

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 539 378**

51 Int. Cl.:

C12N 1/00 (2006.01)

C07H 21/04 (2006.01)

C12N 9/60 (2006.01)

C12N 15/81 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **24.02.2010 E 10746762 (3)**

97 Fecha y número de publicación de la concesión europea: **25.03.2015 EP 2401357**

54 Título: **Células huésped y procedimientos de uso**

30 Prioridad:

26.02.2009 US 155706 P

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

30.06.2015

73 Titular/es:

**GLAXOSMITHKLINE LLC (100.0%)
Corporation Service Company, 2711 Centreville
Road, Suite 400
Wilmington, Delaware 19808, US**

72 Inventor/es:

**JIN, YONG, HWAN;
JOWETT, JAMES, D.;
TAYLOR, ALEXANDER, H. y
ZHU, YUAN**

74 Agente/Representante:

CARPINTERO LÓPEZ, Mario

ES 2 539 378 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

Células huésped y procedimientos de uso

Campo de la invención

5 La invención pertenece al campo de la ingeniería biomédica. Más particularmente, la presente invención se refiere a cepas de *Pichia* modificadas genéticamente y a procedimientos de producción de polipéptidos en las mismas.

Antecedentes de la invención

10 Los polipéptidos y proteínas terapéuticas se pueden expresar en diversas células huésped, incluyendo células bacterianas, células de *E. coli*, células fúngicas o de levaduras, células de un microorganismo, células de insectos y células de mamíferos. Huéspedes fúngicos, tales como la levadura metiltrófica *Pichia pastoris* tiene claras ventajas para la expresión de proteínas terapéuticas, por ejemplo, no secreta grandes cantidades de proteínas endógenas, que tiene un fuerte promotor inducible, se puede cultivar en medios químicos definidos y puede producir altos títulos de proteínas recombinantes (Cregg y col., Mol. Biotech. 16:23-52 (2000)). Las levaduras y hongos filamentosos se han usado con éxito para la producción de proteínas recombinantes, tanto intracelulares y secretadas (Cereghino, J. L. y J. M. Cregg 2000 FEMS Microbiology Reviews 24(1): 45 66; Harkki, A., et al. 1989 Bio-Technology 7(6): 596; Berka, R. M., et al. 1992 Abstr. Papers Amer. Chem.Soc.203: 121-BIOT; Svetina, M., et al. 2000 J. Biotechnol. 76(23): 245-251. *Pichia* es una célula huésped notable para la expresión de seroalbúmina humana recombinante (HSA). Sin embargo, la expresión de otros polipéptidos terapéuticos, incluidos polipéptidos fusionados genéticamente con HSA se enfrenta a los obstáculos técnicos de la proteólisis no deseada y la glicosilación.

20 Las proteínas heterólogas expresadas en *P. pastoris* pueden contener azúcares de manosa adicionales resultantes en glicanos "ricos en manosa", así como grupos manosilfosfato que imparten una carga negativa en una proteína. Las proteínas glicosiladas, ya sea con glicanos ricos en manosa o mananos cargados tienen un alto riesgo de provocar una respuesta inmunitaria en los seres humanos (Takeuchi, Trends in Glycosci. & Glycotech., 9:S29-S35 (1997); Rosenfeld y col., J. Biol. Chem., 249:2319-2321 (1974)). Por consiguiente, sería deseable producir péptidos terapéuticos, polipéptidos y/o proteínas en sistemas de huéspedes fúngicas, de forma que patrón de glicosilación sea idéntico o al menos similar al de los seres humanos.

25 El documento US6051419 describe genes que influyen sobre la actividad proteolítica de *Pichia*. Daly et al. describen la expresión de proteínas heterólogas en *Pichia* (Journal of Molecular Recognition Vol. 18, No. 2, 119-138). Los documentos WO9914347 1 y US6153424 describieron cepas de *Pichia* deficientes en proteasa. Gleeson et al describen la generación de cepas deficientes en proteasa y su uso en la expresión de proteínas heterólogas (Methods in Molecular Biology Vol. 103, 81-94).

30 Por lo tanto, existe la necesidad de cepas de levadura, en particular cepas de *Pichia* que sean capaces de producir péptidos heterólogos, polipéptidos y/o proteínas con proteólisis reducida y/o glicosilación en comparación con las cepas de tipo salvaje. Además, existe la necesidad de identificar genes dentro de cepas de levadura, en particular cepas de *Pichia*, responsables de producir las proteínas implicadas en vías proteolíticas y de glicosilación.

Sumario de la invención

35 La presente invención proporciona una cepa de *Pichia* modificada genéticamente en la que la secuencia de ácido nucleico que codifica:

(i) PEP4, PRB1 e YMP1,

(ii) PEP4, PRB1 e YMP2, o

40 (iii) PEP4, PRB1 e YMP3

es modificada genéticamente para producir deficiencia de proteasa de al menos tres de las siguientes enzimas y/o tipo de enzima: aminopeptidasa de tipo Arg / Ala, aspartil proteasa, serín proteasa, serín proteasa secretada, serín proteasa Prb 1, aminopeptidasa de tipo Leu, en comparación con la cepa de tipo salvaje.

45 Se describen cepas de *Pichia* modificadas genéticamente en las que al menos una secuencia de ácido nucleico que codifica un producto génico funcional y/o al menos un ácido nucleico necesario para la expresión de al menos un producto génico en dicha cepa de *Pichia* está modificada genéticamente, en el que dicho producto génico es responsable de la proteólisis y/o la glicosilación en dicha cepa de *Pichia* modificada genéticamente.

50 También se proporcionan en el presente documento procedimientos de producción de al menos un polipéptido heterólogo que comprende expresar dicho polipéptido heterólogo en una cepas de *Pichia* modificada genéticamente de la presente invención. Se describen péptidos expresados de forma heteróloga, polipéptidos y proteínas expresadas en las células huésped de la presente invención.

Se describe un polipéptido aislado que tiene al menos un 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 % y/o 100 % de identidad de secuencia con la SEC ID N° 4, SEC ID N° 6 y SEC ID N° 8.

También se describe una cepa de células huésped genéticamente modificada en la que dicho padre de tipo salvaje parental de dicha célula huésped comprende un gen que codifica un polipéptido que tiene al menos 60 % de identidad de secuencia con las SEC ID N° 4, 6 y 8, en la que dicho gen se modifica genéticamente en el genoma de dicha célula huésped de tal manera que el producto del gen o su actividad se reduce o se elimina en dicha célula huésped genéticamente modificada en comparación con dicha célula huésped de tipo salvaje.

Breve descripción de las figuras

Figura 1: Nueva proteasa Ymp1 de *Pichia* identificada utilizando la purificación por afinidad y análisis CL/EM

Figura 2: Estructura predicha de *YMP1* (SEC ID N° 4) que muestra los aminoácidos 1-857 como la SEC ID N° 15.

Figura 3: Ensayo de actividad de la proteasa de la cepa mutante *ymp1*

Figura 4: Ensayo zimográfico de la cepa mutante *ymp1*

Figura 5: Efecto de inactivación de *ymp1* sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

Figura 6: Efecto de inactivación de *ysp2* sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

Figura 7: Efecto de inactivación de *ysp1* sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

Figura 8: Efecto de inactivación de *pep4* sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

Figura 9: Estudio de estabilidad en tampón de las cepas mutantes *ymp1*, *ymp3*

Figura 10: Resultados del ensayo de zimoliasa par alas cepas de tipo salvaje (WT) y mutantes *pmt1* y *pmt4* de *Pichia*

Figura 11: Análisis de glicosilación de proteínas heterólogas (SEC ID N° 1) de la cepa mutante *pmt4*

Descripción detallada de la invención

"Célula(s) huésped" como se usa en el presente documento se refiere a una célula que se ha introducido (por ejemplo, transformado, infectado o transfectado) o es capas de introducción (por ejemplo, transformación, infección o transfección) por una secuencia de polinucleótidos aislados. Las células huésped de origen de levadura y/u hongos filamentosos pueden incluir, entre otros, las siguientes familias, géneros y especies: *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia menabranaefaciens*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichi salictaria*, *Pichia guercum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia sp.*, *Saccharomyces castelii*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Candida sp.*, *Aspergillus fumigatus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum*, *Physcomitrella patens* y *Neurospora crassa*.

"Transformado" como se conoce en la técnica, es la modificación dirigida del genoma o apisona de un organismo a través de la introducción de ADN o ARN externo, o a cualquier otra introducción estable de ADN o ARN externo.

"Transfectada" como se conoce en la técnica, es la introducción de ADN o ARN externo en un microorganismo, incluyendo, pero no limitado a, ADN o ARN recombinante.

"Identidad", como se conoce en la técnica, es una relación entre dos o más secuencias polipeptídicas o dos o más secuencias polinucleotídicas, según el caso, determinado comparando las secuencias. En la técnica, "identidad" también significa el grado de relación de la secuencia entre secuencias polipeptídicas o polinucleotídicas, como puede ser el caso según se determine mediante la correspondencia entre cadenas de dichas secuencias. La "identidad" se puede calcular con facilidad mediante procedimientos conocidos, que incluyen pero sin limitación, los descritos en (Computational Molecular Biology, Lesk, A.M. ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., y Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; y Sequence Analysis Printer, Gribskov, M. y Devereux, J., eds., M Stockton Press, New York, 1991; y Carillo, H., y Lipman, D., SIAM J. Applied Match., 48: 1073 (1988). Se diseñan procedimientos para determinar la identidad para dar la mayor coincidencia entre las secuencias analizadas. Además, los procedimientos para determinar la identidad se codifican en programas informáticos disponibles para el público. Los procedimientos de programas informáticos para determinar la identidad entre dos secuencias incluyen, entre otros, el paquete del programa GCG (Devereux, J., y col., Nucleic Acids Research 12(1): 387 (1984)), BLASTP, BLASTN y FASTA (Altschul, S.F. y col., J. Molec. Biol. 215: 403-410 (1990). La familia de programas BLAST X está disponible públicamente en NCBI y otras fuentes BLAST Manual, Altschul, S., y col., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., y col., J. Mol. Biol. 215: 403-410 (1990). También se puede usar el algoritmo bien conocido de SmithWaterman para determinar la identidad.

Los parámetros para la comparación de secuencias de polipéptidos incluyen los siguientes:

Algoritmo: Needleman y Wunsch, J. Mol Biol. 48: 443-453 (1970)

Matriz de comparación: BLOSSUM62 de Henikoff y Henikoff, Proc. Natl. Acad. Sci. USA. 89:10915-10919 (1992)

Sanción por hueco: 12

5 Sanción de longitud de hueco: 4

Un programa útil con estos parámetros está públicamente disponible como el programa de "huecos" de Genetics Computer Group, Madison WI. Los parámetros anteriormente mencionados son los parámetros por defecto para comparaciones de péptidos (junto con ausencia penalización para los huecos en los extremos).

Los parámetros para la comparación de polinucleótidos incluyen los siguientes:

10 Algoritmo: Needleman y Wunsch, J. Mol Biol. 48: 443-453 (1970)

Matriz de comparación: apareamientos = + 10, apareamiento erróneo = 0

Sanción por hueco: 50

Sanción de longitud de hueco: 3

Disponible como: El programa de "huecos" de Genetics Computer Group, Madison WI.

15 Estos son los parámetros por defecto para las comparaciones de los ácidos nucleicos.

Un significado de "identidad" para polinucleótidos y polipéptidos, como puede ser el caso, se proporcionan en (1) y (2) más adelante.

(1) Las realizaciones de polinucleótidos incluyen adicionalmente un polinucleótido aislado que comprende una secuencia de polinucleótidos que tiene al menos 50, 60, 70, 80, 85, 90, 95, 97 o 100 % de identidad con una secuencia de referencia, por ejemplo, la SEC ID N° 3, en la que dicha secuencia de polinucleótidos puede ser idéntica a la secuencia de referencia de la SEC ID N° 3 o puede incluir hasta un cierto número entero de alteraciones de nucleótidos comparada con la secuencia de referencia, en la que dichas alteraciones se seleccionan del grupo constituido por al menos una delección, sustitución de nucleótidos, incluyendo transición y transversión, o inserción, y en la que dichas alteraciones se pueden producir en las posiciones terminales 5' y 3' de la secuencia de nucleótidos de referencia o en cualquier parte entre aquellas posiciones terminales, intercaladas bien individualmente entre los nucleótidos en la secuencia de referencia o bien en uno o más grupos contiguos dentro de la secuencia de referencia, y en la que dicho número de alteraciones de nucleótidos se determina multiplicando el número total de nucleótidos en la SEC ID N° 3 por el número entero que define el porcentaje de identidad dividido por 100 y después restando ese producto de dicho número total de nucleótidos en la SEC ID N° 3, o

$$n_n \leq x_n - (x_n \cdot y),$$

en la que n_n es el número de alteraciones de nucleótidos, x_n es el número total de nucleótidos en la SEC ID N° 3, y es 0,95 para 95 %, 0,97 para 97 %, 1,00 para 100 %, etc., y \cdot es el símbolo del operador de multiplicación, y en la que cualquier producto no entero de x_n e y se redondea hacia abajo al número entero más cercano antes de restarlo de x_n . Las alteraciones de una secuencia de polinucleótidos que codifica un polipéptido pueden crear mutaciones sin sentido, de sentido erróneo o mutaciones de desplazamiento del marco de lectura en esta secuencia codificante y, por lo tanto, alterar el polipéptido codificado por el polinucleótido después de dichas alteraciones.

(2) Las realizaciones de polipéptidos además incluyen un polipéptido aislado que comprende un polipéptido que tiene al menos, 50, 60, 70, 80, 85, 90, 95, 97 o 100 % de identidad a la secuencia de polipéptidos de referencia tal como la SEC ID N° 4, en la que dicha secuencia de polinucleótidos puede ser idéntica a la secuencia de referencia o puede incluir hasta un cierto número entero de alteraciones de aminoácidos comparada con la secuencia de referencia, en la que dichas alteraciones se seleccionan entre el grupo constituido por la menos una delección, sustitución de aminoácidos, incluyendo sustitución, conservadora o y no conservadora, o inserción, y en la que dichas alteraciones se pueden producir en las posiciones terminales amino- o carboxi de la secuencia de polipéptidos de referencia o en cualquier parte entre aquellas posiciones terminales, intercaladas o bien individualmente entre los aminoácidos en la secuencia de referencia o en uno o más grupos contiguos dentro de la secuencia de referencia, y en la que dicho número de alteraciones de aminoácidos se determina multiplicando el número total de aminoácidos por el número entero que define el porcentaje de identidad dividido por 100 y después restando ese producto de dicho número total de aminoácidos, o:

$$n_a \leq x_a - (x_a \cdot y),$$

en la que n_a es el número de alteraciones de aminoácidos, x_a es el número total de aminoácidos en la secuencia, y es 0,95 para 95 %, 0,97 para 97 %, 1,00 para 100 %, $y \cdot$ es el símbolo del operador de multiplicación, y en la que cualquier producto no entero de x_a e y se redondea hacia abajo al número entero más cercano antes de restarlo de x_a .

- 5 "Aislado" significa alterado "por la mano del hombre" a partir de su estado natural, es decir, si se produce en la naturaleza, se ha cambiado o eliminado de su ambiente natural, o ambos. Por ejemplo, un polinucleótido o un polipéptido de origen natural en un organismo vivo no está "aislado", pero el mismo polinucleótido o polipéptido separado de los materiales coexistentes de su estado natural está "aislado", incluyendo, entre otros, cuando dicho polinucleótido o polipéptido se introduce de nuevo en una célula.
- 10 Un ácido nucleico o polinucleótido "aislado" o "sustancialmente puro" (por ejemplo, un ARN, ADN o un polímero mixto) es uno que está sustancialmente separado de otros componentes celulares que acompaña de forma natural al polinucleótido nativo en su célula huésped natural, por ejemplo ribosomas, polimerasas y secuencias genómicas con las que está asociado de forma natural. El término abarca un ácido nucleico o polinucleótido que (1) se ha retirado de su entorno de origen natural, (2) no está asociado con todo o una porción de un polinucleótido en el que
- 15 el "polinucleótido aislado" se encuentra en la naturaleza, (3) está unido operativamente a un polinucleótido al que no está unido en la naturaleza, o (4) no se produce en la naturaleza. El término "aislado" o "sustancialmente puro" se puede utilizar también en referencia a aislados de ADN recombinante o clonado, análogos de polinucleótidos sintetizados químicamente, o análogos de polinucleótidos que son sintetizados biológicamente por sistemas heterólogos.
- 20 Sin embargo, "aislado" no requiere necesariamente que el ácido nucleico o polinucleótido descrito de este modo ha sido eliminado en sí físicamente de su entorno nativo. Por ejemplo, una secuencia de ácido nucleico endógeno en el genoma de un organismo se considera "aislado" en el presente documento si una secuencia heteróloga se coloca adyacente a la secuencia de ácido nucleico endógeno, de manera que la expresión de esta secuencia de ácido nucleico endógeno se altera, por ejemplo, se ha aumentado, disminuido o eliminado. En este contexto, una
- 25 secuencia heteróloga es una secuencia que no está de forma natural adyacente a la secuencia de ácido nucleico endógeno, ya sea o no la secuencia heteróloga es en sí misma endógena (procedente de la misma célula huésped o progenie de la misma) o exógena (procedente de una célula huésped diferente o progenie de la misma). A modo de ejemplo, una secuencia promotora puede sustituirse (por ejemplo, mediante recombinación homóloga) por el promotor nativo de un gen en el genoma de una célula huésped, de tal manera que este gen tiene un patrón de expresión alterado. Este gen se convertiría ahora en "aislado" porque está separado de al menos algunas de las
- 30 secuencias que lo flanquean de forma natural.
- Un ácido nucleico también se considera "aislado" si contiene cualquier modificación que no se produce de forma natural en el ácido nucleico correspondiente en un genoma. Por ejemplo, una secuencia codificante endógena se considera "aislada" si contiene una inserción, delección o una mutación puntual introducida artificialmente, por
- 35 ejemplo, por intervención humana. Un "ácido nucleico aislado" también incluye un ácido nucleico integrado en un cromosoma de la célula huésped en un sitio heterólogo y una construcción de ácido nucleico presente como un episoma. Además, un "ácido nucleico aislado" puede estar sustancialmente libre de otro material celular, o sustancialmente libre de medio de cultivo cuando se produce por técnicas recombinantes, o sustancialmente libre de precursores químicos u otros productos químicos cuando se sintetiza químicamente.
- 40 Tal como se usa en el presente documento, "secuencia de ácido nucleico que codifica un producto génico funcional" se refiere a cualquier porción de una parte codificante de un gen. La secuencia de ácido nucleico que codifica un producto génico funcional puede ser una porción de una enzima que es capaz de hacer al menos una actividad la enzima completa o de la totalidad de la enzima.
- Tal como se usa en el presente documento, "ácido nucleico necesario para la expresión de al menos un producto
- 45 génico" se refiere a una secuencia de ácido nucleico que codifica cualquier porción de un gen y/o está unido operativamente a un ácido nucleico que codifica un producto génico, pero no comprende necesariamente secuencia de codificación. A modo de ejemplo, una secuencia de ácido nucleico necesaria para la expresión de al menos un producto génico incluye, pero no se limita a, potenciadores, promotores, secuencias reguladoras, codones de iniciación, codones de terminación, secuencias de poliadenilación, y/o secuencias de codificación.
- 50 Tal como se usa en el presente documento, "proteolisis" o "producto génico responsable de la proteolisis en una célula" se refiere a cualquier péptido, polipéptido, proteína y/o enzima o porción del mismo capaz de provocar la escisión de al menos un péptido, polipéptido y/o proteína. El producto génico responsable de la proteolisis puede ser directamente responsable de la escisión (es decir, una peptidasa) o puede ser indirectamente responsable como
- 55 parte de una vía de síntesis de peptidasa. Ejemplos de productos génicos que son responsables de la proteolisis en una célula incluyen, pero no se limitan a, aspartil proteasas, serín proteasas, aspartil proteasas secretadas, serín proteasas secretadas, proteasas metiltróficas de levadura, endopeptidasas de tipo DPP IV, metaloendopeptidasas, serín proteasas de tipo Prb1, serín proteasas Prb1 y carboxipeptidasas de tipo CPY. Asimismo se incluyen en esta definición la proteasa que puede ser secretada por una célula, pero todavía mantener parte o toda la actividad de proteolisis, tal como una serín proteasa secretada. Una proteasa secretada puede ser responsable de la proteólisis
- 60 dentro de la célula y/o fuera de la célula.

Tal como se usa en el presente documento, "glicosilación" o "producto génico responsable de la glicosilación en una célula" se refiere a cualquier péptido, polipéptido, proteína y/o enzima o porción de los mismos que participan en la adición de al menos un resto de sacárido a un polipéptido o alargamiento de al menos una cadena de sacárido en la célula. El producto génico responsable de la glicosilación en una célula puede ser directamente responsable de la adición de un sacárido a un polipéptido en una célula, por ejemplo, entre otros, manosiltransferasas. Las manosiltransferasas pueden transferir un residuo de Dol-P-Man a un residuo de serina y/o treonina en un péptido, polipéptido y/o proteína o pueden actuar para transferir un residuo de manosa de GPD-Man a un sacárido, alargando de este modo la cadena de sacárido. Como alternativa, el producto génico responsable de la glicosilación puede ser parte de una ruta de glicosilación y puede ser indirectamente responsable de la adición de polisacárido a un polipéptido en una célula. Ejemplos de productos génicos que son responsables de la glicosilación en una célula incluyen, pero no se limitan a, manosiltransferasas.

Tal como se usa en el presente documento, "actividad de proteasa 1 metiltrófica de levadura (Ymp1)" se refiere a cualquier actividad que la proteína identificada en el presente documento como SEC ID N° 4 pueda realizar. Por ejemplo, la actividad de proteasa 1 metiltrófica de levadura o actividad Ympl incluye, pero no se limita a, la capacidad de una enzima para escindir proteolíticamente un péptido, polipéptido, o proteína. En particular, la actividad Ympl puede referirse a la capacidad de una enzima para escindir un polipéptido que comprende un fragmento y/o variante de un agonista de GLP-1, tal como, pero no limitado a, GLP-1 humano y/o albúmina humana. Al menos una actividad Ymp1 incluye, pero no se limita a, al menos una actividad de serín proteasa.

"Polinucleótido (s)" se refiere generalmente a cualquier polirribonucleótido o polidesoxirribonucleótido que puede ser ARN o ADN no modificado o ARN o ADN modificado. "Polinucleótido (s)" incluyen, sin limitación, ADN mono y bicatenario, ADN que es una mezcla de regiones mono y bicatenarias o regiones mono, bi y tricatenarias, simple, ARN, y ARN mono y bicatenario y ARN que es una mezcla de regiones mono y bicatenarias, moléculas híbridas que comprenden ADN y ARN que pueden ser monocatenarias o, más típicamente, regiones bicatenarias o tricatenarias, o una mezcla de regiones mono y bicatenarias. Además, "polinucleótido" tal como se utiliza en el presente documento se refiere a regiones de triple hebra que comprenden ARN o ADN o ambos ARN y ADN. Las hebras en tales regiones pueden ser de la misma molécula o de diferentes moléculas. Las regiones pueden incluir todas de una o más de las moléculas, pero más típicamente implican solamente una región de algunas de las moléculas. Una de las moléculas de una región de triple hélice a menudo es un oligonucleótido. Tal como se utiliza en el presente documento, el término "polinucleótido (s)" incluye también ADN o ARN como se ha descrito anteriormente que comprenden una o más bases modificadas. Por lo tanto, los ADN o ARN con esqueletos modificados para la estabilidad o por otras razones son "polinucleótido (s)" tal como se pretende con ese término en el presente documento. Además, los ADN o ARN que comprenden bases inusuales, tales como inosina, o bases modificadas, tales como bases tritiladas, por citar sólo dos ejemplos, son polinucleótidos como se usa el término en el presente documento. Se apreciará que se ha realizado una gran variedad de modificaciones se en el ADN y el ARN que sirven para muchos propósitos útiles conocidos por los expertos en la técnica. El término "polinucleótido (s)", como se emplea en el presente documento abarca dichas de polinucleótidos formas químicamente, enzimáticamente o metabólicamente modificadas, así como las formas químicas de ADN y ARN características de virus y células, incluyendo, por ejemplo, células simples y complejas. "Polinucleótido (s)" también abarca polinucleótidos cortos a menudo denominados oligonucleótido (s).

"Polipéptido(s)" se refiere a cualquier péptido o proteína que comprende dos o más aminoácidos unidos entre sí por enlaces peptídicos o enlaces peptídicos modificados. "Polipéptido(s)" se refiere tanto a cadenas cortas, normalmente denominadas péptidos, oligopéptidos y oligómeros, y a cadenas más largas, en general denominadas proteínas. Los polipéptidos pueden comprender aminoácidos distintos a los 20 aminoácidos codificados en los genes. Los polipéptido (s)" incluyen los modificados ya sea mediante procesos naturales, tales como procesamiento y otras modificaciones postraduccionales, pero también mediante técnicas de modificación química. Tales modificaciones están bien descritas en textos básicos y en monografías más detalladas, así como en una voluminosa bibliografía de investigación, y son bien conocidos por los expertos en la técnica. Se apreciará que el mismo tipo de modificación puede estar presente en el mismo grado o en grado variable en varios sitios en un polipéptido dado. Asimismo, un polipéptido dado puede comprender muchos tipos de modificaciones. Las modificaciones se pueden producir en cualquier punto en un polipéptido, incluida la estructura peptídica, las cadenas laterales de aminoácidos y los extremos amino o carboxilo. Las modificaciones incluyen, por ejemplo, acetilación, acilación, ADP-ribosilación, amidación, unión covalente de flavina, unión covalente de un resto hemo, unión covalente de un nucleótido o derivado nucleotídico, unión covalente de un lípido o derivado lipídico, unión covalente de fosfatidilinositol, reticulación, ciclación, formación de puentes disulfuro, desmetilación, formación de reticulaciones covalentes, formación de cisteína, formación de piroglutamato, formilación, carboxilación gamma, glicosilación, formación de anclajes GPI, hidroxilación, yodinación, metilación, miristoilación, oxidación, pegilación, procesamiento proteolítico, fosforilación, prenilación, racemización, glicosilación, unión lipídica, sulfatación, carboxilación gamma de restos de ácido glutámico, hidroxilación y ADP-ribosilación, selenoilación, sulfatación, adición mediada por ARN de transferencia de aminoácidos a proteínas, tales como arginilación y ubiquitinación. Véase, por ejemplo, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2ª Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 y Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pág. 1-12 en POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter y col., Meth. Enzymol. 18: 626-646 (1990) y Rattan y col., Protein Synthesis: Posttranslational

Modifications and Aging, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). Los polipéptidos pueden ser ramificados o cíclicos, con o sin ramificación. Los polipéptidos cíclicos, ramificados y cíclicos ramificados pueden ser el resultado de procesos naturales postraduccionales o pueden fabricarse también mediante procedimientos completamente sintéticos,

5 "Sistema(s) de expresión recombinante" se refiere a sistemas de expresión o porciones de los mismos o polinucleótidos de la invención introducidos, transfectados o transformados en una célula huésped o lisado de célula huésped para la producción de los polinucleótidos y polipéptidos de la invención.

10 "Variante(s)", como el término se usa en el presente documento, se refiere a un polinucleótido o polipéptido que se diferencia de un polinucleótido o polipéptido, respectivamente, de referencia, pero conserva las propiedades esenciales. Una variante típica de un polinucleótido difiere en una secuencia de nucleótidos de otro, polinucleótido de referencia. Los cambios en la secuencia de nucleótidos de la variante pueden o no alterar la secuencia de aminoácidos de un polipéptido codificado por el polinucleótido de referencia. Los cambios de nucleótidos pueden dar como resultado sustituciones, adiciones, deleciones, proteínas de fusión y truncamientos de aminoácidos en el polipéptido codificado por la secuencia de referencia, como se describe más adelante. Una variante típica de un polipéptido difiere en una secuencia de aminoácidos de otro, polipéptido de referencia. Generalmente, las diferencias se limitan de manera que las secuencias del polipéptido de referencia y la variante sean estrechamente similares en conjunto y, en muchas regiones, idénticas. Una variante del polipéptido de referencia se puede diferenciar en la secuencia de aminoácidos en una o más sustituciones, adiciones, deleciones en cualquier combinación. Un residuo de aminoácidos sustituido o insertado puede o no puede estar codificado por el código genético. También se describen variantes de cada uno de los polipéptidos, esto es, polipéptidos que varían de los referentes en sustituciones de aminoácidos conservativas, en las que un residuo es sustituido por otro con características similares. De estas sustituciones las típicas se dan entre Ala, Val, Leu e Ile; entre Ser y Thr; entre los residuos ácidos Asp y Glu; entre Asn y Gln; y entre los residuos básicos Lys y Arg; o los residuos aromáticos Phe y Tyr. Particularmente hay variantes en las que varios, 5-10, 1-5, 1-3, 1-2 i 1 aminoácidos se sustituyen, eliminan, o añaden en cualquier combinación. Una variante de un polinucleótido o polipéptido puede ser de origen natural tal como una variante alélica, o puede ser una variante que no se conoce que sea de origen natural. Variantes no naturales de polinucleótidos y polipéptidos pueden hacerse mediante técnicas de mutagénesis, mediante síntesis directa y mediante otros procedimientos recombinantes conocidos por los expertos en la técnica.

30 "Microorganismo(s)" significa un (i) procarionta, incluyendo, pero no limitado a, un miembro del género *Streptococcus*, *Staphylococcus*, *Bordetella*, *Corynebacterium*, *Mycobacterium*, *Neisseria*, *Haemophilus*, *Actinomycetes*, *Streptomyces*, *Nocardia*, *Enterobacter*, *Yersinia*, *Fancisella*, *Pasturella*, *Moraxella*, *Acinetobacter*, *Erysipelothrix*, *Branhamella*, *Actinobacillus*, *Streptobacillus*, *Listeria*, *Calymmatobacterium*, *Brucella*, *Bacillus*, *Clostridium*, *Treponema*, *Escherichia*, *Salmonella*, *Klebsiella*, *Vibrio*, *Proteus*, *Erwinia*, *Borrelia*, *Leptospira*, *Spirillum*, *Campylobacter*, *Shigella*, *Legionella*, *Pseudomonas*, *Aeromonas*, *Rickettsia*, *Chlamydia*, *Borrelia* y *Mycoplasma*, e incluyendo además, entre otros, un miembro de la especie o grupo Estreptococo del grupo A, Estreptococo del grupo B, Estreptococo del grupo C, Estreptococo del grupo D, Estreptococo del grupo G, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus faecalis*, *Streptococcus faecium*, *Streptococcus durans*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Corynebacterium diphtheriae*, *Gardnerella vaginalis*, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium ulcerans*, *Mycobacterium leprae*, *Actinomyces israelii*, *Listeria monocytogenes*, *Bordetella pertussis*, *Bordetella parapertussis*, *Bordetella bronchiseptica*, *Escherichia coli*, *Shigella dysenteriae*, *Haemophilus influenzae*, *Haemophilus aegyptius*, *Haemophilus parainfluenzae*, *Haemophilus ducreyi*, *Bordetella*, *Salmonella typhi*, *Citrobacter freundii*, *Proteus mirabilis*, *Proteus vulgaris*, *Yersinia pestis*, *Klebsiella pneumoniae*, *Serratia marcescens*, *Serratia liquefaciens*, *Vibrio cholera*, *Shigella dysenteriae*, *Shigella flexneri*, *Pseudomonas aeruginosa*, *Francisella tularensis*, *Brucella abortus*, *Bacillus anthracis*, *Bacillus cereus*, *Clostridium perfringens*, *Clostridium tetani*, *Clostridium botulinum*, *Treponema pallidum*, *Rickettsia rickettsii* Y *Chlamydia trachomatis*, (ii) una arqueobacteria, incluyendo, entre otros, Archaeobacter, y (iii) un eucariota unicelular o filamentoso, incluyendo, entre otros, un protozoo, un hongo, un miembro del género *Saccharomyces*, *Kluveromyces*, o *Candida*, y un miembro de la especie *Saccharomyces cerevisiae*, *Kluveromyces lactis*, o *Candida albicans*.

50 "Bacteria (I)" significa un (i) procarionta, incluyendo, pero no limitado a, un miembro del género *Streptococcus*, *Staphylococcus*, *Bordetella*, *Corynebacterium*, *Mycobacterium*, *Neisseria*, *Haemophilus*, *Actinomycetes*, *Streptomyces*, *Nocardia*, *Enterobacter*, *Yersinia*, *Fancisella*, *Pasturella*, *Moraxella*, *Acinetobacter*, *Erysipelothrix*, *Branhamella*, *Actinobacillus*, *Streptobacillus*, *Listeria*, *Calymmatobacterium*, *Brucella*, *Bacillus*, *Clostridium*, *Treponema*, *Escherichia*, *Salmonella*, *Klebsiella*, *Vibrio*, *Proteus*, *Erwinia*, *Borrelia*, *Leptospira*, *Spirillum*, *Campylobacter*, *Shigella*, *Legionella*, *Pseudomonas*, *Aeromonas*, *Rickettsia*, *Chlamydia*, *Borrelia* y *Mycoplasma*, e incluyendo además, entre otros, un miembro de la especie o grupo Estreptococo del grupo A, Estreptococo del grupo B, Estreptococo del grupo C, Estreptococo del grupo D, Estreptococo del grupo G, *Streptococcus*, *Streptococcus pneumoniae*, *Streptococcus pyogenes*, *Streptococcus agalactiae*, *Streptococcus faecalis*, *Streptococcus faecium*, *Streptococcus durans*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Corynebacterium diphtheriae*, *Gardnerella vaginalis*, *Mycobacterium tuberculosis*, *Mycobacterium bovis*, *Mycobacterium ulcerans*, *Mycobacterium leprae*, *Actinomyces israelii*, *Listeria monocytogenes*, *Bordetella pertussis*, *Bordetella parapertussis*, *Bordetella bronchiseptica*, *Escherichia coli*, *Shigella dysenteriae*, *Haemophilus influenzae*, *Haemophilus aegyptius*, *Haemophilus parainfluenzae*, *Haemophilus ducreyi*,

Bordetella, *Salmonella typhi*, *Citrobacter freundii*, *Proteus mirabilis*, *Proteus vulgaris*, *Yersinia pestis*, *Klebsiella pneumoniae*, *Serratia marcescens*, *Serratia liquefaciens*, *Vibrio cholera*, *Shigella dysenterii*, *Shigella flexneri*, *Pseudomonas aeruginosa*, *Fransciscella tularensis*, *Brucella abortis*, *Bacillus anthracis*, *Bacillus cereus*, *Clostridium perfringens*, *Clostridium tetani*, *Clostridium botulinum*, *Treponema pallidum*, *Rickettsia rickettsii* y *Chlamydia trachomatis*, y (ii) una arqueobacteria incluyendo, entre otras, *Archaeobacter*.

Como se usa en el presente documento, "polipéptido(s) heterólogo (s)" se refiere a un polipéptido no sintetizado de forma natural por una célula huésped transformada o microorganismo de interés e introducido en la célula huésped o microorganismo por ADN recombinante. Por ejemplo, *Pichia* puede actuar como una célula huésped para la expresión de seroalbúmina humana, que no se produce en *Pichia* no transformada o no transfectada. Los polipéptidos heterólogos pueden incluir polipéptidos que han sido modificados para facilitar el aislamiento.

Tal como se usa en el presente documento "marcador de afinidad" se refiere a cualquier resto asociado con una molécula que puede dar a dicha molécula una afinidad selectiva por otra sustancia o molécula. Por ejemplo, un marcador de afinidad puede usarse para facilitar la purificación de una molécula proporcionando a la molécula una afinidad selectiva por el material de empaquetado de la columna. Un ejemplo no limitante de un marcador afinidad es un marcador de His.

Una secuencia de ácido nucleico está "unida operativamente" cuando se coloca en una relación funcional con otra secuencia de ácido nucleico. Por ejemplo, el ADN para una presecuencia o líder secretor está unido operativamente a ADN para un polipéptido si se expresa como una preproteína que participa en la secreción del polipéptido; un promotor o potenciador está unido operativamente a una secuencia codificante si afecta a la transcripción de la secuencia; o un sitio de unión al ribosoma está unido operativamente a una secuencia codificante si está colocado de manera que facilita la traducción. Generalmente, "unido operativamente" significa que las secuencias de ADN que están unidas son contiguas, y, en el caso de un líder secretor, contiguas y en fase de lectura. Sin embargo, los potenciadores no tienen que ser contiguos. La unión se logra mediante uniones en sitios de restricción convenientes. Si tales sitios no existen, los ligadores o adaptadores de oligonucleótidos sintéticos se usan de acuerdo con la práctica convencional.

Como se usa en el presente documento, "cosechado" de células se refiere a la recolección de células del cultivo celular. Las células pueden concentrarse durante el cosechado para separarlos del caldo de cultivo, por ejemplo por centrifugación o filtración. El cosechado de las células pueden comprender además la etapa de lisis de las células para obtener material intracelular, tal como, pero no limitado a, polipéptidos y polinucleótidos. El experto en la técnica debe entender que cierto material celular, incluyendo, pero no limitado a, polipéptido expresado de forma heteróloga, puede liberarse de las células durante el cultivo. Por lo tanto, un producto (por ejemplo, un polipéptido expresado de forma heteróloga) de interés puede permanecer en caldo de cultivo después de cosechar las células.

Además se proporcionan procedimientos en los que la construcción de ADN recombinante codifica un marcador seleccionable. Tal marcador seleccionable proporciona una selección positiva o negativa. También se proporcionan procedimientos que comprenden expresar dicho marcador seleccionable y comparar la cantidad de marcador seleccionable producido por al menos una primera célula transformada de la etapa de selección con la cantidad de marcador seleccionable producido por al menos una segunda célula transformada de la etapa de selección en el que la primera y segunda célula transformada produce el mismo marcador seleccionable. Como se entiende en la técnica, los marcadores seleccionables incluyen, pero no se limitan a, dihidrofolato reductasa (dhfr), β -galactosidasa, proteína fluorescente, forma secretada de la fosfatasa alcalina placentaria humana, beta-glucuronidasa, los marcadores seleccionables de levadura Leu 2 y URA3, genes resistentes a la apoptosis y oligonucleótidos antisentido, así como genes de resistencia a antibióticos que confieren la capacidad de crecer en presencia de antibióticos, incluyendo, neomicina (neo), kanamicina, geneticina, higromicina B, puromicina, zeocina, blasticidina, nourseotricina, bialafos, fleomicina y ampicilina. Como también se entiende en la técnica, las células pueden ordenarse por diversos medios, incluyendo pero no limitados a, inspección visual o un clasificador de células tal como un BD FACS Aria, que puede detectar la expresión de un marcador seleccionable.

El término "tipo salvaje" como se entiende en la técnica se refiere a una célula huésped o una secuencia de polipéptidos o polinucleótidos que se produce en una población nativa sin modificación genética. Por ejemplo, un "padre de tipo salvaje de una célula huésped" se refiere a una cepa no modificada de una célula huésped antes de realizar o que se produzca cualquier modificación genética en el genoma de la célula huésped.

Como se usa en el presente documento, "rendimiento del título" se refiere a la concentración de un producto (por ejemplo, polipéptido, expresado de forma heteróloga) en solución (por ejemplo, caldo de cultivo o mezcla de lisis – células o tampón) y se expresa generalmente como mg/l o g/l. Un aumento en el rendimiento título puede referirse a un aumento absoluto o relativo de la concentración de un producto producido con dos conjunto definidos de condiciones.

Hormona incretina" como se usa en el presente documento significa cualquier hormona que potencia la secreción de insulina o, de otra manera, aumenta el nivel o insulina. Un ejemplo de una hormona incretina es GLP-1. GLP-1 es una incretina secretada por las células L intestinales en respuesta a la ingestión de alimentos. En un individuo sano, GLP-1 desempeña un papel importante en la regulación de los niveles posprandiales de glucosa en sangre mediante la estimulación de la secreción de insulina dependiente de la glucosa por el páncreas lo que da como resultado una

mayor absorción de la glucosa en la periferia. GLP-1 también suprime la secreción de glucagón, que conduce a la reducción de la producción hepática de glucosa. Además, GLP-1 retrasa el tiempo de vaciado gástrico y reduce la motilidad del intestino delgado, de modo que retrasa la absorción de los alimentos. GLP-1 estimula la competencia continuada de las células beta mediante la estimulación de la transcripción de genes implicados en la secreción de insulina dependiente de glucosa y mediante la estimulación de la neogénesis de células beta (Meier, et al. *Biodrugs* 2003; 17 (2): 93-102).

"Actividad de GLP-1" como se usa en el presente documento significa una o más de las actividades del GLP-1 humano de origen natural, incluyendo, pero no limitado a, reducción de la glucosa en sangre y/o en plasma, estimulación de la secreción de insulina dependiente de la glucosa o, de otro modo, elevación del nivel o la insulina, supresión de la secreción de glucagón, reducción de la fructosamina, aumento de la liberación de glucosa y el metabolismo en el cerebro, retraso del vaciado gástrico y estimulación de la competencia de las células beta, y/o neogénesis. Cualquiera de estas actividades y otras actividades asociadas con la actividad de GLP-1 pueden deberse directa o indirectamente a una composición que tiene actividad de GLP-1 o un agonista de GLP-1. A modo de ejemplo, una composición que tiene actividad de GLP-1 puede estimular directa o indirectamente dependiente de la glucosa, mientras que la estimulación de la producción de insulina puede reducir indirectamente los niveles de glucosa en plasma en un mamífero.

Un "mimético de incretina" como se usa en el presente documento es un compuesto capaz de potenciar la secreción de insulina o, de otra manera, aumentar el nivel o insulina. Un mimético de incretina puede ser capaz de estimular la secreción de insulina, aumentar la neogénesis de las células beta, inhibir la apoptosis de las células beta, inhibir la secreción de glucagón, retrasar el vaciado gástrico e inducir de saciedad en un mamífero. Un mimético de incretina puede incluir, pero no se limita a, cualquier polipéptido que tiene actividad de GLP-1, incluyendo, pero no limitado a, exendina 3 y exendina 4, incluyendo cualquier fragmento y/o variante y/o conjugado de los mismos.

Tal como se usa en el presente documento, "fragmento", cuando se utiliza en referencia a un polipéptido, es un polipéptido que tiene una secuencia de aminoácidos que es la misma como parte, pero no toda, la secuencia de aminoácidos de todo el polipéptido de origen natural. Los fragmentos pueden estar "libres" o comprendidos dentro de un polipéptido más grande del cual forman una parte o región como una única región continua en un solo polipéptido más grande. A modo de ejemplo, un fragmento de GLP-1 de origen natural sería incluir los aminoácidos 7 a 36 de los aminoácidos de origen natural 1 a 36. Además, los fragmentos de un polipéptido también pueden ser variantes de la secuencia parcial de origen natural. Por ejemplo, un fragmento de GLP-1 que comprende los aminoácidos 7-30 de GLP-1 de origen natural también puede ser una variante que tiene sustituciones de aminoácidos dentro de su secuencia parcial.

Tal como se usa en el presente documento, "conjugado" o "conjugados" se refiere a dos moléculas que están unidas entre sí. Por ejemplo, un primer polipéptido puede estar unido covalente o no covalentemente a un segundo polipéptido. El primer polipéptido puede estar unido covalentemente mediante un enlazador químico o puede fusionarse genéticamente al segundo polipéptido, en el que el primero y el segundo polipéptido comparten un esqueleto polipeptídico común. Conjugado puede comprender al menos un polipéptido terapéutico conjugado a la seroalbúmina humana. Otros conjugados también incluyen, pero no se limitan a, al menos un polipéptido terapéutico conjugado a la transferrina, un dominio variable de cadena única, y/o al menos una región Fc de un anticuerpo. Los conjugados pueden comprender o no un enlazador.

Tal como se usa en el presente documento "orientado en tándem" se refiere a dos o más polipéptidos que están adyacentes uno a otro como parte de la misma molécula. Pueden estar unidos forma covalente o no covalente. Dos o más polipéptidos orientados en tándem pueden formar parte del mismo esqueleto del polipéptido. Los polipéptidos orientados en tándem pueden tener orientación directa o invertida y/o pueden estar separados por otras secuencias de aminoácidos.

Un "anticuerpo de dominio" o "dAb" se puede considerar lo mismo que un "dominio variable único" que es capaz de unirse a un antígeno. Un dominio variable único puede ser un dominio variable de anticuerpo humano, pero también incluye dominios variables únicos de anticuerpo de otras especies, como dAb V_{HH} de roedores (por ejemplo, tal como se divulga en el documento WO 00/29004), de tiburón nodriza y de *Camelid*. El V_{HH} de camélidos son polipéptidos en el dominio variable único de inmunoglobulina que derivan de especies, entre las que se incluyen camellos, llamas, alpacas, dromedarios y guanacos, que producen anticuerpos de cadena pesada desprovistos de forma natural de las cadenas ligeras. Dichos dominios V_{HH} pueden humanizarse de acuerdo con técnicas estándar disponibles en la técnica y dichos dominios se consideran "anticuerpos de dominio" de acuerdo con la invención. Como se usa en el presente documento V_H incluye dominios V_{HH} de camélidos.

La frase "dominio variable único" se refiere a un dominio variable de la proteína de unión al antígeno (por ejemplo, V_H , V_{HH} , V_L) que se une específicamente a un antígeno o epítipo independientemente de una región o dominio diferente.

La expresión "proteína de unión al antígeno", como se usa en el presente documento se refiere a anticuerpos, fragmentos de anticuerpo y otras construcciones proteicas, tales como dominios, pero no se limitan a dominios variables y anticuerpos de dominios, que pueden unirse a un antígeno.

Como se usa en el presente documento, "cantidad reducida" de una enzima o fragmento de la misma o actividad enzimática en comparación con una célula huésped genéticamente modificado se refiere a una célula huésped modificada genéticamente que produce menos de al menos una enzima o muestra menos de al menos un tipo de actividad enzimática cuando se compara con una célula huésped no modificadas genéticamente. Típicamente, la comparación de la actividad enzimática producida por una célula huésped modificada genéticamente es con la cepa de tipo salvaje de la misma especie antes de la modificación genética. Sin embargo, la comparación también puede ser entre el huésped modificado genéticamente y un huésped de tipo salvaje del género pero diferentes especie o cepa o con otra cepa modificada genéticamente. Una reducción en al menos una enzima o actividad enzimática también incluye una anulación completa de al menos una enzima o actividad enzimática en la que ninguno de al menos una enzima se produce en una célula huésped modificada genéticamente y/o ninguno de al menos una enzima es funcional o muestra actividad. También se incluye dentro de esta definición una cantidad reducida de al menos una actividad enzimática. Es decir, las enzimas que tienen más de una actividad pueden mantener la cantidad de una primera actividad, mientras que se reduce una segunda actividad de la misma enzima.

Como se usan en el presente documento, las expresiones "condiciones rigurosas" y "condiciones de hibridación rigurosas" quieren decir que la hibridación se producirá solo si hay al menos una identidad del 70 % y al menos del 80 %, pero de al menos un 95 %, entre las secuencias. Un ejemplo de condiciones de hibridación rigurosas es la incubación durante toda la noche a 42 °C en una solución que comprende: 50 % de formamida, 5x SSC (NaCl 150 mM, citrato trisódico 15 mM), fosfato sódico 50 mM (pH 7,6), 5x solución de Denhardt, 10 % de sulfato de dextrano y 20 microgramos/ml de ADN de esperma de salmón cortado y desnaturalizado, seguido de lavado de los filtros en 0,1 x SSC a aproximadamente 65 °C. Las condiciones de hibridación y de lavado son muy conocidas y se encuentran ejemplos en Sambrook, y col., Molecular Cloning: A Laboratory Manual, Segunda Edición, Cold Spring Harbor, N.Y., (1989), particularmente en su capítulo 11.

Tal como se usa en el presente documento "modificación genética" o "genéticamente modificado" se refiere a cualquier supresión, sustitución, delección y/o inserción de una o más bases o de un fragmento de una secuencia (s) de ADN celular. Tal modificación genética puede obtenerse *in vitro* (directamente en ADN aislado) o *in situ*, por ejemplo mediante técnicas de ingeniería genética o exponiendo las células a un agente mutagénico. Agentes mutagénicos incluyen, por ejemplo, agentes físicos tales como rayos energéticos (rayos X, rayos γ , UV, etc.) o agentes químicos capaces de reaccionar con diferentes grupos funcionales de ADN, tales como agentes alquilantes (EMS, NQO, etc.) agentes bialquilantes, agentes intercalantes, etc. Las modificaciones genéticas también pueden obtenerse mediante alteración genética, por ejemplo, de acuerdo con el procedimiento divulgado por Rothstein et al. (Meth. Enzymol. 194:281-301(1991)). De acuerdo con este procedimiento, parte o la totalidad de un gen se sustituye a través de recombinación homóloga por una versión modificada *in vitro*. Las modificaciones genéticas pueden obtenerse también por cualquier mutación de tipo inserción en las secuencias de ADN, tales como transposones, fagos, etc. Además, como se usa en el presente documento "genéticamente modificado" se puede referir a un gen que codifica un polipéptido o un polipéptido que tiene al menos una delección, sustitución o supresión de un ácido nucleico o aminoácido, respectivamente. Por ejemplo, un polipéptido en el que al menos un aminoácido está sustituido por la forma de tipo salvaje sería considerado modificado genéticamente.

La modificación genética puede invertirse o atenuarse mediante un mecanismo celular. Como alternativa, las mutaciones pueden no ser reversibles y sin fugas. Las mutaciones "con fugas" incluyen mutaciones que tienen como resultado una inactivación parcial en lugar de completa de la función de tipo salvaje.

Las modificaciones genéticas realizadas por las células huésped de la invención pueden estar localizadas en una región de codificación de la secuencia de ADN de la célula y/o en una región que afecta a la expresión de un gen. Por tanto, las modificaciones generalmente afectarán al producto génico o a la regulación o estimulación del producto génico de proteínas y/o enzimas implicadas en la proteólisis y/o glicosilación. La capacidad reducida de las células para escindir proteolíticamente y/o glicosilar un polipéptido expresado de forma heteróloga puede deberse a cambios estructurales y/o conformacionales, con respecto a la producción de una o más enzimas con propiedades biológicas alteradas, con respecto a la ausencia de producción de dichas una o más enzimas o con respecto a la producción de una o más enzimas en niveles bajos.

Se describen cepas de *Pichia* modificadas genéticamente en las que al menos una secuencia de ácido nucleico que codifica un producto génico funcional y/o al menos un ácido nucleico necesario para la expresión de al menos un producto génico en dicha cepa de *Pichia* está modificada genéticamente, en el que dicho producto génico es responsable de la proteólisis y/o la glicosilación en dicha cepa de *Pichia* modificada genéticamente. Las cepas de *Pichia* modificadas genéticamente de la presente invención incluyen, pero no se limitan a, *Pichia* en la que al menos una secuencia de ácidos nucleicos que codifica al menos uno de los siguientes un producto génico funcional o expresión de dicho producto génico está modificado genéticamente: *PEP4*, *PRB1*, *YPS1*, *YPS2*, *YMP1*, *DAP2*, *GRH1*, *PRD1*, *YSP3*, *PRB3*, *YMP* y/o *YMP3*. También se describen cepas de *Pichia* modificadas genéticamente que producen una cantidad reducida, ninguna y/o al menos una reducción de la actividad de al menos una de las siguientes enzimas y/o tipo de enzima: aspartil proteasas, serín proteasas, aspartil proteasas secretadas, serín proteasas secretadas, proteasas metiltróficas de levadura, endopeptidasas de DPP IV, metaloendopeptidasas, serín proteasas de tipo Prb1, serín proteasas Prb1, carboxipeptidasas de tipo CPY y/o manosiltransferasas en comparación con la cepa de tipo salvaje. Adicionalmente, las cepas modificadas genéticamente pueden producir una enzima seleccionada de: aspartil proteasas, serín proteasas, aspartil proteasas secretadas, serín proteasas secretadas,

proteasas metiltróficas de levadura, endopeptidasas de DPP IV, metaloendopeptidasas, serín proteasas de tipo Prb1, serín proteasas Prb1, carboxipeptidasas de tipo CPY y/o manosiltransferasas, en las que dicha enzima demuestra al menos una actividad reducida de dicha enzima en comparación con una cepa de tipo salvaje y/o enzima.

- 5 Las cepas de *Pichia* genéticamente modificadas también incluyen cepas de *Pichia* en las que al menos una de las siguientes secuencias de ácidos nucleicos que codifican un producto génico funcional o la expresión de dicho producto génico está modificada genéticamente: *OCH1*, *PMT1*, *PMT2*, y/o *PMT4*. La *Pichia* modificada genéticamente producen una cantidad reducida, ninguna actividad o actividad reducida de al menos una manosiltransferasa en comparación con la cepa de tipo salvaje. Adicionalmente, las cepas modificadas
10 genéticamente pueden producir una enzima asociada con la glicosilación tal como al menos una manosiltransferasa que tiene una actividad reducida en comparación con la enzima de tipo salvaje.

Las cepas genéticamente modificadas de *Pichia* incluyen, pero no se limitan a, una forma modificada genéticamente de las cepas salvajes de *Pichia* X-33 o SMD1163. Las cepas de *Pichia* modificadas genéticamente incluyen, pero no se limitan a, una forma modificada genéticamente de tipo salvaje de *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichi salictaria*, *Pichia guercum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia*
15 sp. de tipo salvaje.

Se describen cepas de *Pichia* modificadas en las que la siguiente secuencia de ácido nucleico que codifica un producto génico funcional o la expresión de dicho producto génico está modificada genéticamente: *PEP4*, *PRB1*,
20 *YPS1*, *YPS2*, *YMP1*, *YMP2*, *YMP3* y *PMT4*. La *Pichia* modificada produce una cantidad reducida, ninguna actividad o actividad reducida de al menos uno de los siguientes productos génicos en comparación con una cepa de tipo salvaje: aspartil proteasa, serín proteasa, serín proteasa secretada y manosiltransferasa. Las cepas de *Pichia* modificadas también tienen una actividad reducida de al menos uno de los siguientes productos génicos en comparación con la cepa de tipo salvaje: al menos una aspartil proteasa, al menos una serín proteasa, al menos una
25 serín proteasa secretada y/o al menos una manosiltransferasa.

Las cepas de *Pichia* modificadas genéticamente pueden comprender además un polinucleótido capaz de expresar al menos un polipéptido heterólogo. El polinucleótido capaz de expresar al menos un polipéptido heterólogo incluyen, pero no se limitan a, vectores, ADN transformado en el genoma de la célula huésped, virus o parte de un virus, y/o plásmidos. El polinucleótido capaz de expresar un polipéptido heterólogo puede transformarse en el genoma de
30 *Pichia* y/o puede ser parte de un vector de expresión y/o sistema de expresión episomal.

Como se entiende en la técnica, el ADN puede transformarse en una célula huésped mediante varios procedimientos diferentes. En levaduras, puede usarse cualquier procedimiento conveniente de transferencia de ADN, tal como electroporación, el procedimiento del cloruro de litio o el procedimiento de esferoplasto. Para producir una cepa estable adecuada para la fermentación de alta densidad, es deseable integrar el ADN en el cromosoma huésped. La integración se produce a través de recombinación homóloga, usando técnicas conocidas en la técnica. Por ejemplo, el ADN capaz de expresar al menos una proteína heteróloga se puede proporcionar con secuencias flanqueantes homólogas a las secuencias del organismo huésped. De esta manera, la integración se produce en un sitio definido en el genoma del huésped, sin alteración de genes deseables o esenciales. Como alternativa, el ADN capaz de expresar al menos una proteína heteróloga se integra en el sitio de un gen no deseado en un cromosoma huésped, efectuando la interrupción o delección del gen o la expresión de ese producto génico. Por ejemplo, la integración en los sitios de los genes *YMP1*, *YMP2*, *YMPi*, *PEP4*, *prbl*, *YPSL*, *YPS2*, *DAP2*, *GRHL*, *prdl*, *YSP3*, *PRB3*, SEC ID N° 10, SEC ID N° 12 y/o SEC ID N° 14 permite la expresión de la proteína heteróloga al tiempo que evita la expresión de enzimas implicadas en la proteólisis en levaduras. En otras realizaciones, el ADN puede introducirse en el huésped a través de un cromosoma, plásmido, vector retroviral, o integración aleatoria en el
45 genoma huésped.

En otro aspecto de la presente invención, al menos un polipéptido heterólogo expresado en las células huésped de la presente invención comprende al menos un agonista de GLP-1. En algunos aspectos, el agonista de GLP-1 se selecciona del grupo de: hormona incretina y/o fragmento, variante y/o conjugado del mismo y mimético de incretina y/o fragmento, variante y/o conjugado del mismo. En algunos aspectos, al menos un polipéptido heterólogo tiene al menos una actividad de GLP-1.
50

Los polipéptidos que tienen actividad de GLP-1 pueden comprender al menos un fragmento y/o variante de GLP-1 humano. Los dos fragmentos de origen natural de GLP-1 humano se representan en la SEC ID N° 2

```

7  8  9  10 11 12 13 14 15 16 17
His-Ala-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-
18 19 20 21 22 23 24 25 26 27 28
Ser-Tyr-Leu-Glu-Gly-Gln-Ala-Ala-Lys-Glu-Phe-
29 30 31 32 33 34 35 36 37
Ile-Ala-Trp-Leu-Val-Lys-Gly-Arg-Xaa (SEQ ID NO:2)

```

5 en la que: Xaa en la posición 37 es Gly (en lo sucesivo designado como "GLP-1 (7-37)"), o -NH₂ (en lo sucesivo designado como "GLP-1 (7-36)"). Los fragmentos de GLP-1 pueden incluir, pero no se limitan a, moléculas de GLP-1 que comprenden, o como alternativa que consisten en, los aminoácidos 7 a 36 del GLP-1 humano (GLP-1 (7-36)). Las variantes de GLP-1 o fragmentos de los mismos pueden incluir, pero no se limitan a, uno, dos, tres, cuatro, cinco o más sustituciones de aminoácidos de GLP-1 tipo salvaje en o en los fragmentos de origen natural de GLP-1 se muestra en la SEC ID N° 2. Las variantes de GLP-1 o fragmentos de GLP-1 pueden incluir, pero no se limitan a, sustituciones de un residuo de alanina análogo a la alanina 8 del GLP-1 de tipo salvaje, estando dicha alanina mutada a una glicina (en lo sucesivo designado "A8G") (véase, por ejemplo, los mutantes divulgados en la patente de EE.UU. N° 5.545.618).

10 En otro aspecto, el al menos un polipéptido que tiene actividad de GLP-1 comprende al menos un fragmento y/o variante de GLP-1 humano fusionado con seroalbúmina humana. En otro aspecto, al menos un fragmento y la variante de GLP-1 comprenden GLP-1(7-36(A8G)). El al menos un fragmento y la variante de GLP-1 está fusionado genéticamente a seroalbúmina humana.. En otro aspecto, el polipéptido heterólogo de la presente invención
15 comprende al menos dos GLP-1 (7-36 (A8G)) en tándem y genéticamente fusionados a la seroalbúmina humana. Los dos GLP-1 (7-36 (A8G)) se fusionan genéticamente en el extremo N de la seroalbúmina humana. En algunos casos, el polipéptido heterólogo comprende la SEC ID N° 1.

```

HGEGTFTSDVSSYLEGQAAKEFIAWLKGRHGEGTFTSDVSSYLEGQAAKEFIAWLKGR 60
DAHKSEVAHRFKDLGEENFKALVLI AFAQYLQCCPFEDHVKLVNEVTEFAKTCVADESAE 120
NCDKLSHTL FGDKLC TVATLRETYGEMADCCAKQEPERNECF LQHKDDNPNL PRLVRPE 180
VDVMCTAFHDNEETFLKKYLYE IARRHPYFYAPELLFFAKRYKAAFT ECCQAADKAA CLL 240
PKLDEL RDEGKASSAKQRLK CASLQKFGERAFKAWAVARLSQRFPKAEFAEVSKLVTDLT 300
KVHTECC HGDLL ECADDRADLAKY ICENQDSISSK LKECCEKPLLEKSHCIAEVENDEMP 360
ADLPSLAADFVESKDVCKNYAEAKD VFLGMFLYEYARRHPDYSV VLLRLAKTYETLEK 420
CCAAADPHECYAKVFDEFKPLVEEPQNL IKQNCELFEQLGEYKFN ALLVRYTKKVPQVS 480
TPTLVEVSRNLGKVGSKCCKHPEAKRMP CAEDYLSVVLNQLCVLHEKTPVSDRVTKCCTE 540
SLVNR RPFCSALEVDETYVPKEFNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKA 600
TKEQLKAVMDDFAAFVEKCKADDKETCF AEKGKLVAAASQAALGL 674
(SEQ ID NO:1)

```

20 En todavía otra realización, al menos un polipéptido heterólogo expresado en las células huésped de la invención comprende uno o más de los siguientes: al menos una proteína de unión a antígeno, al menos un dominio variable sencillo, y/o al menos un anticuerpo de dominio. Los polipéptidos que comprenden al menos un dominio de unión a antígeno pueden comprender también al menos un polipéptido y/o agonista y/o antagonista del receptor del péptido. En algunos casos, el agonista del polipéptido puede ser un agonista del receptor de GLP-1. Como se entiende en la técnica, más de un polipéptido heterólogo puede expresarse en la misma célula. A modo de ejemplo, un polipéptido
25 heterólogo que tiene actividad de GLP-1 se puede expresar en la misma célula como una proteína de unión a antígeno. El polipéptido que tiene actividad de GLP-1 puede expresarse a partir el mismo polinucleótido que la proteína de unión a antígeno, unido operativamente a la secuencia de ácido nucleico necesaria para la expresión. Como alternativa, y a modo de ejemplo, un polipéptido que tiene actividad de GLP-1 puede expresarse independientemente de un segundo polipéptido heterólogo tal como una proteína de unión a antígeno, ya sea desde
30 el mismo ADN episomal o genoma pero unido operativamente a diferentes secuencias de polinucleótidos necesarias para la expresión o a partir de secuencias de ADN localizadas en vectores separados.

También, en el presente documento se proporcionan cepas de *Pichia* modificadas genéticamente, en las que dicha cepa muestra una proteólisis reducida de dicho al menos un polipéptido heterólogo en dicha cepa en comparación con *Pichia* de tipo salvaje. Además, *Pichia* modificada genéticamente puede mostrar una glicosilación reducida o
35 ausencia de glicosilación de dicho al menos un polipéptido heterólogo en dicha cepa en comparación con *Pichia* de tipo salvaje.

En otro aspecto se proporcionan procedimientos para producir un polipéptido heterólogo que comprenden expresar dicho polipéptido heterólogo en una *Pichia* modificada genéticamente de la invención. También se describen polipéptidos heterólogos producidos en una cualquiera de las *Pichia* modificadas genéticamente de la presente
40 invención.

Se describen polinucleótidos aislados que tienen una identidad de secuencia de al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 %, y/o 100 % con las SEC ID N° 3, 5, y 7. Se describen polipéptidos aislados que tienen una identidad de secuencia de al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90

%, 95 %, 99 %, y/o 100 % con las SEC ID N° 4, 6, y 8. Se describen células huésped modificadas genéticamente en las que el padre de tipo silvestre de dicha célula huésped modificada genéticamente comprende un gen que codifica un polipéptido que tiene una identidad de secuencia de al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 %, y/o 100 % con las SEC ID N° 4, 6, 8, en las que dicho gen está modificado genéticamente en el genoma de dicha célula huésped de tal manera que el producto génico se reduce o elimina en dicha célula huésped modificada genéticamente en comparación con dicha célula huésped de tipo salvaje. También se describen células huésped genéticamente modificadas en las que dicho tipo silvestre parental de dicha célula huésped modificada genéticamente comprende un gen que tiene una identidad de secuencia de al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 %, y/o 100 % con las SEC ID N° 3, 5, y 7, en las que dicho gen está modificado genéticamente en el genoma de dicha célula huésped genéticamente modificada. La célula huésped modificada genéticamente de la presente invención puede tener una actividad proteasa reducida o eliminada en comparación con dicha célula huésped parental de tipo salvaje. La célula huésped modificada genéticamente de la presente invención tiene actividad proteasa metiltrófica de levadura reducida o eliminada (Ympl). En algunos casos, la célula huésped es *Pichia*.

Se describen polinucleótidos aislados que tienen una identidad de secuencia de al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 %, y/o 100 % con las SEC ID N° 9, 11, y 13. Se describen polipéptidos aislados que tienen una identidad de secuencia de al menos 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 99 %, y/o 100 % con las SEC ID N° 10, 12, y 14.

Como se entiende en la técnica, la actividad enzimática de una o más enzimas producidas por células huésped en cultivo puede verse afectada por las condiciones de crecimiento del cultivo. Por ejemplo, la actividad proteolítica de una proteasa producida por una célula huésped en cultivo podría disminuirse mediante la alteración de una o más de las siguientes condiciones: pH, oxígeno disuelto, temperatura, osmolaridad, uno o más componentes de los medios, inhibidores específicos de la proteasa, tiempo y/o velocidad de crecimiento, concentración celular, duración del cultivo y/o la tasa de alimentación de glucosa (por ejemplo, alimentación discontinua). La adición de los hidrolizados de proteínas complejas al cultivo puede ser especialmente eficaz en la inhibición de la proteólisis. Además, las condiciones pueden alterarse en uno o más puntos específicos durante el cultivo de una manera tal que se maximice el efecto. Del mismo modo, la glicosilación de las proteínas producidas en cultivo puede verse afectada por factores similares. Por lo tanto, las condiciones de crecimiento para reducir la actividad enzimática de una célula huésped, tal como la actividad proteolítica o de glicosilación, en cultivo pueden optimizarse mediante el ajuste de uno o más de los factores no limitantes mencionados anteriormente.

También, como se entiende en la técnica de producción de la proteína heteróloga en una célula huésped puede aumentarse mediante el control de muchos de los mismos factores indicados anteriormente. Además, la adición de los factores que aumentan el número de copias del vector, incluyendo, pero no limitado a, la adición de rapamicina a los medios de cultivo, también puede aumentar la producción. Otros factores que pueden aumentar la producción incluyen, pero no se limitan a, la coexpresión de una o más proteínas chaperonas, tal como la proteína disulfuro isomerasa (PDI). Además, la hemoglobina (Hb) se puede coexpresar con al menos un polipéptido heterólogo en una célula huésped para potenciar la disponibilidad de oxígeno para el metabolismo oxidativo, aumentando por lo tanto, la producción de polipéptido.

Las proteínas heterólogas que se secretan a partir de una célula huésped durante la producción pueden comprender una secuencia líder que facilita la secreción. Las secuencias líder pueden modificarse para mejorar la secreción y, por tanto, la producción y recuperación global de la proteína expresada de forma heteróloga; por ejemplo diferentes secuencias líder de varias proteínas secretadas pueden estar unidas operativamente a la proteína heteróloga y evaluarse para determinar la expresión potenciada. Como alternativa, una secuencia líder dada puede modificarse por mutagénesis dirigida al sitio, o por medio de un enfoque de biblioteca combinatoria para identificar una variante mejorada de la secuencia líder. Se pueden encontrar secuencias líder quiméricas, que comprenden regiones de dos o más péptidos líder, se pueden para mejorar el nivel de expresión de la proteína heteróloga.

Ejemplos

Los ejemplos siguientes ilustran diversos aspectos no limitantes de la presente invención.

Ejemplo 1 – Nuevas proteasas Ympl

Mediante purificación por afinidad y análisis CL/EM se identificó una nueva proteasa en el sobrenadante del cultivo de *Pichia*. La figura 1 muestra una nueva proteasa metiltrófica de levadura (Ympl) de *Pichia* identificada en el gel. El sobrenadante filtrado clarificado de B378 (cepa nula de *Pichia*) se incubó durante la noche con aprotinina biotinilada. El sobrenadante se cargó en una columna de 2 ml de avidina (Pierce) y se lavó con PBS. La columna se eluyó secuencialmente con Tris 0,1 M, cloruro sódico 2 M, pH 7,5, después con acetato sódico 0,1 M, cloruro de sodio 2 M, pH 4,5 y, por último, con tampón de elución de biotina (Pierce). Los eluidos se concentraron y pasaron por SDS-PAGE. Las bandas más prominentes, las bandas 1 y 2, se escindieron del gel, se redujeron, alquilaron y digirieron con tripsina *in situ*. Los péptidos tripticos de cada banda se analizaron mediante espectrometría de masas en tándem - cromatografía líquida. Se realizó una búsqueda de los datos de la secuencia no interpretados frente a la base de datos de contig-péptido de *Pichia* (Integrated Genetics) utilizando software de identificación de proteínas Mascot.

5 Los péptidos generados a partir de las dos bandas 1 y 2 coincidieron con un marco de lectura abierto (ORF) en una base de datos del genoma de *Pichia*. El gen se aisló y secuenció, y contenía un marco de lectura abierto de 5115 pb de ácidos nucleótidos (SEC ID N° 3) que codifica una nueva proteína de 1.704 aminoácidos (SEC ID N° 4). El peso molecular predicho de la proteína entera es 185,8 kDa. La secuencia de polinucleótidos de este gen de *Pichia* (que se muestra en SEC ID N° 3) no comparte una homología significativa con ninguna secuencia en GenBank. La secuencia de aminoácidos de codificación contiene dos dominios de estructura: el dominio en N-terminal (AA 1-865) comparte identidades del 30,2 % con la serín endopeptidasa de *Hypocrea lixii*, y el dominio en C-terminal (AA 1305-1449) comparte identidades del 48 % con el motivo LPXTG de la proteína del dominio de anclaje a la pared celular de *Lactobacillus reuteri*. Sin embargo, la comparación de la longitud completa de la SEC ID N° 4 y la serina endopeptidasa de *Hypocrea lixii* muestra solo una identidad del 14,8 % usando el programa LALIGN de William Pearson que calcula un alineamiento global de dos secuencias, versión 2.2u (Myers y Miller, CABIOS (1989) 4:11-17). La estructura predicha para esta proteasa se muestra en la Figura 2. Su estructura y la secuencia en C-terminal son únicas entre todas las serín proteasas de todas las especies. El gen se denominó proteasa metiltrófica de levadura (*YMP1*) ya que está presente en el cultivo de *Pichia* metiltrófica. Es una nueva clase de proteasa basada en su estructura única. Ymp1 es uno de las proteínas secretadas más abundantes en medio de cultivo metiltrófico de *Pichia* sobre la base de un análisis global del sobrenadante de cultivo de *Pichia* usando CL / EM (datos no mostrados).

```

ATGTTCTCTCA AAAGTCTCCT TAGTTTTGCG TCTATCCTAA CGCTTTGCAA GGCCTGGGAT
CTGGAAGATG TACAAGATGC ACCAAAGATC AAAGGTAATG AAGTACCCGG TCGCTATATC
ATTGAGTATG AAGAAGCTTC CACTTCAGCA TTTGCTACCC AACTGAGAGC TGGGGGATAT
GACTTTAACA TCCAATACGA CTACTCAACT GGTTCCCTTT TCAACGGAGC ATCTGTTCAA
ATCAGCAACG ATAACAAAAC CACTTCCAG GATTTGCAA GTTTGCGTGC AGTCAAAAAT
GTTTACCCAG CTACTCTCAT TACATTAGAT GAAACATTTG AGCTTGCTGA CACGAAGCCA
TGGAACCCCT ATGGAATTAC CGGTGTCGAT TCTTTGCATG AGCAAGGATA TACTGGTAGT
GGTGTGTGTTA TTGCAGTTAT CGATACTGGT GTTGACTATA CACACCCTGC TCTGGGTGGT
GGTATCGGAG ATAATTTCCC TATCAAAGCT GGTTATGATT TGTCTTCCGG TGATGGTGTC
ATCAGCAATG ATCCTATGGA TTGTGACGGT CATGGTACCT TTGTATCCTC CATCATTTGTT
GCAAATAACA AAGATATGGT TGGTGTGCA CCAGATGCTC AGATTGTCAT GTACAAAGTG
TTCCCTTGTT CTGATAGTAC TTCGACTGAC ATAGTTATGG CGGGTATGCA AAAGGCCAT
GATGATGGTC ACAAGATTAT TTCGCTATCA CTGGGATCTG ACTCGGGGTT TTCCAGTACT
CCAGCTTCCT TAATGGCCAG CAGGATTGCT CAAGACAGAG TTGTTTTGGT GGCTGCTGGT
AACTCTGGAG AACTTGGTCC ATTCTATGCC TCCTCCCTG CTTCTGGGAA ACAAGTCATT
TCAGTTGGAT CTGTTCAAAA CGAACAATGG ACAACCTTTC CAGTAACCTT TACCTCTTCA
AACGGTGAAT CAAGGGTTTT TCCTTACCTC GCTTACAATG GTGCACAGAT TGGATTTGAT
GCCGAGCTTG AGGTTGATTT TACCGAAGAA AGAGGATGCG TCTATGAACC AGAGATCTCC
GCAGATAATG CGAATAAAGC TATTTTGTTA AGAAGGGGCG TCGGCTGTGT TGAAAACCTG

```

ES 2 539 378 T3

GAATTC AATT TATTGCTCTST GGCTGGTTAC AAGGCTTACT TCTTGTACAA CTCATTTTCA
 AGACCATGGA GTCTCTTGAA TATTTCTCCA CTGATTGAGC TAGACAACGC TTACTCTCTT
 GTTGAAGAGG AAGTTGGAAT ATGGGTGAAA ACCCAAATCG ACGCCGGTAA CACCGTCAAG
 TTAAAGGTGA GCACGAGTGA CCAAATGTTG CCATCTGATA AAGAGTATTT GGGAGTTGGA
 AAGATGGATT ATIACTCCTC TCAAGGACCT GCTTATGAGC TTGAATTTTT CCCAACGATA
 TCCGCTCCAG GTGGAGACAG TTGGGGCGCT TGGCCCGGTG GGCAATACGG TGTGCCTCA
 GGAACAAGTT TTGCTTGCCC CTATGTTGCA GGTCCTACAG CTCCTTATGA ATCGCAGTTT
 GGAATTC AAG ATCCCCAGGA CTATGTGAGA AAATTAGTCT CCACAGCTAC CGATCTTCAA
 TTATTTGACT GGAACGCAGT GAAACTGAG ACCTCTATGA ATGCTCCACT TATTCAACAG
 GGAGCTGGTC TAGTGAACGC TCTTGGTTTG TTTGAGACTA AGACTGTGAT CGTGTCTGCT
 CCTTATTTGG AGCTCAATGA CACCATCAAT AGAGCCAGTG AGTATACCAT TCAAATTAAG
 AATGAGA ACT CTGAGACTAT TACCTATCAA GTTGTTCCAG TTCCGGGAAC TACTGTCTAC
 TCTAGATCAG CTTCTGGGAA CATCCCATAC CTGGTCAATC AAGATTTTGC ACCTTACGGT
 GATAGTGATG CTGCGACAGT TGCTCTATCC ACAGAAGAGT TGGTTTTGGG ACCAGGAGAA
 GTTGGTGAAG TCACTGTGAT CTTCTCTACA GAAGAAATTG ATCAAGAAAC TGCTCCAATT
 ATTCAGGGTA AGATTACATT TTATGGTGAT GTCATACCGA TTGCTGTTC TTATATGGGA
 GTTGAAGTTG ATATTCATTC CTGGGAGCCT CTCATTGAGA GGCCTTTATC AGTGAGAATG
 TATTTGGATG ATGGTTCCTT AGCATATGTT GATGATGATC CTGATTATGA GTTCAATGTG
 TATGACTGGG ATTCTCCTAG ATTTTATTTT AACCTGAGAT ATGCAACCAA AGAAGTATCG
 ATTGACTTGG TGCACCCTGA TTATAGCATT GAGAACGACT ACGAATGGCC TTTAGTTTCC
 GGACACAACA ACTATTATGG TCCCGTGGGA TACGACTACG ATTATACCTC GGGTCAAGCC
 TTTTTCCTC GTTACTTTCA ACAACGTATT AACGAACTTG GATACTTTTC TTTTCCAGA
 TTTGCTAACT TTCTGTAGT TCCTGCTGGT GAATACAAAG CTCATTTTAG AGTTTGTCTA
 CCATATGGAG ACTTTTGGAA CAAAGAAGAC TGGCAATTGT TTGAATCCCC AGTGTTTAAC
 GTCCTCGCTC CACCGAATGA AGAAAACACT ACTGAAGAGC CAACTGAGGA ATCCAGCGAG
 GAGCCTACCG AAGAGTCAAC GTCTGAGTCA ACTGAAGAGC CCTCTTCTGA GTCAACTGAG
 AAATCTAGCG AGGTGCCAAC TGAAGAAATT ACTGAAGATG CAACATCCAC AATTGATGAT
 GATGAAGCAT CCACCGAAAG CTCTACTGAA GAACCAAGTG CTCAGCCCAC CGGTCTTAC
 TCTGATTTGA CTGTCCGTGA GGCCATTACC GACGTTAGTG TCACCAGTTT GAGGACAAC T
 GAAGCATTG GATACACTTC CGACTGGTTG GTTGTGCTT TCACTTTCAA CACTACTGAC
 AGAGATATTA CTCTCCACC TTACGCTGTT GTACAAGTAA CTATCCAAA TGAACTTCAA
 TTCATTGCTC ATCCAGAATA CGCCCATAC CTTGAGCCCT CATTGCAAGT TTTCTACACT
 AAGAATGAAA GATTAATTAT GACTAGTCAG TTCAACTACG ACACCAGAGT CATCGACTTC
 AAGTTTGACA ATCGAGACCA AGTAATAACT CAAGTGGAGG GAGTGTGTTA TTTACCGATG
 AACTAGAAC AAGATTTTCA TTCTGCATTG GCCCCAGGTG AATACGATTT TGAATTTTCA
 ACATCCGTTG ATTCTTATGC TTCGACCTTT GACTTTATTC CATTGATTAG ATCCGAGCCA
 ATCAAATTGA TAGCAGGTGC ACCAGACGAA GTTGAATGGT TTATTGATAT TCCAAGTGCA
 TACAGCGATT TGGCAACGAT AGATATTAGT TCTGATATCG ATACTAATGA TAAATTTGAG
 CAGTACTTCT ATGATTGCTC AAAGCTCAAG TACACTATTG GAAAAGAGTT TGATCAGTGG
 GGTAAATTTA CAGCTGGATC AGATGGTAAC CAATACAGCA ATACCACCGA TGGGTATGTT
 CCAATTA CTG ATTCTACCGG CTCTCCAGTA GCTGAAGTTC AATGTTAAT GGAAGTATC

ES 2 539 378 T3

TCATTGAGTT TCACAAATAC TCTTGCTGAG GATGAAGTAT TGAGAGTTGT TCTTCACTCT
TCTGCGTTTA GACGTGGTTC ATTCACCATG GCCAACGTGG TAAACGTTGA CATTACAGCT
GGTGGATTGG CAAAAAGAGA ACTCTTCTCT TATATATTGG ATGAAAATTA CTATGCTAGT
ACTGGATCTG AGGGGTTGGC ATTTGACGTA TTTGAAGTTG CTGATCAGGT CGAGGAGCCA
ACTGAGGAGT CAACCTCAGA GGAATCTACT GAACAGGAAA CTTCCACCGA GGAACCTACC
GAGGAATCAA CTGAACCTAC TGAGGAATCT ACCCAGGAAC CTAAGTGAAGA GCCCACCACG
GAGCCTACTT CTGAGTCAAC TGAGGAACCT TCTGAGGAGC CAACTTCTGA CGATCTCTCA
ATTGACCCAA CTGCTGTACC TACCGATGAA CCTACTGAAG AGCCAACTGA GGAGCCTACT
TCTGAGTCAA CTGAGGAACC TTCTGAGGAG CCAACTTCTG ACGATCTCTC AATTGACCCA
ACTGCTGTAC CTACCGATGA ACCTACTGAA GAGCCAACTG AGGAGCCTAC TTCTGAGTCA
ACTGAGGAAC CTTCTGAGGA GCCAACTTCT GACGATCTCT CAATTGACCC AACTGCTGTA
CCTACCGATG AACCTACTGA AGAGCCAACT GAGGAGCCGA CCTCTGAGAC TACCGATGAT
CCATCGATAG CACCTACTGC TGTGCCAACT TCCGACACAT CTTCTGGACA ATCGGTGGTT
ACTCAAACA CTACAGTAC TCAGACTACC ATCACTTCAG TCTGTAATGT TTGTGCTGAG
ACCCCTGTAA CAATCACTTA CACTGCACCA GTTGTGACTA AGCCAGTTT TTACACCACC
GTTACTTCAG TTTGCCATGT ATGTGCAGAG ACACCAATCA CAGTTACCTT GACGTTGCCA
TGTGAAACCG AAGACGTGAC AAAGACTGCC GGCCCTAAGA CTGTCACCTA CACCGAAGTT
TGCAACTCCT GTGCTGACAA GCCTATCACT TACACCTACA TCGCTCCAGA GTACACTCAA
GGTGCCGAAC GTACAACAGT TACATCGGTT TGCAACGTTT GTGCTGAGAC ACCTGTAACG
CTAACATACT CTGCGCCGAA AGCCAGTCGT CATAAGTTT CTTACAAATA TTCAAGTGCC
GGAGAGCTCA TTTCATCCAA GGGGATCACG ATTCCTACTG TTCCTGCCCC TCCAAGTGGT
ACTTATAGTA AGTCTGTTGA CACTAGCCAA CGTACACTCG CTACCATTAC AAAATCTTCA
GATGAGTCTA AACTGTTAC CACTACTCAA GCCACACAAG TTTTGAGCGG TGAATCCAGT
GGAATTCAAG CTGCTTCAA CAGCAGGAG ATCTCAGCTC CAACTGTAC TACAGCTGGG
AACGAGAACT CTGGATCTAG ATTTTCGTTT GCTGGACTAT TCACAGTTCT GCCTCTTATC
TTGTTCTGTTA TATAA (SEQ ID NO:3)

La secuencia polipeptídica codificada por la secuencia génica de la SEC ID N° 3 se presenta a continuación como SEC ID N° 4.

MFLKSLLSFA SILTLCKAWD LEDVQDAPKI KGNEVPGRYI IEYEEASTSA FATQLRAGGY
DFNIQYDYST GSLFNGASVQ ISNDNKTTFQ DLQSLRAVKN VYPATLITLD ETFELADTKP
WNPHGITGVD SLHEQGYTGS GVVIIVIDTG VDYTHPALGG GIGDNFPIKA GYDLSSGDGV
ITNDPMDCDG HGTFVSSIIV ANNKDMVGA PDAQIVMYKV FPCSDSTSTD IVMAGMQKAY
DDGHKIISLS LGSDSGFSST PASLMASRIA QDRVVLVAAG NSGELGPFYA SSPASGKQVI
SVGSVQNEQW TTFPVFTSS NGESRVFPYL AYNGAQIGFD AELEVDFTTE RGCVYEPEIS
ADNANKAILL RRGVGCVENL EFNLLSVAGY KAYFLYNSFS RPWSLLNISP LIELDNAYSL
VEEEVGIWVK TQIDAGNTVK LKVSTSDQML PSDKEYLGVG KMDYYSSQGP AYELEFFPTI
SAPGGDSWGA WPGGQYGVAS GTSFACPYVA GLTALYESQF GIQDPQDYVR KLVSTATDLQ
LFDWNAVKLE TSMNAPLIQQ GAGLVNALGL FETKTVIVSA PYLELNDTIN RASEYTIQIK

NENSETITYQ VVHVPGTTVY SRSASGNIPY LVNQDFAPYG DSDAATVALS TEELVLGPGE
 VGEVTVIFST EEIDQETAPI IQGKITFYGD VIPIAVPYMG VEVDIHSWEP LIERPLSVRM
 YLDDGSLAYV DDDPDYEFNV YDWDSRPFYF NLRYATKEVS IDLVHPDYSI ENDYEWPLVS
 GHNNYYGPGV YDYDYTSGQA FLPRYFQORI NELGYLSFSR FANFSVVPAG EYKALFRVLL
 PYGDFWNKED WQLFESPWFN VLAPPNEENT TEEPTEESSE EPTEESTSES TEEFSSSESTE
 KSSEVPTEEI TEDATSTIDD DEASTESSTE EPSAQPTGPY SDLTVGEAIT DVSVTSLRIT
 EAFGYTSDWL VVSFTFNTTD RDITLPPYAV VQVTIPNELQ FIAHPEYAPY LEPSLQVIFYT
 KNERLIMTSQ FNYDTRVIDF KFDNRDQVIT QVEGVVYFTM KLEQDFISAL APGEYDFEFH
 TSVDSYASTF DFIPLIRSEP IKLIAGAPDE VEFWIDIPSA YSDLATIDIS SDIDTNDNLQ
 QYFYDCSKLK YTIGKEFDQW GNFTAGSDGN QYSNTTDGYV PITDSTGSPV AEVQCLMESI
 SLSFTNTLAE DEVLRVVLHS SAFRRGSFTM ANVVNVDITA GGLAKRELFS YILDENNYAS
 TGSEGLAFDV FEVADQVEEP TEESTSEEST EQETSTEEPT EESTEPTEES TQEPTEEPTD
 EPTSESTEPP SEEPTSDDL S IDPTAVPTDE PTEEPTEEPT SESTEPEPSEE TSSDLSIDPT
 AVPTDEPTEE PTEEPTSEST EEPSEPTSD DLSIDPTAVP TDEPTEEPT EPTSETTDDP
 SIAPTAVPTS DTSSGQSVVT QNTTVTQTTI TSVCNVCAET PVTITYTAPV VTKPVSYTIV
 TSVCHVCAET PITVTLTLPC ETEDVTKTAG PKTVTYTEVC NSCADKPITY TYIAPEYTOG
 AERTTVTSVC NVCAETPVTI TYTAPKASRH TVPSQYSSAG ELISSKGITI PTVPARPTGT
 YKSVVDTSQR TLATITKSSD ESNTVTTTQA TQVLSGESSG IQAASNSTSI SAPTVTTAGN
 ENSGSRFSFA GLFTVLPLIL FVI. (SEQ ID NO:4)

Ejemplo 2 – Nuevas proteasas

Se identificaron dos nuevas proteasas adicionales metiltróficas de levadura (*Ymp2* y *Ymp3*) utilizando purificación por afinidad similar y el análisis CL / EM como se describe en el Ejemplo 1 a partir del sobrenadante de la cepa defectiva *ymp1* de *Pichia* (véase el Ejemplo 3). El sobrenadante filtrado clarificado de B580 (cepa de *Pichia ymp1* con delección) se incubó durante la noche con AEBSF biotinilado (4- (2-aminoetil) bencenosulfoniilfluoruro clorhidrato). El sobrenadante se cargó en una columna de 10 ml de avidina (Pierce) y se lavó con PBS. La columna se eluyó secuencialmente con Tris 0,1 M, cloruro sódico 2 M, pH 7,5, después con acetato sódico 0,1 M, cloruro de sodio 2 M, pH 4,5 y, por último, con tampón de elución de biotina (Pierce). Los eluidos se desnaturalizaron mediante la adición de urea 8 M y los residuos de Cys se redujeron mediante reacción con ditioneitol (DTT) 20 mM y después se alquilaron mediante reacción con yodoacetamida 20 mM. Después de diluir la muestra a urea 2 M con bicarbonato amónico 100 mM (pH 8,5), las proteínas se digirieron con tripsina (tripsina modificada de calidad para secuenciación, Promega, Madison, WI). La mezcla de péptidos resultante de la digestión se desaló con una columna de centrifugación PepClean (Pierce, Rockford, IL), y se analizó mediante un sistema de CL-EM/EM, en el que en una cromatografía líquida de alta presión (HPLC) con una columna C18 de fase inversa con un diámetro interior de 75 micrómetros se acopló en línea con un espectrómetro de masas de trampa de iones (Thermo, Palo Alto, CA). Los datos de espectrometría de masas adquiridos a partir del análisis CL-EM / EM se utilizaron para buscar contra la base de datos de proteínas redundantes de GenBank, así como en la base de datos de *Pichia* (Integrated Genetics) usando el software de identificación de proteínas Mascot.

A través de las coincidencias de péptidos se identificaron dos genes, llamados *YMP2* y *YMP3* que compartían homologías con proteasas conocidas. *YMP2* contiene un marco de lectura abierto de 618 pb de ácidos nucleótidos (SEC ID N° 5) que codifica una nueva proteína de 205 aminoácidos (SEC ID N° 6). El peso molecular predicho de la proteína entera es 22,9 kDa. La secuencia de polinucleótidos de este gen de *Pichia* (que se muestra en SEC ID N° 5) no comparte una homología significativa con ninguna secuencia en GenBank. No obstante, la secuencia de aminoácidos de codificación comparte identidades del 29 % con la Arginina / alanina aminopeptidasa de *Pichia pastoris*.

ATGGCTCCCA GAACACTACC AGAAGACTTA ATTCCCTCCC TATACGACTT GCACATCTAC
 AACTTCCAAC CCGAAAAAAA GACTTATGAT GGAGACATTG TCATCCACTT GGAGGTGAAG
 GAGCCCACTG ATGAAGTGGT CTTCAATGCC AAGGATTTGG AATTGAAAGA CGTACATGTC
 TTCCACAATG TCAACAAGTC TGAAAACGAA ATCCCGGTTA AGGAGATTGT TGATAACGAG
 CTCATCACAA TTAAGCTCAA AGAGAAGGTT ACTTCCGGAA CGTTGCTGGT GAATATTTC
 TTCACCGGTA ACATTCAATC TGATAAAATT GGATTTTACA AGGGAGACAC AGATGTGGAA
 GGAAGAGTCA CATACTACTAC AAACCTTACC ACTCCAAATG CCAGGTTGGC ATCCCATGTT
 CTTGATAACA TATTGTTGAA AGCTCCATTC AAGTTCGGAG TAACTGCCAA TCCAGGACAA
 TTAGTGAGTT CCATTTTGGG TCTAAGCTCT GAGGCTGACG TCTTGAATGA CAATGACGAT
 GTGATTGGTA CGAGATACCA ATACCAAGTG AGTGAGCCAA TAGCCCCAGC TTTACTGGAG
 TGGACCATTC ATATTTAA (SEQ ID NO:5)

La secuencia polipeptídica codificada por la secuencia génica de la SEC ID N° 5 se presenta a continuación como SEC ID N° 6.

MAPRTLPEDL IPSLYDLHIY NFQPEKKTYD GDIVIHLEVK EPTDEVVFNA KDLELKDVBV
 FHNVNKSENE IPVKEIVDNE LITIKLKEKV TSGTLLVNIS FTGNIQSDKI GFYKGDTDVE
 GRVYTTNLT TPNARLAFPC LDNILLKAPF KFGVTANPGQ LVSSILDLS EADVLNDND
 VIGTRYQYQV SEPIAPALLE WTIHI (SEQ ID NO:6)

- 5 YMP3 contiene un marco de lectura abierto de 1233 pb de ácidos nucleótidos (SEC ID N° 7) que codifica una nueva proteína de 410 aminoácidos (SEC ID N° 8). El peso molecular predicho de la proteína entera es 46,0 kDa. Su secuencia de aminoácidos comparte identidades del 47 % con la leucina aminopeptidasa de *Pichia stipitis*.

ATGGTCAAAC TCATATCAAT TATAGCCCTA GTTCAACTTG TCTCTGCGAC AATTGTACCT
 TGGAAATCTCC AGAACGTCTT ATCTGACGTC CATCACCCCTT CTCTCCATCT CTTGGATTAT
 ATCAATCCCT TGAAGAACGA GGTAATGTTT GATGGCGACG ATCGCAGAA AATCAAGTTA
 GGCCCCAAG AATAACCGTAT TATCACTGAA AAAGAGAAAT ACCAGTTGAA AACAGAGGGG
 ATATCATTTA TCGATGTCAC CTATCAGCAT GGAGACAATG TAGAGCTGCT CTATTCCAGT
 GCGCCAGTTA CCGTCCAGTA CTATCTTTAT CCGTCCAATG ATACTTTCCA TTCAAACAA
 GTAATTCTT TGATAGGTGA GATTGACATT GGCAGAATGC AGGCGTTTTT GGGAAGGTTT
 TCTAGCTTCT TTACAAGATT TTACAAATCT GACAAGGGGT TGCAGAGTTC TATCTGGTTA

 CAAGGTGAAT TGGTTC AAT GGCCTTGAAA GATCCATCGA GGTTC AATGT TACTACTGTG
 GAACACCCTT GGAAGCAGAA TTCTGCCATC TTTACGATAT ACGGTGAAAA TGITGATCCT
 TCGAAAGGAA AAGGGGACAT TGTAGTAGTG GGATGCCATC AAGATTCCAT AAACCTTGCTT
 TTCCCAACA TTCTCCGTGC TCCAGGGGCT GATGATGATG GATCTGGTGT AACTTCCAAC
 CTTGAAGCGC TCAGAATCAT AGTTGAAAGT GGCCTCAAGT TTCACAATAC AGTAGAGTTT
 CACTTTTATT CTGCCGAAGA AGGAGGACTA CTTGGCTCCC AGCAAATTT CAGCTCGTAT
 AGAGCTGCAG AAGAGACTGT TGTGCTATG CTACACAGG ACATGACTGG ATACATCCAA
 AAAGCTTTAG ACCACGGGGA ATCCGACCAC TTCGGGCTAA TCACTGACCA TACAAACGCA
 AATCTGAATA GCTTCCTTGC ACTTTAATC GATGCATACA CTTCAATPCC CTACAAAGAA
 ACCGAATGTG GGTATGCCTG CTCAGATCAT AGTTCTGCCT TGGAACATGG TTATCCATCT
 GCCATGGTCT TTGAAAGTAG TTTTGCCTAC ACAAATCCT TCATCCATAG CACCCAAGAC
 ACAATTGACA AGATCAATTT TCCACATATG GCAGAGCAT GTCAAGTTGG TCCTGGGTTA
 CGTTGTAGAG TTGGGATTAG AACATTTAG GTGA (SEQ ID NO:7)

- 10 La secuencia polipeptídica codificada por la secuencia génica de la SEC ID N° 7 se presenta a continuación como SEC ID N° 8.

MVKLISIIAL VQLVSATIVP WNLQNVLSDV HHPSLHLLDY IQSLKNEVMF DGDDRRRIKL
 GPQEYRIITE KEKYQLKTEG ISFIDVTYQH GDNVELLYSS APVTVDPDLY PSNDTFHFKQ
 VNSLIGEIDI GRMQAFLGRF SFFTRFYKS DKGLQSSIWL QGELVQLALK DPSRFNVTTV
 EHPWKQNSAI FTIYGENVDP SKGKGDIVV GCHQDSINLL FPNILRAPGA DDDGSGVTSN
 LEALRIIVES GLKFHNTVEF HFYSAEEGGL LGSQQIFSSY RAAEETVVM LQQDMTGYIQ
 KALDHGESDH FGLITDHTNA NLNSFLALLI DAYTSIPYKE TECGYACSDH SSALEHGYP
 AMVFESSFAY TNPFIHSTQD TIDKINFPHM AEHVKLVLGY VVELGLEHFR (SEQ ID NO:8)

Adicionalmente, otras tres proteasas potenciales se identificaron utilizando todas las proteasas de levadura conocidas contra la base de datos de *Pichia*. Las secuencias de polinucleótidos identificadas que codifican estas proteasas identificadas y sus respectivas secuencias de polipéptidos se presentan en las SEC ID N° 9-14 siguientes.

ES 2 539 378 T3

ATGAAATCGG TTATTTGGAG CTTCTATCT TTGCTAGCAT TGTCGCAGGC ATTGACTATT
 CCATTGCTGG AAGAGCTTCA ACAGCAAACA TTTTITAGCA AGAAAACCGT TCCTCAACAA
 GTTGCTGAAT TGGTGGGCAC CCATTACTCT AAGGATGAGA TAATCAGTCT ATGGAAGGAC
 ATTGAGCTGG ATGTACCCAG GGAAAAGATC CAAGAGGCCT TCGATAAGTT CGTAAAACAA
 TCAACTGCCA CTTCCCCCGT TAGAAATGAA TTTCCCTTGT CTCAGCAAGA TTGGGTGACA
 GTGACCAACA CCAAGTTTGA TAATTATCAA TTGAGGGTTA AAAAAATCCCA CCCTGAAAAG
 CTAACATTG ATAAGGTAAA GCAATCTTCG GGATACCTGG ATATCATTGA TCAAGATAAG
 CATCTTTTCT ATTGGTTTTT TGAATCCCGA AATGATCCGT CCACAGACCC AATCATCCTA
 TGGTTGAATG GTGGACCCGG CTGCTCTTCT ATTACAGGGT TGCTATTCTGA AAAGATTGGC
 CCCAGTTACA TCACCAAAGA GATTAAGCCG GAACATAATC CTTATTCTATG GAACAACAAT
 GCTAGTGTTA TCTTCCTTGA GCAACCGGTT GGAGTAGGAT TTTCTTACTC TTCTAAGAAA
 GTCGGTGATA CTGCAACTGC TGCCAAAGAT ACATATGTGT TTTTGGAGCT TTTCTTCCAA
 AAGTTTCCCTC AGTTCCTGAC CTCTAATCTG CACATTCGCTG GGGAAATCGTA TGCTGGCCAT
 TATTTGCCCA AGATTGCTTC TGAGATTGTG TCTCACGCAG ACAAGACGTT TGACCTTTCA
 GGAGTCATGA TCGGTAATGG TCTTACTGAT CCTCTAATTC AGTATAAGTA CTATCAGCCA
 ATGGCCTGTG GAAAAGGTGG CTACAAGCAG GTCATTTCCG ACGAGGAATG TGATGAATTG
 GATAGGGTCT ATCCAAGATG TGAACGTTTA ACGCGGGCAT GTTATGAGTT CCAAAATTCA
 GTTACTTGTG TTCCGGCAAC ACTTATTGC GACCAAAAGC TACTGAAGCC GTACACTGAC
 ACTGGCTTGA ATGTCTATGA TATTTCGTACA ATGTGCGATG AAGGGACTGA TTTGTGTTAC
 AAAGAACTGG AATACGTGGA GAAGTACATG AACCAGCCTG AAGTGCAGGA AGCCGTGGGC
 TCTGAAGTCA GTTCTTACAA AGGTGTGAC GATGATGTCT TCTTAAGATT TTTGTACTCT
 GGCATGGAT CTAAGCCTTT CCACCAGTAT ATCACGGATG TTCTCAATGC AAGTATTCCG
 GTTCTGATTT ACGCAGGTGA TAAAGATTAT ATCTGTAATT GGCTAGGAAA CCAAGCTTGG
 GTCATGAGC TAGAATGGAA CTTGTCTGAG GAATTCAGG CAACTCCGAT TCGACCGTGG

ES 2 539 378 T3

TTCACITTTGG ACAATAACGA TTATGCAGGA AACGTACAAA CTTATGGAAA CTTTTCTTTT
 CTAAGAGTAT TTGATGCTGG TCACATGTT CCTTACAATC AACCAGTCAA CGCACITTGAC
 ATGTTTGTCA GATGGACACA CCGTGATTTC TCATTTGGTT AITAA (SEQ ID:9)

MKSVIWSLLS LLALSQALTI PLLEELQQOT FFSKKTVPQQ VAELVGTHTS KDEIISLWKD
 IELDVPREKI QEAFDKFVKQ STATSPPVRNE FPLSQDQWVT VINTKFDNYQ LRVKKSHPK
 LNIIDKVKQSS GYLDIIDQDK HLFYWFESR NDPSTDEIIL WLNCGGPGCSS ITGLLEFKIG
 PSYITKEIKP EHNPPYSWNNM ASVIFLEQPV GVGFSYSSKK VGDTATAAKD TYVFLLEFFQ
 KFPQFLTSLN HIAGESYAGH YLPKIASIIV SHADKTFDLS GVMIGNGLTD PLIQYKYYQP
 MACGKGGYKQ VISDEECDL DRVYPCRERL TRACYEFQNS VTCVPATLYC DQKLLKPYTD
 TGLNVYDIRT MCDEGTDLCY KELEYVEKYM NQPEVQEAUV SEVSSYKCCD DDVFLRFLYS
 GDGSKPFHQY ITDVLNASIP VLIYAGDKDY ICNWLGNQAV VNFLEWNLSE EFQATPIRPW
 FTLDNDYDAG NVQTYGNFSF LRVFDAGHMV PYNQPVNALD MVRVWTHGDF SFGY (SEQ ID
 No:10)

ATGATATTAC ACACCTATAT TATTCTCTCG TTATTGACTA TATTTCTTAA AGCTATTGGT
 CTGTCTTGGC AGATGCCAAT GGCCTTGGAA GCTAGTTATG CCTCATTAGT GGAGAAAGCA
 ACCCTCGCTG TTGACAAAGA AATTGATGCC ATACAAAAGG GTATTGAGCA AGGTGGTTG
 GAAGTAGAGA CAAGATTTC AACTATAGTG TCACAGTTAT CCTATAGTAC TGGCCAAAAA
 TTTGCGATCA AGAAGAAAGA TGCAACTTTT TGGGATTTCT ATGTTGAAAG TCAAGAGTTG
 CCAAACTACC GAATTAATGA AAATCTGAAA CCAATTTTCA ACCCTATTTC GTGGAAATGGT
 AATGCTTCAA TCATCTACTT AGATCAACCG GTCATGTTG GGTTCCTTA TTCTTCATCA
 TCGTGGAGTA ACACGTGTGT TGCGGGAGAA GATGTGTATG CATTTCITCA GCTTTTTTTT
 CAACACTTCC CGGAATATCA AACTAATGAC TTTTATATTG CCGGTGAATC TTATGCAGGA
 CATTACATTC CGGTGTTTGC AGACGAAATT TTGAGTCAAA AGAACAGAAA TTTCAATCTT
 ACTTCAGTCT TGATCGGAAA TGGATTAAT GACCCTTTGA CTCAATACCG ATATTACGAG
 CCAATGGCTT GTGGTGAAGG TGGTGGCCCG TCAGTACTGC CTGCCGATGA GTGCCAAAAT
 ATGCTAGTTA CCCAAGATAA ATGTTTGTCT TTAATCAAG CATGCTACGA CTCACAGTCG
 GCATTCACAT GCGCACCAGC TGCCATTAT TGTAAATACG CTCAGATGGG ACCCTATCAG
 AGAACTGGGA AGAATGTGTA TGATATTCGT AAGGAATGTG ATGGTGGATC CTTGTGCTAT
 AAGGACTTG AATTCATCGA TACCTACTTA AATCAAAAGT TTGTTCAAGA TGCTTTGGGC
 GCCGAGGTCG ATACCTATGA ATCTTGCAAT TTTGAAATCA ACAGAAACTT TTTATTTGCT
 GGAGATTGGA TGAACCTTA TCATGAACAT GTCAGCAGTC TCTTGAAACA AGGTTTGCCC
 GTTTTGATTT ACGCAGGGGA CAAAGATTTT ATTTGCAACT GGTGTTGGTAA TCGAGCATGG
 ACTGATGTCT TGCCGTGGGT TGATGCTGAT GGTTTTGA AAAGCCGAAGT CCAAGATTGG
 TTGGTTAATG GAAGGAAGGC TGGTGAATTT AAGAACTATA GCAACTTCAC CTACCTAAGG
 GTTTATGATG CTGGTCATAT GGCCCATAT GATCAGCCAG AGAATTCTCA TGAATGGTC
 AATAGATGGA TATCCGGAGA CTTTAGCTTT CACTAG (SEQ ID No:11)

MILHTYIILS LLTIFPKAIG LSLQMPMALE ASYASLVEKA TLAVGQEIIDA IQKGIQOGLW
 EVETRFPTIV SLSYSTGPK FAIKKKDATF WDFYVESQEL ENYRINENLK PIFNPYSWNG
 NASIIYLDQP VNVGFSYSS SVSNTVVAGE DVYAFLLQFF QHFPEYQTNF FHIAGESYAG
 HYI PVFADEI LSQKNRNFNL TSVLIGNGLT DPLTQYRYYE PMACGEGGAP SVLPADECEN
 MLVTQDKCLS LIQACYDSQS AFTCAPAAIY CNNAQMGPIY RTGKNVYDIR KECDDGSLCY
 KDLEFIDTYL NQKRVQDALC AEVDTYESCN FEINRNFLFA GDMKPYHEH VSSLINKGLP
 VLIYAGDKDF ICNWLGNRAW TDVLPWVDAD GFKAQEVQDW LVNKRKAGEF KNYSNFTYLR
 VYDAGHMAMP DQPENSHEMV NRWISGDFSF H (SEQ ID No:12)

ATGCAATTGC GTCAATCCGT TGGATTGGCT ATCTTATCTG CCATAGCAGT CCAAGGATTG
 CTAATTCCTA ACATTGAGTC ATTACCCAGC CAGTTTGGTG CTAATGGTGA CAGTGAACAA
 GGTGTATTAG CCCACCATGG TAAACATCCT AAAGTTGATA TGGCTACCA TGGAAAGCAT
 CCTAAAATCG CTAAGGATTC CAAGGGACAC CCTAAGCTTT GCCCTGAAGC TTTGAAAGAG
 ATGAAAGAAG GCCACCCCTC GGCTCCAGTC ATTACTACCC ATTCGCTTC TAAAACTTA
 ATCCCTTACT CTTATATTAT AGTCTTCAAG AAGGGTGTCA CTTGAGAGGA TATCGACTTC
 CACCGTGACC TTATCTCCAC TCTTCATGAA GAGTCTGTGA GCAAATTAAG AGAGTCAGAT
 CCAATCACT CATTCTTCTG TTCTAATGAG AATGGCGAAA CAGGTACAC CGGTGACTTC
 TCCGTTGGTG ACTTGCTCAR GGGTTACACC GGATACTTCA CGGATGACAC TTTAGAGCTT
 ATCAGTAAGC ATCCAGCAGT TGCTTTCAAT GAAAGGGATT CGAGAGTATT TGCCACCGAT
 TTTGAAACTC AAAACGGTGC TCCTTGGGGT TTGGCCAGAG TCTCTCACAG AAAGCCTCTT
 TCCCTAGGCA GCTTCAACAA GTACTTATAT GATGGAGCTG GTGGTGAAGG TGTACTTCC
 TATGTTATCG ATACAGGTAT CCACGTCAC CACAAAGAAT TCCAGGGTAG AGCATCTTGG

GGTAAGACCA TTCCAGCTGG AGACGTTGAT GACGATGGAA ACGGTCACGG AACTCACTGT
 GCTGGTACCA TTGCTTCTGA AAGCTACGGT GTTGCCAAGA AGGCTAATGT TGTGCCATC
 AAGTCTTTGA GATCTAATGG TTCTGGTTCG ATGTCAAGATG TTCTGAAGGG TGTGAGTAT
 GCCACCCAAT CCCACTTGGA TGCTGTAA AAGGGCAACA AGAAATTTAA GGGCTCTACC
 GCTAACATGI CACTGGGTGG TGGTAAATCT CCTGCTTGG ACCTTGCAGT CAATGCTGCT
 GTTAAGAATG GTATCACTT TGCCCTTGCA GCAGGTAACG AAAACCAAGA TGCTTGTAAC
 ACCTCGCCAG CAGCTGCTGA GAATGCCATC ACCGTCGGTG CATCAACCTT ATCAGACGCT
 AGAGCTTACT TTTCTAACTA CGGTAAATGT GTTGACATTT TCGCTCCAGG TTTAAACATT
 CTTTCTACCT ACACTGGTTC GGATGACGCA ACTGCTACCT TGCTTGSTAC TTCAATGGCC
 TCTCTCACA TTGCTGGTCT GTTGACTTAC TTCTATCAT TGCAGCCTGC TGCTGGATCT
 CTGTACTCTA ACGGAGGATC TGAGGGTGTG ACACCTGCTC AATTGAAAAA GAACCTCCTC
 AAGTATGCAT CTGTCGGAGT ATTAGAGGAT GTTCCAGAAG ACACTCCAAA CCTCTGGTT
 TACAATGGTG GTGGACAAA CCTTCTTCT TTCTGGGGAA AGGAGACAGA AGACAATGTT
 GCTTCTCCG ACGATACTGG TGAGTTTAC TCTTTTGTGA ACAAGCTTGA ATCAGCTGTT
 GAAAACCTGG CCCAAGAGTT TGCACATTCA GTGAAGGAGC TGGCTTCTGA ACTATTTAG
 (SEQ ID No:13)

MQLRHSVGLA ILSAIVQGL LIPNIESLPS QFGANGDSEQ GVLAHHGKHP KVDMAHHGKH
 PKIAKDSKGH PKLCPEALKK MKEGHPSAPV ITTHSASKNL IPYSYIIVFK KGVTSEDI DF
 HRDLISTLHE ESVSKLRES DPNHSFFVSNE NGETGYTGDF SVGDLLKGYT GYFTDDTLEL
 ISKHFAVAFI ERDSRVFAD FETQNGAPWG LARVSHRKPL SLGSFNKYLY DGAGGEGVTS
 YVIDTGIHVT HKEFQGRASW GKTI PAGDVD DDGNHGHGTHC AGTIASESYG VAKKANVVAI
 KVLRSNGSGS MSDVLKGV EY ATQSHLDAVK KGNKKFKGST ANMSLGGGKS PALDLAVNAA
 VKNGIHFAVA AGNENQDACN TSPAAENAI TVGASTLSDA RAYFSNYGKC VDFAPGLNI
 LSTYTGSDDA TATLSGTSMA SPHIAGLLTY FLQLQPAAGS LYSNGGSEGV TPAQLKKLLL
 KYASVGVLED VPEDTPNLLV YNGGGQNLSS FWGKETEDNV ASSDDTGEFH SFVNKLESVA
 ENLAQEF AHS VKELASELI (SEQ ID No:14)

Ejemplo 3 Proteasas defectivas

Se generaron varias cepas inactivadas de *Pichia* mediante los procedimientos siguientes. El extremo 5' del gen que se va a inactivar (200-500 pb) se amplificó a partir de ADN genómico de *Pichia* mediante PCR usando un cebador 5' específico del gen y un cebador 3' específico del gen con una secuencia de 20 pb complementaria al extremo 5' del casete de expresión del marcador de selección, por ejemplo KanMX. El extremo 3' del gen que se va a inactivar (200-500 pb) también se amplificó a partir del ADN genómico de *Pichia* mediante PCR usando un cebador 5' específico del gen con una secuencia de 20 pb complementaria al extremo 3' del casete de expresión KanMX y un cebador 3' específico del gen. En una reacción de PCR separada, una cantidad adecuada del producto 5', producto 3', el ADN que contiene el casete de expresión KanMX, y los cebadores específicos del gen 5' y 3' utilizados antes se añadieron para hacer un casete de inactivación que tiene el casete de expresión KanMX flanqueando regiones homólogas en 5' y 3' del gen inactivado. El fragmento de ADN se purificó y se transformó en *Pichia* usando un procedimiento de electroporación estándar. Las células recombinantes se seleccionaron en las placas YPD que contienen un antibiótico, por ejemplo KAN, y se efectuó detección selectiva con PCR utilizando cebadores específicos. Las cepas defectivas resultantes se confirmaron mediante secuenciación genómica y análisis de transferencia de tipo Southern.

Las cepas defectivas individuales y múltiples se transformaron después con ADN capaz de expresar un polipéptido heterólogo. Cada cepa defectiva se transformó con al menos un vector que codifica la proteína terapéutica descrita mediante la SEC ID N° 1. Los cultivos de cepas defectivas transformadas se cultivaron y se recogieron los sobrenadantes. Los productos secretados se pasaron por geles NuPAGE MOPS para la detección de degradación del producto por proteólisis. Las múltiples bandas de la proteína heteróloga semipurificada se interpretó que indicaban actividad proteasa, mientras que una única banda de proteína heteróloga en el peso molecular apropiado se interpretó como ausencia de actividad proteasa. Véase la figura 5-8.

La actividad proteasa en las muestras de sobrenadante de la fermentación se midió utilizando un péptido sintético con un marcador fluorescente e inactivador con longitudes de onda de excitación y emisión de 340 y 405 nm, respectivamente. El sustrato peptídico se incubó con muestras de sobrenadante del cultivo y la tasa de degradación del sustrato se controló usando HPLC fluorescente. Los resultados se expresan como aumento en el área del pico / h. En la Figura 3, la actividad proteasa en el sobrenadante de la fermentación fue de aproximadamente 68 a 285 en la cepa mutante ymp1, mientras que la actividad proteasa en la cepa de control fue de aproximadamente 1541 a 1909. En consecuencia, la cepa mutante demostró una reducción espectacular de la actividad proteasa sobre el péptido.

El gel del zimograma es otro procedimiento para medir la actividad proteasa utilizando caseína o gelatina como sustratos. En este experimento, las muestras de fermentación se separaron en geles de acrilamida, ya sea con caseína o con gelatina incorporada en la matriz en condiciones reductoras no desnaturalizantes. Después de la separación, las bandas de proteínas se renaturalizaron y se aplicó un tampón de desarrollo para proporcionar los cofactores necesarios para la actividad proteasa. El gel se incubó a 37 °C durante un período prolongado y después se tiñó con azul de Coomassie. Las áreas de actividad proteolítica se identificaron como áreas claras sobre el gel sobre un fondo azul. Las proteasas activas degradaron el sustrato, no permitiendo que el área sobre el gel se uniera

al azul de Coomassie. Las áreas en el gel que carecen de proteasa activa permiten que la mancha se una al sustrato, de modo que se genera el fondo azul. Figura 4 mostró las áreas activas de la proteasa utilizando sobrenadantes de fermentación de las cepas mutantes *ymp1* y dos cepas de control. Las cepas de control tenían proteasas activas que migraron en las zonas de peso molecular tanto alto como bajo. Sin embargo, las señales en la zona de peso molecular alto estaban muy disminuidas o desaparecidas en las cepas mutantes *ymp1*. Estos datos muestran que las actividades en las zonas de alto peso molecular podrían deberse a la acción de *ymp1*. El resumen de los resultados para las cepas defectivas se presentan en la Tabla 5.

Otros estudios utilizando un ensayo de estabilidad en tampón demostraron la eficacia de la escisión y el sitio cambia cuando *YMP1* y *YPS1* se eliminaron de las cepas huésped *Pichia*. En este ensayo, el patrón proteína descrito por la SEC ID N° 1 se incubó con los sobrenadantes del cultivo y los productos proteicos escindidos se analizaron mediante espectrofotometría de masas para indicar los sitios de escisión y el porcentaje de productos de degradación (Tabla 1-4).

Usando el sobrenadante del caldo de fermentación B804 (cepa parental SMD 1163 con delección *pmt4*, véase el Ejemplo 4), se detectaron 5 sitios de escisión principales: Lys28/Gly29, Phe36/Thr37, Ser38/Asp39, Tyr43/Leu44 y Lys50/Glu51 (Tabla 1).

Tabla 1: Caldo de fermentación BS04 de SMD1163_pmt4 (01:40)

Secuencia	0 h	24 h	48 h	5 días
His1-Leu645	100	73,4	33,9	4,6
(Trp)Leu26 - Leu645	0	1,4	1,8	0
(Leu)Val27 - Leu645	0	0	0	0
(Lys)Gly29 - Leu645	0	5,9	6,3	1,9
(Phe)Thr37 - Leu645	0	8,7	28,4	31,7
(Thr)Ser r38 - Leu645	0	1	0	0
(Ser)Asp39 - Leu645	0	4,6	12,9	18
(Val)Ser41 - Leu645	0	0	0	0
(Ser)Ser42 - Leu645	0	0	2,3	2,2
(Tyr)Leu44 - Leu645	0	1,9	5,8	5,3
(Lys)Glu51 - Leu645	0	4,1	8,8	28
(Ala)Trp55 - Leu645	0	0	0	3,4
(Trp)Leu56 - Leu645	0	0	0	4,9

La delección de *YMP1* de SMD1163_pmp4 (B688) disminuyó los sitios de escisión de Phe36 / Thr37, Ser38 / Asp39 y Tyr43 / Leu44, pero mantuvo o aumentó la escisión en Lys50 / Glu51 y Lys28 / Gly29 (Tabla 2).

Tabla 2: Caldo de fermentación B688 de la cepa defectiva en *ymp1* de SMD1163_pmt4 (01:40)

Secuencia	0 h	24 h	48 h
His1-Leu645	93,1	75,1	64,9
(Trp)Leu26 - Leu645	0	0	0
(Leu)Val27 - Leu645	0	0	0
(Lys)Gly29 - Leu645	1,4	18,5	26
(Phe)Thr37 - Leu645	0	0,8	1
(Thr)Ser r38 - Leu645	0	0	0
(Ser)Asp39 - Leu645	0	0	0
(Val)Ser41 - Leu645	0	0	0
(Ser)Ser42 - Leu645	0	0	0
(Tyr)Leu44 - Leu645	0	0	0
(Lys)Glu51 - Leu645	5,5	4,6	7
(Ala)Trp55 - Leu645	0	0	0
(Trp)Leu56 - Leu645	0	0	0

La delección de la proteasa *yps1* de SMD1163_ *pmt4* (B805) no eliminó ningún sitio de escisión pero disminuyó la cantidad escindida en el sitio Lys50 / Glu51 (Tabla 3)

Tabla 3: Caldo de fermentación B805 de la cepa defectiva en *yps1* de SMD1163_ *pmt4* (01:40)

Secuencia	0 h	24 h	48 h	5 días
His1-Leu645	100	73,6	37,3	6
(Trp)Leu26 - Leu645	0	1,6	1,9	2,3
(Leu)Val27 – Leu645	0	0	0	0
(Lys)Gly29 – Leu645	0	4,9	5,6	2,7
(Phe)Thr37 – Leu645	0	10	27,1	41
(Thr)Ser38 – Leu645	0	0	0	0
(Ser)Asp39 – Leu645	0	5,5	16	28
(Val)Ser41 – Leu645	0	0	0	0
(Ser)Ser42 – Leu645	0	1	2,3	4,2
(Tyr)Leu44 – Leu645	0	2,3	5,6	9
(Lys)Glu51 – Leu645	0	1,1	2	3,1
(Ala)Trp55 – Leu645	0	0	1,9	3,8
(Trp)Leu56 – Leu645	0	0	0	0

5

Cuando se deleccionaron las proteasas *Ymp1* y *Yps1* de SMD1163_ *pmt4* (B803), la escisión se redujo significativamente y solo se produjo en Lys28/Gly29 (Tabla 4).

Tabla 4: Caldo de fermentación B803 de la cepa defectiva en *ymp1/yps1* de SMD1163_ *pmt4* (01:40)

Secuencia	0 h	24 h	48 h	5 días
His1-Leu645	100	94,5	89,4	73,1
(Trp)Leu26 - Leu645	0	0	0	0
(Leu)Val27 – Leu645	0	0	0	0
(Lys)Gly29 – Leu645	0	5,5	10,6	26,9
(Phe)Thr37 – Leu645	0	0	0	0
(Thr)Ser – Leu645r38	0	0	0	0
(Ser)Asp39 – Leu645	0	0	0	0
(Val)Ser41 – Leu645	0	0	0	0
(Ser)Ser42 – Leu645	0	0	0	0
(Tyr)Leu44 – Leu645	0	0	0	0
(Lys)Glu51 – Leu645	0	0	0	0
(Ala)Trp55 – Leu645	0	0	0	0
(Trp)Leu56 – Leu645	0	0	0	0

10 La inactivación individual de *ymp2*, *ymp3* se realizó en SMD1163. Se realizaron inducciones en matraz de agitación con estas cepas defectivas y se obtuvieron muestras de los sobrenadantes control SMD1163 y del cultivo a las 24 horas de la inducción. El patrón de proteína descrito por la SEC ID N° 1 se incubó con estas muestras a 30 °C durante 24 horas y se analizó mediante gel de SDS PAGE.

15 A partir del análisis PAGE, la delección de *ymp3* disminuyó significativamente la escisión del patrón de proteína, pero la delección de *ymp2* tiene menos efecto sobre la proteólisis (Véase la Figura 9).

Tabla 5

<u>Gen</u>	<u>Descripción</u>	<u>Huésped</u>	<u>Proteasa def.</u>	<u>Evaluación</u>
------------	--------------------	----------------	----------------------	-------------------

<u>Gen</u>	<u>Descripción</u>	<u>Huésped</u>	<u>Proteasa def.</u>	<u>Evaluación</u>
<u>PEP4</u>	<u>Aspartil proteasa</u>	<u>X-33</u>	<u>pep4</u>	<u>Proteolisis heteróloga reducida</u>
<u>PRB1</u>	<u>Serín proteasa</u>	<u>X-33 pep4</u>	<u>pep4, prb1</u>	<u>Proteolisis heteróloga reducida</u>
<u>YPS1</u>	<u>Aspartil proteasa</u>	<u>X-33 pep4</u>	<u>pep4, ysp1</u>	<u>Proteolisis heteróloga reducida</u>
<u>YPS2</u>	<u>Aspartil proteasa secretada</u>	<u>SMD1163</u>	<u>pep4, prb1, yps2</u>	<u>Proteolisis heteróloga reducida</u>
<u>YMP1 (SEC ID N° 4)</u>	<u>Serín proteasa secretada</u>	<u>SMD1163</u>	<u>pep4, prb1, ymp1</u>	<u>Proteolisis heteróloga reducida</u>
<u>DAP2</u>	<u>Endoproteasa de tipo DPP IV</u>	<u>SMD1163</u>	<u>pep4, prb1, dap2</u>	<u>Sin cambios en la proteolisis</u>
<u>GRH1</u>	<u>Serín proteasa</u>	<u>SMD1163 clon6</u>	<u>pep4, prb1, grh1</u>	<u>Sin cambios en la proteolisis</u>
<u>PRD1</u>	<u>metaloendopeptasa</u>	<u>SMD1163 c16 [PDI]</u>	<u>pep4, prb1, prd1</u>	<u>Sin cambios en la proteolisis</u>
<u>YSP3</u>	<u>Serín proteasa secretada</u>	<u>SMD1163 c16 [PDI]</u>	<u>pep4, prb1, vsp3</u>	<u>Sin cambios en la proteolisis</u>
<u>PRB2</u>	<u>Serín proteasa de tipo Prb1</u>	<u>SMD1163 c16 [PDI]</u>	<u>pep4, prb1, prb2</u>	=
<u>PRB3</u>	<u>Serín proteasa de tipo Prb1</u>	<u>SMD1163 clon6</u>	<u>pep4, prb1, prb3</u>	<u>Sin cambios en la proteolisis</u>
<u>YMP2 (SEC ID N° 6)</u>	<u>De tipo Arg/Ala aminopeptidasa</u>	<u>SMD1163</u>	<u>pep4, prb1, ymp2</u>	<u>Reducción moderada de la proteolisis heteróloga</u>
<u>YMP3 (SEC ID N° 8)</u>	<u>De tipo Leu aminopeptidasa</u>	<u>SMD1163</u>	<u>pep4, prb1, ymp3</u>	<u>Proteolisis heteróloga reducida</u>
<u>PPC1 (SEC ID N° 10)</u>	<u>Carboxipeptidasa de tipo CPY</u>	<u>SMD1163 c16</u>	<u>pep4, prb1, ppc1</u>	=
<u>PPC2 (SEC ID N° 12)</u>	<u>Carboxipeptidasa de tipo CPY</u>	<u>SMD1163 c16</u>	<u>pep4, prb1, ppc2</u>	=
<u>PPB1 (SEC ID N° 14)</u>	<u>De tipo proteasa B vacuolar</u>	<u>SMD1163 c16</u>	<u>pep4, prb1, ppb1</u>	=

Aquí se identifican seis polipéptidos, Pep4, Prb1, Yps1, Yps2 y los recién descubiertos Ymp1, Ymp3, en *Pichia* que son responsables de la degradación del polipéptido heterólogo (SEC ID N° 1). La inactivación genética de estos genes puede mejorar significativamente la productividad de la expresión de la proteína heteróloga.

5 Ejemplo 4 Manosiltransferasas defectivas

Cuando el polipéptido heterólogo (SEC ID N° 1) producido a partir de la cepa de *Pichia* de tipo salvaje GS 115 se analizó mediante CL / EM se observaron múltiples isoformas de masa incrementada de 162 Da. Un análisis adicional usando digestión con varias enzimas y CL / EM demostró que la glicosilación se producía en el resto HSA, y los

resultados se confirmaron usando secuenciación de Edman y análisis de la composición de hidratos de carbono (datos no mostrados). El análisis de la composición de hidratos de carbono confirmó que estas modificaciones se debían a la glicosilación con O-manosa, que es una modificación postraduccional frecuente de las proteínas producidas en *Pichia pastoris*.

5 Las cepas defectivas en manosiltransferasa de *Pichia* se realizaron utilizando los mismos procedimientos descritos en el Ejemplo 2 para las proteasas defectivas. Es decir, se insertó un casete de expresión KanMX en uno de estos genes responsables de la glicosilación en levadura: *OCH1*, *PMT1*, *PMT2*, y *PMT4*. Las cepas defectivas se transformaron con un vector capaz de expresar la SEC ID N° 1. Los cambios fenotípicos de estas cepas defectivas se demostraron usando el ensayo de sensibilidad zimoliasa (Figura 10). En este experimento, el cultivo de levadura en crecimiento exponencial se diluyó hasta DO600 = 0,3 en YPD suplementado con 5U / ml de zimoliasa. Los cultivos se incubaron a temperatura ambiente con agitación suave. Se recogieron muestras cada 15 minutos durante 10 2 horas y la medición de la DO600 se realizó como un indicador de la muerte celular.

Los genes de la manosiltransferasa son esenciales para la integridad de la pared celular completa. Las cepas defectivas en *pmt1* y *pmt4* pueden tener paredes celulares parciales o incompletas, que pueden resultar en una mayor sensibilidad al tratamiento con zimoliasa. Los datos mostraron que las mutantes *pmt1* y *pmt4* murieron rápidamente con el tratamiento con zimoliasa. 15

La proteína heteróloga se evaluó para determinar la glicosilación comparando la producción heteróloga en cepas modificadas genéticamente con cepas de tipo salvaje. El resumen de los resultados para las cepas defectivas se presentan en la Tabla 6 a continuación. Asimismo, véase la Figura 11. Los resultados mostraron que Pmt4 es el único responsable de la glicosilación de HSA en *Pichia*. Una cepa defectiva solo en *pmt4* puede producir HSA humana sin glicosilar. 20

Tabla 6

<u>Gen</u>	<u>Descripción</u>	<u>Huésped</u>	<u>Proteasa def.</u>	<u>Evaluación</u>
<u>OCH1</u>	<u>manosiltransferasa</u>	<u>SMD1163</u>	<u>pep4, prb1, och1</u>	<u>sin cambios en la glicosilación</u>
<u>PMT1</u>	<u>manosiltransferasa</u>	<u>SMD1163</u>	<u>pep4, prb1, pmt1</u>	<u>sin cambios en la glicosilación</u>
<u>PMT2</u>	<u>manosiltransferasa</u>	<u>SMD1163</u>	<u>pep4, prb1, pmt2</u>	<u>ausencia de inactivación limpia</u>
<u>PMT4</u>	<u>manosiltransferasa</u>	<u>SMD1163</u>	<u>pep4, prb1, pmt4</u>	<u>sin glicosilación</u>

Ejemplo 5 – Coexpresión con proteínas chaperonas

25 Varias proteínas chaperonas pudieron coexpresarse con la proteína heteróloga en una célula huésped para aumentar la producción de proteínas heterólogas. Ejemplos de proteínas chaperonas que pueden coexpresarse con la proteína heteróloga en levaduras para aumentar la producción heteróloga se presentan en la Tabla 7 a continuación.

Tabla 7

GEN	Nombre de la proteína	Especie de origen
<i>HAC1</i>	Factor de transcripción bZip	<i>S. cerevisiae</i>
<i>KAR2</i>	Proteína de unión BiP	<i>S. cerevisiae</i>
<i>EUG</i>	Homólogo de PDI	<i>S. cerevisiae</i>
<i>JEM1</i>	Proteína de tipo DnaJ de la membrana del RE 1	<i>Pichia</i> y <i>S. cerevisiae</i>
<i>CUP5</i>	Subunidad proteolipídica de la V-ATPasa	<i>Pichia</i> y <i>S. cerevisiae</i>
<i>KIN2</i>	serina/treonina proteína quinasa	<i>Pichia</i> y <i>S. cerevisiae</i>
<i>SSA4</i>	Proteína del shock térmico	<i>Pichia</i> y <i>S. cerevisiae</i>
<i>SSE1</i>	Proteína del shock térmico	<i>Pichia</i> y <i>S. cerevisiae</i>

HSP26	Proteína del shock térmico de 26 kDa	<i>S. cerevisiae</i>
-------	--------------------------------------	----------------------

Listado de secuencias

- <110> JIN, Yong Hwan
 JOWETT, James D.
 5 TAYLOR, Alexander H.
 ZHU, Yuan
- <120> Células huésped y procedimientos de uso
- 10 <130> PU63510
- <140> tbd
 <141> 2010-02-24
- 15 <150> 61.155.706
 <151> 2009-02-26
- <160> 15
- 20 <170> FastSEQ para Windows Versión 4.0
- <210> 1
 <211> 645
 25 <212> PRT
 <213> Secuencia artificial
- <220>
 30 <223> Homo Sapiens
- <400> 1
- 35 His Gly Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly

ES 2 539 378 T3

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

ES 2 539 378 T3

Ala Trp Ala Val Ala Arg Leu Ser Gln Arg Phe Pro Lys Ala Glu Phe
 275 280 285

Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys Val His Thr Glu
 290 295 300

Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu
 305 310 315 320

Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys
 325 330 335

Glu Cys Cys Glu Lys Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu
 340 345 350

Val Glu Asn Asp Glu Met Pro Ala Asp Leu Pro Ser Leu Ala Ala Asp
 355 360 365

Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala Glu Ala Lys Asp
 370 375 380

Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp
 385 390 395 400

Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr
 405 410 415

Leu Glu Lys Cys Cys Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys
 420 425 430

Val Phe Asp Glu Phe Lys Pro Leu Val Glu Glu Pro Gln Asn Leu Ile
 435 440 445

Lys Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu Tyr Lys Phe Gln
 450 455 460

Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr
 465 470 475 480

Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys
 485 490 495

Cys Cys Lys His Pro Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr
 500 505 510

Leu Ser Val Val Leu Asn Gln Leu Cys Val Leu His Glu Lys Thr Pro
 515 520 525

Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser Leu Val Asn Arg

ES 2 539 378 T3

530 535 540
 Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys
 545 550 555 560
 Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu
 565 570 575
 Ser Glu Lys Glu Arg Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu
 580 585 590
 Val Lys His Lys Pro Lys Ala Thr Lys Glu Gln Leu Lys Ala Val Met
 595 600 605
 Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys Ala Asp Asp Lys
 610 615 620
 Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Gln
 625 630 635 640
 Ala Ala Leu Gly Leu
 645

<210> 2
 <211> 31
 5 <212> PRT
 <213> Homo Sapiens

<220>
 <221> VARIANTE
 10 <222> 31
 <223> Xaa = Cualquier aminoácido

<400> 2

His Ala Glu Gly Thr Phe Thr Ser Asp Val Ser Ser Tyr Leu Glu Gly
 1 5 10 15
 Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Xaa

15 20 25 30

<210> 3
 <211> 5115
 <212> ADN
 20 <213> *Pichia Pastoris*

<400> 3

ES 2 539 378 T3

atgttctca aaagtctct tagtttgcg tctatcctaa cgcttfgcaa ggcctgggat 60
ctggaagatg tacaagatgc accaaagatc aaaggtaatg aagtaccggg tcgctatctc 120
attgagatg aagaagcttc cacttcagca tttgctaccc aactgagagc tgggggat 180
gactttaaca tccaatcaga ctactcaact ggttccctt tcaacggagc atctgttcaa 240
atcagcaacg ataacaaaac cactttccag gatttgcaaa gtttgcgtgc agtcaaaaat 300
gtttaccag ctactctcat tacattagat gaaacattg agcttctga cacgaagcca 360
tggaaccctc atggaattac cgggtgat tcttgcag agcaaggata tactgtagt 420
ggtgttfta ttgcagttat cgatactggt gttgactata cacacctgc tctgggtggt 480
ggtatcggag ataattccc tatcaaagct ggttatgatt tgtctccgg tgatggtgc 540
atcacgaatg atcctatgga ttgtacggt catggtacct ttgatctc catcattgt 600
gcaaataaca aagatatggt tgggttga ccagatgctc agattgcat gtacaaagt 660
ttcccctgt ctgatagiac ttcgactgac atagtatgg cgggatgca aaaggcctat 720
gatgatggtc acaagattat ttcgctatca ctgggatctg actcggggtt ttccagta 780
ccagcttctc taatggccag caggattgct caagacagag ttgtttggt ggctgctggt 840
aactctggag aacttggtcc attctatgcc tctcccctg ctctgggaa acaagtcatt 900
tcagttgat ctgtcaaaa cgaacaatgg acaaccttc cagtaacct tacctctca 960
aacggtgaat caagggttt tcttacctc gcttacaatg gtcacagat tggattgat 1020
gccgagctg aggttgatt taccgaagaa agaggatgcg tctatgaacc agagatctc 1080
gcagataatg cgaataaagc tatttgta agaagggcg tcggctgtgt tgaactg 1140
gaattcaati tattgtctgt ggctggttac aaggcttact tctgtacaa ctcaattca 1200
agaccatgga gtctctgaa tattctcca ctgattgagc tagacaacgc ttactctt 1260
gttgaagagg aagttggaat atgggtgaaa acccaatcg acgccggtaa caccgtcaag 1320
ftaaaggatg gcacagatga ccaatgttg ccatctgata aagagtattt gggagttgga 1380
aagatggatt attactctc tcaaggacct gcttatgagc ttgaatttt cccaacgata 1440
tccgctccag gtggagacag ttggggcgt tggcccgtg ggcaatcgg tgttgcctca 1500
ggaacaagt ttgctgcc ctatgttga ggtcttacag ctcttatga atcgagtt 1560

ES 2 539 378 T3

ggaattcaag atccccagga ctatgtgaga aaattagtct ccacagctac cgatcttcaa 1620
ttatttgact ggaacgcagt gaaacttgag acctctatga atgctccact tattcaacag 1680
ggagctggtc tagtgaacgc tcttggttg tttgagacta agactgtgat cgtgtctgct 1740
ccttatttgg agctcaatga cacatcaat agagccagtg agtataccat tcaaattaag 1800
aatgagaact ctgagactat tacctatcaa gttgttcacg ttccgggaac tactgtctac 1860
tctagatcag cttctgggaa catcccatac ctggtaac aagatttgc accttacggt 1920
gatagtgat ctgcgacagt tgctctatcc acagaagagt tggtttggg accaggagaa 1980
gttggtgaag tcaactgtgat cttcttaca gaagaaattg atcaagaac tgctccaatt 2040
attcagggtg agattacatt ttatgggat gtcataccga ttgctgtcc ttatatggga 2100
gttgaagtg atattcattc ctgggagcct ctcatgaga ggccttacc agtgagaatg 2160
tatttgatg atggttcctt agcatatgtt gatgatgac ctgattatga gtccaatgtg 2220
tatgactggg attctcctag attttattt aacctgagat atgcaaccaa agaagtatcg 2280
attgacttgg tgcacctga ttatagcatt gagaacgact acgaatggcc tttagttcc 2340
ggacacaaca actattatgg tcccgtggga tacgactacg attatacctc gggtaagcc 2400
ttttgctc gttacttca acaactatt aacgaactg gatacttcc ttttccaga 2460
tttctaact ttctgtagt tctgtgtgt gaatacaaag ctctatttag agttttgcta 2520
ccatattggg acttttggaa caaagaagac tggcaattgt tgaatcccc agtgtttaa 2580
gtcctcgctc caccgaatga agaaaacact actgaagagc caactgagga atccagcgag 2640
gagcctaccg aagagtcaac gtctgagta actgaagagc cctctctga gtaactgag 2700
aatctagcg aggtccaac tgaagaaatt actgaagatg caacatccac aattgatgat 2760
gatgaagcat ccaccgaaag ctctactgaa gaaccaagtg ctgagccac cggctctac 2820
tctgatttga ctgtcgggta ggccattacc gacgttagtg tcaccagttt gaggacaact 2880
gaagcatttga gatacctc cgactgggtg gttgtgtctt tcaatttcaa cactactgac 2940
agagatatta ctctcccacc ttacgtgtt gtacaagtaa ctatccaaa tgaacttcaa 3000
ttcattgctc atccagaata cggcccatac ctgagccct cattgcaagt ttctacact 3060
aagaatgaaa gattaattat gactagttag ttcaactacg acaccagagt catcgactc 3120
aagtttgaca atcgagacca agtaataact caagtggagg gaggtttga tttcacgatg 3180
aaactagaac aagatttcat ttctgcattg gcccaggtg aatacgattt tgaatttcat 3240
acatcgggtg attcttatgc ttgacctt gactttatc cattgattag atccgagcca 3300
atcaaattga tagcaggtgc accagacgaa gttgaatggt ttattgatat tccaagtga 3360
tacagcgatt tggcaacgat agatattagt tctgatatcg atactaatga taatttgcag 3420
cagtactct atgattgctc aaagctcaag tacactattg gaaaagagt ttgatcagtg 3480
ggtaatttca cagctggatc agatggtaac caatacagca ataccaccga tgggtatgtt 3540

ES 2 539 378 T3

ccaattactg attctaccgg ctctccagta gctgaagttc aatgtttaat ggaaagtatc 3600
tcattgagtt tcacaaatac tcttgctgag gatgaagtat tgagagttgt tcttactct 3660
tctgcgttta gacgtggttc atccacatg gccaacgtgg taaacgttga cattacagct 3720
gggggattgg caaaaagaga actcttctct tatatattgg atgaaaatta ctatgctagt 3780
actggatctg agggggttggc attgacgta tttgaagttg ctgatcaggt cgaggagcca 3840
actgaggagt caacctcaga ggaatctact gaacaggaaa ctccaccga ggaacctacc 3900
gaggaatcaa ctgaacctac tgaggaatct acccaggaac ctactgaaga gccaccgac 3960
gagcctactt ctgagtcaac tgaggaacct tctgaggagc caacttctga cgatctctca 4020
attgacccaa ctgctgtacc taccgatgaa cctactgaag agccaactga ggagcctact 4080
tctgagtcaa ctgaggaacc ttctgaggag ccaactctg acgatctctc aattgaccca 4140
actgctgtac ctaccgatga acctactgaa gagccaactg aggagcctac ttctgagtea 4200
actgaggaac cttctgagga gccaaactct gacgatctct caattgacct aactgctgta 4260
cctaccgatg aacctactga agagccaact gaggagccga cctctgagac taccgatgat 4320
ccatcgatag cacctactgc tgtgccaact tccgacacat ctctggaca atcggtggtt 4380
actcaaaaca ctacagtcac tcagactacc atcaactcag tctgtaatgt ttgtgctgag 4440
accctgtaa caatcacta cactgcacca gttgtgacta agccagttc ttacaccacc 4500
gttacttcag tttgcatgt atgtgcagag acaccaatca cagttacctt gacgttgcca 4560
tgtgaaaccg aagacgtgac aaagactgcc ggcctaaga ctgtcactta caccgaagt 4620
tgcaactct gtgctgacaa gcctatcact tacacctaca tgctccaga gtacactcaa 4680
ggtgccgaac gtacaacagt tacatcggtt tgcaacgttt gtgctgagac acctgtaacg 4740
ctaacataca ctgcccgaag agccagctgt catacagttc ctcaacaata ttcaagtgcc 4800
ggagagctca ttcatccaa ggggatcacg attcctactg ttctgcccg tccaactggt 4860
acttatagta agtctgtga cactagccaa cgtacactcg ctaccattac aaaatcttca 4920
gatgagtcta aactgttac cactactcaa gccacacaag tttgagcgg tgaatccagt 4980
ggaattcaag ctgcttcaaa cagcagcagc atctcagctc caactgtcac tacagctggg 5040
aacgagaact ctggatctag atttctggtt gctggactat tcacagttct gcctcttacc 5100
ttgtcgtta tataa 5115

<210> 4
<211> 1703
<212> PRT
<213> *Pichia Pastoris*

5

<400> 4

ES 2 539 378 T3

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

ES 2 539 378 T3

Asp Asp Gly His Lys Ile Ile Ser Leu Ser Leu Gly Ser Asp Ser Gly
 245 250 255
 Phe Ser Ser Thr Pro Ala Ser Leu Met Ala Ser Arg Ile Ala Gln Asp
 260 265 270
 Arg Val Val Leu Val Ala Ala Gly Asn Ser Gly Glu Leu Gly Pro Phe
 275 280 285
 Tyr Ala Ser Ser Pro Ala Ser Gly Lys Gln Val Ile Ser Val Gly Ser
 290 295 300
 Val Gln Asn Glu Gln Trp Thr Thr Phe Pro Val Thr Phe Thr Ser Ser
 305 310 315 320
 Asn Gly Glu Ser Arg Val Phe Pro Tyr Leu Ala Tyr Asn Gly Ala Gln
 325 330 335
 Ile Gly Phe Asp Ala Glu Leu Glu Val Asp Phe Thr Glu Glu Arg Gly
 340 345 350
 Cys Val Tyr Glu Pro Glu Ile Ser Ala Asp Asn Ala Asn Lys Ala Ile
 355 360 365
 Leu Leu Arg Arg Gly Val Gly Cys Val Glu Asn Leu Glu Phe Asn Leu
 370 375 380
 Leu Ser Val Ala Gly Tyr Lys Ala Tyr Phe Leu Tyr Asn Scr Phe Ser
 385 390 395 400
 Arg Pro Trp Ser Leu Leu Asn Ile Ser Pro Leu Ile Glu Leu Asp Asn
 405 410 415
 Ala Tyr Ser Leu Val Glu Glu Glu Val Gly Ile Trp Val Lys Thr Gln
 420 425 430
 Ile Asp Ala Gly Asn Thr Val Lys Leu Lys Val Ser Thr Ser Asp Gln
 435 440 445
 Met Leu Pro Ser Asp Lys Glu Tyr Leu Gly Val Gly Lys Met Asp Tyr
 450 455 460
 Tyr Ser Ser Gln Gly Pro Ala Tyr Glu Leu Glu Phe Phe Pro Thr Ile
 465 470 475 480
 Ser Ala Pro Gly Gly Asp Ser Trp Gly Ala Trp Pro Gly Gly Gln Tyr
 485 490 495
 Gly Val Ala Ser Gly Thr Ser Phe Ala Cys Pro Tyr Val Ala Gly Leu

ES 2 539 378 T3

500 505 510
 Thr Ala Leu Tyr Glu Ser Gln Phe Gly Ile Gln Asp Pro Gln Asp Tyr
 515 520 525
 Val Arg Lys Leu Val Ser Thr Ala Thr Asp Leu Gln Leu Phe Asp Trp
 530 535 540
 Asn Ala Val Lys Leu Glu Thr Scr Met Asn Ala Pro Leu Ile Gln Gln
 545 550 555 560
 Gly Ala Gly Leu Val Asn Ala Leu Gly Leu Phe Glu Thr Lys Thr Val
 565 570 575
 Ile Val Ser Ala Pro Tyr Leu Glu Leu Asn Asp Thr Ile Asn Arg Ala
 580 585 590
 Ser Glu Tyr Thr Ile Gln Ile Lys Asn Glu Asn Ser Glu Thr Ile Thr
 595 600 605
 Tyr Gln Val Val His Val Pro Gly Thr Thr Val Tyr Ser Arg Ser Ala
 610 615 620
 Ser Gly Asn Ile Pro Tyr Leu Val Asn Gln Asp Phe Ala Pro Tyr Gly
 625 630 635 640
 Asp Ser Asp Ala Ala Thr Val Ala Leu Ser Thr Glu Glu Leu Val Leu
 645 650 655
 Gly Pro Gly Glu Val Gly Glu Val Thr Val Ile Phe Ser Thr Glu Glu
 660 665 670
 Ile Asp Gln Glu Thr Ala Pro Ile Ile Gln Gly Lys Ile Thr Phe Tyr
 675 680 685
 Gly Asp Val Ile Pro Ile Ala Val Pro Tyr Met Gly Val Glu Val Asp
 690 695 700
 Ile His Ser Trp Glu Pro Leu Ile Glu Arg Pro Leu Ser Val Arg Met
 705 710 715 720
 Tyr Leu Asp Asp Gly Ser Leu Ala Tyr Val Asp Asp Asp Pro Asp Tyr
 725 730 735
 Glu Phe Asn Val Tyr Asp Trp Asp Ser Pro Arg Phe Tyr Phe Asn Leu
 740 745 750
 Arg Tyr Ala Thr Lys Glu Val Ser Ile Asp Leu Val His Pro Asp Tyr
 755 760 765

ES 2 539 378 T3

Ser Ile Glu Asn Asp Tyr Glu Trp Pro Leu Val Ser Gly His Asn Asn
 770 775 780
 Tyr Tyr Gly Pro Val Gly Tyr Asp Tyr Asp Tyr Thr Ser Gly Gln Ala
 785 790 795 800
 Phe Leu Pro Arg Tyr Phe Gln Gln Arg Ile Asn Glu Leu Gly Tyr Leu
 805 810 815
 Ser Phe Ser Arg Phe Ala Asn Phe Ser Val Val Pro Ala Gly Glu Tyr
 820 825 830
 Lys Ala Leu Phe Arg Val Leu Leu Pro Tyr Gly Asp Phe Trp Asn Lys
 835 840 845
 Glu Asp Trp Gln Leu Phe Glu Ser Pro Val Phe Asn Val Leu Ala Pro
 850 855 860
 Pro Asn Glu Glu Asn Thr Thr Glu Glu Pro Thr Glu Glu Ser Ser Glu
 865 870 875 880
 Glu Pro Thr Glu Glu Ser Thr Ser Glu Ser Thr Glu Glu Pro Ser Ser
 885 890 895
 Glu Ser Thr Glu Lys Ser Ser Glu Val Pro Thr Glu Glu Ile Thr Glu
 900 905 910
 Asp Ala Thr Ser Thr Ile Asp Asp Asp Glu Ala Ser Thr Glu Ser Ser
 915 920 925
 Thr Glu Glu Pro Ser Ala Gln Pro Thr Gly Pro Tyr Ser Asp Leu Thr
 930 935 940
 Val Gly Glu Ala Ile Thr Asp Val Ser Val Thr Ser Leu Arg Thr Thr
 945 950 955 960
 Glu Ala Phe Gly Tyr Thr Ser Asp Trp Leu Val Val Ser Phe Thr Phe
 965 970 975
 Asn Thr Thr Asp Arg Asp Ile Thr Leu Pro Pro Tyr Ala Val Val Gln
 980 985 990
 Val Thr Ile Pro Asn Glu Leu Gln Phe Ile Ala His Pro Glu Tyr Ala
 995 1000 1005
 Pro Tyr Leu Glu Pro Ser Leu Gln Val Phe Tyr Thr Lys Asn Glu Arg
 1010 1015 1020
 Leu Ile Met Thr Ser Gln Phe Asn Tyr Asp Thr Arg Val Ile Asp Phe

ES 2 539 378 T3

1025 1030 1035 1040
 Lys Phe Asp Asn Arg Asp Gln Val Ile Thr Gln Val Glu Gly Val Val
 1045 1050 1055
 Tyr Phe Thr Met Lys Leu Glu Gln Asp Phe Ile Ser Ala Leu Ala Pro
 1060 1065 1070
 Gly Glu Tyr Asp Phe Glu Phe His Thr Ser Val Asp Ser Tyr Ala Ser
 1075 1080 1085
 Thr Phe Asp Phe Ile Pro Leu Ile Arg Ser Glu Pro Ile Lys Leu Ile
 1090 1095 1100
 Ala Gly Ala Pro Asp Glu Val Glu Trp Phe Ile Asp Ile Pro Ser Ala
 1105 1110 1115 1120
 Tyr Ser Asp Leu Ala Thr Ile Asp Ile Ser Ser Asp Ile Asp Thr Asn
 1125 1130 1135
 Asp Asn Leu Gln Gln Tyr Phe Tyr Asp Cys Ser Lys Leu Lys Tyr Thr
 1140 1145 1150
 Ile Gly Lys Glu Phe Asp Gln Trp Gly Asn Phe Thr Ala Gly Ser Asp
 1155 1160 1165
 Gly Asn Gln Tyr Ser Asn Thr Thr Asp Gly Tyr Val Pro Ile Thr Asp
 1170 1175 1180
 Ser Thr Gly Ser Pro Val Ala Glu Val Gln Cys Leu Met Glu Ser Ile
 1185 1190 1195 1200
 Ser Leu Ser Phe Thr Asn Thr Leu Ala Glu Asp Glu Val Leu Arg Val
 1205 1210 1215
 Val Leu His Ser Ser Ala Phe Arg Arg Gly Ser Phe Thr Met Ala Asn
 1220 1225 1230
 Val Val Asn Val Asp Ile Thr Ala Gly Gly Leu Ala Lys Arg Glu Leu
 1235 1240 1245
 Phe Ser Tyr Ile Leu Asp Glu Asn Tyr Tyr Ala Ser Thr Gly Ser Glu
 1250 1255 1260
 Gly Leu Ala Phe Asp Val Phe Glu Val Ala Asp Gln Val Glu Glu Pro
 1265 1270 1275 1280
 Thr Glu Glu Ser Thr Ser Glu Glu Ser Thr Glu Gln Glu Thr Ser Thr
 1285 1290 1295

ES 2 539 378 T3

Glu Glu Pro Thr Glu Glu Ser Thr Glu Pro Thr Glu Glu Ser Thr Gln
 1300 1305 1310
 Glu Pro Thr Glu Glu Pro Thr Asp Glu Pro Thr Ser Glu Ser Thr Glu
 1315 1320 1325
 Glu Pro Ser Glu Glu Pro Thr Ser Asp Asp Leu Ser Ile Asp Pro Thr
 1330 1335 1340
 Ala Val Pro Thr Asp Glu Pro Thr Glu Glu Pro Thr Glu Glu Pro Thr
 1345 1350 1355 1360
 Ser Glu Ser Thr Glu Glu Pro Ser Glu Glu Thr Ser Asp Asp Leu Ser
 1365 1370 1375
 Ile Asp Pro Thr Ala Val Pro Thr Asp Glu Pro Thr Glu Glu Pro Thr
 1380 1385 1390
 Glu Glu Pro Thr Ser Glu Ser Thr Glu Glu Pro Ser Glu Glu Pro Thr
 1395 1400 1405
 Ser Asp Asp Leu Ser Ile Asp Pro Thr Ala Val Pro Thr Asp Glu Pro
 1410 1415 1420
 Thr Glu Glu Pro Thr Glu Glu Pro Thr Ser Glu Thr Thr Asp Asp Pro
 1425 1430 1435 1440
 Ser Ile Ala Pro Thr Ala Val Pro Thr Ser Asp Thr Ser Ser Gly Gln
 1445 1450 1455
 Ser Val Val Thr Gln Asn Thr Thr Val Thr Gln Thr Thr Ile Thr Ser
 1460 1465 1470
 Val Cys Asn Val Cys Ala Glu Thr Pro Val Thr Ile Thr Tyr Thr Ala
 1475 1480 1485
 Pro Val Val Thr Lys Pro Val Ser Tyr Thr Thr Val Thr Ser Val Cys
 1490 1495 1500
 His Val Cys Ala Glu Thr Pro Ile Thr Val Thr Leu Thr Leu Pro Cys
 1505 1510 1515 1520
 Glu Thr Glu Asp Val Thr Lys Thr Ala Gly Pro Lys Thr Val Thr Tyr
 1525 1530 1535
 Thr Glu Val Cys Asn Ser Cys Ala Asp Lys Pro Ile Thr Tyr Thr Tyr
 1540 1545 1550
 Ile Ala Pro Glu Tyr Thr Gln Gly Ala Glu Arg Thr Thr Val Thr Ser

ES 2 539 378 T3

1555 1560 1565
Val Cys Asn Val Cys Ala Glu Thr Pro Val Thr Leu Thr Tyr Thr Ala
1570 1575 1580
Pro Lys Ala Ser Arg His Thr Val Pro Ser Gln Tyr Ser Ser Ala Gly
1585 1590 1595 1600
Glu Leu Ile Ser Ser Lys Gly Ile Thr Ile Pro Thr Val Pro Ala Arg
1605 1610 1615
Pro Thr Gly Thr Tyr Ser Lys Ser Val Asp Thr Ser Gln Arg Thr Leu
1620 1625 1630
Ala Thr Ile Thr Lys Ser Ser Asp Glu Ser Asn Thr Val Thr Thr Thr
1635 1640 1645
Gln Ala Thr Gln Val Leu Ser Gly Glu Ser Ser Gly Ile Gln Ala Ala
1650 1655 1660
Ser Asn Ser Thr Ser Ile Ser Ala Pro Thr Val Thr Thr Ala Gly Asn
1665 1670 1675 1680
Glu Asn Ser Gly Ser Arg Phe Ser Phe Ala Gly Leu Phe Thr Val Leu
1685 1690 1695
Pro Leu Ile Leu Phe Val Ile
1700

<210> 5
<211> 618
5 <212> ADN
<213> *Pichia* Pastoris

<400> 5

atggctceca gaacactacc agaagaetta attcctccc tatacgactt gcacatctac 60
aactccaac ccgaaaaaaaa gacttatgat ggagacattg tcatccactt ggagggtgaag 120
gagcccactg atgaagtggc ttcaatgcc aaggatttgg aattgaaaga cgtacatgtc 180
ttccacaatg tcaacaagtc tgaaaacgaa atccccgta aggagattgt tgataacgag 240
ctcatcaaa ttaagctcaa agagaagggt acttccggaa cgttgctggt gaatatttc 300
10 ttaccggta acattcaatc tgataaaatt ggattttaca agggagacac agatgtggaa 360
ggaagagtca catacactac aaaccttacc actcctaatg ccagggtggc attccatgt 420
cttgataaca tattgtttaa agctccattc aagttcggag taactgcaa tccaggacaa 480
ttagtgagt ccatttggg tctaagctct gaggctgacg tcttgaatga caatgacgat 540
gtgattgga cgagatacca ataccaagt agtgagccaa tagccccagc ttactggag 600
tggaccattc atatttaa 618

<210> 6
<211> 205
15 <212> PRT
<213> *Pichia* Pastoris

<400> 6

ES 2 539 378 T3

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

<210> 7
 <211> 1233
 <212> ADN
 <213> *Pichia* Pastoris

<400> 7

5

10

atggtaaac tcataatcaat tatagcccta gttcaacttg tctctgcgac aattgtacct 60
 tggaaatccc agaactgttt atctgacgtc catcaccctt ctctccatct cttggattat 120
 attcaatcct tgaagaacga ggtaatgttc gatggcgacg atcgcagaat aatcaagtta 180
 ggcccccaag aataccgtat tatcactgaa aaagagaaat accagttgaa aacagagggg 240
 atatcattta tcatgtcac ctatcagcat ggagacaatg tagagctgct ctattccagt 300
 gcgccagtta ccgtccaga ctatctttat ccgtccaatg atactttcca ttcaaacaa 360
 gtaaattctt tgataggtga gattgacatt ggcagaatgc aggcgttttt gggaaggttc 420
 tctagcttct ttacaagatt ttacaaatct gacaaggggt tgcagagttc tatctggtta 480
 caaggtgaat tggftcaatt ggccttgaaa gatccatga ggttcaatgt tactactgtg 540
 gaacaccctt ggaagcagaa ttctgccatc ttacgatat acggtgaaaa tgttgatcct 600
 tcgaaaggaa aaggggacat tgtagtagtg ggatgccatc aagattccat aaacttgctt 660
 ttccccaaca ttctccgtgc tccaggggct gatgatgatg gatctggtgt aactccaac 720
 cttgaagcgc tcagaatcat agttgaaagt ggcctcaagt ttacaatac agtagagttt 780
 cacttttatt ctgccgaaga aggaggacta cttggctccc agcaaatttt cagctcgtat 840
 agagctgcag aagagactgt tgttgctatg ctacaacagg acatgactgg atacatcaa 900
 aaagctttag accacgggga atccgaccac ttcgggctaa tcaactgacca tacaacgca 960
 aatctgaata gcttccttgc acttttaate gatgcataca ctcaattcc ctacaagaa 1020
 accgaatgtg ggtatgctg ctcatatcat agttctgcct tggaaatgg ttatccatct 1080
 gccatggtct ttgaaagtag ttttgctac acaaaccct tcacatag cacccaagac 1140
 acaattgaca agatcaattt tccacatag gcagagcatg tcaagttggt cctgggttac 1200
 gttgtagagt tgggattaga acattttagg tga 1233

- <210> 8
- 5 <211> 410
- <212> PRT
- <213> *Pichia Pastoris*
- <400> 8
- 10

ES 2 539 378 T3

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

ES 2 539 378 T3

195 200 205
 Val Val Gly Cys His Gln Asp Ser Ile Asn Leu Leu Phe Pro Asn Ile
 210 215 220
 Leu Arg Ala Pro Gly Ala Asp Asp Asp Gly Ser Gly Val Thr Ser Asn
 225 230 235 240
 Leu Glu Ala Leu Arg Ile Ile Val Glu Ser Gly Leu Lys Phe His Asn
 245 250 255
 Thr Val Glu Phe His Phe Tyr Ser Ala Glu Glu Gly Gly Leu Leu Gly
 260 265 270
 Ser Gln Gln Ile Phe Ser Ser Tyr Arg Ala Ala Glu Glu Thr Val Val
 275 280 285
 Ala Met Leu Gln Gln Asp Met Thr Gly Tyr Ile Gln Lys Ala Leu Asp
 290 295 300
 His Gly Glu Ser Asp His Phe Gly Leu Ile Thr Asp His Thr Asn Ala
 305 310 315 320
 Asn Leu Asn Ser Phe Leu Ala Leu Leu Ile Asp Ala Tyr Thr Ser Ile
 325 330 335
 Pro Tyr Lys Glu Thr Glu Cys Gly Tyr Ala Cys Ser Asp His Ser Ser
 340 345 350
 Ala Leu Glu His Gly Tyr Pro Ser Ala Met Val Phe Glu Ser Ser Phe
 355 360 365
 Ala Tyr Thr Asn Pro Phe Ile His Ser Thr Gln Asp Thr Ile Asp Lys
 370 375 380
 Ile Asn Phe Pro His Met Ala Glu His Val Lys Leu Val Leu Gly Tyr
 385 390 395 400
 Val Val Glu Leu Gly Leu Glu His Phe Arg
 405 410

- <210> 9
- <211> 1605
- 5 <212> ADN
- <213> *Pichia Pastoris*
- <400> 9

ES 2 539 378 T3

atgaaatcgg ttattggag ccttctatct ttgctagcat tgtcgcaggc attgactatt 60
 ccattgctgg aagagcttca acagcaaaca ttttttagca agaaaaccgt toctcaacaa 120
 gttgctgaat tggtagggcac ccattactct aaggatgaga taatcagtct atggaaggac 180
 attgagctgg atgtaccag gaaaaagatc caagaggcct tcgataagtt cgtaaaaaa 240
 tcaactgcca ctccccctg tagaaatgaa ttcccttgt ctacagcaaga ttgggtgaca 300
 gtgaccaaca ccaagtttga taattatcaa ttgagggtta aaaaatccca ccctgaaaag 360
 ctaaacattg ataaggtaaa gcaatcttcg ggatacctgg atatcattga tcaagataag 420
 catctttct attggtttt tgaatccga aatgatccgt ccacagacc aatcatccta 480
 tggttgaatg gtaggaccgg ctgctcttct attacagggt tgctattcga aaagattggc 540
 ccagttaca tcaccaaga gattaagccg gaacataatc ctattcatg gaacaacaat 600
 gctagtgtta tcttcttga gcaaccggtt ggagtaggat ttcttactc ttctaagaaa 660
 gtcggtgata ctgcaactgc tgccaaagat acatatgtgt tttggagct ttcttccaa 720
 aagttctct agttctgac ctctaactg cacattgctg gggaaatcga tcttgccat 780
 tattgcccc agattgcttc tgagattgtg tctcacgcag acaagacgtt tgaccttca 840
 ggagtcata tcggtaatgg tcttactgat cctctaatic agtataagta ctatcagcca 900
 atggcctgtg gaaaaggtgg ctacaagcag gtcatttcgg acgaggaatg tgatgaattg 960
 gatagggtct atccaagatg tgaacgttta acgcgggcat gttatgagtt ccaaaattca 1020
 gttacttgtg ttccggcaac actttattgc gacaaaagc tactgaagcc gtacactgac 1080
 actggcttga atgtctatga tattcgtaca atgtgcgatg aagggactga tttgtgttac 1140
 aaagaactgg aatacgtgga gaagtacatg aaccagcctg aagtgcagga agccgtgggc 1200
 tctgaagtca gttcttaca aggttgtgac gatgatgtct tcttaagatt tttgtactct 1260
 ggcgatggat ctaagccttt ccaccagtat atcacggatg ttctcaatgc aagtattccg 1320
 gttctgattt acgcaggtga taaagattat atctgtaatt ggctaggaaa ccaagcttgg 1380
 gtcaatgagc tagaatggaa ctgtctgag gaattccagg caactccgat tcgaccgtgg 1440
 ttcactttgg acaataacga ttatgcagga aacgtacaaa ctatggaaa ctttctctt 1500
 ctaagagtat ttgatgctgg tcacatggtt cttacaatc aaccagtcaa cgcacttgac 1560
 atggttgcga gatggacaca cggtgatttc tcatttggtt attaa 1605

<210> 10
 <211> 534
 <212> PRT
 <213> *Pichia* Pastoris

5

<400> 10

ES 2 539 378 T3

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

ES 2 539 378 T3

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

ES 2 539 378 T3

<212> ADN
<213> *Pichia* Pastoris

<400> 11

5

atgatattac acacctatat tattctctcg ttattgacta tatttcttaa agctattggt 60
ctgtccttgc agatgccaat ggccttgga gctagttag cctcattagt ggagaaagca 120
accctcgctg ttggacaaga aattgatgcc atacaaaagg gtattcagca aggttggttg 180
gaagtagaga caagattcc aactatagtg tcacagttag cctatagtac tggcccaaaa 240
tttgcgatca agaagaaaga tgcaactttt tgggatttct atgttgaaag tcaagagttg 300
ccaaactacc gaattaatga aaatctgaaa ccaattttca acccctattc gtggaatggt 360
aatgcttcaa tcatctactt agatcaaccg gtcaatgttg ggttttctta ttcttcatca 420
tcggtgagta aactgttgtg tgcgggagaa gatgtgatg cattttctca gcttttttt 480
caacacttcc cggaatatca aactaatgac ttcatattg ccggtgaatc ttatgcagga 540
cattacattc cgggtgttgc agacgaaatt ttgagtcaaa agaacagaaa ttcaatctt 600
acttcagtct tgatcggaaa tggattaact gacccttga ctcaataccg atattacgag 660
ccaatggctt gtggtgaagg tggtgccccg tcagtactgc ctgccgatga gtgcgaaaat 720
atgctagtta cccaagataa atgtttgtct ttaattcaag catgctacga ctcacagtcg 780
gcattcatat gcgcaccggc tgccatttat tgtaataacg ctcatagggg accctatcag 840
agaactggga agaattgtga tgatattcgt aaggaatgtg atggtggtgc cttgtgctat 900
aaggacctg aattcatcga tacctactta aatcaaaagt ttgttcaaga tgctttgggc 960
gccgaggtcg atacctatga atcttgcaat ttgaaatca acagaaactt tttatttgc 1020
ggagattgga tgaaacctta tcatgaacat gtcagcagtc tcttgaacaa aggtttgccc 1080
gttttgattt acgcagggga caaagatttc atttgcaact ggttgggtaa tcgagcatgg 1140
actgatgtct tgccgtgggt tgatgctgat ggttttgaaa aagccgaagt ccaagattgg 1200
ttggttaatg gaaggaaggc tgggaattt aagaactata gcaactcac ctacctaagg 1260
gtttatgatg ctggcatat ggcccatat gatcagccag agaatttca tgaatggtc 1320
aatagatgga tatccggaga ctttagcttt cactag 1356

<210> 12
<211> 451
<212> PRT
<213> *Pichia* Pastoris

10

<400> 12

15

ES 2 539 378 T3

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

ES 2 539 378 T3

180 185 190
 Gln Lys Asn Arg Asn Phe Asn Leu Thr Ser Val Leu Ile Gly Asn Gly
 195 200 205
 Leu Thr Asp Pro Leu Thr Gln Tyr Arg Tyr Tyr Glu Pro Met Ala Cys
 210 215 220
 Gly Glu Gly Gly Ala Pro Ser Val Leu Pro Ala Asp Glu Cys Glu Asn
 225 230 235 240
 Met Leu Val Thr Gln Asp Lys Cys Leu Ser Leu Ile Gln Ala Cys Tyr
 245 250 255
 Asp Ser Gln Ser Ala Phe Thr Cys Ala Pro Ala Ala Ile Tyr Cys Asn
 260 265 270
 Asn Ala Gln Met Gly Pro Tyr Gln Arg Thr Gly Lys Asn Val Tyr Asp
 275 280 285
 Ile Arg Lys Glu Cys Asp Gly Gly Ser Leu Cys Tyr Lys Asp Leu Glu
 290 295 300
 Phe Ile Asp Thr Tyr Leu Asn Gln Lys Phe Val Gln Asp Ala Leu Gly
 305 310 315 320
 Ala Glu Val Asp Thr Tyr Glu Ser Cys Asn Phe Glu Ile Asn Arg Asn
 325 330 335
 Phe Leu Phe Ala Gly Asp Trp Met Lys Pro Tyr His Glu His Val Ser
 340 345 350
 Ser Leu Leu Asn Lys Gly Leu Pro Val Leu Ile Tyr Ala Gly Asp Lys
 355 360 365
 Asp Phe Ile Cys Asn Trp Leu Gly Asn Arg Ala Trp Thr Asp Val Leu
 370 375 380
 Pro Trp Val Asp Ala Asp Gly Phe Glu Lys Ala Glu Val Gln Asp Trp
 385 390 395 400
 Leu Val Asn Gly Arg Lys Ala Gly Glu Phe Lys Asn Tyr Ser Asn Phe
 405 410 415
 Thr Tyr Leu Arg Val Tyr Asp Ala Gly His Met Ala Pro Tyr Asp Gln
 420 425 430
 Pro Glu Asn Ser His Glu Met Val Asn Arg Trp Ile Ser Gly Asp Phe
 435 440 445

Ser Phe His

450

- <210> 13
- 5 <211> 1680
- <212> ADN
- <213> *Pichia Pastoris*

<400> 13

ES 2 539 378 T3

atgcaattgc gtcattccgt tggattggct atcttatctg ccatagcagt ccaaggattg 60
ctaattccta acattgagtc attaccacgc cagtttgggtg ctaatgggtga cagtgaacaa 120
gggttattag cccaccatgg taaacatcct aaagttgata tggctcacca tggaaagcat 180
cctaaaatcg ctaaggattc caagggacac cctaagcttt gccctgaagc tttgaagaag 240
atgaaagaag gccacccttc ggctccagtc attactaccc attccgcttc taaaaactta 300
atcccctact cttatattat agtcttcaag aaggggtgtca cttcagagga tatcgacttc 360
cacctgacc ttatctccac tcttcatgaa gagtctgtga gcaaattaag agagtcagat 420
ccaaatcact cattttctgt ttctaagag aatggcgaaa caggttacac cggtgacttc 480
tccgttgggtg acttgctcaa gggttacacc ggatacttca cggatgacac tttagagctt 540
atcagtaagc atccagcagt tgccttcatt gaaagggatt cgagagtatt tgccaccgat 600
tttgaactc aaaacgggtc tcctgggggt ttggccagag tctctcacag aaagcctctt 660
tcctagga gcttcaacaa gtacttatat gatggagctg gtggtgaagg tgttacttcc 720
tatgtatcg atacaggtat ccacgtcact cacaaagaat tccagggtag agcatcttgg 780
ggtaagacca ttccagctgg agacgttgat gacgatgaa acggtcacgg aactcactgt 840
gctggtagca ttgcttctga aagctacggg gttccaaga aggctaattgt tgttccatc 900
aaggtcttga gatctaattg ttctgggtcg atgtcagatg ttctgaaggg tgttgagtat 960
gccaccaat ccacttggga tgctgtaaa aagggcaaca agaaattaa gggctctacc 1020
gctaacatgt cactgggtgg tggtaaatct cctgcttgg accttgagc caatgctgct 1080
gftaagaatg gtattcactt tgccgttgcg gcaggtaacg aaaaccaaga tgcttgaac 1140
acctgccag cagctgctga gaatgccatc accgtcgggtg catcaacctt atcagacgct 1200
agagcttact ttctaacta cggtaaatgt gttgacattt tgcctccagg tttaacatt 1260
ctttctacct aactgggtc ggatgacgca actgctacct tgtctggtac ttcaatggc 1320
tctctcaca ttgctgggtc gttgacttac ttctatcat tgcagcctgc tgctggatct 1380
ctgtactcta acggaggatc tgagggtgac acacctgctc aattgaaaaa gaacctctc 1440
aagtatgcat ctgtcggagt attagaggat gttccagaag aactccaaa cctcttgggt 1500
tacaatgggtg gtggacaaaa cctttctct tctggggaa aggagacaga agacaatgtt 1560
gcttctccg acgatactgg tgagttcac tcttttga acaagcttga atcagctgtt 1620
gaaaacttgg cccaagagt tgcacattca gtgaaggagc tggcttctga acttattag 1680

5 <210> 14
<211> 559
<212> PRT
<213> *Pichia* Pastoris

10 <400> 14

ES 2 539 378 T3

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

ES 2 539 378 T3

Ser Val Gly Asp Leu Leu Lys Gly Tyr Thr Gly Tyr Phe Thr Asp Asp
 165 170 175

Thr Leu Glu Leu Ile Ser Lys His Pro Ala Val Ala Phe Ile Glu Arg
 180 185 190

Asp Ser Arg Val Phe Ala Thr Asp Phe Glu Thr Gln Asn Gly Ala Pro
 195 200 205

Trp Gly Leu Ala Arg Val Ser His Arg Lys Pro Leu Ser Leu Gly Ser
 210 215 220

Phe Asn Lys Tyr Leu Tyr Asp Gly Ala Gly Gly Glu Gly Val Thr Ser
 225 230 235 240

Tyr Val Ile Asp Thr Gly Ile His Val Thr His Lys Glu Phe Gln Gly
 245 250 255

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

Gly Asn Gly His Gly Thr His Cys Ala Gly Thr Ile Ala Ser Glu Ser
 275 280 285

Tyr Gly Val Ala Lys Lys Ala Asn Val Val Ala Ile Lys Val Leu Arg
 290 295 300

Ser Asn Gly Ser Gly Ser Met Ser Asp Val Leu Lys Gly Val Glu Tyr
 305 310 315 320

Ala Thr Gln Ser His Leu Asp Ala Val Lys Lys Gly Asn Lys Lys Phe
 325 330 335

Lys Gly Ser Thr Ala Asn Met Ser Leu Gly Gly Gly Lys Ser Pro Ala
 340 345 350

Leu Asp Leu Ala Val Asn Ala Ala Val Lys Asn Gly Ile His Phe Ala
 355 360 365

Val Ala Ala Gly Asn Glu Asn Gln Asp Ala Cys Asn Thr Ser Pro Ala
 370 375 380

Ala Ala Glu Asn Ala Ile Thr Val Gly Ala Ser Thr Leu Ser Asp Ala
 385 390 395 400

Arg Ala Tyr Phe Ser Asn Tyr Gly Lys Cys Val Asp Ile Phe Ala Pro
 405 410 415

Gly Leu Asn Ile Leu Ser Thr Tyr Thr Gly Ser Asp Asp Ala Thr Ala

ES 2 539 378 T3

420 425 430
 Thr Leu Ser Gly Thr Ser Met Ala Ser Pro His Ile Ala Gly Leu Leu
 435 440 445
 Thr Tyr Phe Leu Ser Leu Gln Pro Ala Ala Gly Ser Leu Tyr Ser Asn
 450 455 460
 Gly Gly Ser Glu Gly Val Thr Pro Ala Gln Leu Lys Lys Asn Leu Leu
 465 470 475 480
 Lys Tyr Ala Ser Val Gly Val Leu Glu Asp Val Pro Glu Asp Thr Pro
 485 490 495
 Asn Leu Leu Val Tyr Asn Gly Gly Gly Gln Asn Leu Ser Ser Phe Trp
 500 505 510
 Gly Lys Glu Thr Glu Asp Asn Val Ala Ser Ser Asp Asp Thr Gly Glu
 515 520 525
 Phe His Ser Phe Val Asn Lys Leu Glu Ser Ala Val Glu Asn Leu Ala
 530 535 540
 Gln Glu Phe Ala His Ser Val Lys Glu Leu Ala Ser Glu Leu Ile
 545 550 555

<210> 15
 <211> 865
 <212> PRT
 <213> *Pichia Pastoris*

5

<400> 15

Met Phe Leu Lys Ser Leu Leu Ser Phe Ala Ser Ile Leu Thr Leu Cys

1 5 10 15

Lys Ala Trp Asp Leu Glu Asp Val Gln Asp Ala Pro Lys Ile Lys Gly

20 25 30

Asn Glu Val Pro Gly Arg Tyr Ile Ile Glu Tyr Glu Glu Ala Ser Thr

35 40 45

Ser Ala Phe Ala Thr Gln Leu Arg Ala Gly Gly Tyr Asp Phe Asn Ile

10

50 55 60

ES 2 539 378 T3

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

ES 2 539 378 T3

325 330 335
 Ile Gly Phe Asp Ala Glu Leu Glu Val Asp Phe Thr Glu Glu Arg Gly
 340 345 350
 Cys Val Tyr Glu Pro Glu Ile Ser Ala Asp Asn Ala Asn Lys Ala Ile
 355 360 365
 Leu Leu Arg Arg Gly Val Gly Cys Val Glu Asn Leu Glu Phe Asn Leu
 370 375 380
 Leu Ser Val Ala Gly Tyr Lys Ala Tyr Phe Leu Tyr Asn Ser Phe Ser
 385 390 395 400
 Arg Pro Trp Ser Leu Leu Asn Ile Ser Pro Leu Ile Glu Leu Asp Asn
 405 410 415
 Ala Tyr Ser Leu Val Glu Glu Glu Val Gly Ile Trp Val Lys Thr Gln
 420 425 430
 Ile Asp Ala Gly Asn Thr Val Lys Leu Lys Val Ser Thr Ser Asp Gln
 435 440 445
 Met Leu Pro Ser Asp Lys Glu Tyr Leu Gly Val Gly Lys Met Asp Tyr
 450 455 460
 Tyr Ser Ser Gln Gly Pro Ala Tyr Glu Leu Glu Phe Phe Pro Thr Ile
 465 470 475 480
 Ser Ala Pro Gly Gly Asp Ser Trp Gly Ala Trp Pro Gly Gly Gln Tyr
 485 490 495
 Gly Val Ala Ser Gly Thr Ser Phe Ala Cys Pro Tyr Val Ala Gly Leu
 500 505 510
 Thr Ala Leu Tyr Glu Ser Gln Phe Gly Ile Gln Asp Pro Gln Asp Tyr
 515 520 525
 Val Arg Lys Leu Val Ser Thr Ala Thr Asp Leu Gln Leu Phe Asp Trp
 530 535 540
 Asn Ala Val Lys Leu Glu Thr Ser Met Asn Ala Pro Leu Ile Gln Gln
 545 550 555 560
 Gly Ala Gly Leu Val Asn Ala Leu Gly Leu Phe Glu Thr Lys Thr Val
 565 570 575
 Ile Val Ser Ala Pro Tyr Leu Glu Leu Asn Asp Thr Ile Asn Arg Ala
 580 585 590

ES 2 539 378 T3

Ser Glu Tyr Thr Ile Gln Ile Lys Asn Glu Asn Ser Glu Thr Ile Thr
 595 600 605

Tyr Gln Val Val His Val Pro Gly Thr Thr Val Tyr Ser Arg Ser Ala
 610 615 620

Ser Gly Asn Ile Pro Tyr Leu Val Asn Gln Asp Phe Ala Pro Tyr Gly
 625 630 635 640

Asp Ser Asp Ala Ala Thr Val Ala Leu Ser Thr Glu Glu Leu Val Leu
 645 650 655

Gly Pro Gly Glu Val Gly Glu Val Thr Val Ile Phe Ser Thr Glu Glu
 660 665 670

Ile Asp Gln Glu Thr Ala Pro Ile Ile Gln Gly Lys Ile Thr Phe Tyr
 675 680 685

Gly Asp Val Ile Pro Ile Ala Val Pro Tyr Met Gly Val Glu Val Asp
 690 695 700

Ile His Ser Trp Glu Pro Leu Ile Glu Arg Pro Leu Ser Val Arg Met
 705 710 715 720

Tyr Leu Asp Asp Gly Ser Leu Ala Tyr Val Asp Asp Asp Pro Asp Tyr
 725 730 735

Glu Phe Asn Val Tyr Asp Trp Asp Ser Pro Arg Phe Tyr Phe Asn Leu
 740 745 750

Arg Tyr Ala Thr Lys Glu Val Ser Ile Asp Leu Val His Pro Asp Tyr
 755 760 765

Ser Ile Glu Asn Asp Tyr Glu Trp Pro Leu Val Ser Gly His Asn Asn
 770 775 780

Tyr Tyr Gly Pro Val Gly Tyr Asp Tyr Asp Tyr Thr Ser Gly Gln Ala
 785 790 795 800

Phe Leu Pro Arg Tyr Phe Gln Gln Arg Ile Asn Glu Leu Gly Tyr Leu
 805 810 815

Ser Phe Ser Arg Phe Ala Asn Phe Ser Val Val Pro Ala Gly Glu Tyr
 820 825 830

Lys Ala Leu Phe Arg Val Leu Leu Pro Tyr Gly Asp Phe Trp Asn Lys
 835 840 845

Glu Asp Trp Gln Leu Phe Glu Ser Pro Val Phe Asn Val Leu Ala Phe
 850 855 860

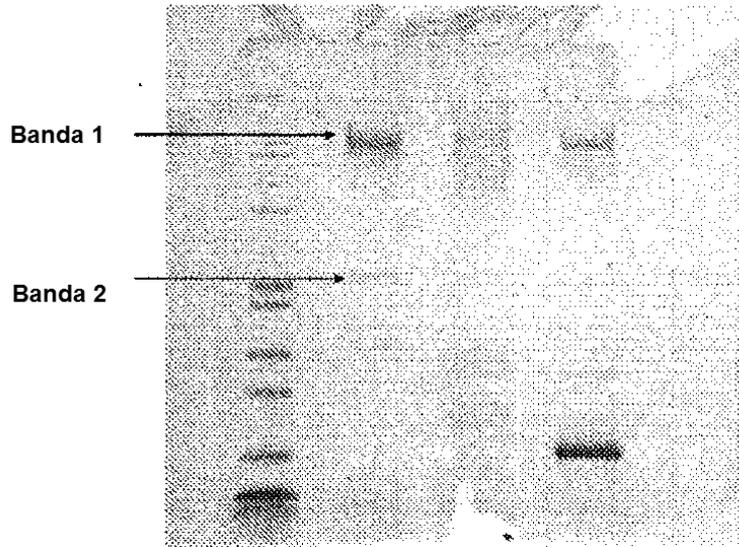
Phe

REIVINDICACIONES

1. Una cepa de *Pichia* modificada genéticamente, en la que la secuencia de ácido nucleico que codifica:
- (i) *PEP4*, *PRB1* y *YMP1*,
 - (ii) *PEP4*, *PRB1* y *YMP2*, o
 - 5 (iii) *PEP4*, *PRB1* y *YMP3*
- es modificada genéticamente para producir deficiencia de proteasa de al menos tres de las siguientes enzimas y/o tipo de enzima: aminopeptidasa de tipo Arg / Ala, aspartil proteasa, serín proteasa, serín proteasa secretada, serín proteasa Prb 1, aminopeptidasa de tipo Leu, en comparación con la cepa de tipo salvaje.
- 10 2. La cepa de *Pichia* modificada genéticamente de la reivindicación 1, en la que dicha cepa de *Pichia* modificada genéticamente es una forma modificada genéticamente de X-33 o SMD1163 de tipo salvaje.
3. La cepa de *Pichia* modificada genéticamente de la reivindicación 1 o 2, en la que dicha cepa de *Pichia* modificada genéticamente es una forma modificada genéticamente de *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minute*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia sp.* de tipo salvaje.
- 15 4. La cepa de *Pichia* modificada genéticamente de una cualquiera de las reivindicaciones 1 a 3, en la que el genoma de una cepa de tipo salvaje de dicha *Pichia* comprende un gen que tiene al menos un 85 % de identidad de secuencia con un gen seleccionado de las SEC ID N° 3, 5, 7 y en la que cuando dicho gen que tiene al menos un 85 % de identidad de secuencia con las SEC ID N° 3, 5 y 7 se deleciona, muta o rompe, dicha cepa de *Pichia* modificada genéticamente produce menos producto génico de dicho gen alterado que dicha cepa de tipo salvaje.
- 20 5. La cepa modificada genéticamente de una cualquiera de las reivindicaciones 1 a 3, en la que al menos una secuencia de ácidos nucleicos que codifican al menos uno del producto génico funcional siguiente o expresión de dicho producto génico está modificado genéticamente: SEC ID N° 4, SEC ID N° 6, y/o SEC ID N° 8.
- 25 6. La cepa de *Pichia* modificada genéticamente de una cualquiera de las reivindicaciones precedentes que comprende además al menos un polinucleótido capaz de expresar al menos un polipéptido heterólogo.
7. La cepa de *Pichia* modificada genéticamente de la reivindicación 6, en la que dicho polinucleótido capaz de expresar al menos un polipéptido heterólogo es un vector y/o expresión episomal de un plásmido.
8. La *Pichia* modificada genéticamente de la reivindicación 7, en la que dicho polinucleótido capaz de expresar al menos un polipéptido heterólogo es transformada en el genoma de la *Pichia*.
- 30 9. La *Pichia* modificada genéticamente de cualquiera de las reivindicaciones 6 a 8 en la que dicho al menos un polipéptido heterólogo comprende al menos un agonista de GLP-1; o uno o más de los siguientes: al menos una proteína de unión a antígeno, al menos un dominio variable sencillo, y/o al menos un anticuerpo de dominio.
10. La *Pichia* modificada genéticamente de cualquiera de las reivindicaciones 6 a 8, en la que el al menos un polipéptido heterólogo tiene al menos una actividad de GLP-1.
- 35 11. La *Pichia* modificada genéticamente de cualquiera de las reivindicaciones 6 a 8, en la que el al menos un polipéptido heterólogo comprende al menos un fragmento y/o variante de GLP-1 humano fusionado con seroalbúmina humana.
12. La cepa de *Pichia* modificada genéticamente de una cualquiera de las reivindicaciones 6 a 9, en la que dicha cepa muestra una proteólisis reducida de dicho al menos un polipéptido heterólogo en dicha cepa en comparación con *Pichia* de tipo salvaje.
- 40 13. La cepa de *Pichia* modificada genéticamente de una cualquiera de las reivindicaciones 6 a 10, en la que dicha cepa muestra una glicosilación reducida o ausencia de glicosilación de dicho al menos un polipéptido heterólogo en dicha cepa en comparación con *Pichia* de tipo salvaje.
- 45 14. Un procedimiento de producción de un polipéptido heterólogo que comprende expresar dicho polipéptido heterólogo en una *Pichia* modificada genéticamente como se describe en una cualquiera de las reivindicaciones precedentes.

FIG 1

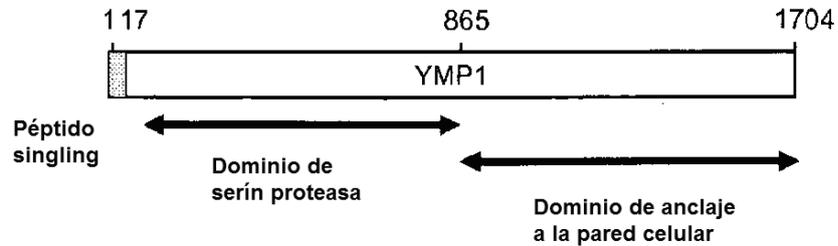
Nueva proteasa de *Pichia* (Ymp1) identificada usando purificación por afinida y análisis CL/EM



Las bandas 1 y 2 se escindieron del gel, se redujeron, alquilaron y digirieron con tripsina *in situ*. Los péptidos trípticos de cada banda se analizaron mediante espectrometría de masas en tándem - cromatografía líquida. Se realizó una búsqueda de los datos de la secuencia no interpretados frente a la base de datos de contig-péptido de *Pichia* (Integrated Genetics) utilizando software de identificación de proteínas Mascot.

FIG 2

Estructura prevista de YMP1 (SEC ID N° 4)



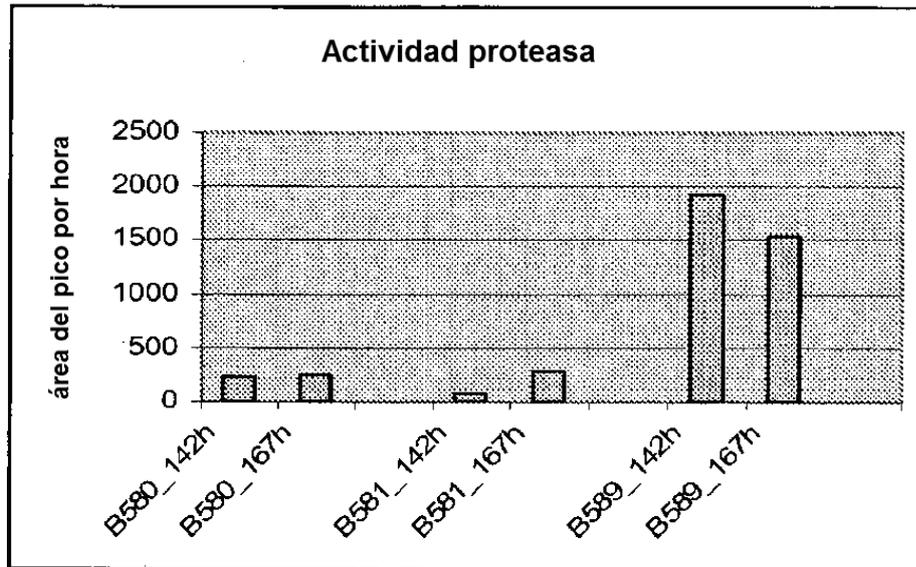
Dominio de proteasa homóloga { **MFLKSLLSFASILTLC**KAWDLEDVQDAPKIKGNEVPGRYIIIEEASTSAFATQLRAGGYDFNIQDYDSTGSLFNGASVQ
 ISNDNKTTFQDLQSLRAVKNVYPATLITLDETFELADTKPWNPHGITGVDSLHEQGYTGSGVVIIVIDTGVDTYHPALGG
 GIGDNFPIKAGYDLSGGDGVITNDPMDCDGHTFVSSIIIVANNKDMVGVA PDAQIVMYKV FPCSDSTSTD IVMACMQKAY
 DDGHKIIISLS LGSDSGFSS PASLMASRIA QDRVVLVAAG NSGELGPFYA SSPASGKQVI
 SVGSVQNEQW TTFPVIFTSS NGESRVFPYL AYNGAQIGFD AELEVDFTEE RGCVYEPEIS
 ADNANKAILL RRGVGCVENL EFNLLSVAGY KAYFLYNSFS RWSLLNISP LIELDNAYSL
 VEEVGIWVK TQIDAGNTVK LKVSTSDQML PSDKEYLGVG KMDYYSSQGP AYELEFFFTI
 SAPGGDSWGA WPGGQYGVAS GTSFACPYVA GLTALYESQF GIQDPQDYVR KLVSTATDLQ
 LFDWNAVKLE TSMNAPLIQQ GAGLVNALGL FETKTVIVSA PYLELNDTIN RASEYTIQIK
 NENSETITYQ VVHVPGTTVY SRSASGNIPY LVNQDFAPYG DSDAATVALS TEELVLPGE
 VGEVTIVFST EEIDQETAPI IQGKITFYGD VIPIAVPYMG VEVDIHSWEP LIERPLSVRM
 YLDDGSLAYV DDDPDYEFNV YDWSPRFYF NLRATKEVS IDLVHPDYSI ENDYEWPLVS
 GHNNYGPVG YDYDTSQQA FLPRYFQORI NELGYLSFSR FANFSVVPAG EYKALFRVLL
 PYGDFWNKED WQLFESPVFN VLAPP (SEC ID N° 15)

Dominio trans membrana {
 EAFGYTSOWLVSFTNTDRDITLPPYAVVQVTPHNLQFAHFEYAPYLEPSLQVFTKNERLIMTSQFNQYTRWICF
 KFDNRDQVITQVEGVVYITMKLEQDFISALAPQEVDFEPIHYSVDSYAGTFDFPLIRSEPIKLIAGAPDEVEWFDIPIA
 YSELAIDISSIDTNDNLOGYFYCCSKLKYTIKKEFDQWGNFTAGSDGNQYSNTTDDYVPIIDSTGSPVAEVOCLVESI
 SLSFTN ILARDEVLRVVLHSAFRRGSFTMANVYVNCIACGLAKRELFYILDENYASTGSEGLAFDVEVADQVEEP
 TEESTSEESTEQETSTEETEESTEPTTEESTGEPTEPTDEPTSEESTEEPSEPTSDLSIDPTAVPTDEPTTEPTTEPT
 SESTEPESEPTSDLSIDPTAVPTDEPTTEPTTEPTSEESTEEPSEPTSDLSIDPTAVPTDEPTTEPTTEPTTEPT
 PSIAPTAVPTSDTSSGQSIVTQNTTITVTITVSNVCAETPVITVYAPVVKPVSYTITVSVCHVCAETPITVTLPLP
 CETEDVTKTAGPKVTVYEVNCSADKPIYTYIAPEVYTGAEERTVTVSNVCAETPVITVYAPVVKPVSYTITVSVCHVCAETPITVTLPLP
 GELISSKGIPTVFAFPYSGYKQVTSQRTLATITKSSDESNTVTTIQTQVLSGESSGHOAASNSTSIAPVTVTAG
 NENSSRSFAGLFTVLLPLVLI

MFLKSLLSFA SILTLC KAWD LEDVQDAPKI KGNEVPGRYI IEYEEASTSA FATQLRAGGY
 DFNIQDYDST GSLFNGASVQ ISNDNKTTFQ DLQSLRAVKN VYPATLITLDETFELADTKP
 WNPHGITGVDSLHEQGYTGS GVVIAVIDTG VDYTHPALGG GIGDNFPIKA GYDLSGGDGV
 ITNDPMDCDG HGTFVSSIIIV ANNKDMVGVA PDAQIVMYKV FPCSDSTSTD IVMACMQKAY
 DDGHKIIISLS LGSDSGFSS PASLMASRIA QDRVVLVAAG NSGELGPFYA SSPASGKQVI
 SVGSVQNEQW TTFPVIFTSS NGESRVFPYL AYNGAQIGFD AELEVDFTEE RGCVYEPEIS
 ADNANKAILL RRGVGCVENL EFNLLSVAGY KAYFLYNSFS RWSLLNISP LIELDNAYSL
 VEEVGIWVK TQIDAGNTVK LKVSTSDQML PSDKEYLGVG KMDYYSSQGP AYELEFFFTI
 SAPGGDSWGA WPGGQYGVAS GTSFACPYVA GLTALYESQF GIQDPQDYVR KLVSTATDLQ
 LFDWNAVKLE TSMNAPLIQQ GAGLVNALGL FETKTVIVSA PYLELNDTIN RASEYTIQIK
 NENSETITYQ VVHVPGTTVY SRSASGNIPY LVNQDFAPYG DSDAATVALS TEELVLPGE
 VGEVTIVFST EEIDQETAPI IQGKITFYGD VIPIAVPYMG VEVDIHSWEP LIERPLSVRM
 YLDDGSLAYV DDDPDYEFNV YDWSPRFYF NLRATKEVS IDLVHPDYSI ENDYEWPLVS
 GHNNYGPVG YDYDTSQQA FLPRYFQORI NELGYLSFSR FANFSVVPAG EYKALFRVLL
 PYGDFWNKED WQLFESPVFN VLAPP (SEC ID N° 15)

FIG 3

Ensayo de actividad proteasa de la cepa mutante *ymp1*

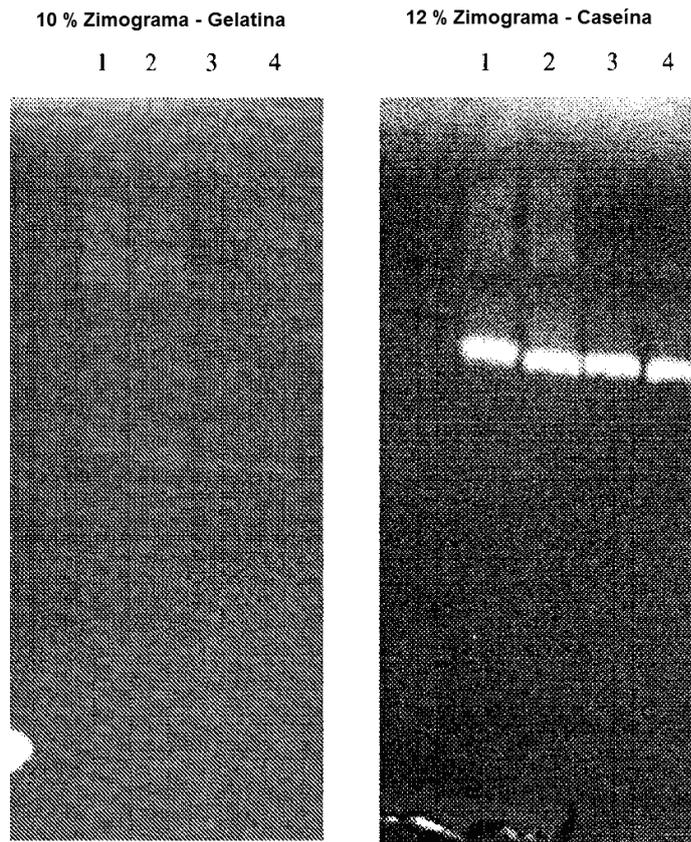


B580 y B581 fueron lotes de fermentación para SMD1163_ *ymp1*

B589 fue lote de fermentación para la cepa control SMD1163

FIG 4

Análisis zimográfico de la cepa mutante *ymp1*



Muestra 1 - SMD1163
Muestra 2 - X33_ *pep4*
Muestra 3 - SMD1163_ *ymp1*_clon 1
Muestra 4 - SMD 1163_ *ymp1*_clon 2

FIG 5

Efecto de inactivación de *ymp1* sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

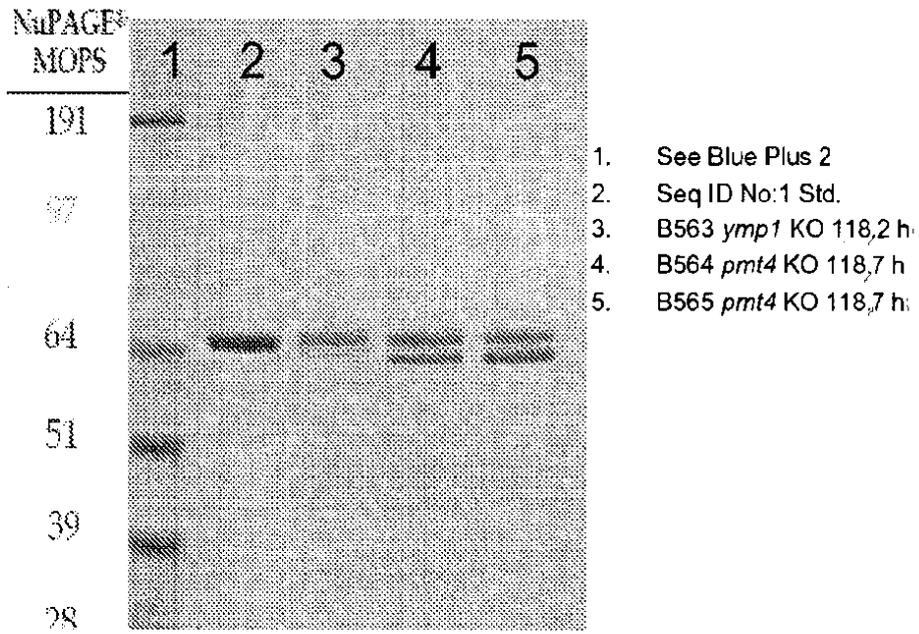
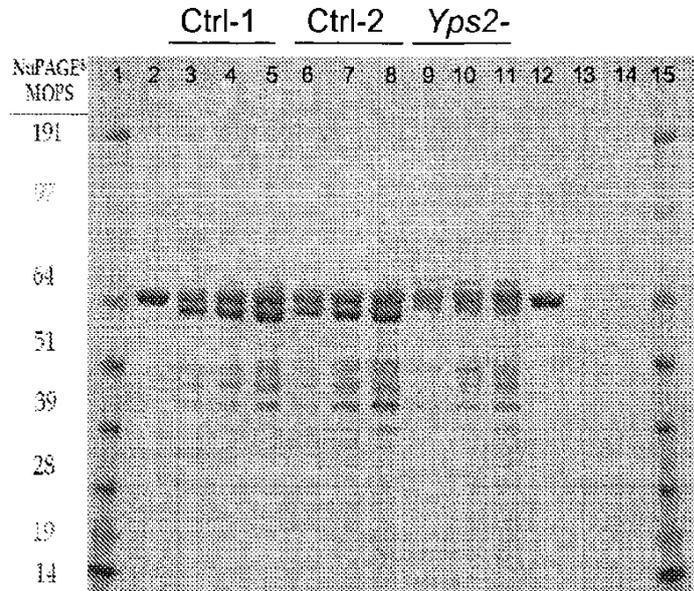


FIG 6

Efecto de la inactivación de ysp2 sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)



1-See Blue Plus 2

2- SEC ID N° 1 Ptr.

3-B542 125 h

4-B542 148 h

5-B542 169 h

6-B544 125 h

7-B544 148 h

8-B544 169 h

9-B545 125 h

10-B545 148 h

11-B545 168 h

12- SEC ID N° 1 Ptr.

13- Blanco

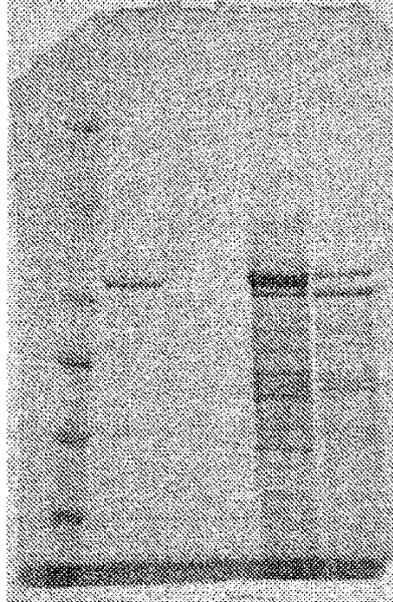
14- Blanco

15-See Blue Plus 2

FIG 7

Efecto de la inactivación de *usp1* sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

1 2 3 4 5



- 1 -- SeeBlue Plus Marcador
- 2 -- SEC ID N° 1 ptr.
- 3 -- X33pep4 yps1
- 4 -- X33pep4 yps1 / SEC ID N° 1
- 5 -- X33pep4 / SEC ID N° 1

FIG 8

Efecto de la inactivación de pep4 sobre la proteólisis de la proteína heteróloga (SEC ID N° 1)

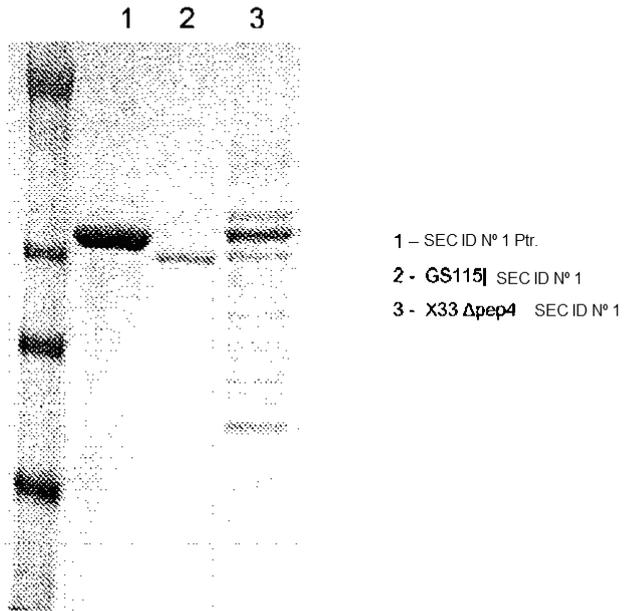


FIG 9

Estudio de estabilidad en tampón de las cepas mutantes de *ymp2* e *ymp3*

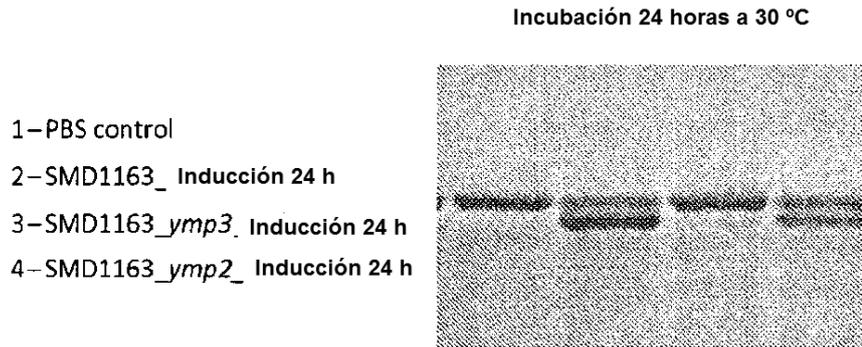


FIG 10

Resultados del ensayo con zimoliasa para *Pichia* de tipo salvaje (WT), y mutantes *pmt1* y *pmt4*

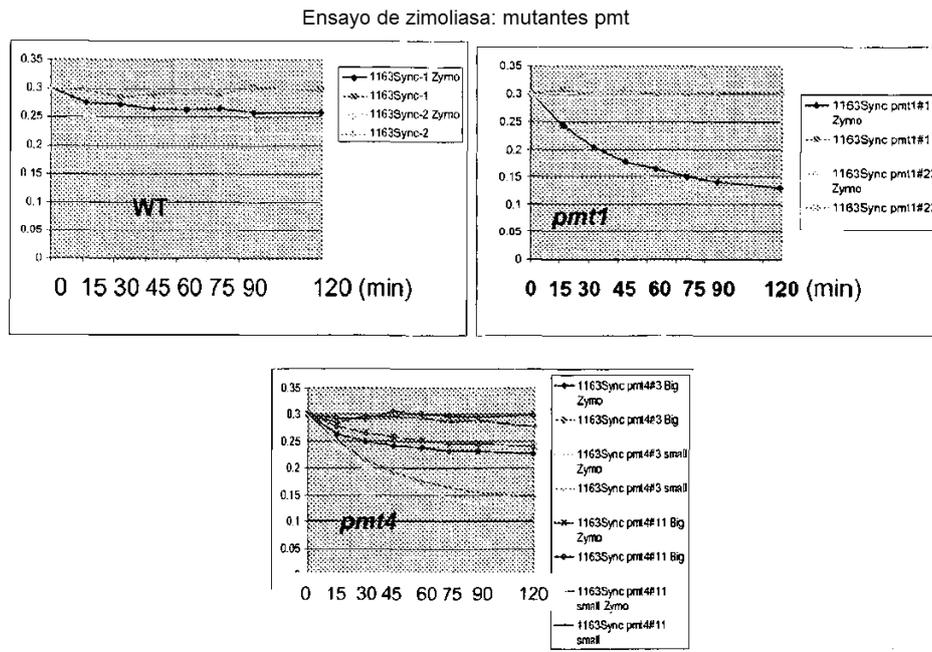


FIG 11

Análisis de glicosilación de la cepa mutante *pmt4* mediante CL/EM

Comparación de los espectros de desconvolución del pico principal del patrón de referencia y las muestras de Pichia

