaggatcc tccgtgttcg
3961 agaccaacat tcggtacctc ggtggcctgc tttctgccta tgacctgttg cgaggtcctt
4021 tcagctcctt ggcgacaaac cagaccctgg taaacagcct tctgaggcag gctcaaacac
4081 tggccaaccg cctcaaggtt gcgttcacca ctcccagcgg tgtcccggac cctaccgtct
4141 tcttcaaccc tactgtccgg agaagtggtg catctagcaa caacgtcgct gaaattggsa
4201 gcctggtgct cgagtggaca cggttgagcg acctgacggg aaacccgcag tatgcccagc
4261 ttgcgcagaa gggcgagtcg tatctcctga atccaaaggg aagcccggag gcatggcctg
4321 gcctgattgg Qacgtttgtc agcacgagca acggtacctt tcaggatagc agcggcagct
4381 ggtccggcct catcgacagc ttctacgagt acctgatcaa gatgtacctg tacgacccgg
4441 ttgcgtttgc acactacaag gatcgctggg tccttggtgc cgactcgacc attgggcatc
4501 tcggctctca cccgtcgacg agcaaggact tgaccttttt gtcttcgtac aacggacagt
4561 ctacgtcgcc aQQCtcagga catttggcca gttttggcgg tggcaacttc atcttgggag
4621 gcattctcct gaacgagcaa aagtacattg actttggaat caagcttgcc agctcgtact
4681 ttggcctcgta cacccagacg gcttctggaa tcggccccga aggcttcgcg tgggtggaca
4741 gcgtgacggg cgccggcggc txgccgcccfc cgtcccagtc cgggttctac tcgtcggcag
4801 gattctgggt gacggcaccg tattacatcc tgcggccgga gacgctggag agcttgtact
4861 acgcataccg cgtcacgggc gactccaagt ggtaggacct ggcgtgggaa gcgttgagtg
4921 ccattgagga cgcatgtcgc gccggcagcg cgtactcgtc catcaatgat gtgacgcagg
4981 ccaacggcgg gggtgcctxt gacgatatgg agagcttctg gtttgccgag gcgctcaagt
5041 atgcgtacct gatctttgcg gaggagtcgg atgtgcaggt gcaggccacc ggcgggaaca
5101 aatttgtctt taacacggag gcgcacccct ttagcatccg ttcatcatca cgacggggcg
5161 gccaccttgc tcacgacgag ttgtaatctn gggcGGCCGC CAGCTTGGGC CCGAACAAAA
5221 ACTCATCTCA GAAGAGGATC TGAATAGtGC CGTCGACCAT CATCATCATC ATCATTGAGT
5281 TTTAGCCTTA GAGATGACTG TTCCTCAGTT CAAGTTGGGC AGTTACGAGA AGACCGGTCT
5341 TGCTAGATTC TAATCAAGAG GATGTCAGAA TGCCATTTGC CTGAGAGATG CAGGCTTCAT
5401 TTTTGATACT TTTTTATTTG TMCGTATAT AGTATAGGAT TTTTTTTGTC ATTTTGTTTC
5461 TTCTCGTACG AGCTTGGTCG TCATCAGCCT ATCTCGCAGC TGATGAATAT CTTGTGGTAG
5521 GGGTTTGGGA AAATCATTCG AGTTTGATGT TTTTCTTGGT ATTTCCCACT CCTCTTCAGA
5581 GTACAGAAGA TTAAGTGACA CCTTCGTTTG TGCGGATCCC CGACACACCA TAGCTTCAAA
5641 ATGTTTCTAC TCCTTTTTTA CTCTTCCAGA TTTTCTCGGA CTCCGCGCAT CGCCGTACGA
5701 CTTCAAAACA CCCAAGCACA GCATACTAAA TTTCCCCTCT TTCTTCCTCT AGGGTGTCGT
5761 TAATTACCCG TACTAAAGGT TTGGAAAAGA AAAAAGAGAC CGGCTCGTTT CTTTTTCTTC
5821 GTCGAAAAAG GCAATAAAAA TTTTTATCAC GTTTCTTTTT CTTGAAAATT TTTTTTTTTG
5881 ATTTTTTTCT CTTTCGATGA CCTCCCATTG ATATTTAAGT TAATAAACGG TCTTCAATTT
5941 CTCAAGTTTC AGTTTCATTT TTCTTGTTCT ATTACAACTT TTTTTACTTC TTGCTCATTA
6001 GAAAGAAAGC ATAGCAATCT AATCTAAGGG CGGTGTTGAC AATTAATCAT CGGCATAGTA
6061 TATCGGCATA GTATAATACG ACAAGGTGAG GAACTAAACC ATGGCCAAGC CTTTSTCTCA
6121 AGAAGAATCC ACCCTCATTG AAAGAGCAAC GGCTACAATC AACAGCATCC CCATCTCTGA

6181 AGACTACAGC GTCGCGAGCG CAGCTCTCTC TAGCGACGGC CGCATCTTCA CTGGTGTCAA
6241 TGTATATCAT TTTACTGGGG GACCTTGTGC AGAAGTGGTG GTGCTGGGCA CTGCTGCTGC
6301 TGCGGCAGCT GGGAACCTGA CTTGTATCGT CGCGATCGGA AATGAGAACA GGGGCATCTT
6361 GAGCCCCTGC GGACGGTGCC GACAGGTGCT TCTCGATCTG CATCCTGGGA TCAAAGCCAT
6421 AGTGAAGGAC AGTGATGGAC AGCCGACGGC AGTTGGGATT CGTGAATTGC TGCCCTCTGG
6481 TTATGTGTGG GAGGGCTAAG CACTTCGTGG CCGAGGAGCA GGACTGACAC GTCCGACGCG
6541 GCCCGACGGG TCCGAGGCCT CGGAGATCCG TCCCCCTTTT CCTTTGTCGA Tatcatgtaa
6601 ttagttatgt cacccttaca ttcacgccct ccccccacat ccgctctaac cgaaaaagaa
6661 ggagttagac aacctgaagt ctaggtccct atttattttt ttatagttat gttagtatta
6721 agaacgttat ttatatttca aatttttctt ttttttctgt acagacgcgt gtacgcatgt
6781 aacctttatac tgaaaacctt gcttgagaag gttttgggac gctcgaaggc tttaatttgc
6841 aagctggaga ccaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc
6901 gcgttgctgg cgtttttcca taggctccgc ccccctgacc agcatcacaa aaatcgacgc
6961 tcaaglcaga ggtggcgaaa cccgacagga ctataaagat accacgcgtt tccccctgga
7021 agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt
7081 ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct gtaggtatct cagttcggtg
7141 taggTcgttc gctccaagct gggctgtgtg cacgaacccc cccttcagcc cgaccgctgc
7201 gccttatccg gtaactatcg tcttgagtcc aacccggtaa cacacgactt atcgccattg
7261 gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtcc tacagagttc
7321 ttgaagtggt ggcctaacta cggctacact agaaggacag tatttgctat ctgcgttctg
7331 ctgaagccag ttaccttcgg aaaaagagtt cagcagatta gatccgacaa acaaaccacc
7441 gctggtagcg gtggtttttt tgtttgcaag tctcacgctc cgcgcagaaa aaaaggatct
7501 caagaagatc ctttgatctt ttctacgggg acatccataa agtgcaacca aaactcacgt
7561 taagggattt tggtcatgag atcagatcta ttatcatctc tcctattcgc cgtttctgtc
7621 atttgcgttt tgtacggacc ctcacaacaa ttggaagctc caaaaataga ctatgatcca
7681 ttgacgctcc gatcacttga tttgaagact gatagctctt cttcacagtt gagtccaggc
7741 accgtagaag ataatcttCG AAGACAATTG GAGTTTCATT TTCCTTACCG CAGTTACGAA
7801 CCTTTTCCCC AACATATTT6 GCAAACGTGG AAAGTTTCTC CCTCTGATAG TTCCTTTCCG
7861 AAAAACTTCA AAGACTTAGG T&AAAGTTGG CTGCAAAGGT CCCCAAATTA TGATCATTTT
7921 GTGATACCCG ATGATGCAGC ATGGGAACTT ATTCACCATG AATACGAACG TGTACCAGAA
7981 GTCTTGGAAG CTTTCCACCT GCTACCAGAG CCCATTCTAA AGGCCGATTT TTTCAGGTAT
8041 TTGATTCTTT TTGCCCGTGG AGGACTGTAT GCTGACATGG ACACTATGTT ATTAAAACCA
8101 ATAGAATCGT GGCTGACTTT CAATGAAACT ATTGGTGGAG TAAAAAACAA TGCTGGGTTG
8161 GTCATTGGTA TTGAGGCTGA TCCT6ATA6A CCTGATTGGC ACGACTGGTA TGCTAGAAGG
8221 ATACAATTTT GCCAATGGGC AATTCAGTCC AAACGAGGAC ACCCAGCACT GCGTGAACTG
8281 ATTGTAAGAG TTGTCAGCAC GACTTTACGG AAAGAGAAAA GCGGTTACTT GAACATGGTG
8341 GAAGGAAAGG ATCGTG6AA6 T&ATGT6ATG GACTGGACGG GTCCAGGAAT ATTTACAGAC
8401 ACTCTATTTG ATTATATGAC TAAT6TCAAT ACAACAGGCC ACTCAGGCCA AGGAATTGGA
8461 GCTGGCTCAG CGTATTACAA TGCCTTATCG TTGGAAGAAC GTGATGCCCT CTCTGCKGC
8521 CCGAACGGAG AGATGTTAAA AGAGAAAGTC CCAGGTAAAT ATGCACAGCA GGTTGTTTTA
8581 TGGGAACAAT TTACCAACCT GC6CTCCCCC AAATTAATCG ACGATATTCT TATTCTTCCG
8641 ATCACCAGCT TCAGTCCAGG 6ATT6GCCAC AGTGGAGCTG GAGATTTGAA CCATCACCTT
8701 GCATATATTA GGCATACATT TGAAGGAAGT TGGAAGGACT AAAGAAAGCT AGAGTAAAAT
8761 AGATATAGCG AGATTAGAGA ATGAATACCT TCTTCTAAGC GATCCTCCGT CATCATAGAA
8821 TATCATGGAC TGTATAGTTT TTTTTTTGTA CATATAATGA TTAAACGGTC ATCCAACATC
8881 TCGTTGACAG ATCTCTCAGT ACGCGAAATC CCTGACTATC AAAGCAAGAA CCGATGAAGA
8941 AAAAAACAAC AGTAACCCAA ACACCACAAC AAACACTTTA TCTTCTCCCC CCCAACACCA
9001 ATCATCAAAG AGATGTCGGA ACCAAACACC AAGAAGCAAA AACTAACCCC ATATAAAAAC
9061 ATCCTGGTAG ATAATGCTGG TAACCCGCTC TCCTTCCATA TTCTGGGCTA CTTCACGAAG
9121 TCTGACCGGT CTCAGTTGAT CAACATGATC CTCGAAATGG GTGGGAAGAT CGTTCCAGAC
9181 CTGCCTCCTC TGGTAGATGG AGTGTTGTTT TTGACAGGGG ATTACAAGTC TATTGATGAA
9241 GATACCCTAA AGCAACTGGG GGACGTTCCA ATATACAGAG ACTCCTTCAT CTAGCAGTGT
9301 TTTGTGCACA AGACATCTCT TCCCATTGAC ACTTTCCGAA TTGACAAGAA CGTCGACTTG
9361 GCTCAAGATT TGATCAATAG GGCCCTTCAA GAGTCTGTGG ATCATGTCAC TTCTGCCAGC
9421 ACAGCTGCAG CTGCTGCTGT T6TTGTCGCT ACCAACGGCC TGTCTTCTAA ACCAGACGCT
9481 CGTACTAGCA AAATACAGTT CACTCCCGAA GAAGATCGTT TTATTCTTGA CTTTGTTAGG
9541 AGAAATCCTA AACGAAGAAA CACACATCAA CTCTACACTG AGCTCGCTCA GCACATGAAA
9601 AACCATACGA ATCATTCTAT CCGCCACAGA TTTCGTCGTA ATCTTTCCGC TCAACTTGAT
9661 TGGGTTTATG ATATCGATCC ATTGACCAAC CAACCTCGAA AAGATGAAAA CGGGAACTAG
9721 ATCAAGGTAC AAGATCTTCC ACAA6GAATT CGTGGTCATT ATTCTGCCCA AGATGATTAC
9781 AATTTGTGTT TATC6GTTCA ACCTTTCATT GAATCTGTAG ATGAGACAAC AGGCCAAGAA
9841 TTTTTCAAAC CTCTGAAAG6 TGTATTTGAT GACTTGGAAT CTCGCTTTCC TCACCATACA
9901 AAGACTTCCT GGAGAGACA6 ATTGA&AAAG TTTGCCTCTA AATACGGTGT TCGTCAGTAC
9961 ATCGCGTATT ATGAAAAGAC TGTTGAACTC AATGGTGTTC CTAATCCGAT GACGAACTTT
10021 ACCTCAAAGG CTTCCATTGA AAAATTTAGA GAAAGACGCG GGACTTCACG TAACAGTGGC
10081 CTTCCAGGCC CGGTTGGTGT AGAAGCTGTA AGCTCTTTGG ACCACATATC CCCATTGGTC
10141 ACATCTAATT CCAATTCTGC AGCTGCTGCA GCTGCTGCCG CAGCAGTTGC AGCCTCTGCC

```
10201 TCTGCTTCTT  CAGCTCCTAA  TACTTCMCT   ACCAATTTCT TTGAACAGGA GAATATTGCC
10261 CAAGTTCTCT  CTGCACATAA  CAACGAGCAG  TCTATTGCAG AAGTTATTGA GTCCGCACAG
10321 AATGTCAACA  CCCATGAAAG  TGAACCTATA  GCTGATCATG TTCGAAAAAA TCTTACAGAC
10381 GATGAATTGC  TTGACAAAAT  GGATGATATT  TTAAGCTCCA GAAGTCTAGG CGGACTAGAT
10441 GACTTGATAA  AGATCCTCTA  CACTGAGCTG  GGATTTGCTC ATCGTTATAC CGAATTTCTT
10501 TTTACCTCAT  GTTCTGGTGA  TGTGATTTTC  TTCCGACCAT TAGTGGAACA TTTCCTTCTT
10561 ACTGGTGAGT  GGGAGCTGGA  GAATACTCGT  GGCATCTGGA CCGGTCGTCA AGACGAAATG
18621 CTACGTGCTA  GCAATCTAGA  TGACCTGCAC  AAGTTAATTG ACCTGCATGG GAAAGAACGT
10681 GTTGAGACCA  GAAGAAAAGC  CATCAAGGGA  GAATGATCAT AAGAAATGAA AAACGTATAA
10741 GT
```

Tabla 5. SEQ ID NO: 54 (arriba) y SEQ ID NO: 55 (abajo)

SECUENCIA DE AMINOÁCIDOS

MAKADGSLLY YNPHNPPRRY YFYMAIFAVS VICVLYGPSQ QLSSPKIDYD

PLTLRSLDLKTLEAPSQLSP GTVEDNLRRQ LEFHFPYRSY EPFPQHIWQT

WKVSPSDSSF PKNFKDLGES WLQRSPNYDH FVIPDDAAWE LIHHEYERVP

EVLEALDAHR NAVKVRMEKL GLI

SECUENCIA DE ADN

ATGGCGAAGG CAGATGGCAG TTTGCTCTAC TATAATCCTC ACAATCCACC

CAGAAGGTAT TACTTCTACA TGGCTATATT CGCCGTTTCT GTCATTTGCG

TTTTGTACGG ACCCTCACAA CAATTATCAT CTCCAAAAAT AGACTATGAT

CCATTGACGC TCCGATCACT TGATTTGAAG ACTTTGGAAG CTCCTTCACA

GTTGAGTCCA GGCACCGTAG AAGATAATCT TCGAAGACAA

TTGGAGTTTC ATTTTCCTTA CCGCAGTTAC GAACCTTTTC CCCAACATAT

TTGGCAAACG TGGAAAGTTT CTCCCTCTGA TAGTTCCTTT CCGAAAAACT

TCAAAGACTT AGGTGAAAGT TGGCTGCAAA GGTCCCCAAA TTATGATCAT

TTTGTGATAC CCGATGATGC AGCATGGGAA CTTATTCACC ATGAATACGA

ACGTGTACCA GAAGTCTTGG AAGCTCTAGA TGCTCACCGC

AATGCTGTTA AGGTTCGTAT GGAGAAACTG GGACTTATTT AA

Tabla 6. SEQ ID NO: 56 (arriba) y SEQ ID NO: 57 (abajo)

SECUENCIA DE AMINOÁCIDOS

MRSDLTSIIV FAVSVICVLY GPSQQLSSPK IDYDPLTLRS LDLKTLEAPS
QLSPGTVEDN LRRQLEFHFP YRSYEPFPQH IWQTWKVSPS DSSFPKNFKD
LGESWLQRSP NYDHFVIPDD AAWELIHHEY ERVPEVLEAF HLLPEPILKA
DFFRYLILFA RGGLYADMDT MLLKPIESWLTFNETIGGVK NNAGLVIGIE
ADPDRPDWHD WYARRIQFCQ WAIQSKRGHP ALRELIVRVV
STTLRKEKSG YLNMVEGKDR GSDVMDWTGP GIFTDTLFDY
MTNVNTTGHS GQGIGAGSAY YNALSLEERD ALSARPNGEM LKEKVPGKYA
QQVVLWEQFT NLRSPKLIDD ILILPITSFS PGIGHSGAGD LNHHLAYIRH
TFEGSWKD

SECUENCIA DE ADN

```
  1 atgagatcag atctaacatc cataatcgta ttcgccgttt ctgtcatttg cgttttgtac
 61 ggaccctcac aacaattatc atctccaaaa atagactatg atccattgac gctccgatca
121 cttgatttga agactttgga agctccttca cagttgagtc caggcaccgt agaagataat
```

181 CTTCGAAGAC AATTGGAGTT TCATTTTCCT TACCGCAGTT ACGAACCTTT TCCCCAACAT
241 ATTTGGCAAA CGTGGAAAGT TTCTCCCTCT GATAGTTCCT TTCCGAAAAA CTTCAAAGAC
301 TTAGGTGAAA GTTGGCTGCA AAGGTCCCCA AATTATGATC ATTTTGTGAT ACCCGATGAT
361 GCAGCATGGG AACTTATTCA CCATGAATAC GAACGTGTAC CAGAAGTCTTGGAAGCTTTC
421 CACCTGCTAC CAGAGCCCAT TCTAAAGGCC GATTTTTTCA GGTATTTGAT TCTTTTTGCC
481 CGTGGAGGAC TGTATGCTGA CATGGACACT ATGTTATTAA AACCAATAGA ATCGTGGCTG
541 ACTTTCAATG AAACTATTGG TGGAGTAAAA AACAATGCTG GGTTGGTCAT TGGTATTGAG
601 GCTGATCCTG ATAGACCTGA TTGGCACGAC TGGTATGCTA GAAGGATACA ATTTTGCCAA
661 TGGGCAATTC AGTCCAAACG AGGACACCCA GCACTGCGTG AACTGATTGT AAGAGTTGTC
721 AGCACGACTT TACGGAAAGA GAAAAGCGGT TACTTGAACA TGGTGGAAGG AAAGGATCGT
781 GGAAGTGATG TGATGGACTG GACGGGTCCA GGAATATTTA CAGACACTCT ATTTGATTAT
841 ATGACTAATG TCAATACAAC AGGCCACTCA GGCCAAGGAA TTGGAGCTGG CTCAGCGTAT
901 TACAATGCCT TATCGTTGGA AGAACGTGAT GCCCTCTCTG CCCGCCCGAA CGGAGAGATG
961 TTAAAAGAGA AAGTCCCAGG TAAATATGCA CAGCAGGTTG TTTTATGGGA ACAATTTACC
1021 AACCTGCGCT CCCCCAAATT AATCGACGAT ATTCTTATTC TTCCGATCAC CAGCTTCAGT
1081 CCAGGGATTG GCCACAGTGG AGCTGGAGAT TTGAACCATC ACCTTGCATA TATTAGGCAT
1141 ACATTTGAAG GAAGTTGGAA GGACTAA

LISTADO DE SECUENCIAS

<110> Research Corporation Technologies, Inc.
Gehlsen, Kurt R.
Chappell, Thomas G.

<120> CEPAS DE *PICHIA PASTORIS* PARA LA PRODUCCIÓN DE UNA ESTRUCTURA DE GLICANO PREDOMINANTEMENTE HOMOGÉNEA

<130> 30272

<140> A asignar
<141> 23-10-2013

<150> 61/717.423
<151> 23-10-2012

<160> 93

<170> PatentIn versión 3.5

<210> 1
<211> 9865
<212> ADN
<213> Secuencia artificial

<220>
<223> Nucleótido Sintético

<400> 1

```
aacgtcaaag acagcaatgg agtcaatatt gataacacca ctggcagagc ggttcgtacg      60
tcgttttgga gccgatatga ggctcagcgt gctaacagca cgattgacaa gaagactctc     120
gagtgacagt aggttgagta aagtattcgc ttagattccc aaccttcgtt ttattctttc     180
gtagacaaag aagctgcatg cgaacatagg gacaactttt ataaatccaa ttgtcaaacc     240
aacgtaaaac cctctggcac cattttcaac atatatttgt gaagcagtac gcaatatcga     300
taaatactca ccgttgtttg taacagcccc aacttgcata cgccttctaa tgacctcaaa     360
tggataagcc gcagcttgtg ctaacatacc agcagcaccg cccgcggtca gctgcgccca     420
cacatataaa ggcaatctac gatcatggga ggaattagtt ttgaccgtca ggtcttcaag     480
agttttgaac tcttcttctt gaactgtgta accttttaaa tgacgggatc taaatacgtc     540
atggatgaga tcatgtgtgt aaaaactgac tccagcatat ggaatcattc caaagattgt     600
aggagcgaac ccacgataaa agtttcccaa ccttgccaaa gtgtctaatg ctgtgacttg     660
aaatctgggt tcctcgttga agaccctgcg tactatgccc aaaaactttc ctccacgagc     720
cctattaact tctctatgag tttcaaatgc caaacggaca cggattaggt ccaatgggta     780
agtgaaaaac acagagcaaa ccccagctaa tgagccggcc agtaaccgtc ttggagctgt     840
ttcataagag tcattaggga tcaataacgt tctaatctgt tcataacata caaattttat     900
ggctgcatag ggaaaaattc tcaacagggt agccgaatga ccctgatata gacctgcgac     960
accatcatac ccatagatct gcctgacagc cttaaagagc ccgctaaaag acccggaaaa    1020
```

ccgagagaac tctggattag cagtctgaaa aagaatcttc actctgtcta gtggagcaat 1080
taatgtctta gcggcacttc ctgctactcc gccagctact cctgaataga tcacatactg 1140
caaagactgc ttgtcgatga ccttggggtt atttagcttc aagggcaatt tttgggacat 1200
tttggacaca ggagactcag aaacagacac agagcgttct gagtcctggt gctcctgacg 1260
taggcctaga acaggaatta ttggctttat ttgtttgtcc atttcatagg cttggggtaa 1320
tagatagatg acagagaaat agagaagacc taatattttt tgttcatggc aaatcgcggg 1380
ttcgcggtcg ggtcacacac ggagaagtaa tgagaagagc tggtaatctg gggtaaaagg 1440
gttcaaaaga aggtcgcctg gtagggatgc aatacaaggt tgtcttggag tttacattga 1500
ccagatgatt tggctttttc tctgttcaat tcacattttt cagcgagaat cggattgacg 1560
gagaaatggc ggggtgtggg gtggatagat ggcagaaatg ctcgcaatca ccgcgaaaga 1620
aagactttat ggaatagaac tactgggtgg tgtaaggatt acatagctag tccaatggag 1680
tccgttggaa aggtaagaag aagctaaaac cggctaagta actagggaag aatgatcaga 1740
ctttgatttg atgaggtctg aaaatactct gctgcttttt cagttgcttt ttccctgcaa 1800
cctatcattt tccttttcat aagcctgcct tttctgtttt cacttatatg agttccgccg 1860
agacttcccc aaattctctc ctggaacatt ctctatcgct ctccttccaa gttgcgcccc 1920
ctggcactgc ctagtaatat taccacgcga cttatattca gttccacaat ttccagtgtt 1980
cgtagcaaat atcatcagcc taccgttcgt atagcataca ttatacgaac ggtacttttt 2040
tgtagaaatg tcttggtgtc ctcgtccaat caggtagcca tctctgaaat atctggctcc 2100
gttgcaactc cgaacgacct gctggcaacg taaaattctc cggggtaaaa cttaaatgtg 2160
gagtaatgga accagaaacg tctcttccct tctctctcct tccaccgccc gttaccgtcc 2220
ctaggaaatt ttactctgct ggagagcttc ttctacggcc cccttgcagc aatgctcttc 2280
ccagcattac gttgcgggta aaacggaggt cgtgtacccg acctagcagc ccagggatgg 2340
aaaagtcccg gccgtcgctg gcaataatag cgggcggacg catgtcatga gattattgga 2400
aaccaccaga atcgaatata aaaggcgaac acctttccca attttggttt ctcctgaccc 2460
aaagacttta aatttaattt atttgtccct atttcaatca attgaacaac tatttcgcga 2520
aacgatgaga tttccttcaa tttttactgc tgttttattc gcagcatcct ccgcattagc 2580
tgctccagtc aacactacaa cagaagatga aacggcacaa attccggctg aagctgtcat 2640
cggttactca gatttagaag gggatttcga tgttgctgtt ttgccatttt ccaacagcac 2700
aaataacggg ttattgttta taaatactac tattgccagc attgctgcta aagaagaagg 2760
ggtatctctc gagaaaagag aggctgaagc tgaattcgcc acaaaacgtg gatctcccaa 2820
ccctacgagg gcggcagcag tcaaggccgc attccagacg tcgtggaacg cttaccacca 2880

ttttgccttt ccccatgacg acctccaccc ggtcagcaac agctttgatg atgagagaaa 2940
cggctggggc tcgtcggcaa tcgatggctt ggacacggct atcctcatgg gggatgccga 3000
cattgtgaac acgatccttc agtatgtacc gcagatcaac ttcaccacga ctgcggttgc 3060
caaccaaggc atctccgtgt tcgagaccaa cattcggtac ctcggtggcc tgctttctgc 3120
ctatgacctg ttgcgaggtc ctttcagctc cttggcgaca aaccagaccc tggtaaacag 3180
ccttctgagg caggctcaaa cactggccaa cggcctcaag gttgcgttca ccactcccag 3240
cggtgtcccg gaccctaccg tcttcttcaa ccctactgtc cggagaagtg gtgcatctag 3300
caacaacgtc gctgaaattg gaagcctggt gctcgagtgg acacggttga gcgacctgac 3360
gggaaacccg cagtatgccc agcttgcgca gaagggcgag tcgtatctcc tgaatccaaa 3420
gggaagcccg gaggcatggc ctggcctgat tggaacgttt gtcagcacga gcaacggtac 3480
ctttcaggat agcagcggca gctggtccgg cctcatggac agcttctacg agtacctgat 3540
caagatgtac ctgtacgacc cggttgcgtt tgcacactac aaggatcgct gggtccttgc 3600
tgccgactcg accattgcgc atctcgcctc tcacccgtcg acgcgcaagg acttgacctt 3660
tttgtcttcg tacaacggac agtctacgtc gccaaactca ggacatttgg ccagttttgc 3720
cggtggcaac ttcatcttgg gaggcattct cctgaacgag caaaagtaca ttgactttgg 3780
aatcaagctt gccagctcgt actttgccac gtacaaccag acggcttctg gaatcggccc 3840
cgaaggcttc gcgtgggtgg acagcgtgac gggcgccggc ggctcgccgc cctcgtccca 3900
gtccgggttc tactcgtcgg caggattctg ggtgacggca ccgtattaca tcctgcggcc 3960
ggagacgctg gagagcttgt actacgcata ccgcgtcacg ggcgactcca agtggcagga 4020
cctggcgtgg gaagcgttca gtgccattga ggacgcatgc cgcgccggca gcgcgtactc 4080
gtccatcaac gacgtgacgc aggccaacgg cgggggtgcc tctgacgata tggagagctt 4140
ctggtttgcc gaggcgctca agtatgcgta cctgatcttt gcggaggagt cggatgtgca 4200
ggtgcaggcc aacggcggga acaaatttgt ctttaacacg gaggcgcacc cctttagcat 4260
ccgttcatca tcacgacggg gcggccacct tgctcacgac gagttgtaat ctagggcggc 4320
cgccagcttg ggcccgaaca aaaactcatc tcagaagagg atctgaatag cgccgtcgac 4380
catcatcatc atcatcattg agttttagcc ttagacatga ctgttcctca gttcaagttg 4440
ggcacttacg agaagaccgg tcttgctaga ttctaatcaa gaggatgtca gaatgccatt 4500
tgcctgagag atgcaggctt catttttgat actttttttat ttgtaaccta tatagtatag 4560
gatttttttt gtcattttgt ttcttctcgt acgagcttgc tcctgatcag cctatctcgc 4620
agctgatgaa tatcttgtgg taggggtttg ggaaaatcat tcgagtttga tgtttttctt 4680
ggtatttccc actcctcttc agagtacaga agattaagtg agaccttcgt ttgtgcggat 4740
cccccacaca ccatagcttc aaaatgtttc tactcctttt ttactcttcc agattttctc 4800

ggactccgcg catcgccgta ccacttcaaa acacccaagc acagcatact aaatttcccc 4860

tctttcttcc tctagggtgt cgttaattac ccgtactaaa ggtttggaaa agaaaaaaga 4920

gaccgcctcg tttctttttc ttcgtcgaaa aaggcaataa aaatttttat cacgtttctt 4980

tttcttgaaa atttttttt ttgatttttt tctctttcga tgacctccca ttgatattta 5040

agttaataaa cggtcttcaa tttctcaagt ttcagtttca tttttcttgt tctattacaa 5100

ctttttttac ttcttgctca ttagaaagaa agcatagcaa tctaatctaa gggcggtgtt 5160

gacaattaat catcggcata gtatatcggc atagtataat acgacaaggt gaggaactaa 5220

accatggcca agcctttgtc tcaagaagaa tccaccctca ttgaaagagc aacggctaca 5280

atcaacagca tccccatctc tgaagactac agcgtcgcca gcgcagctct ctctagcgac 5340

ggccgcatct tcactggtgt caatgtatat cattttactg ggggaccttg tgcagaactc 5400

gtggtgctgg gcactgctgc tgctgcggca gctggcaacc tgacttgtat cgtcgcgatc 5460

ggaaatgaga acaggggcat cttgagcccc tgcggacggt gccgacaggt gcttctcgat 5520

ctgcatcctg ggatcaaagc catagtgaag gacagtgatg gacagccgac ggcagttggg 5580

attcgtgaat tgctgccctc tggttatgtg tgggagggct aagcacttcg tggccgagga 5640

gcaggactga cacgtccgac gcggcccgac gggtccgagg cctcggagat ccgtccccct 5700

tttcctttgt cgatatcatg taattagtta tgtcacgctt acattcacgc cctcccccca 5760

catccgctct aaccgaaaag gaaggagtta gacaacctga agtctaggtc cctatttatt 5820

tttttatagt tatgttagta ttaagaacgt tatttatatt tcaaattttt ctttttttc 5880

tgtacagacg cgtgtacgca tgtaacatta tactgaaaac cttgcttgag aaggttttgg 5940

gacgctcgaa ggctttaatt tgcaagctgg agaccaacat gtgagcaaaa ggccagcaaa 6000

aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgcccccctg 6060

acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa 6120

gataccaggc gtttccccct ggaagctccc tcgtgcgctc tcctgttccg accctgccgc 6180

ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac 6240

gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac 6300

cccccgttca gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg 6360

taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt 6420

atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagaa 6480

cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct 6540

cttgatccgg caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga 6600

ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg 6660

ctcagtggaa cgaaaactca cgttaaggga ttttggtcat gagatcagat ctaacatcca 6720
taatcgtatt cgccgtttct gtcatttgcg ttttgtacgg accctcacaa caattatcat 6780
ctccaaaaat agactatgat ccattgacgc tccgatcact tgatttgaag actttggaag 6840
ctccttcaca gttgagtcca ggcaccgtag aagataatct tcgaagacaa ttggagtttc 6900
attttcctta ccgcagttac gaacctttc cccaacatat ttggcaaacg tggaaagttt 6960
ctccctctga tagttccttt ccgaaaaact tcaaagactt aggtgaaagt tggctgcaaa 7020
ggtccccaaa ttatgatcat tttgtgatac ccgatgatgc agcatgggaa cttattcacc 7080
atgaatacga acgtgtacca gaagtcttgg aagctttcca cctgctacca gagcccattc 7140
taaaggccga ttttttcagg tatttgattc tttttgcccg tggaggactg tatgctgaca 7200
tggacactat gttattaaaa ccaatagaat cgtggctgac tttcaatgaa actattggtg 7260
gagtaaaaaa caatgctggg ttggtcattg gtattgaggc tgatcctgat agacctgatt 7320
ggcacgactg gtatgctaga aggatacaat tttgccaatg ggcaattcag tccaaacgag 7380
gacacccagc actgcgtgaa ctgattgtaa gagttgtcag cacgacttta cggaaagaga 7440
aaagcggtta cttgaacatg gtggaaggaa aggatcgtgg aagtgatgtg atggactgga 7500
cgggtccagg aatatttaca gacactctat ttgattatat gactaatgtc aatacaacag 7560
gccactcagg ccaaggaatt ggagctggct cagcgtatta caatgcctta tcgttggaag 7620
aacgtgatgc cctctctgcc cgcccgaacg gagagatgtt aaaagagaaa gtcccaggta 7680
aatatgcaca gcaggttgtt ttatgggaac aatttaccaa cctgcgctcc cccaaattaa 7740
tcgacgatat tcttattctt ccgatcacca gcttcagtcc agggattggc cacagtggag 7800
ctggagattt gaaccatcac cttgcatata ttaggcatac atttgaagga agttggaagg 7860
actaaagaaa gctagagtaa aatagatata gcgagattag agaatgaata ccttcttcta 7920
agcgatcgtc cgtcatcata gaatatcatg gactgtatag tttttttttt gtacatataa 7980
tgattaaacg gtcatccaac atctcgttga cagatctctc agtacgcgaa atccctgact 8040
atcaaagcaa gaaccgatga agaaaaaaac aacagtaacc caaacaccac aacaaacact 8100
ttatcttctc ccccccaaca ccaatcatca aagagatgtc ggaaccaaac accaagaagc 8160
aaaaactaac cccatataaa aacatcctgg tagataatgc tggtaacccg ctctccttcc 8220
atattctggg ctacttcacg aagtctgacc ggtctcagtt gatcaacatg atcctcgaaa 8280
tgggtggcaa gatcgttcca gacctgcctc ctctggtaga tggagtgttg tttttgacag 8340
gggattacaa gtctattgat gaagataccc taaagcaact gggggacgtt ccaatataca 8400
gagactcctt catctaccag tgttttgtgc acaagacatc tcttcccatt gacactttcc 8460
gaattgacaa gaacgtcgac ttggctcaag atttgatcaa tagggccctt caagagtctg 8520
tggatcatgt cacttctgcc agcacagctg cagctgctgc tgttgttgtc gctaccaacg 8580

gcctgtcttc taaaccagac gctcgtacta gcaaaataca gttcactccc gaagaagatc 8640

gttttattct tgactttgtt aggagaaatc ctaaacgaag aaacacacat caactgtaca 8700

ctgagctcgc tcagcacatg aaaaaccata cgaatcattc tatccgccac agatttcgtc 8760

gtaatctttc cgctcaactt gattgggttt atgatatcga tccattgacc aaccaacctc 8820

gaaaagatga aaacgggaac tacatcaagg tacaagatct tccacaagga attcgtggtc 8880

attattctgc ccaagatgat tacaatttgt gtttatcggt tcaacctttc attgaatctg 8940

tagatgagac aacaggccaa gaatttttca aacctctgaa aggtgtattt gatgacttgg 9000

aatctcgctt tcctcaccat acaaagactt cctggagaga cagattcaga aagtttgcct 9060

ctaaatacgg tgttcgtcag tacatcgcgt attatgaaaa gactgttgaa ctcaatggtg 9120

ttcctaatcc gatgacgaac tttacctcaa aggcttccat tgaaaaattt agagaaagac 9180

gcgggacttc acgtaacagt ggccttccag gcccggttgg tgtagaagct gtaagctctt 9240

tggaccacat atccccattg gtcacatcta attccaattc tgcagctgct gcagctgctg 9300

ccgcagcagt tgcagcctct gcctctgctt cttcagctcc taatacttca actaccaatt 9360

tctttgaaca ggagaatatt gcccaagttc tctctgcaca taacaacgag cagtctattg 9420

cagaagttat tgagtccgca cagaatgtca acacccatga aagtgaacct atagctgatc 9480

atgttcgaaa aaatcttaca gacgatgaat tgcttgacaa aatggatgat attttaagct 9540

ccagaagtct aggcggacta gatgacttga taaagatcct ctacactgag ctgggatttg 9600

ctcatcgtta taccgaattt ctttttacct catgttctgg tgatgtgatt ttcttccgac 9660

cattagtgga acatttcctt cttactggtg agtgggagct ggagaatact cgtggcatct 9720

ggaccggtcg tcaagacgaa atgctacgtg ctagcaatct agatgacctg cacaagttaa 9780

ttgacctgca tgggaaagaa cgtgttgaga ccagaagaaa agccatcaag ggagaatgat 9840

cataagaaat gaaaaacgta taagt 9865

<210> 2
<211> 404
<212> PRT
<213> *Pichia pastoris*

<400> 2

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1 5 10 15

Pro Arg Arg Tyr Tyr Phe Tyr Met Ala Ile Phe Ala Val Ser Val Ile
20 25 30

Cys Val Leu Tyr Gly Pro Ser Gln Gln Leu Ser Ser Pro Lys Ile Asp
35 40 45

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

Pro Ser Gln Leu Ser Pro Gly Thr Val Glu Asp Asn Leu Arg Arg Gln
65 70 75 80

Leu Glu Phe His Phe Pro Tyr Arg Ser Tyr Glu Pro Phe Pro Gln His
85 90 95

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

Asn Phe Lys Asp Leu Gly Glu Ser Trp Leu Gln Arg Ser Pro Asn Tyr
115 120 125

Asp His Phe Val Ile Pro Asp Asp Ala Ala Trp Glu Leu Ile His His
130 135 140

Glu Tyr Glu Arg Val Pro Glu Val Leu Glu Ala Phe His Leu Leu Pro
145 150 155 160

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

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

Glu Ser Trp Leu Thr Phe Asn Glu Thr Ile Gly Gly Val Lys Asn Asn
195 200 205

Ala Gly Leu Val Ile Gly Ile Glu Ala Asp Pro Asp Arg Pro Asp Trp
210 215 220

His Asp Trp Tyr Ala Arg Arg Ile Gln Phe Cys Gln Trp Ala Ile Gln
225 230 235 240

Ser Lys Arg Gly His Pro Ala Leu Arg Glu Leu Ile Val Arg Val Val
245 250 255

Ser Thr Thr Leu Arg Lys Glu Lys Ser Gly Tyr Leu Asn Met Val Glu
260 265 270

Gly Lys Asp Arg Gly Ser Asp Val Met Asp Trp Thr Gly Pro Gly Ile
275 280 285

Phe Thr Asp Thr Leu Phe Asp Tyr Met Thr Asn Val Asn Thr Thr Gly

290 295 300

His Ser Gly Gln Gly Ile Gly Ala Gly Ser Ala Tyr Tyr Asn Ala Leu
305 310 315 320

Ser Leu Glu Glu Arg Asp Ala Leu Ser Ala Arg Pro Asn Gly Glu Met
325 330 335

Leu Lys Glu Lys Val Pro Gly Lys Tyr Ala Gln Gln Val Val Leu Trp
340 345 350

Glu Gln Phe Thr Asn Leu Arg Ser Pro Lys Leu Ile Asp Asp Ile Leu
355 360 365

Ile Leu Pro Ile Thr Ser Phe Ser Pro Gly Ile Gly His Ser Gly Ala
370 375 380

Gly Asp Leu Asn His His Leu Ala Tyr Ile Arg His Thr Phe Glu Gly
385 390 395 400

Ser Trp Lys Asp

<210> 3
<211> 388
<212> PRT
<213> *Pichia pastoris*

<400> 3

Met Arg Ser Asp Leu Thr Ser Ile Ile Val Phe Ala Val Ser Val Ile
1 5 10 15

Cys Val Leu Tyr Gly Pro Ser Gln Gln Leu Ser Ser Pro Lys Ile Asp
20 25 30

Tyr Asp Pro Leu Thr Leu Arg Ser Leu Asp Leu Lys Thr Leu Glu Ala
35 40 45

Pro Ser Gln Leu Ser Pro Gly Thr Val Glu Asp Asn Leu Arg Arg Gln
50 55 60

Leu Glu Phe His Phe Pro Tyr Arg Ser Tyr Glu Pro Phe Pro Gln His
65 70 75 80

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

Asn Phe Lys Asp Leu Gly Glu Ser Trp Leu Gln Arg Ser Pro Asn Tyr

100 105 110

Asp His Phe Val Ile Pro Asp Asp Ala Ala Trp Glu Leu Ile His His
115 120 125

Glu Tyr Glu Arg Val Pro Glu Val Leu Glu Ala Phe His Leu Leu Pro
130 135 140

Glu Pro Ile Leu Lys Ala Asp Phe Phe Arg Tyr Leu Ile Leu Phe Ala
145 150 155 160

Arg Gly Gly Leu Tyr Ala Asp Met Asp Thr Met Leu Leu Lys Pro Ile
165 170 175

Glu Ser Trp Leu Thr Phe Asn Glu Thr Ile Gly Gly Val Lys Asn Asn
180 185 190

Ala Gly Leu Val Ile Gly Ile Glu Ala Asp Pro Asp Arg Pro Asp Trp
195 200 205

His Asp Trp Tyr Ala Arg Arg Ile Gln Phe Cys Gln Trp Ala Ile Gln
210 215 220

Ser Lys Arg Gly His Pro Ala Leu Arg Glu Leu Ile Val Arg Val Val
225 230 235 240

Ser Thr Thr Leu Arg Lys Glu Lys Ser Gly Tyr Leu Asn Met Val Glu
245 250 255

Gly Lys Asp Arg Gly Ser Asp Val Met Asp Trp Thr Gly Pro Gly Ile
260 265 270

Phe Thr Asp Thr Leu Phe Asp Tyr Met Thr Asn Val Asn Thr Thr Gly
275 280 285

His Ser Gly Gln Gly Ile Gly Ala Gly Ser Ala Tyr Tyr Asn Ala Leu
290 295 300

Ser Leu Glu Glu Arg Asp Ala Leu Ser Ala Arg Pro Asn Gly Glu Met
305 310 315 320

Leu Lys Glu Lys Val Pro Gly Lys Tyr Ala Gln Gln Val Val Leu Trp
325 330 335

Glu Gln Phe Thr Asn Leu Arg Ser Pro Lys Leu Ile Asp Asp Ile Leu
340 345 350

Ile Leu Pro Ile Thr Ser Phe Ser Pro Gly Ile Gly His Ser Gly Ala
355 360 365

Gly Asp Leu Asn His His Leu Ala Tyr Ile Arg His Thr Phe Glu Gly
370 375 380

Ser Trp Lys Asp
385

<210> 4
<211> 5
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 4

Ala Lys Ala Asp Gly
1 5

<210> 5
<211> 6
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 5

Ala Lys Ala Asp Gly Ser
1 5

<210> 6
<211> 7
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 6

Ala Lys Ala Asp Gly Ser Leu
1 5

<210> 7
<211> 8
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 7

Ala Lys Ala Asp Gly Ser Leu Leu
1 5

<210> 8
<211> 9
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 8

Ala Lys Ala Asp Gly Ser Leu Leu Tyr
1 5

<210> 9
<211> 10
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 9

Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr
1 5 10

<210> 10
<211> 6
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 10

Met Ala Lys Ala Asp Gly
1 5

<210> 11
<211> 7
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 11

Met Ala Lys Ala Asp Gly Ser
1 5

<210> 12
<211> 8
<212> PRT
<213> Secuencia artificial

<220>

<223> Péptido sintético

<400> 12

Met Ala Lys Ala Asp Gly Ser Leu
1 5

<210> 13
<211> 9
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 13

Met Ala Lys Ala Asp Gly Ser Leu Leu
1 5

<210> 14
<211> 10
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 14

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr
1 5 10

<210> 15
<211> 11
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 15

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr
1 5 10

<210> 16
<211> 12
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 16

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn
1 5 10

<210> 17
<211> 13

<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 17

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro
1               5                   10
```

<210> 18
<211> 14
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 18

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His
1               5                   10
```

<210> 19
<211> 15
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 19

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn
1               5                   10                  15
```

<210> 20
<211> 16
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 20

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1               5                   10                  15
```

<210> 21
<211> 17
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 21

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1 5 10 15

Pro

<210> 22
<211> 18
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 22

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1 5 10 15

Pro Arg

<210> 23
<211> 19
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 23

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1 5 10 15

Pro Arg Arg

<210> 24
<211> 20
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 24

Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro

1 5 10 15

Pro Arg Arg Tyr
20

<210> 25
<211> 21
<212> PRT

<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 25

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1               5                   10                  15

Pro Arg Arg Tyr Tyr
            20
```

<210> 26
<211> 24
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 26

```
Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro Pro
1               5                   10                  15

Arg Arg Tyr Tyr Phe Tyr Met Ala
            20
```

<210> 27
<211> 25
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 27

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1               5                   10                  15

Pro Arg Arg Tyr Tyr Phe Tyr Met Ala
            20                  25
```

<210> 28
<211> 9
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 28

```
Arg Ser Asp Leu Thr Ser Ile Ile Val
1               5
```

<210> 29
<211> 15
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 29
gcgaaggcag atggc 15

<210> 30
<211> 18
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 30
gcgaaggcag atggcagt 18

<210> 31
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 31
gcgaaggcag atggcagttt g 21

<210> 32
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 32
gcgaaggcag atggcagttt gctc 24

<210> 33
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 33
gcgaaggcag atggcagttt gctctac 27

<210> 34
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 34
gcgaaggcag atggcagttt gctctactat 30

<210> 35
<211> 18
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 35
atggcgaagg cagatggc 18

<210> 36
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 36
atggcgaagg cagatggcag t 21

<210> 37
<211> 24
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 37
atggcgaagg cagatggcag tttg 24

<210> 38
<211> 27
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 38
atggcgaagg cagatggcag tttgctc 27

<210> 39
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 39
atggcgaagg cagatggcag tttgctctac 30

<210> 40
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 40
atggcgaagg cagatggcag tttgctctac 30

<210> 41
<211> 36
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 41
atggcgaagg cagatggcag tttgctctac tataat 36

<210> 42
<211> 39
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 42
atggcgaagg cagatggcag tttgctctac tataatcct 39

<210> 43
<211> 42
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 43
atggcgaagg cagatggcag tttgctctac tataatcctc ac 42

<210> 44
<211> 42
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 44
atggcgaagg cagatggcag tttgctctac tataatcctc ac 42

<210> 45
<211> 48
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 45
atggcgaagg cagatggcag tttgctctac tataatcctc acaatcca 48

<210> 46
<211> 51
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 46
atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc c 51

<210> 47
<211> 54
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 47
atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc caga 54

<210> 48
<211> 57
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 48
atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc cagaagg 57

<210> 49
<211> 60
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 49
atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc cagaaggtat 60

<210> 50
<211> 63
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 50
atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc cagaaggtat 60

tac 63

<210> 51
<211> 73
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 51
gcgaaggcag atggcagttt gctctactat aatcctcaca atccacccag aaggtattac 60

ttctacatgg cta 73

<210> 52
<211> 76
<212> ADN
<213> Secuencia artificial

<220>

<223> Oligonucleótido sintético

<400> 52

atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc cagaaggtat 60

tacttctaca tggcta 76

<210> 53
<211> 10742
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 53

aacgtcaaag acagcaatgg agtcaatatt gataacacca ctggcagagc ggttcgtacg 60
tcgttttgga gccgatatga ggctcagcgt gctaacagca cgattgacaa gaagactctc 120
gagtgacagt aggttgagta aagtattcgc ttagattccc aaccttcgtt ttattctttc 180
gtagacaaag aagctgcatg cgaacatagg gacaactttt ataaatccaa ttgtcaaacc 240
aacgtaaaac cctctggcac cattttcaac atatatttgt gaagcagtac gcaatatcga 300
taaatactca ccgttgtttg taacagcccc aacttgcata cgccttctaa tgacctcaaa 360
tggataagcc gcagcttgtg ctaacatacc agcagcaccg cccgcggtca gctgcgccca 420
cacatataaa ggcaatctac gatcatggga ggaattagtt ttgaccgtca ggtcttcaag 480
agttttgaac tcttcttctt gaactgtgta accttttaaa tgacgggatc taaatacgtc 540
atggatgaga tcatgtgtgt aaaaactgac tccagcatat ggaatcattc caaagattgt 600
aggagcgaac ccacgataaa agtttcccaa ccttgccaaa gtgtctaatg ctgtgacttg 660
aaatctgggt tcctcgttga agaccctgcg tactatgccc aaaaactttc ctccacgagc 720
cctattaact tctctatgag tttcaaatgc caaacggaca cggattaggt ccaatgggta 780
agtgaaaaac acagagcaaa ccccagctaa tgagccggcc agtaaccgtc ttggagctgt 840
ttcataagag tcattaggga tcaataacgt tctaatctgt tcataacata caaattttat 900
ggctgcatag ggaaaaattc tcaacagggt agccgaatga ccctgatata gacctgcgac 960
accatcatac ccatagatct gcctgacagc cttaaagagc ccgctaaaag acccggaaaa 1020
ccgagagaac tctggattag cagtctgaaa aagaatcttc actctgtcta gtggagcaat 1080
taatgtctta gcggcacttc ctgctactcc gccagctact cctgaataga tcacatactg 1140
caaagactgc ttgtcgatga ccttggggtt atttagcttc aagggcaatt tttgggacat 1200
tttggacaca ggagactcag aaacagacac agagcgttct gagtcctggt gctcctgacg 1260
taggcctaga acaggaatta ttggctttat ttgtttgtcc atttcatagg cttggggtaa 1320
tagatagatg acagagaaat agagaagacc taatattttt tgttcatggc aaatcgcggg 1380
ttcgcggtcg ggtcacacac ggagaagtaa tgagaagagc tggtaatctg gggtaaaagg 1440
gttcaaaaga aggtcgcctg gtagggatgc aatacaaggt tgtcttggag tttacattga 1500
ccagatgatt tggctttttc tctgttcaat tcacattttt cagcgagaat cggattgacg 1560
gagaaatggc ggggtgtggg gtggatagat ggcagaaatg ctcgcaatca ccgcgaaaga 1620
aagactttat ggaatagaac tactgggtgg tgtaaggatt acatagctag tccaatggag 1680
tccgttggaa aggtaagaag aagctaaaac cggctaagta actagggaag aatgatcaga 1740
ctttgatttg atgaggtctg aaaatactct gctgcttttt cagttgcttt ttccctgcaa 1800
cctatcattt tccttttcat aagcctgcct tttctgtttt cacttatatg agttccgccg 1860
agacttcccc aaattctctc ctggaacatt ctctatcgct ctccttccaa gttgcgcccc 1920

ctggcactgc ctagtaatat taccacgcga cttatattca gttccacaat ttccagtgtt 1980
cgtagcaaat atcatcagcc atggcgaagg cagatggcag tttgctctac tataatcctc 2040
acaatccacc cagaaggtat tacttctaca tggctatatt cgccgtttct gtcatttgcg 2100
ttttgtacgg accctcacaa caattatcat ctccaaaaat agactatgat ccattgacgc 2160
tccgatcact tgatttgaag actttggaag ctccttcaca gttgagtcca ggcaccgtag 2220
aagataatct tcgaagacaa ttggagtttc attttcctta ccgcagttac gaacctttc 2280
cccaacatat ttggcaaacg tggaaagttt ctccctctga tagttccttt ccgaaaaact 2340
tcaaagactt aggtgaaagt tggctgcaaa ggtccccaaa ttatgatcat tttgtgatac 2400
ccgatgatgc agcatgggaa cttattcacc atgaatacga acgtgtacca gaagtcttgg 2460
aagctctaga tgctcaccgc aatgctgtta aggttcgtat ggagaaactg ggacttattt 2520
aattatttag agattttaac ttacatttag attcgataga tccacaggac gggtgtggtc 2580
gccatgatcg cgtagtcgat agtggctcca agtagcgaag cgagcaggac tgggcggcgg 2640
ccaaagcggt cggacagtgc tccgagaacg ggtgcgcata gaaattgcat caacgcatat 2700
agcgctagca gcacgccata gtgactggcg atgctgtcgg aatggacgat atcccgcaag 2760
aggcccggca gtaccggcat aaccaagcct atgcctacag catccagggt gacggtgccg 2820
aggatgacga tgagcgcatt gttagatttc atacacggtg cctgactgcg ttagcaattt 2880
aactgtgata aactaccgca ttaaagctga tctttttgt agaaatgtct tggtgtcctc 2940
gtccaatcag gtagccatct ctgaaatatc tggctccgtt gcaactccga acgacctgct 3000
ggcaacgtaa aattctccgg ggtaaaactt aaatgtggag taatggaacc agaaacgtct 3060
cttcccttct ctctccttcc accgcccgtt accgtcccta ggaaatttta ctctgctgga 3120
gagcttcttc tacggccccc ttgcagcaat gctcttccca gcattacgtt gcgggtaaaa 3180
cggaggtcgt gtacccgacc tagcagccca gggatggaaa agtcccggcc gtcgctggca 3240
ataatagcgg gcggacgcat gtcatgagat tattggaaac caccagaatc gaatataaaa 3300
ggcgaacacc tttcccaatt ttggtttctc ctgacccaaa gactttaaat ttaatttatt 3360
tgtccctatt tcaatcaatt gaacaactat ttcgcgaaac gatgagattt ccttcaattt 3420
ttactgctgt tttattcgca gcatcctccg cattagctgc tccagtcaac actacaacag 3480
aagatgaaac ggcacaaatt ccggctgaag ctgtcatcgg ttactcagat ttagaagggg 3540
atttcgatgt tgctgttttg ccattttcca acagcacaaa taacgggtta ttgtttataa 3600
atactactat tgccagcatt gctgctaaag aagaaggggt atctctcgag aaaagagagg 3660
ctgaagctga attcgccaca aaacgtggat ctcccaaccc tacgagggcg gcagcagtca 3720
aggccgcatt ccagacgtcg tggaacgctt accaccattt tgcctttccc catgacgacc 3780

tccacccggt cagcaacagc tttgatgatg agagaaacgg ctggggctcg tcggcaatcg 3840
atggcttgga cacggctatc ctcatggggg atgccgacat tgtgaacacg atccttcagt 3900
atgtaccgca gatcaacttc accacgactg cggttgccaa ccaaggatcc tccgtgttcg 3960
agaccaacat tcggtacctc ggtggcctgc tttctgccta tgacctgttg cgaggtcctt 4020
tcagctcctt ggcgacaaac cagaccctgg taaacagcct tctgaggcag gctcaaacac 4080
tggccaacgg cctcaaggtt gcgttcacca ctcccagcgg tgtcccggac cctaccgtct 4140
tcttcaaccc tactgtccgg agaagtggtg catctagcaa caacgtcgct gaaattggaa 4200
gcctggtgct cgagtggaca cggttgagcg acctgacggg aaacccgcag tatgcccagc 4260
ttgcgcagaa gggcgagtcg tatctcctga atccaaaggg aagcccggag gcatggcctg 4320
gcctgattgg aacgtttgtc agcacgagca acggtacctt tcaggatagc agcggcagct 4380
ggtccggcct catggacagc ttctacgagt acctgatcaa gatgtacctg tacgacccgg 4440
ttgcgtttgc acactacaag gatcgctggg tccttggtgc cgactcgacc attgggcatc 4500
tcggctctca cccgtcgacg cgcaaggact tgaccttttt gtcttcgtac aacggacagt 4560
ctacgtcgcc aaactcagga catttggcca gttttggcgg tggcaacttc atcttgggag 4620
gcattctcct gaacgagcaa aagtacattg actttggaat caagcttgcc agctcgtact 4680
ttggcacgta cacccagacg gcttctggaa tcggccccga aggcttcgcg tgggtggaca 4740
gcgtgacggg cgccggcggc tcgccgccct cgtcccagtc cgggttctac tcgtcggcag 4800
gattctgggt gacggcaccg tattacatcc tgcggccgga gacgctggag agcttgtact 4860
acgcataccg cgtcacgggc gactccaagt ggcaggacct ggcgtgggaa gcgttgagtg 4920
ccattgagga cgcatgccgc gccggcagcg cgtactcgtc catcaacgac gtgacgcagg 4980
ccaacggcgg gggtgcctct gacgatatgg agagcttctg gtttgccgag gcgctcaagt 5040
atgcgtacct gatctttgcg gaggagtcgg atgtgcaggt gcaggccacc ggcgggaaca 5100
aatttgtctt taacacggag gcgcacccct ttagcatccg ttcatcatca cgacggggcg 5160
gccaccttgc tcacgacgag ttgtaatcta gggcggccgc cagcttgggc ccgaacaaaa 5220
actcatctca gaagaggatc tgaatagcgc cgtcgaccat catcatcatc atcattgagt 5280
tttagcctta gacatgactg ttcctcagtt caagttgggc acttacgaga agaccggtct 5340
tgctagattc taatcaagag gatgtcagaa tgccatttgc ctgagagatg caggcttcat 5400
ttttgatact tttttatttg taacctatat agtataggat tttttttgtc attttgtttc 5460
ttctcgtacg agcttgctcc tgatcagcct atctcgcagc tgatgaatat cttgtggtag 5520
gggtttggga aaatcattcg agtttgatgt ttttcttggt atttcccact cctcttcaga 5580
gtacagaaga ttaagtgaga ccttcgtttg tgcggatccc ccacacacca tagcttcaaa 5640
atgtttctac tccttttttta ctcttccaga ttttctcgga ctccgcgcat cgccgtacca 5700

cttcaaaaca cccaagcaca gcatactaaa tttcccctct ttcttcctct agggtgtcgt 5760

taattacccg tactaaaggt ttggaaaaga aaaaagagac cgcctcgttt ctttttcttc 5820

gtcgaaaaag gcaataaaaa tttttatcac gtttcttttt cttgaaaatt tttttttttg 5880

atttttttct ctttcgatga cctcccattg atatttaagt taataaacgg tcttcaattt 5940

ctcaagtttc agtttcattt ttcttgttct attacaactt tttttacttc ttgctcatta 6000

gaaagaaagc atagcaatct aatctaaggg cggtgttgac aattaatcat cggcatagta 6060

tatcggcata gtataatacg acaaggtgag gaactaaacc atggccaagc ctttgtctca 6120

agaagaatcc accctcattg aaagagcaac ggctacaatc aacagcatcc ccatctctga 6180

agactacagc gtcgccagcg cagctctctc tagcgacggc cgcatcttca ctggtgtcaa 6240

tgtatatcat tttactgggg gaccttgtgc agaactcgtg gtgctgggca ctgctgctgc 6300

tgcggcagct ggcaacctga cttgtatcgt cgcgatcgga aatgagaaca ggggcatctt 6360

gagcccctgc ggacggtgcc gacaggtgct tctcgatctg catcctggga tcaaagccat 6420

agtgaaggac agtgatggac agccgacggc agttgggatt cgtgaattgc tgccctctgg 6480

ttatgtgtgg gagggctaag cacttcgtgg ccgaggagca ggactgacac gtccgacgcg 6540

gcccgacggg tccgaggcct cggagatccg tccccctttt cctttgtcga tatcatgtaa 6600

ttagttatgt cacgcttaca ttcacgccct ccccccacat ccgctctaac cgaaaaggaa 6660

ggagttagac aacctgaagt ctaggtccct atttattttt ttatagttat gttagtatta 6720

agaacgttat ttatatttca aatttttctt ttttttctgt acagacgcgt gtacgcatgt 6780

aacattatac tgaaaacctt gcttgagaag gttttgggac gctcgaaggc tttaatttgc 6840

aagctggaga ccaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc 6900

gcgttgctgg cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc 6960

tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga 7020

agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt 7080

ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct gtaggtatct cagttcggtg 7140

taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc 7200

gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg 7260

gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc 7320

ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg 7380

ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc 7440

gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct 7500

caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt 7560

taagggattt tggtcatgag atcagatcta acatccataa tcgtattcgc cgtttctgtc 7620
atttgcgttt tgtacggacc ctcacaacaa ttatcatctc caaaaataga ctatgatcca 7680
ttgacgctcc gatcacttga tttgaagact ttggaagctc cttcacagtt gagtccaggc 7740
accgtagaag ataatcttcg aagacaattg gagtttcatt ttccttaccg cagttacgaa 7800
ccttttcccc aacatatttg gcaaacgtgg aaagtttctc cctctgatag ttcctttccg 7860
aaaaacttca aagacttagg tgaaagttgg ctgcaaaggt ccccaaatta tgatcatttt 7920
gtgatacccg atgatgcagc atgggaactt attcaccatg aatacgaacg tgtaccagaa 7980
gtcttggaag ctttccacct gctaccagag cccattctaa aggccgattt tttcaggtat 8040
ttgattcttt ttgcccgtgg aggactgtat gctgacatgg acactatgtt attaaaacca 8100
atagaatcgt ggctgacttt caatgaaact attggtggag taaaaaacaa tgctgggttg 8160
gtcattggta ttgaggctga tcctgataga cctgattggc acgactggta tgctagaagg 8220
atacaatttt gccaatgggc aattcagtcc aaacgaggac acccagcact gcgtgaactg 8280
attgtaagag ttgtcagcac gactttacgg aaagagaaaa gcggttactt gaacatggtg 8340
gaaggaaagg atcgtggaag tgatgtgatg gactggacgg gtccaggaat atttacagac 8400
actctatttg attatatgac taatgtcaat acaacaggcc actcaggcca aggaattgga 8460
gctggctcag cgtattacaa tgccttatcg ttggaagaac gtgatgccct ctctgcccgc 8520
ccgaacggag agatgttaaa agagaaagtc ccaggtaaat atgcacagca ggttgtttta 8580
tgggaacaat ttaccaacct gcgctccccc aaattaatcg acgatattct tattcttccg 8640
atcaccagct tcagtccagg gattggccac agtggagctg gagatttgaa ccatcacctt 8700
gcatatatta ggcatacatt tgaaggaagt tggaaggact aaagaaagct agagtaaaat 8760
agatatagcg agattagaga atgaatacct tcttctaagc gatcgtccgt catcatagaa 8820
tatcatggac tgtatagttt tttttttgta catataatga ttaaacggtc atccaacatc 8880
tcgttgacag atctctcagt acgcgaaatc cctgactatc aaagcaagaa ccgatgaaga 8940
aaaaaacaac agtaacccaa acaccacaac aaacacttta tcttctcccc cccaacacca 9000
atcatcaaag agatgtcgga accaaacacc aagaagcaaa aactaacccc atataaaaac 9060
atcctggtag ataatgctgg taacccgctc tccttccata ttctgggcta cttcacgaag 9120
tctgaccggt ctcagttgat caacatgatc ctcgaaatgg gtggcaagat cgttccagac 9180
ctgcctcctc tggtagatgg agtgttgttt ttgacagggg attacaagtc tattgatgaa 9240
gataccctaa agcaactggg ggacgttcca atatacagag actccttcat ctaccagtgt 9300
tttgtgcaca agacatctct tcccattgac actttccgaa ttgacaagaa cgtcgacttg 9360
gctcaagatt tgatcaatag ggcccttcaa gagtctgtgg atcatgtcac ttctgccagc 9420
acagctgcag ctgctgctgt tgttgtcgct accaacggcc tgtcttctaa accagacgct 9480

```
cgtactagca aaatacagtt cactcccgaa gaagatcgtt ttattcttga ctttgttagg     9540
agaaatccta aacgaagaaa cacacatcaa ctgtacactg agctcgctca gcacatgaaa     9600
aaccatacga atcattctat ccgccacaga tttcgtcgta atctttccgc tcaacttgat     9660
tgggtttatg atatcgatcc attgaccaac caacctcgaa aagatgaaaa cgggaactac     9720
atcaaggtac aagatcttcc acaaggaatt cgtggtcatt attctgccca agatgattac     9780
aatttgtgtt tatcggttca acctttcatt gaatctgtag atgagacaac aggccaagaa     9840
tttttcaaac ctctgaaagg tgtatttgat gacttggaat ctcgctttcc tcaccataca     9900
aagacttcct ggagagacag attcagaaag tttgcctcta aatacggtgt tcgtcagtac     9960
atcgcgtatt atgaaaagac tgttgaactc aatggtgttc ctaatccgat gacgaacttt    10020
acctcaaagg cttccattga aaaatttaga gaaagacgcg ggacttcacg taacagtggc    10080
cttccaggcc cggttggtgt agaagctgta agctctttgg accacatatc cccattggtc    10140
acatctaatt ccaattctgc agctgctgca gctgctgccg cagcagttgc agcctctgcc    10200
tctgcttctt cagctcctaa tacttcaact accaatttct ttgaacagga gaatattgcc    10260
caagttctct ctgcacataa caacgagcag tctattgcag aagttattga gtccgcacag    10320
aatgtcaaca cccatgaaag tgaacctata gctgatcatg ttcgaaaaaa tcttacagac    10380
gatgaattgc ttgacaaaat ggatgatatt ttaagctcca gaagtctagg cggactagat    10440
gacttgataa agatcctcta cactgagctg ggatttgctc atcgttatac cgaatttctt    10500
tttacctcat gttctggtga tgtgattttc ttccgaccat tagtggaaca tttccttctt    10560
actggtgagt gggagctgga gaatactcgt ggcatctgga ccggtcgtca agacgaaatg    10620
ctacgtgcta gcaatctaga tgacctgcac aagttaattg acctgcatgg gaaagaacgt    10680
gttgagacca gaagaaaagc catcaaggga gaatgatcat aagaaatgaa aaacgtataa    10740
gt                                                                   10742
```

<210> 54
<211> 173
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 54

```
Met Ala Lys Ala Asp Gly Ser Leu Leu Tyr Tyr Asn Pro His Asn Pro
1               5                   10                  15

Pro Arg Arg Tyr Tyr Phe Tyr Met Ala Ile Phe Ala Val Ser Val Ile
            20                  25                  30
```

Cys Val Leu Tyr Gly Pro Ser Gln Gln Leu Ser Ser Pro Lys Ile Asp
35 40 45

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

Pro Ser Gln Leu Ser Pro Gly Thr Val Glu Asp Asn Leu Arg Arg Gln
65 70 75 80

Leu Glu Phe His Phe Pro Tyr Arg Ser Tyr Glu Pro Phe Pro Gln His
85 90 95

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

Asn Phe Lys Asp Leu Gly Glu Ser Trp Leu Gln Arg Ser Pro Asn Tyr
115 120 125

Asp His Phe Val Ile Pro Asp Asp Ala Ala Trp Glu Leu Ile His His
130 135 140

Glu Tyr Glu Arg Val Pro Glu Val Leu Glu Ala Leu Asp Ala His Arg
145 150 155 160

Asn Ala Val Lys Val Arg Met Glu Lys Leu Gly Leu Ile
165 170

<210> 55
<211> 522
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 55

```
atggcgaagg cagatggcag tttgctctac tataatcctc acaatccacc cagaaggtat      60
tacttctaca tggctatatt cgccgtttct gtcatttgcg ttttgtacgg accctcacaa     120
caattatcat ctccaaaaat agactatgat ccattgacgc tccgatcact tgatttgaag     180
actttggaag ctccttcaca gttgagtcca ggcaccgtag aagataatct tcgaagacaa     240
ttggagtttc attttcctta ccgcagttac gaaccttttc cccaacatat ttggcaaacg     300
tggaaagttt ctccctctga tagttccttt ccgaaaaact tcaaagactt aggtgaaagt     360
tggctgcaaa ggtccccaaa ttatgatcat tttgtgatac ccgatgatgc agcatgggaa     420
cttattcacc atgaatacga acgtgtacca gaagtcttgg aagctctaga tgctcaccgc     480
aatgctgtta aggttcgtat ggagaaactg ggacttattt aa                        522
```

<210> 56
<211> 388
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 56

Met Arg Ser Asp Leu Thr Ser Ile Ile Val Phe Ala Val Ser Val Ile
1 5 10 15

Cys Val Leu Tyr Gly Pro Ser Gln Gln Leu Ser Ser Pro Lys Ile Asp
20 25 30

Tyr Asp Pro Leu Thr Leu Arg Ser Leu Asp Leu Lys Thr Leu Glu Ala
35 40 45

Pro Ser Gln Leu Ser Pro Gly Thr Val Glu Asp Asn Leu Arg Arg Gln
50 55 60

Leu Glu Phe His Phe Pro Tyr Arg Ser Tyr Glu Pro Phe Pro Gln His
65 70 75 80

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

Asn Phe Lys Asp Leu Gly Glu Ser Trp Leu Gln Arg Ser Pro Asn Tyr
100 105 110

Asp His Phe Val Ile Pro Asp Asp Ala Ala Trp Glu Leu Ile His His
115 120 125

Glu Tyr Glu Arg Val Pro Glu Val Leu Glu Ala Phe His Leu Leu Pro
130 135 140

Glu Pro Ile Leu Lys Ala Asp Phe Phe Arg Tyr Leu Ile Leu Phe Ala
145 150 155 160

Arg Gly Gly Leu Tyr Ala Asp Met Asp Thr Met Leu Leu Lys Pro Ile
165 170 175

Glu Ser Trp Leu Thr Phe Asn Glu Thr Ile Gly Gly Val Lys Asn Asn
180 185 190

Ala Gly Leu Val Ile Gly Ile Glu Ala Asp Pro Asp Arg Pro Asp Trp
195 200 205

His Asp Trp Tyr Ala Arg Arg Ile Gln Phe Cys Gln Trp Ala Ile Gln
210 215 220

Ser Lys Arg Gly His Pro Ala Leu Arg Glu Leu Ile Val Arg Val Val
225 230 235 240

Ser Thr Thr Leu Arg Lys Glu Lys Ser Gly Tyr Leu Asn Met Val Glu
245 250 255

Gly Lys Asp Arg Gly Ser Asp Val Met Asp Trp Thr Gly Pro Gly Ile
260 265 270

Phe Thr Asp Thr Leu Phe Asp Tyr Met Thr Asn Val Asn Thr Thr Gly
275 280 285

His Ser Gly Gln Gly Ile Gly Ala Gly Ser Ala Tyr Tyr Asn Ala Leu
290 295 300

Ser Leu Glu Glu Arg Asp Ala Leu Ser Ala Arg Pro Asn Gly Glu Met
305 310 315 320

Leu Lys Glu Lys Val Pro Gly Lys Tyr Ala Gln Gln Val Val Leu Trp
325 330 335

Glu Gln Phe Thr Asn Leu Arg Ser Pro Lys Leu Ile Asp Asp Ile Leu
340 345 350

Ile Leu Pro Ile Thr Ser Phe Ser Pro Gly Ile Gly His Ser Gly Ala
355 360 365

Gly Asp Leu Asn His His Leu Ala Tyr Ile Arg His Thr Phe Glu Gly
370 375 380

Ser Trp Lys Asp
385

<210> 57
<211> 1167
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 57

atgagatcag atctaacatc cataatcgta ttcgccgttt ctgtcatttg cgttttgtac 60

ggaccctcac aacaattatc atctccaaaa atagactatg atccattgac gctccgatca 120

cttgatttga agactttgga agctccttca cagttgagtc caggcaccgt agaagataat 180

cttcgaagac aattggagtt tcattttcct taccgcagtt acgaaccttt tccccaacat 240

atttggcaaa cgtggaaagt ttctccctct gatagttcct ttccgaaaaa cttcaaagac 300

ttaggtgaaa gttggctgca aaggtcccca aattatgatc attttgtgat acccgatgat 360

gcagcatggg aacttattca ccatgaatac gaacgtgtac cagaagtctt ggaagctttc 420

cacctgctac cagagcccat tctaaaggcc gatttttca ggtatttgat tctttttgcc 480

cgtggaggac tgtatgctga catggacact atgttattaa aaccaataga atcgtggctg 540

actttcaatg aaactattgg tggagtaaaa aacaatgctg ggttggtcat tggtattgag 600

gctgatcctg atagacctga ttggcacgac tggtatgcta gaaggataca attttgccaa 660

tgggcaattc agtccaaacg aggacaccca gcactgcgtg aactgattgt aagagttgtc 720

agcacgactt tacggaaaga gaaaagcggt tacttgaaca tggtggaagg aaaggatcgt 780

ggaagtgatg tgatggactg gacgggtcca ggaatattta cagacactct atttgattat 840

atgactaatg tcaatacaac aggccactca ggccaaggaa ttggagctgg ctcagcgtat 900

tacaatgcct tatcgttgga agaacgtgat gccctctctg cccgcccgaa cggagagatg 960

ttaaaagaga aagtcccagg taaatatgca cagcaggttg ttttatggga acaatttacc 1020

aacctgcgct cccccaaatt aatcgacgat attcttattc ttccgatcac cagcttcagt 1080

ccagggattg gccacagtgg agctggagat ttgaaccatc accttgcata tattaggcat 1140

acatttgaag gaagttggaa ggactaa 1167

<210> 58
<211> 4311
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 58

```
ttcgcggtcg ggtcacacac ggagaagtaa tgagaagagc tggtaatctg gggtaaaagg      60
gttcaaaaga aggtcgcctg gtagggatgc aatacaaggt tgtcttggag tttacattga     120
ccagatgatt tggctttttc tctgttcaat tcacattttt cagcgagaat cggattgacg     180
gagaaatggc ggggtgtggg gtggatagat ggcagaaatg ctcgcaatca ccgcgaaaga     240
aagactttat ggaatagaac tactgggtgg tgtaaggatt acatagctag tccaatggag     300
tccgttggaa aggtaagaag aagctaaaac cggctaagta actagggaag aatgatcaga     360
ctttgatttg atgaggtctg aaaatactct gctgcttttt cagttgcttt ttccctgcaa     420
cctatcattt tccttttcat aagcctgcct tttctgtttt cacttatatg agttccgccg     480
agacttcccc aaattctctc ctggaacatt ctctatcgct ctccttccaa gttgcgcccc     540
```

ctggcactgc ctagtaatat taccacgcga cttatattca gttccacaat ttccagtgtt 600
cgtagcaaat atcatcagcc taccgttcgt atagcataca ttatacgaag ttatggatct 660
aacatccaaa gacgaaaggt tgaatgaaac ctttttgcca tccgacatcc acaggtccat 720
tctcacacat aagtgccaaa cgcaacagga ggggatacac tagcagcaga ccgttgcaaa 780
cgcaggacct ccactcctct tctcctcaac acccactttt gccatcgaaa aaccagccca 840
gttattgggc ttgattggag ctcgctcatt ccaattcctt ctattaggct actaacacca 900
tgactttatt agcctgtcta tcctggcccc cctggcgagg ttcatgtttg tttatttccg 960
aatgcaacaa gctccgcatt acacccgaac atcactccag atgagggctt tctgagtgtg 1020
gggtcaaata gtttcatgtt ccccaaatgg cccaaaactg acagtttaaa cgctgtcttg 1080
gaacctaata tgacaaaagc gtgatctcat ccaagatgaa ctaagtttgg ttcgttgaaa 1140
tgctaacggc cagttggtca aaaagaaact tccaaaagtc ggcataccgt ttgtcttgtt 1200
tggtattgat tgacgaatgc tcaaaaataa tctcattaat gcttagcgca gtctctctat 1260
cgcttctgaa ccccggtgca cctgtgccga aacgcaaatg gggaaacacc cgctttttgg 1320
atgattatgc attgtctcca cattgtatgc ttccaagatt ctggtgggaa tactgctgat 1380
agcctaacgt tcatgatcaa aatttaactg ttctaacccc tacttgacag caatatataa 1440
acagaaggaa gctgccctgt cttaaacctt tttttttatc atcattatta gcttactttc 1500
ataattgcga ctggttccaa ttgacaagct tttgatttta acgactttta acgacaactt 1560
gagaagatca aaaaacaact aattattcga aacgatggta agccgatacg tacccgatat 1620
gggcgatctg atttgggttg attttgaccc gacaaaaggt agcgagcaag ctggacatcg 1680
tccagctgtt gtcctgagtc ctttcatgta caacaacaaa acaggtatgt gtctgtgtgt 1740
tccttgtaca acgcaatcaa aaggatatcc gttcgaagtt gttttatccg gtcaggaacg 1800
tgatggcgta gcgttagctg atcaggtaaa aagtatcgcc tggcgggcaa gaggagcaac 1860
gaagaaagga acagttgccc cagaggaatt acaactcatt aaagccaaaa ttaacgtact 1920
gattgggtaa tcaagaggat gtcagaatgc catttgcctg agagatgcag gcttcatttt 1980
tgatactttt ttatttgtaa cctatatagt ataggatttt ttttgtcatt ttgtttcttc 2040
tcgtacgagc ttgctcctga tcagcctatc tcgcagctga tgaatatctt gtggtagggg 2100
tttgggaaaa tcattcgagt ttgatgtttt tcttggtatt tcccactcct cttcagagta 2160
cagaagatta agtgacacgt tcgtttgtgc aagcttcaac gatgccaaaa gggtataata 2220
agcgtcattt gcagcattgt gaagaaaact atgtggcaag ccaagcctgc gaagaatgta 2280
gtcgagaatt gagcttgcct cgtccccgcc gggtcacccg gccagcgaca tggaggccca 2340
gaataccctc cttgacagtc ttgacgtgcg cagctcaggg gcatgatgtg actgtcgccc 2400
gtacatttag cccatacatc cccatgtata atcatttgca tccatacatt ttgatggccg 2460

cacggcgcga agcaaaaatt acggctcctc gctgcagacc tgcgagcagg gaaacgctcc 2520
cctcacagac gcgttgaatt gtccccacgc cgcgcccctg tagagaaata taaaaggtta 2580
ggatttgcca ctgaggttct tctttcatat acttcctttt aaaatcttgc taggatacag 2640
ttctcacatc acatccgaac ataaacaaaa atgaccactt tggatgatac tgcttacaga 2700
tacagaactt ctgttccagg tgatgctgaa gctattgaag ctttggatgg atctttcacc 2760
actgatactg ttttcagagt cactgctact ggtgatggat tcactttgag agaagttcct 2820
gttgatcctc ctttgaccaa agtttttcct gatgatgaat ctgatgatga atctgatgct 2880
ggtgaagatg gtgatccaga ttctagaact tttgttgctt atggtgatga tggtgatttg 2940
gctggatttg ttgttgtttc ttattctgga tggaacagaa gattgactgt tgaagatatt 3000
gaagttgctc cagaacatag aggtcatggt gttggaagag ctttgatggg attggcaact 3060
gagtttgcca gagaaagagg tgctggtcat ctttggttgg aagtcaccaa tgtcaatgct 3120
ccagctattc atgcttacag aagaatggga ttcactcttt gtggattgga tactgctttg 3180
tatgatggaa ctgcttctga tggagaacaa gctttgtaca tgtccatgcc atgtccttaa 3240
agtaactgac aataaaaaga ttcttgtttt caagaacttg tcatttgtat agttttttta 3300
tattgtagtt gttctatttt aatcaaatgt tagcgtgatt tatatttttt ttcgcctcga 3360
catcatctgc ccagatgcga agttaagtgc gcagaaagta atatcatgcg tcaatcgtat 3420
gtgaatgctg gtcgctatac tgctgtcgat tcgatactaa cgccgccatc cagtgtcata 3480
acttcgtata gcatacatta tacgaacggt actttttttgt agaaatgtct tggtgtcctc 3540
gtccaatcag gtagccatct ctgaaatatc tggctccgtt gcaactccga acgacctgct 3600
ggcaacgtaa aattctccgg ggtaaaactt aaatgtggag taatggaacc agaaacgtct 3660
cttcccttct ctctccttcc accgcccgtt accgtcccta ggaaatttta ctctgctgga 3720
gagcttcttc tacggccccc ttgcagcaat gctcttccca gcattacgtt gcgggtaaaa 3780
cggaggtcgt gtacccgacc tagcagccca gggatggaaa agtcccggcc gtcgctggca 3840
ataatagcgg gcggacgcat gtcatgagat tattggaaac caccagaatc gaatataaaa 3900
ggcgaacacc tttcccaatt ttggtttctc ctgacccaaa gactttaaat ttaatttatt 3960
tgtccctatt tcaatcaatt gaacaactat ttcgcgaaac gatgagattt ccttcaattt 4020
ttactgctgt tttattcgca gcatcctccg cattagctgc tccagtcaac actacaacag 4080
aagatgaaac ggcacaaatt ccggctgaag ctgtcatcgg ttactcagat ttagaagggg 4140
atttcgatgt tgctgttttg ccattttcca acagcacaaa taacgggtta ttgtttataa 4200
atactactat tgccagcatt gctgctaaag aagaaggggt atctctcgag aaaagagagg 4260
ctgaagctga attcgccaca aaacgtggat ctcccaaccc tacgagggcg g 4311

<210> 59
<211> 12722
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 59

aacgtcaaag acagcaatgg agtcaatatt gataacacca ctggcagagc ggttcgtacg 60

tcgttttgga gccgatatga ggctcagcgt gctaacagca cgattgacaa gaagactctc 120

gagtgacagt aggttgagta aagtattcgc ttagattccc aaccttcgtt ttattctttc 180

gtagacaaag aagctgcatg cgaacatagg gacaactttt ataaatccaa ttgtcaaacc 240

aacgtaaaac cctctggcac cattttcaac atatatttgt gaagcagtac gcaatatcga 300

taaatactca ccgttgtttg taacagcccc aacttgcata cgccttctaa tgacctcaaa 360

tggataagcc gcagcttgtg ctaacatacc agcagcaccg cccgcggtca gctgcgccca 420

cacatataaa ggcaatctac gatcatggga ggaattagtt ttgaccgtca ggtcttcaag 480

agttttgaac tcttcttctt gaactgtgta accttttaaa tgacgggatc taaatacgtc 540

atggatgaga tcatgtgtgt aaaaactgac tccagcatat ggaatcattc caaagattgt 600

aggagcgaac ccacgataaa agtttcccaa ccttgccaaa gtgtctaatg ctgtgacttg 660

aaatctgggt tcctcgttga agaccctgcg tactatgccc aaaaactttc ctccacgagc 720

cctattaact tctctatgag tttcaaatgc caaacggaca cggattaggt ccaatgggta 780

agtgaaaaac acagagcaaa ccccagctaa tgagccggcc agtaaccgtc ttggagctgt 840

ttcataagag tcattaggga tcaataacgt tctaatctgt tcataacata caaattttat 900

ggctgcatag ggaaaaattc tcaacagggt agccgaatga ccctgatata gacctgcgac 960

accatcatac ccatagatct gcctgacagc cttaaagagc ccgctaaaag acccggaaaa 1020

ccgagagaac tctggattag cagtctgaaa aagaatcttc actctgtcta gtggagcaat 1080

taatgtctta gcggcacttc ctgctactcc gccagctact cctgaataga tcacatactg 1140

caaagactgc ttgtcgatga ccttggggtt atttagcttc aagggcaatt tttgggacat 1200

tttggacaca ggagactcag aaacagacac agagcgttct gagtcctggt gctcctgacg 1260

taggcctaga acaggaatta ttggctttat ttgtttgtcc atttcatagg cttggggtaa 1320

tagatagatg acagagaaat agagaagacc taatattttt tgttcatggc aaatcgcggg 1380

ttcgcggtcg ggtcacacac ggagaagtaa tgagaagagc tggtaatctg gggtaaaagg 1440

gttcaaaaga aggtcgcctg gtagggatgc aatacaaggt tgtcttggag tttacattga 1500

ccagatgatt tggctttttc tctgttcaat tcacattttt cagcgagaat cggattgacg 1560

gagaaatggc ggggtgtggg gtggatagat ggcagaaatg ctcgcaatca ccgcgaaaga 1620

aagactttat ggaatagaac tactgggtgg tgtaaggatt acatagctag tccaatggag 1680

tccgttggaa aggtaagaag aagctaaaac cggctaagta actagggaag aatgatcaga 1740

ctttgatttg atgaggtctg aaaatactct gctgcttttt cagttgcttt ttccctgcaa 1800

cctatcattt tccttttcat aagcctgcct tttctgtttt cacttatatg agttccgccg 1860

agacttcccc aaattctctc ctggaacatt ctctatcgct ctccttccaa gttgcgcccc 1920

ctggcactgc ctagtaatat taccacgcga cttatattca gttccacaat ttccagtgtt 1980

cgtagcaaat atcatcagcc taccgttcgt atagcataca ttatacgaag ttatggatct 2040

aacatccaaa gacgaaaggt tgaatgaaac ctttttgcca tccgacatcc acaggtccat 2100

tctcacacat aagtgccaaa cgcaacagga ggggatacac tagcagcaga ccgttgcaaa 2160

cgcaggacct ccactcctct tctcctcaac acccactttt gccatcgaaa aaccagccca 2220

gttattgggc ttgattggag ctcgctcatt ccaattcctt ctattaggct actaacacca 2280

tgactttatt agcctgtcta tcctggcccc cctggcgagg ttcatgtttg tttatttccg 2340

aatgcaacaa gctccgcatt acacccgaac atcactccag atgagggctt tctgagtgtg 2400

gggtcaaata gtttcatgtt ccccaaatgg cccaaaactg acagtttaaa cgctgtcttg 2460

gaacctaata tgacaaaagc gtgatctcat ccaagatgaa ctaagtttgg ttcgttgaaa 2520

tgctaacggc cagttggtca aaaagaaact tccaaaagtc ggcataccgt ttgtcttgtt 2580

tggtattgat tgacgaatgc tcaaaaataa tctcattaat gcttagcgca gtctctctat 2640

cgcttctgaa ccccggtgca cctgtgccga aacgcaaatg gggaaacacc cgctttttgg 2700

atgattatgc attgtctcca cattgtatgc ttccaagatt ctggtgggaa tactgctgat 2760

agcctaacgt tcatgatcaa aatttaactg ttctaacccc tacttgacag caatatataa 2820

acagaaggaa gctgccctgt cttaaacctt tttttttatc atcattatta gcttactttc 2880

ataattgcga ctggttccaa ttgacaagct tttgatttta acgactttta acgacaactt 2940

gagaagatca aaaaacaact aattattcga aacgatggta agccgatacg tacccgatat 3000

gggcgatctg atttgggttg attttgaccc gacaaaaggt agcgagcaag ctggacatcg 3060

tccagctgtt gtcctgagtc ctttcatgta caacaacaaa acaggtatgt gtctgtgtgt 3120

tccttgtaca acgcaatcaa aaggatatcc gttcgaagtt gttttatccg gtcaggaacg 3180

tgatggcgta gcgttagctg atcaggtaaa aagtatcgcc tggcgggcaa gaggagcaac 3240

gaagaaagga acagttgccc cagaggaatt acaactcatt aaagccaaaa ttaacgtact 3300

gattgggtaa tcaagaggat gtcagaatgc catttgcctg agagatgcag gcttcatttt 3360

tgatactttt ttatttgtaa cctatatagt ataggatttt ttttgtcatt ttgtttcttc 3420

tcgtacgagc ttgctcctga tcagcctatc tcgcagctga tgaatatctt gtggtagggg 3480

tttgggaaaa tcattcgagt ttgatgtttt tcttggtatt tcccactcct cttcagagta 3540

cagaagatta agtgacacgt tcgtttgtgc aagcttcaac gatgccaaaa gggtataata 3600

agcgtcattt gcagcattgt gaagaaaact atgtggcaag ccaagcctgc gaagaatgta 3660

gtcgagaatt gagcttgcct cgtccccgcc gggtcacccg gccagcgaca tggaggccca 3720

gaataccctc cttgacagtc ttgacgtgcg cagctcaggg gcatgatgtg actgtcgccc 3780

gtacatttag cccatacatc cccatgtata atcatttgca tccatacatt ttgatggccg 3840

cacggcgcga agcaaaaatt acggctcctc gctgcagacc tgcgagcagg gaaacgctcc 3900

cctcacagac gcgttgaatt gtccccacgc cgcgcccctg tagagaaata taaaaggtta 3960

ggatttgcca ctgaggttct tctttcatat acttcctttt aaaatcttgc taggatacag 4020

ttctcacatc acatccgaac ataaacaaaa atgaccactt tggatgatac tgcttacaga 4080

tacagaactt ctgttccagg tgatgctgaa gctattgaag ctttggatgg atctttcacc 4140

actgatactg ttttcagagt cactgctact ggtgatggat tcactttgag agaagttcct 4200

gttgatcctc ctttgaccaa agtttttcct gatgatgaat ctgatgatga atctgatgct 4260

ggtgaagatg gtgatccaga ttctagaact tttgttgctt atggtgatga tggtgatttg 4320

gctggatttg ttgttgtttc ttattctgga tggaacagaa gattgactgt tgaagatatt 4380

gaagttgctc cagaacatag aggtcatggt gttggaagag ctttgatggg attggcaact 4440

gagtttgcca gagaaagagg tgctggtcat ctttggttgg aagtcaccaa tgtcaatgct 4500

ccagctattc atgcttacag aagaatggga ttcactcttt gtggattgga tactgctttg 4560

tatgatggaa ctgcttctga tggagaacaa gctttgtaca tgtccatgcc atgtccttaa 4620

agtaactgac aataaaaaga ttcttgtttt caagaacttg tcatttgtat agtttttta 4680

tattgtagtt gttctatttt aatcaaatgt tagcgtgatt tatatttttt ttcgcctcga 4740

catcatctgc ccagatgcga agttaagtgc gcagaaagta atatcatgcg tcaatcgtat 4800

gtgaatgctg gtcgctatac tgctgtcgat tcgatactaa cgccgccatc cagtgtcata 4860

acttcgtata gcatacatta tacgaacggt actttttgt agaaatgtct tggtgtcctc 4920

gtccaatcag gtagccatct ctgaaatatc tggctccgtt gcaactccga acgacctgct 4980

ggcaacgtaa aattctccgg ggtaaaactt aaatgtggag taatggaacc agaaacgtct 5040

cttcccttct ctctccttcc accgcccgtt accgtcccta ggaaatttta ctctgctgga 5100

gagcttcttc tacggccccc ttgcagcaat gctcttccca gcattacgtt gcgggtaaaa 5160

cggaggtcgt gtacccgacc tagcagccca gggatggaaa agtcccggcc gtcgctggca 5220

ataatagcgg gcggacgcat gtcatgagat tattggaaac caccagaatc gaatataaaa 5280

ggcgaacacc tttcccaatt ttggtttctc ctgacccaaa gactttaaat ttaatttatt 5340

tgtccctatt tcaatcaatt gaacaactat ttcgcgaaac gatgagattt ccttcaattt 5400

ttactgctgt tttattcgca gcatcctccg cattagctgc tccagtcaac actacaacag 5460
aagatgaaac ggcacaaatt ccggctgaag ctgtcatcgg ttactcagat ttagaagggg 5520
atttcgatgt tgctgttttg ccattttcca acagcacaaa taacgggtta ttgtttataa 5580
atactactat tgccagcatt gctgctaaag aagaaggggt atctctcgag aaaagagagg 5640
ctgaagctga attcgccaca aaacgtggat ctcccaaccc tacgagggcg gcagcagtca 5700
aggccgcatt ccagacgtcg tggaacgctt accaccattt tgcctttccc catgacgacc 5760
tccacccggt cagcaacagc tttgatgatg agagaaacgg ctggggctcg tcggcaatcg 5820
atggcttgga cacggctatc ctcatggggg atgccgacat tgtgaacacg atccttcagt 5880
atgtaccgca gatcaacttc accacgactg cggttgccaa ccaaggatcc tccgtgttcg 5940
agaccaacat tcggtacctc ggtggcctgc tttctgccta tgacctgttg cgaggtcctt 6000
tcagctcctt ggcgacaaac cagaccctgg taaacagcct tctgaggcag gctcaaacac 6060
tggccaacgg cctcaaggtt gcgttcacca ctcccagcgg tgtcccggac cctaccgtct 6120
tcttcaaccc tactgtccgg agaagtggtg catctagcaa caacgtcgct gaaattggaa 6180
gcctggtgct cgagtggaca cggttgagcg acctgacggg aaacccgcag tatgcccagc 6240
ttgcgcagaa gggcgagtcg tatctcctga atccaaaggg aagcccggag gcatggcctg 6300
gcctgattgg aacgtttgtc agcacgagca acggtacctt tcaggatagc agcggcagct 6360
ggtccggcct catggacagc ttctacgagt acctgatcaa gatgtacctg tacgacccgg 6420
ttgcgtttgc acactacaag gatcgctggg tccttggtgc cgactcgacc attgggcatc 6480
tcggctctca cccgtcgacg cgcaaggact tgaccttttt gtcttcgtac aacggacagt 6540
ctacgtcgcc aaactcagga catttggcca gttttggcgg tggcaacttc atcttgggag 6600
gcattctcct gaacgagcaa aagtacattg actttggaat caagcttgcc agctcgtact 6660
ttggcacgta cacccagacg gcttctggaa tcggccccga aggcttcgcg tgggtggaca 6720
gcgtgacggg cgccggcggc tcgccgccct cgtcccagtc cgggttctac tcgtcggcag 6780
gattctgggt gacggcaccg tattacatcc tgcggccgga gacgctggag agcttgtact 6840
acgcataccg cgtcacgggc gactccaagt ggcaggacct ggcgtgggaa gcgttgagtg 6900
ccattgagga cgcatgccgc gccggcagcg cgtactcgtc catcaacgac gtgacgcagg 6960
ccaacggcgg gggtgcctct gacgatatgg agagcttctg gtttgccgag gcgctcaagt 7020
atgcgtacct gatctttgcg gaggagtcgg atgtgcaggt gcaggccacc ggcgggaaca 7080
aatttgtctt taacacggag gcgcacccct ttagcatccg ttcatcatca cgacggggcg 7140
gccaccttgc tcacgacgag ttgtaatcta gggcggccgc cagcttgggc ccgaacaaaa 7200
actcatctca gaagaggatc tgaatagcgc cgtcgaccat catcatcatc atcattgagt 7260

tttagcctta gacatgactg ttcctcagtt caagttgggc acttacgaga agaccggtct 7320
tgctagattc taatcaagag gatgtcagaa tgccatttgc ctgagagatg caggcttcat 7380
ttttgatact tttttatttg taacctatat agtataggat tttttttgtc attttgtttc 7440
ttctcgtacg agcttgctcc tgatcagcct atctcgcagc tgatgaatat cttgtggtag 7500
gggtttggga aaatcattcg agtttgatgt ttttcttggt atttcccact cctcttcaga 7560
gtacagaaga ttaagtgaga ccttcgtttg tgcggatccc ccacacacca tagcttcaaa 7620
atgtttctac tccttttta ctcttccaga ttttctcgga ctccgcgcat cgccgtacca 7680
cttcaaaaca cccaagcaca gcatactaaa tttcccctct ttcttcctct agggtgtcgt 7740
taattacccg tactaaaggt ttggaaaaga aaaaagagac cgcctcgttt ctttttcttc 7800
gtcgaaaaag gcaataaaaa tttttatcac gtttcttttt cttgaaaatt tttttttttg 7860
atttttttct ctttcgatga cctcccattg atatttaagt taataaacgg tcttcaattt 7920
ctcaagtttc agtttcattt ttcttgttct attacaactt tttttacttc ttgctcatta 7980
gaaagaaagc atagcaatct aatctaaggg cggtgttgac aattaatcat cggcatagta 8040
tatcggcata gtataatacg acaaggtgag gaactaaacc atggccaagc ctttgtctca 8100
agaagaatcc accctcattg aaagagcaac ggctacaatc aacagcatcc ccatctctga 8160
agactacagc gtcgccagcg cagctctctc tagcgacggc cgcatcttca ctggtgtcaa 8220
tgtatatcat tttactgggg gaccttgtgc agaactcgtg gtgctgggca ctgctgctgc 8280
tgcggcagct ggcaacctga cttgtatcgt cgcgatcgga aatgagaaca ggggcatctt 8340
gagcccctgc ggacggtgcc gacaggtgct tctcgatctg catcctggga tcaaagccat 8400
agtgaaggac agtgatggac agccgacggc agttgggatt cgtgaattgc tgccctctgg 8460
ttatgtgtgg gagggctaag cacttcgtgg ccgaggagca ggactgacac gtccgacgcg 8520
gcccgacggg tccgaggcct cggagatccg tccccctttt cctttgtcga tatcatgtaa 8580
ttagttatgt cacgcttaca ttcacgccct ccccccacat ccgctctaac cgaaaaggaa 8640
ggagttagac aacctgaagt ctaggtccct atttattttt ttatagttat gttagtatta 8700
agaacgttat ttatatttca aatttttctt ttttttctgt acagacgcgt gtacgcatgt 8760
aacattatac tgaaaacctt gcttgagaag gttttgggac gctcgaaggc tttaatttgc 8820
aagctggaga ccaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc 8880
gcgttgctgg cgtttttcca taggctccgc ccccctgacg agcatcacaa aaatcgacgc 8940
tcaagtcaga ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga 9000
agctccctcg tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt 9060
ctcccttcgg gaagcgtggc gctttctcaa tgctcacgct gtaggtatct cagttcggtg 9120
taggtcgttc gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc 9180

gccttatccg gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg 9240
gcagcagcca ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc 9300
ttgaagtggt ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg 9360
ctgaagccag ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc 9420
gctggtagcg gtggtttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct 9480
caagaagatc ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt 9540
taagggattt tggtcatgag atcagatcta acatccataa tcgtattcgc cgtttctgtc 9600
atttgcgttt tgtacggacc ctcacaacaa ttatcatctc caaaaataga ctatgatcca 9660
ttgacgctcc gatcacttga tttgaagact ttggaagctc cttcacagtt gagtccaggc 9720
accgtagaag ataatcttcg aagacaattg gagtttcatt ttccttaccg cagttacgaa 9780
ccttttcccc aacatatttg gcaaacgtgg aaagtttctc cctctgatag ttcctttccg 9840
aaaaacttca aagacttagg tgaaagttgg ctgcaaaggt ccccaaatta tgatcatttt 9900
gtgatacccg atgatgcagc atgggaactt attcaccatg aatacgaacg tgtaccagaa 9960
gtcttggaag ctttccacct gctaccagag cccattctaa aggccgattt tttcaggtat 10020
ttgattcttt ttgcccgtgg aggactgtat gctgacatgg acactatgtt attaaaacca 10080
atagaatcgt ggctgacttt caatgaaact attggtggag taaaaaacaa tgctgggttg 10140
gtcattggta ttgaggctga tcctgataga cctgattggc acgactggta tgctagaagg 10200
atacaatttt gccaatgggc aattcagtcc aaacgaggac acccagcact gcgtgaactg 10260
attgtaagag ttgtcagcac gactttacgg aaagagaaaa gcggttactt gaacatggtg 10320
gaaggaaagg atcgtggaag tgatgtgatg gactggacgg gtccaggaat atttacagac 10380
actctatttg attatatgac taatgtcaat acaacaggcc actcaggcca aggaattgga 10440
gctggctcag cgtattacaa tgccttatcg ttggaagaac gtgatgccct ctctgcccgc 10500
ccgaacggag agatgttaaa agagaaagtc ccaggtaaat atgcacagca ggttgtttta 10560
tgggaacaat ttaccaacct gcgctccccc aaattaatcg acgatattct tattcttccg 10620
atcaccagct tcagtccagg gattggccac agtggagctg gagatttgaa ccatcacctt 10680
gcatatatta ggcatacatt tgaaggaagt tggaaggact aaagaaagct agagtaaaat 10740
agatatagcg agattagaga atgaatacct tcttctaagc gatcgtccgt catcatagaa 10800
tatcatggac tgtatagttt tttttttgta catataatga ttaaacggtc atccaacatc 10860
tcgttgacag atctctcagt acgcgaaatc cctgactatc aaagcaagaa ccgatgaaga 10920
aaaaaacaac agtaacccaa acaccacaac aaacacttta tcttctcccc cccaacacca 10980
atcatcaaag agatgtcgga accaaacacc aagaagcaaa aactaacccc atataaaaac 11040

atcctggtag ataatgctgg taacccgctc tccttccata ttctgggcta cttcacgaag 11100
tctgaccggt ctcagttgat caacatgatc ctcgaaatgg gtggcaagat cgttccagac 11160
ctgcctcctc tggtagatgg agtgttgttt ttgacagggg attacaagtc tattgatgaa 11220
gataccctaa agcaactggg ggacgttcca atatacagag actccttcat ctaccagtgt 11280
tttgtgcaca agacatctct tcccattgac actttccgaa ttgacaagaa cgtcgacttg 11340
gctcaagatt tgatcaatag ggcccttcaa gagtctgtgg atcatgtcac ttctgccagc 11400
acagctgcag ctgctgctgt tgttgtcgct accaacggcc tgtcttctaa accagacgct 11460
cgtactagca aaatacagtt cactcccgaa gaagatcgtt ttattcttga ctttgttagg 11520
agaaatccta aacgaagaaa cacacatcaa ctgtacactg agctcgctca gcacatgaaa 11580
aaccatacga atcattctat ccgccacaga tttcgtcgta atctttccgc tcaacttgat 11640
tgggtttatg atatcgatcc attgaccaac caacctcgaa aagatgaaaa cgggaactac 11700
atcaaggtac aagatcttcc acaaggaatt cgtggtcatt attctgccca agatgattac 11760
aatttgtgtt tatcggttca acctttcatt gaatctgtag atgagacaac aggccaagaa 11820
tttttcaaac ctctgaaagg tgtatttgat gacttggaat ctcgctttcc tcaccataca 11880
aagacttcct ggagagacag attcagaaag tttgcctcta aatacggtgt tcgtcagtac 11940
atcgcgtatt atgaaaagac tgttgaactc aatggtgttc ctaatccgat gacgaacttt 12000
acctcaaagg cttccattga aaaatttaga gaaagacgcg ggacttcacg taacagtggc 12060
cttccaggcc cggttggtgt agaagctgta agctctttgg accacatatc cccattggtc 12120
acatctaatt ccaattctgc agctgctgca gctgctgccg cagcagttgc agcctctgcc 12180
tctgcttctt cagctcctaa tacttcaact accaatttct ttgaacagga gaatattgcc 12240
caagttctct ctgcacataa caacgagcag tctattgcag aagttattga gtccgcacag 12300
aatgtcaaca cccatgaaag tgaacctata gctgatcatg ttcgaaaaaa tcttacagac 12360
gatgaattgc ttgacaaaat ggatgatatt ttaagctcca gaagtctagg cggactagat 12420
gacttgataa agatcctcta cactgagctg ggatttgctc atcgttatac cgaatttctt 12480
tttacctcat gttctggtga tgtgattttc ttccgaccat tagtggaaca tttccttctt 12540
actggtgagt gggagctgga gaatactcgt ggcatctgga ccggtcgtca agacgaaatg 12600
ctacgtgcta gcaatctaga tgacctgcac aagttaattg acctgcatgg gaaagaacgt 12660
gttgagacca gaagaaaagc catcaaggga gaatgatcat aagaaatgaa aaacgtataa 12720
gt 12722

<210> 60
<211> 19
<212> ADN

<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 60
caagttgcgc cccctggca 19

<210> 61
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 61
tggagcagct aatgcggagg a 21

<210> 62
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 62
agttccgccg agacttcccc a 21

<210> 63
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 63
ttcagccgga atttgtgccg t 21

<210> 64
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 64
atccagggtg acggtgccga 20

<210> 65
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 65
gcaagaggcc cggcagtacc 20

<210> 66
<211> 21
<212> ADN

<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 66
ccgccctcgt agggttggga g 21

<210> 67
<211> 19
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 67
ttcgcggtcg ggtcacaca 19

<210> 68
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 68
aactgccatc tgccttcgcc 20

<210> 69
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 69
caaatcgcgg gttcgcggtc 20

<210> 70
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 70
gagcaaactg ccatctgcct tcg 23

<210> 71
<211> 75
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 71

gtgttcgtag caaatatcat cagcctaccg ttcgtatagc atacattata cgaagttatg 60

gatctaacat ccaaa 75

<210> 72
<211> 75
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 72

tttggatgtt agatccataa cttcgtataa tgtatgctat acgaacggta ggctgatgat 60

atttgctacg aacac 75

<210> 73
<211> 75
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 73

gccgccatcc agtgtcataa cttcgtatag catacattat acgaacggta ctttttgta 60

gaaatgtctt ggtgt 75

<210> 74
<211> 75
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 74

acaccaagac atttctacaa aaaagtaccg ttcgtataat gtatgctata cgaagttatg 60

acactggatg gcggc 75

<210> 75
<211> 30
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 75
gtgttcgtag caaatatcat cagcctaccg 30

<210> 76
<211> 31
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 76
acaccaagac atttctacaa aaaagtaccg t 31

<210> 77
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 77
ttcgcggtcg ggtcacacac 20

<210> 78
<211> 23
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 78
ggagcagcta atgcggagga tgc 23

<210> 79
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 79
cggtcgggtc acacacggag 20

<210> 80
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 80
tggagcagct aatgcggagg a 21

<210> 81
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 81
tgagtcctgg tgctcctgac g 21

<210> 82
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 82
cccctcctgt tgcgtttggc 20

<210> 83
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 83
agcgttctga gtcctggtgc t 21

<210> 84
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 84
ggtcctgcgt ttgcaacggt 20

<210> 85
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 85
actaacgccg ccatccagtg tc 22

<210> 86
<211> 22
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 86
gcttcagccg gaatttgtgc cg 22

<210> 87
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 87
cgcctcgaca tcatctgccc 20

<210> 88
<211> 20
<212> ADN
<213> Secuencia artificial

<220>
<223> Oligonucleótido sintético

<400> 88
tcagccggaa tttgtgccgt 20

<210> 89
<211> 23
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 89

```
Phe Tyr Met Ala Ile Phe Ala Val Ser Val Ile Cys Val Leu Tyr Gly
1               5                   10                  15

Pro Ser Gln Gln Leu Ser Ser
            20
```

<210> 90
<211> 5
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 90

```
Phe Tyr Met Ala Ile
1               5
```

<210> 91
<211> 4
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 91

```
His Asp Glu Leu
1
```

<210> 92
<211> 14
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 92

```
Leu Phe Ala Arg Gly Gly Leu Tyr Ala Asp Met Asp Thr Met
1               5                   10
```

<210> 93

<211> 14
<212> PRT
<213> Secuencia artificial

<220>
<223> Péptido sintético

<400> 93

Gly Lys Pro Ile Pro Asn Pro Leu Leu Gly Leu Asp Ser Thr
1 5 10

REIVINDICACIONES

1. Una cepa modificada estable de *Pichia pastoris*, que comprende:

un alelo mutante que se transcribe en un ARNm que codifica una proteína OCH1 mutante, donde la proteína OCH1 mutante comprende un dominio catalítico al menos un 95 % idéntico a los aminoácidos 45-404 de la SEQ ID NO: 2 y que mantiene sustancialmente la actividad catalítica de la proteína OCH1 de tipo silvestre que comprende la SEQ ID NO: 2, donde la proteína OCH1 mutante carece de una secuencia en el extremo N para dirigir la proteína OCH1 mutante al aparato de Golgi y que carece de un dominio de anclaje a la membrana en la región del extremo N, y donde dicha cepa produce N-glicanos sustancialmente homogéneos.

2. La cepa de la reivindicación 1, donde la falta de un dominio de anclaje a la membrana en la proteína OCH1 mutante es resultado de la eliminación de una parte del extremo N de la proteína OCH1 de tipo silvestre, donde la parte eliminada comprende el dominio de anclaje a la membrana de la proteína OCH1 de tipo silvestre.

3. La cepa de la reivindicación 2, donde la parte eliminada comprende además uno o más aminoácidos de la cola citoplasmática de la proteína OCH1 de tipo silvestre.

4. La cepa de la reivindicación 1, donde dicha proteína OCH1 mutante comprende la secuencia que se expone en SEQ ID NO: 3.

5. La cepa de la reivindicación 1, donde dicho alelo OCH1 mutante está presente en un cromosoma.

6. La cepa de la reivindicación 5, donde dicho alelo OCH1 mutante sustituye al alelo OCH1 de tipo silvestre en el locus OCH1.

7. La cepa de la reivindicación 1, donde dicho alelo OCH1 mutante se mantiene en un plásmido, y donde el alelo OCH1 de tipo silvestre en el cromosoma se ha destruido.

8. La cepa de la reivindicación 1, donde dicha cepa produce N-glicanos sustancialmente homogéneos, siendo $Man_8GlcNAc_2$ la forma de N-glicano predominante.

9. La cepa de la reivindicación 1, donde dicha cepa comprende adicionalmente un ácido nucleico que codifica y expresa una α-1,2-manosidasa o un fragmento funcional de la misma.

10. La cepa de la reivindicación 9, donde dicho ácido nucleico que codifica y expresa dicha α-1,2-manosidasa o dicho fragmento funcional de la misma se integra en el locus OCH1 de la cepa.

11. La cepa de la reivindicación 10, donde el locus OCH1 comprende la secuencia de nucleótidos que se expone en la SEQ ID NO: 1.

12. La cepa de la reivindicación 9, donde dicha cepa produce N-glicanos sustancialmente homogéneos siendo $Man_5GlcNAc_2$ la forma de N-glicano predominante.

13. La cepa de la reivindicación 1 o la reivindicación 9, que comprende adicionalmente un ácido nucleico que codifica y expresa una proteína heteróloga.

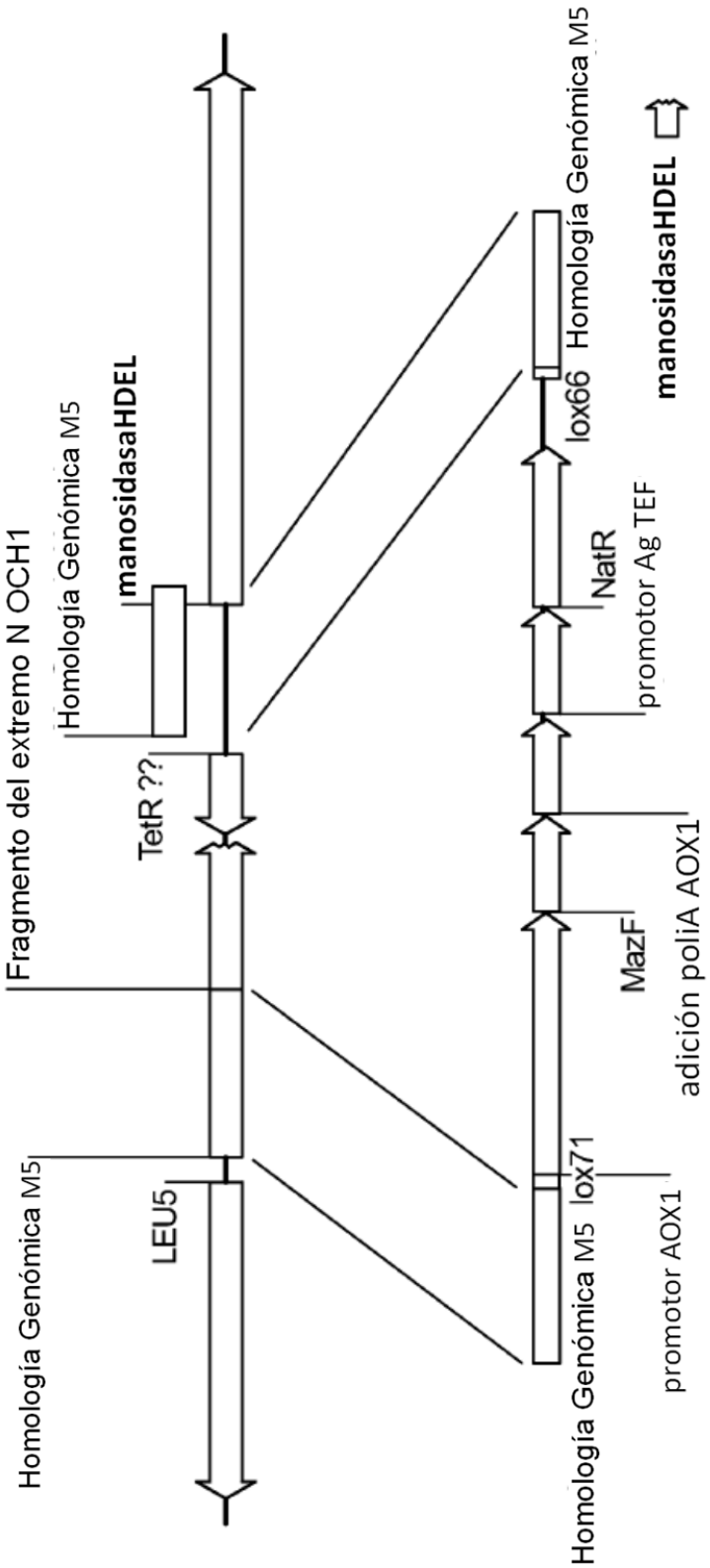

Homología Genómica M5
LEU5
Fragmento del extremo N OCH1
TetR ??
Homología Genómica M5
manosidasaHDEL
Homología Genómica M5
lox71
promotor AOX1
MazF
adición poliA AOX1
NatR
promotor Ag TEF
lox66
Homología Genómica M5
manosidasaHDEL

Figura 1

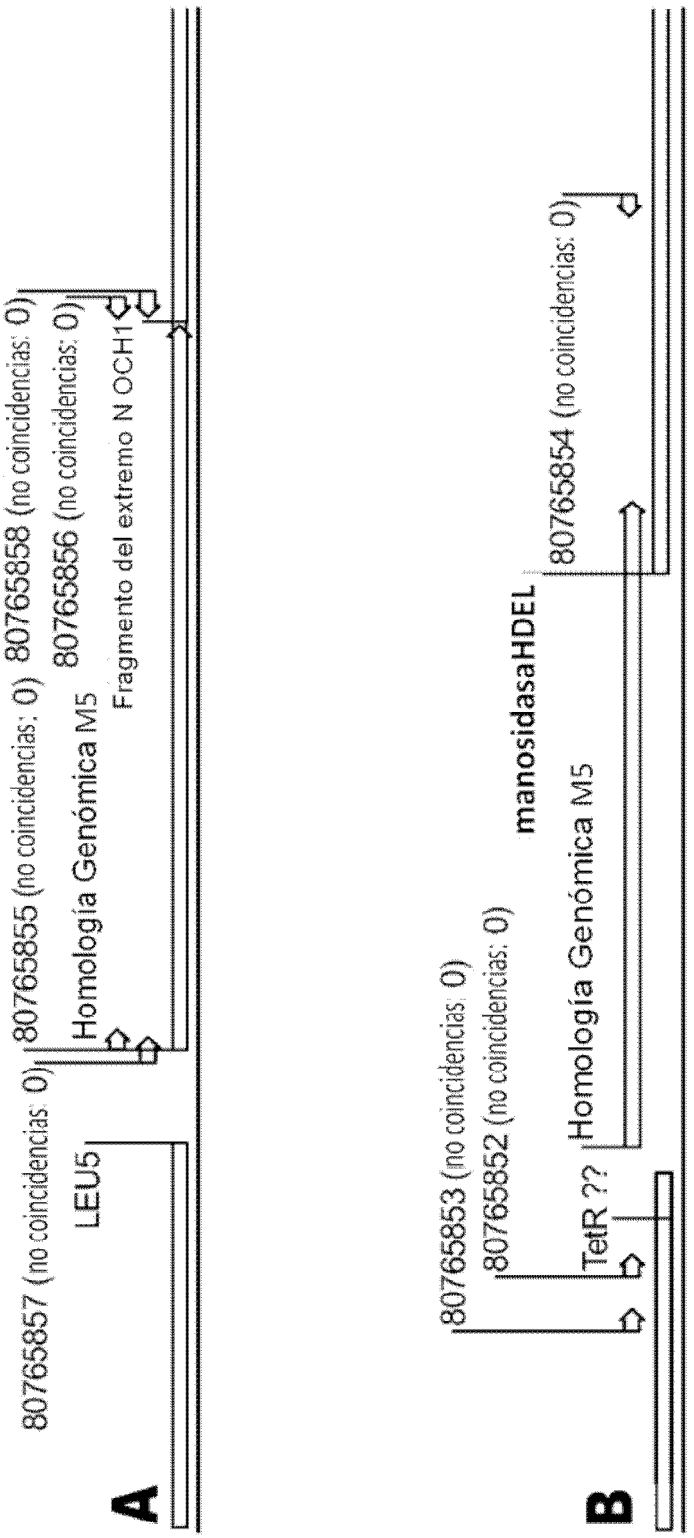

A
80765857 (no coincidencias: 0)
LEU5
80765855 (no coincidencias: 0)
Homología Genómica M5
80765858 (no coincidencias: 0)
80765856 (no coincidencias: 0)
Fragmento del extremo N OCH1
B
80765853 (no coincidencias: 0)
80765852 (no coincidencias: 0)
TetR ??
Homología Genómica M5
manosidasaHDEL
80765854 (no coincidencias: 0)

Figura 2

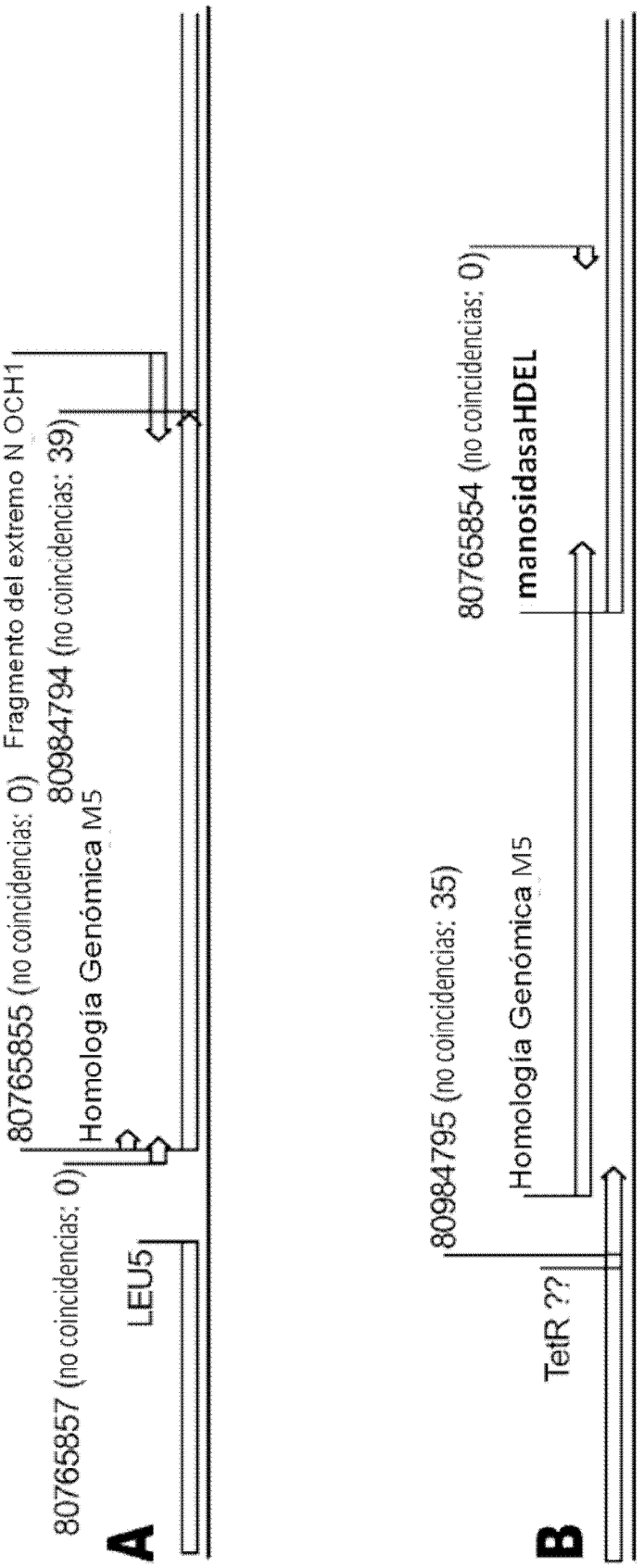

A
80765857 (no coincidencias: 0)
LEU5
80765855 (no coincidencias: 0)
Homología Genómica M5
80984794 (no coincidencias: 39)
Fragmento del extremo N OCH1
B
TetR ??
80984795 (no coincidencias: 35)
Homología Genómica M5
80765854 (no coincidencias: 0)
manosidasaHDEL

Figura 3

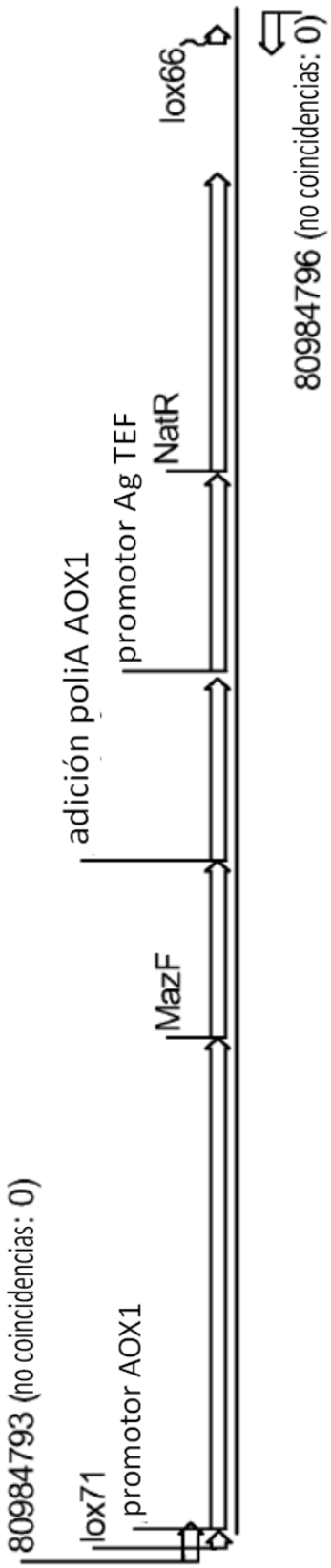


Figura 4

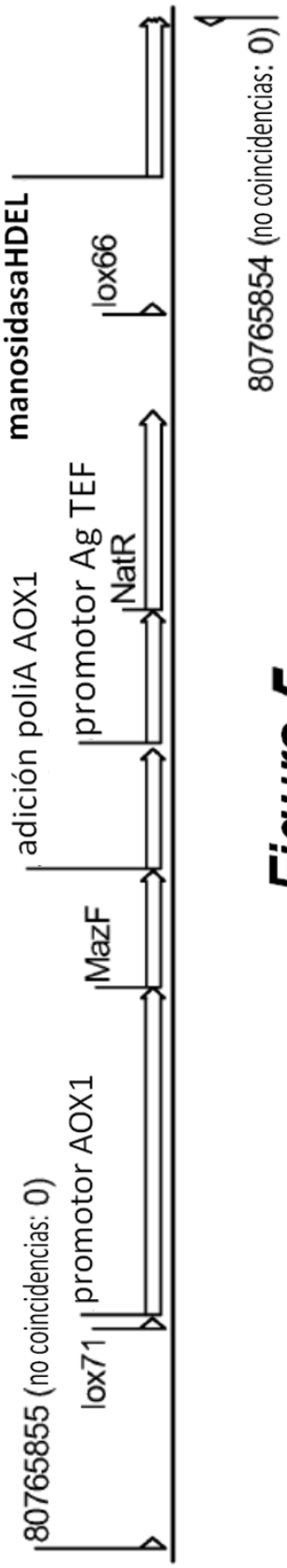


Figura 5

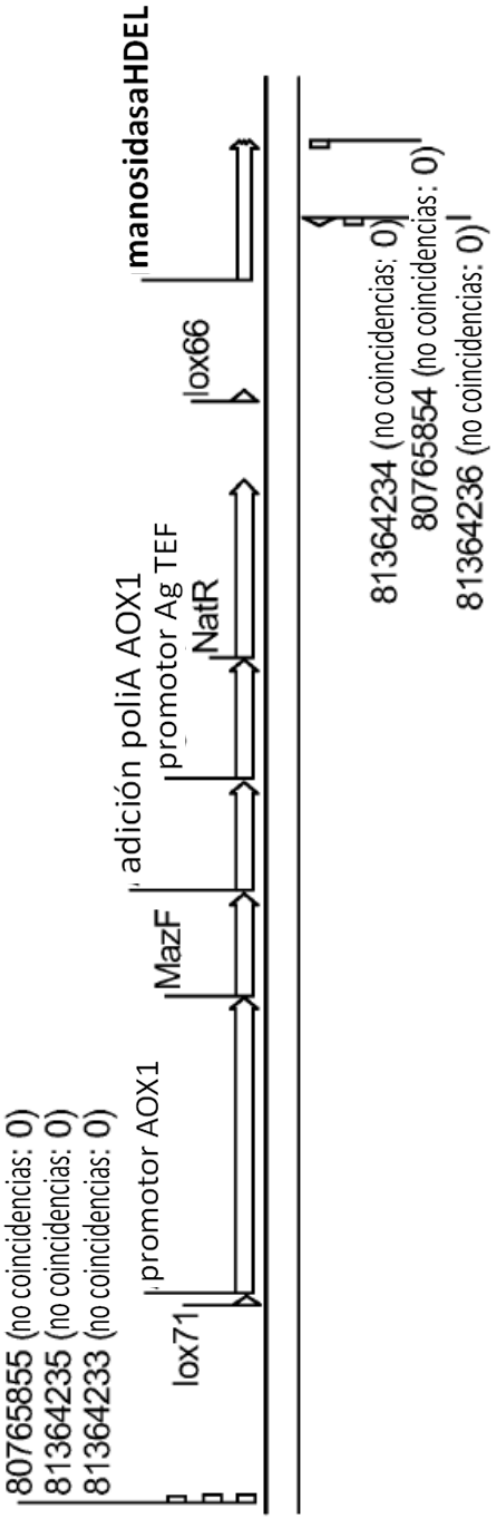


Figura 6

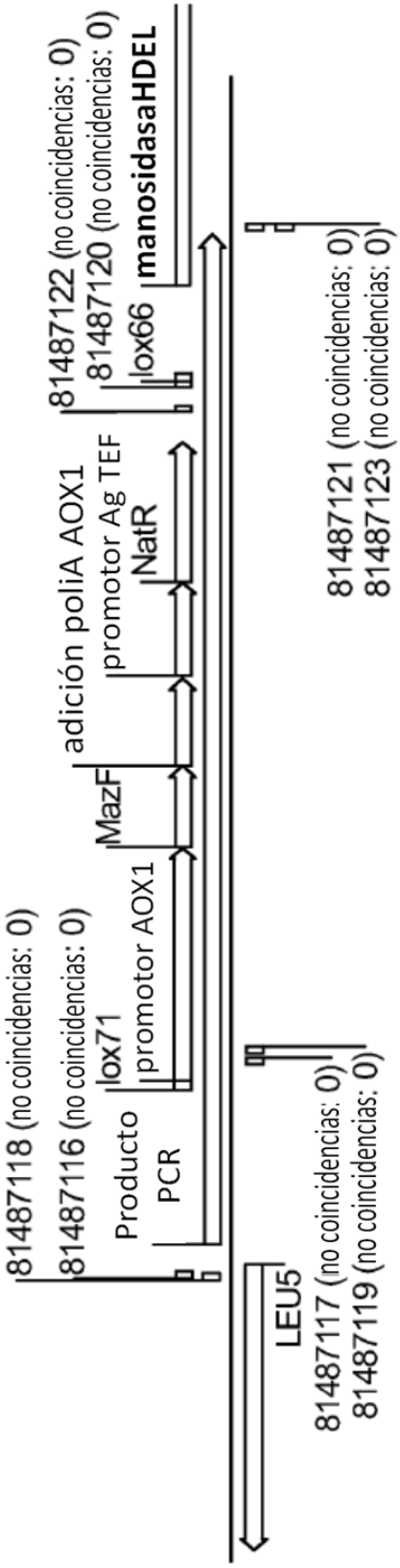


Figura 7

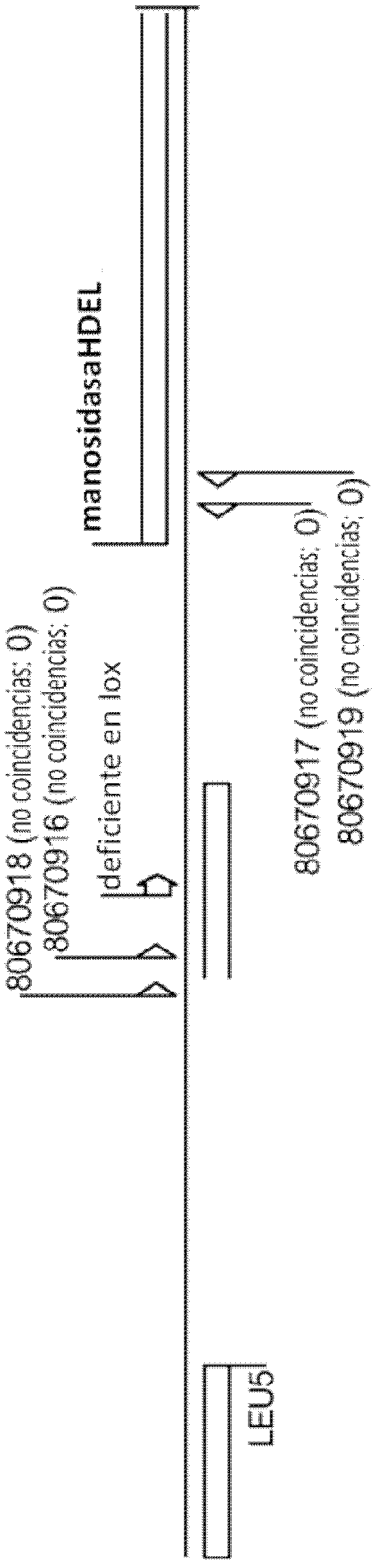


Figura 8

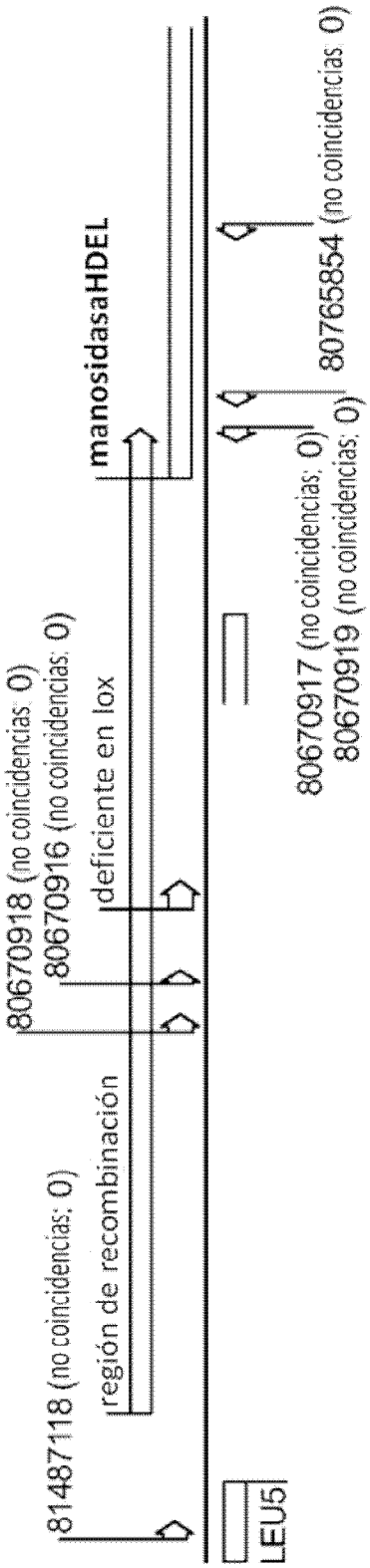


Figura 9

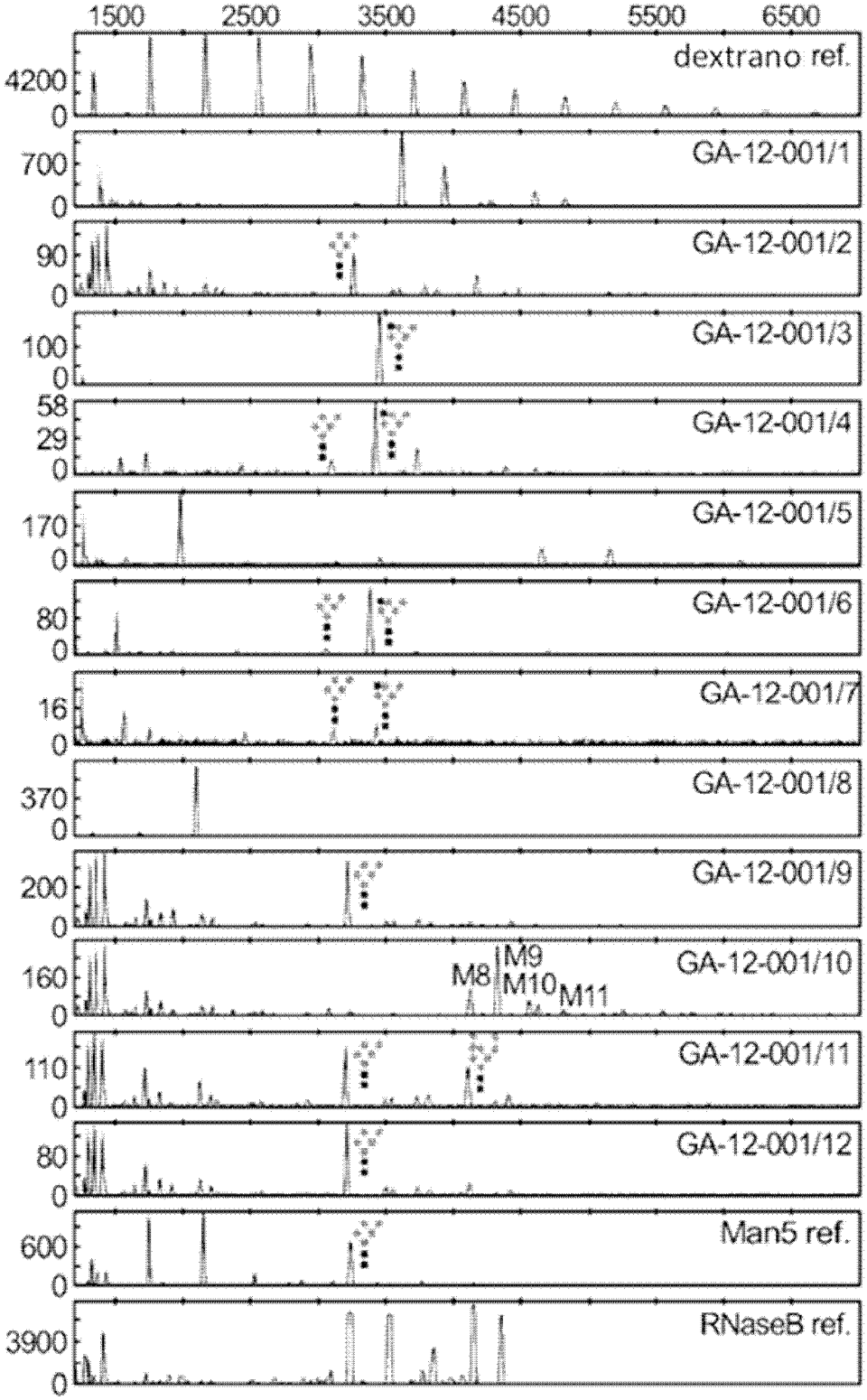
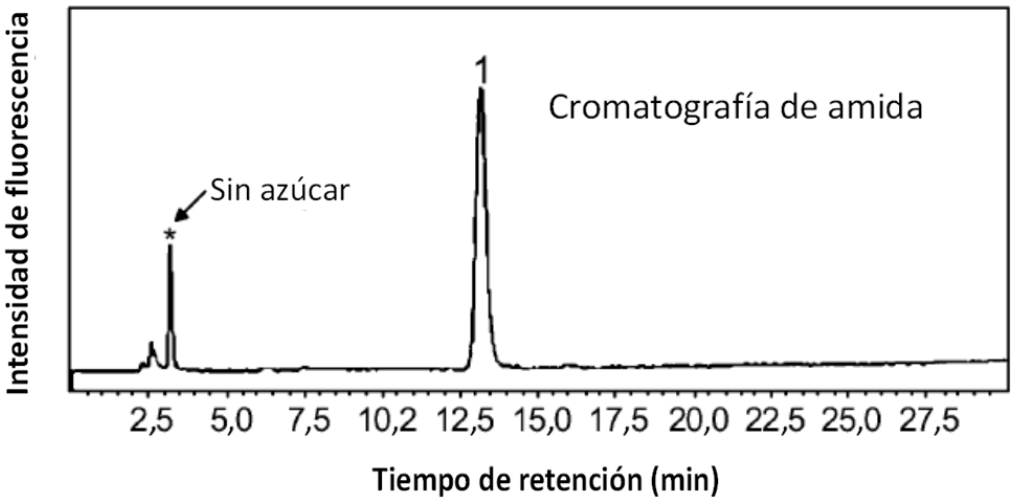
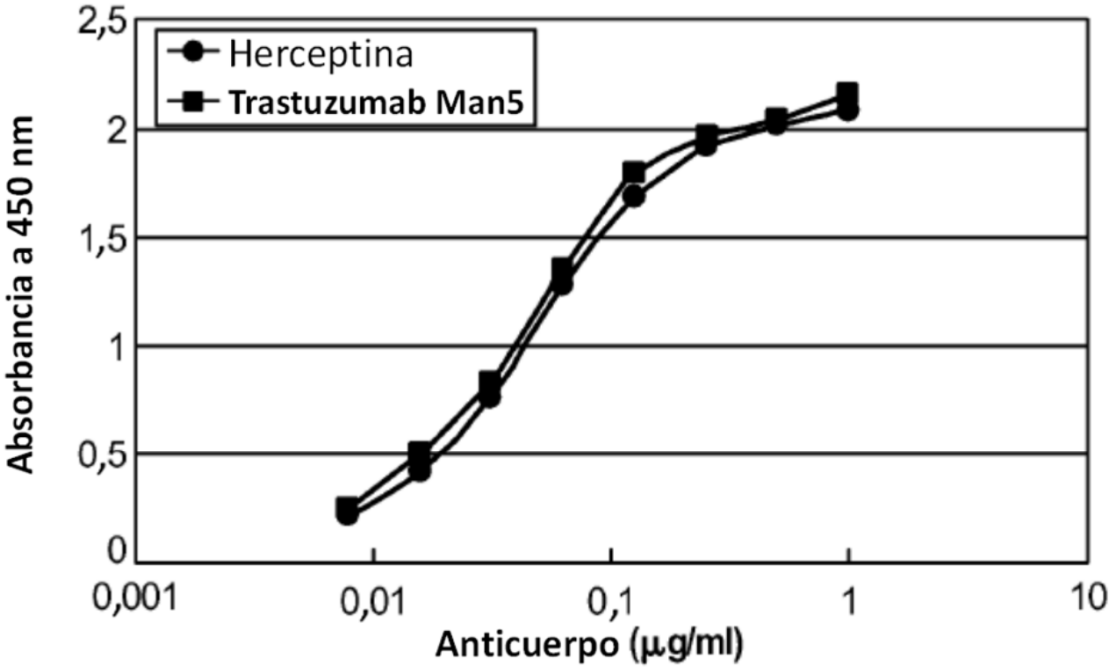

1500
2500
3500
4500
5500
6500
dextrano ref.
4200
0
GA-12-001/1
700
0
GA-12-001/2
90
0
GA-12-001/3
100
0
GA-12-001/4
58
29
0
GA-12-001/5
170
0
GA-12-001/6
80
0
GA-12-001/7
16
0
GA-12-001/8
370
0
GA-12-001/9
200
0
GA-12-001/10
160
0
M8
M9
M10
M11
GA-12-001/11
110
0
GA-12-001/12
80
0
Man5 ref.
600
0
RNaseB ref.
3900
0

Figura 10



	Estructura de N-glicano estimada
N1	Manα1-6 \ Manα1-6 \ Manβ1-4GlcNAcβ1-4G1cNAc-PA Manα1-3 / Manα1-3 /

Figura 11



Los círculos sólidos (●) se refieren a Herceptina comercial producida en células CHO, los cuadrados sólidos (■) se refieren a trastuzumab tipo Man5.

Figura 12

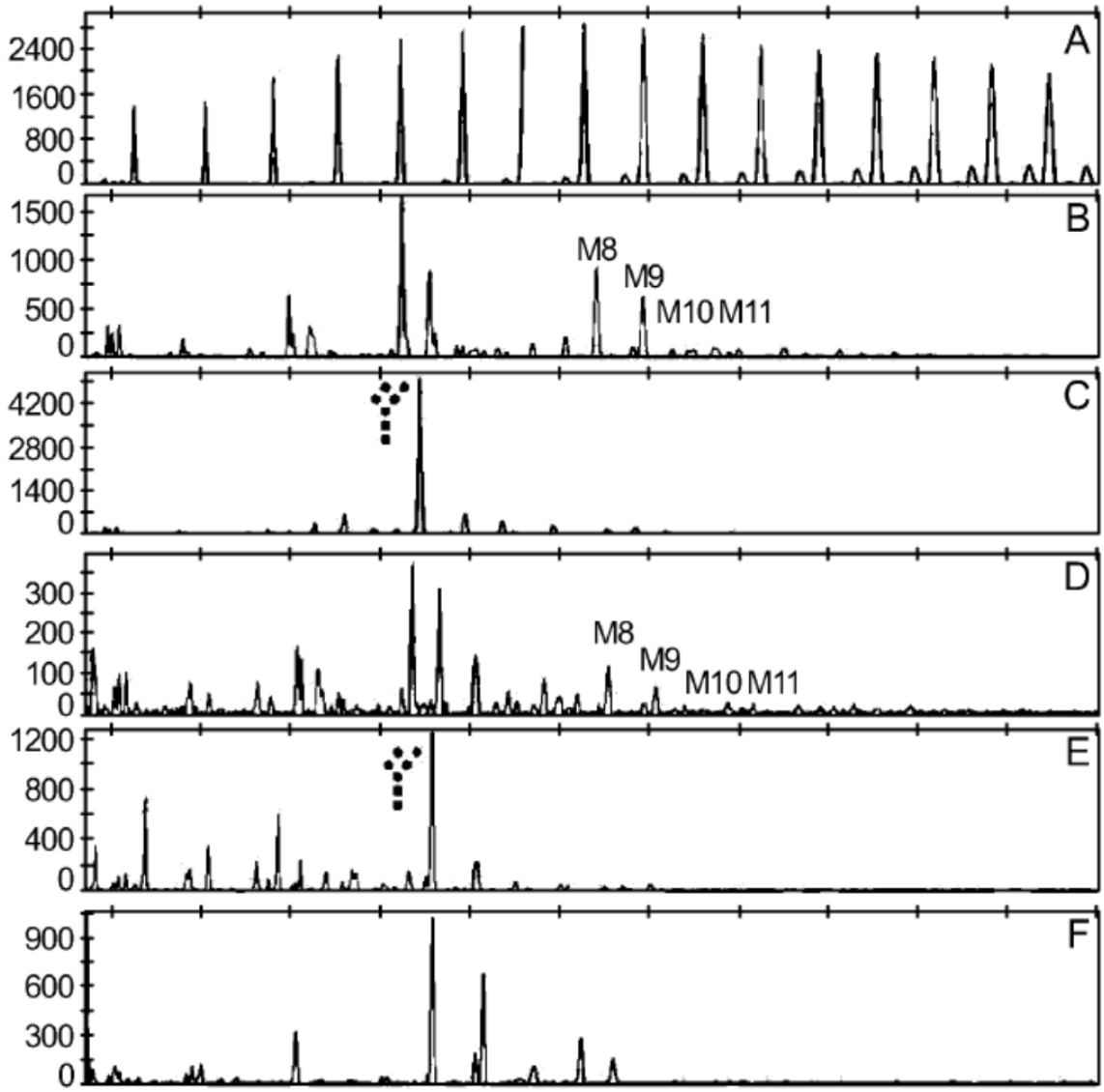

A
2400
1600
800
0
B
1500
1000
500
0
M8
M9
M10 M11
C
4200
2800
1400
0
D
300
200
100
0
M8
M9
M10 M11
E
1200
800
400
0
F
900
600
300
0

Figura 13

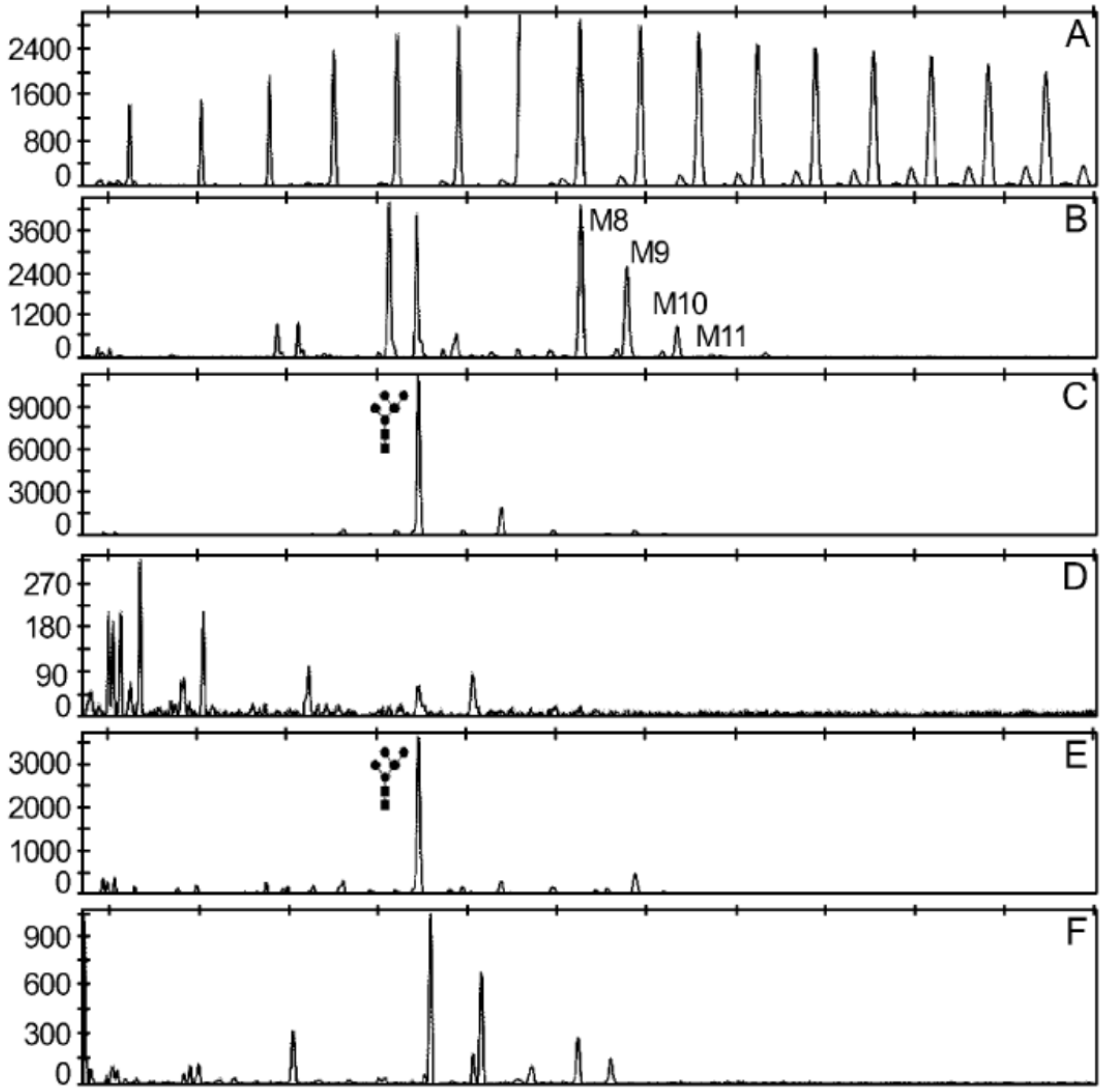

A
2400
1600
800
0
B
3600
2400
1200
0
M8
M9
M10
M11
C
9000
6000
3000
0
D
270
180
90
0
E
3000
2000
1000
0
F
900
600
300
0

Figura 14

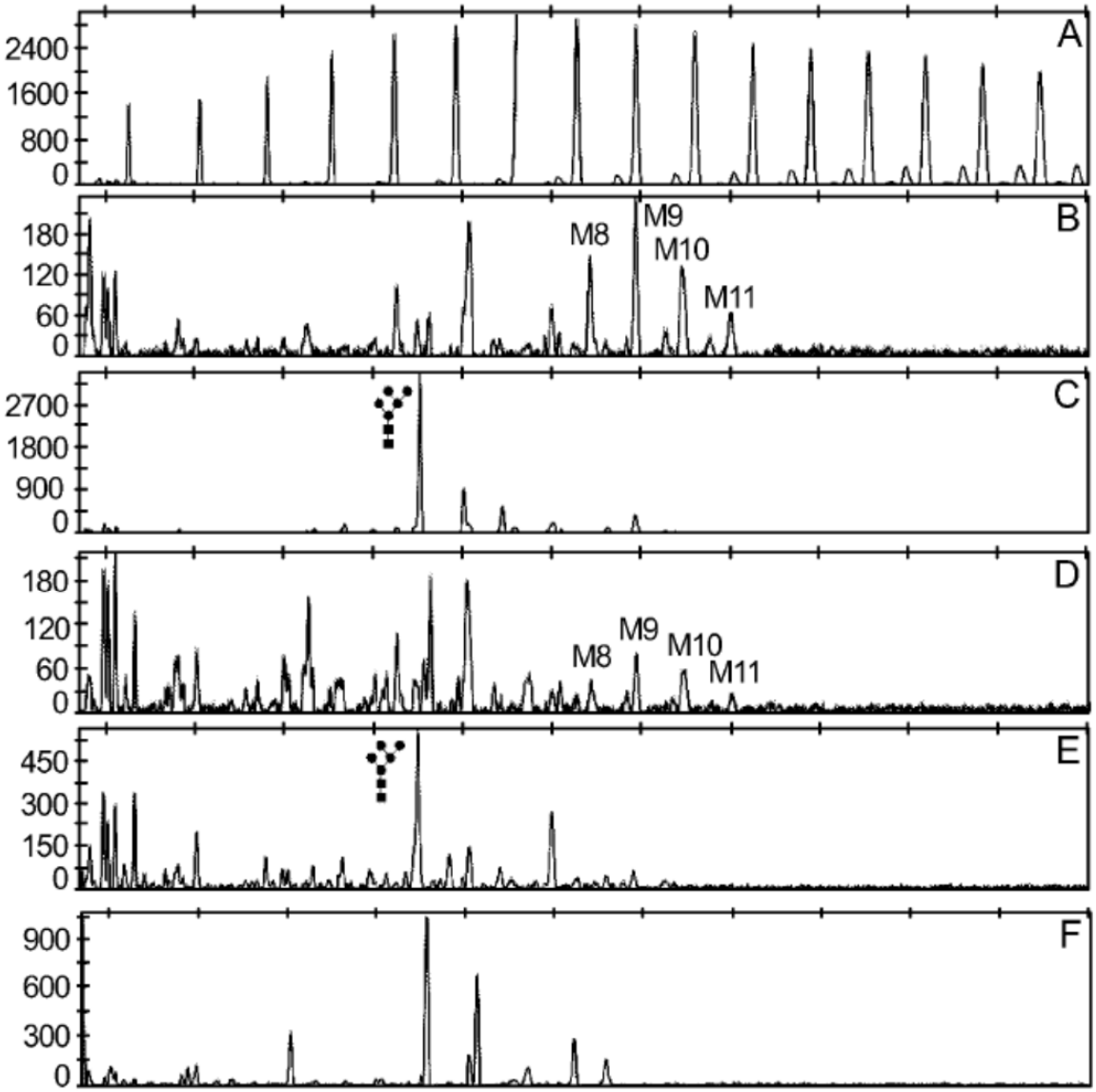

A
2400
1600
800
0
B
180
120
60
0
M8
M9
M10
M11
C
2700
1800
900
0
D
180
120
60
0
M8
M9
M10
M11
E
450
300
150
0
F
900
600
300
0

Figura 15