

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 549 328**

51 Int. Cl.:

C12N 9/42 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **11.02.2010 E 10153355 (2)**

97 Fecha y número de publicación de la concesión europea: **15.07.2015 EP 2357227**

54 Título: **Enzimas celulasas optimizadas**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.10.2015

73 Titular/es:

**Süd-Chemie IP GmbH & Co. KG (100.0%)
Lenbachplatz 6
80333 München, DE**

72 Inventor/es:

**REISINGER, CHRISTOPH, DR.;
BRÜCK, THOMAS, DR.;
KOLTERMANN, ANDRÉ, DR.;
GERLACH, JOCHEN, DR.;
UNTERSTRASSER, ISABEL;
RÖCHER, LUTZ;
RARBACH, MARKUS, DR.;
CLAREN, JÖRG, DR.;
KOHL, ANDREAS, DR.;
PIECK, JAN, CARSTEN, DR.;
SCHLOSSER, DOMINIK**

74 Agente/Representante:

LEHMANN NOVO, María Isabel

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

ES 2 549 328 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

Enzimas celulasas optimizadas

5 Campo del invento

El invento divulga unas enzimas celulasas con unas propiedades optimizadas para el tratamiento de unos substratos, que contienen celulosas y lignocelulosas. En particular, se divulgan unas enzimas celobiohidrolasas con unas características preferidas. La presente divulgación proporciona unas variantes de fusión, inserción, supresión (deleción) y/o sustitución de tales enzimas. Las variantes de enzimas tienen unas elevadas propiedades de estabilidad térmica, estabilidad proteolítica, actividad específica y/o una estabilidad frente a un pH extremo. Se divulgan unas moléculas de ácidos nucleicos que codifican tales enzimas, una composición que comprende tales enzimas, un método para su producción, y el uso para el tratamiento de una celulosa y/o para la producción de combustibles biológicos.

15

Antecedentes del invento

El desarrollo de unos procedimientos de producción basados en materias primas renovables es muy deseado, por ejemplo, para la producción de etanol a partir de materiales celulósicos y lignocelulósicos.

20

Un material celulósico en una forma pura o en combinación con una hemicelulosa y/o una lignina es una materia prima valiosa y que está fácilmente disponible para la producción de productos químicos y combustibles. Una etapa clave en el tratamiento de una celulosa y una lignocelulosa es la hidrólisis de la celulosa, que es un polímero de glucosa con enlaces beta-1,4, y la subsiguiente liberación de los monómeros de glucosa y de unos cortos oligómeros de glucosa tales como una celobiosa, una celotriosa, etc. Unas enzimas que catalizan esta reacción son encontradas en diversos organismos, especialmente en ciertos hongos filamentosos y en ciertas bacterias, que son capaces de degradar e hidrolizar a una celulosa.

25

Son conocidos unos procedimientos continuos para convertir a una biomasa lignocelulósica sólida en unos productos carburantes combustibles. Un tratamiento para hacer a los substratos celulósicos más susceptibles frente a una degradación enzimática comprende una molienda, un tratamiento químico y/o un tratamiento hidrotérmico. Unos ejemplos de éstos son una oxidación en húmedo y/o un reventamiento con vapor. Tales tratamientos aumentan la accesibilidad de las fibras de celulosa y las separan con respecto de una hemicelulosa y una lignina, lo que es requerido para la degradación de los polímeros de celulosa. Entre éstas, unas enzimas celobiohidrolasas (CBH), y más específicamente una celobiohidrolasa I (CBHI) desempeñan un cometido especial en la etapa de hidrólisis y proporcionan la actividad enzimática más favorable para el procedimiento. Las enzimas CBHI catalizan la liberación hidrolítica progresiva de celobiosa desde el extremo reductor de los polímeros de celulosa (Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology [Utilización de la celulosa por microbios: fundamentos y biotecnología]. Microbiol Mol Biol Rev. 2002 Sep;66(3):506-77).

30

Unos materiales celulósicos hidrolizados contienen diversas y valiosas moléculas de hidratos de carbono, que pueden ser aisladas a partir de las mezclas. Unos materiales hidrolizados de materiales celulósicos, que contienen azúcares, pueden ser utilizados para una producción microbiana de una variedad de productos químicos finos o biopolímeros, tales como unos ácidos orgánicos, etanol o unos alcoholes superiores (es decir unos dioles o polioles) o unos poli(hidroxialcanoatos) (PHAs). Uno de los usos principales de los materiales hidrolizados de azúcares se encuentra en la producción de combustibles biológicos.

45

Kurabi y colaboradores (2005) describen unos preparados de celulasas procedentes de *Trichoderma reesei* y de otros hongos, tales como *Penicillium* sp. El rendimiento se ha analizado en abetos Douglas, que habían sido reventados con vapor y tratados previamente con un disolvente orgánico de etanol. Un mejor rendimiento de unas mezclas de enzimas parece ser el resultado de las propiedades mejoradas de unos componentes individuales de enzimas así como del efecto de cada compuesto en la mezcla, especialmente la presencia de una beta-glucosidasa (Kurabi A, Berlin A, Gilkes N, Kilburn D, Bura R, Robinson J, Markov A, Skomarovsky A, Gusakov A, Okunev O, Sinitsyn A, Gregg D, Xie D, Saddler J.(2005) Enzymatic hydrolysis of steam-exploded and ethanol organosolv-pretreated Douglas-Fir by novel and commercial fungal cellulases [Hidrólisis enzimática de abetos de Douglas reventados con vapor y tratados previamente con un disolvente orgánico de etanol mediante unas celulasas fúngicas y adquiribles comercialmente]. Appl Biochem Biotechnol. 121-124: 219-30).

55

Unas secuencias de celobiohidrolasas de la clase de una glucohidrolasa 7 (cel7) son conocidas en la especialidad a partir de varias fuentes fúngicas. Una celobiohidrolasa Cel7 de *Talaromayces emersonii* es conocida y ya se ha informado acerca de su expresión en *Escherichia coli*. Grassick y colaboradores han presentado un informe sobre la purificación y la determinación de la estructura en 3D de una proteína CBH de núcleo natural, y de la clonación y sobreexpresión de los correspondientes genes, procedentes de una fuente fúngica termófila. Se encontró que una CBH 16 es extremadamente estable térmicamente con un valor óptimo de la temperatura de 68 °C a pH 5,0 y un

60

período de tiempo de vida media ($t_{1/2}$) de 68,0 min a 80 °C y un pH de 5,0. (Grassick A, Murray PG, Thompson R, Collins CM, Byrnes L, Birrane G, Higgins TM, Tuohy MG. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus, *Talaromyces emersonii* [Estructura tridimensional de una celobiohidrolasa natural termoestable, CBH IB, y caracterización molecular del gen cel7 del hongo filamentoso, *Talaromyces emersonii*]. Eur J Biochem. 2004 Nov;271 (22):4495-4506) y en *Saccharomyces cerevisiae* (Voutilainen SP, Murray PG, Tuohy MG, Koivula A. Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity (Expresión de una celobiohidrolasa de *Talaromyces emersonii* Cel7A en *Saccharomyces cerevisiae* y mutagénesis racional para mejorar su estabilidad térmica y su actividad]. Protein Eng Des Sel. 2010 Feb;23(2):69-79), sin embargo, la proteína o bien era producida en una forma inactiva o en unos rendimientos bastante bajos (de menos que o iguales a 5 mg/l). Una celobiohidrolasa I de *Hypocrea jecorina* puede ser producida a partir de unas cepas de tipo silvestre (wt) o modificadas genéticamente del género *Hypocrea* o *Trichoderma* en unos altos rendimientos. Unas secuencias mejoradas de una Cel7A de *Hypocrea jecorina* han sido divulgadas en los documentos de patentes de los EE.UU. US7459299B2, US7452707B2, en los documentos de solicitudes de patentes internacionales WO2005/030926, WO01/04284A1 o en el documento de solicitud de patente de los EE.UU.: US2009/0162916 A1.

Unas posiciones que conducen a mejoramientos fueron deducidas a partir de unas alineaciones con unas secuencias procedentes de unas enzimas termoestables reseñadas, y que se han propuesto a partir de una información estructural y una barajadura de unas posiciones identificadas, seguida por unos escrutinios limitados. Se ha informado sobre un escrutinio de unas grandes bibliotecas en ciertos organismos transformables tales como *Saccharomyces cerevisiae* mediante el uso de unos substratos fluorescentes muy sensibles, que se asemejan a unos substratos naturales de un modo muy restrictivo. (Percival Zhang YH, Himmel ME, Mielenz JR. Outlook for cellulase improvement: screening and selection strategies [Perspectivas del mejoramiento de celulastas: escrutinio y estrategias de selección]. Biotechnol Adv. 2006 Sep-Oct;24(5):452-81).

Se ha informado sobre la producción de unas celobiohidrolasas procedentes de otros sistemas fúngicos tales como los de *Thermoascus aurantiacus*, *Chrysosporium lucknowense* o *Phanerochaete chrysosporium*. Se ha informado sobre la expresión de una celobiohidrolasa Cel7 a partir de levaduras, pero los rendimientos enzimáticos o las propiedades enzimáticas siguen siendo insatisfactorios/as. (Penttilä ME, André L, Lehtovaara P, Bailey M, Teeri TT, Knowles JK. Efficient secretion of two fungal cellobiohydrolases by *Saccharomyces cerevisiae* [Secreción eficiente de dos celobiohidrolasas fúngicas por *Saccharomyces cerevisiae*]. Gene. 1988;63(1):103-12).

El documento WO03/000941 divulga un cierto número de CBHs y sus correspondientes secuencias génicas. Sin embargo, no se divulgaron propiedades fisiológicas ni aplicaciones. Se ha informado de que la fusión de unos dominios de fijación de celulosa con unas subunidades catalíticas de celobiohidrolasas mejora las propiedades hidrolíticas de unas proteínas sin ningún dominio natural.

El documento US 2009042266 (A1) divulga unas fusiones de Cel7A de *Thermoascus aurantiacus* con unos dominios de fijación de celulosa procedentes de una celobiohidrolasa I de *Chaetomium thermophilum* y *Hypocrea jecorina*.

El documento US5686593 informa acerca de la fusión de unas regiones engarzadoras y de unos dominios de fijación que se han diseñado especialmente, con ciertas celobiohidrolasas.

Hong y colaboradores (2003) describen la producción de una CBHI de *Thermoascus aurantiacus* en levaduras y su caracterización (Hong J, Tamaki H, Yamamoto K, Kumagai H Cloning of a gene encoding thermostable cellobiohydrolase from *Thermoascus aurantiacus* and its expression in yeast [Clonación de un gen que codifica una celobiohidrolasa termoestable procedente de *Thermoascus aurantiacus* y su expresión en una levadura]. Appl Microbiol Biotechnol. 2003 Nov;63(1):42-50).

Tuohy y colaboradores (2002) informan acerca de la expresión y la caracterización de una CBH de *Talaromyces emersonii*. (Tuohy MG, Walsh DJ, Murray PG, Claeyssens M, Cuffe MM, Savage AV, Coughlan MP.: Kinetic parameters and mode of action of the cellobiohydrolases produced by *Talaromyces emersonii* [Parámetros cinéticos y modo de acción de las celobiohidrolasas producidas por *Talaromyces emersonii*]. Biochim Biophys Acta. 2002 Apr 29;1596(2):366-80).

Nevoigt y colaboradores (2008) informan acerca de la expresión de ciertas enzimas celulolíticas en levaduras. (Nevoigt E. Progress in metabolic engineering of *Saccharomyces cerevisiae* [Progresos en la modificación metabólica de *Saccharomyces cerevisiae*]. Microbiol Mol Biol Rev. 2008 Sep;72(3):379-412).

Fujita y colaboradores (2004) informan acerca de una cepa de *Saccharomyces cerevisiae* que expresa una combinación de una endoglucanasa, una beta glucosidasa y una CBHIII presentada gráficamente sobre la superficie celular. No se usó una celobiohidrolasa I (Cel7) en este planteamiento. (Fujita Y, Ito J, Ueda M, Fukuda H, Kondo A. Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered

yeast strain codisplaying three types of cellulolytic enzyme [Sacarificación sinérgica, y fermentación directa para dar etanol de una celulosa amorfa mediando utilización de una cepa de una levadura modificada que presenta concomitantemente tres tipos de una enzima celulolítica]. Appl Environ Microbiol. 2004 Feb;70(2):1207-12).

5 **Boer H y colaboradores (2000)** describen la expresión de unas enzimas clasificadas como GH7 en diferentes anfitriones de levaduras, pero los niveles de las proteínas expresadas fueron bajos. (Boer H, Teeri TT, Koivula A. Characterization of *Trichoderma reesei* cellobiohydrolase Cel7A secreted from *Pichia pastoris* using two different promoters [Caracterización de una celobiohidrolasa Cel7A de *Trichoderma reesei* que ha sido segregada por *Pichia pastoris* usando dos promotores diferentes]. Biotechnol Bioeng. 2000 Sep 5;69(5):486-94).

10 **Godbole y colaboradores (1999)** y **Hong y colaboradores (2003)** encontraron que unas proteínas de esta clase de enzimas expresadas a partir de levaduras frecuentemente estaban mal plegadas e hiperglicosiladas y tenían unas capacidades hidrolíticas disminuidas en comparación con la proteína expresada por el anfitrión homólogo. (Godbole S, Decker SR, Nieves RA, Adney WS, Vinzant TB, Baker JO, Thomas SR, Himmel ME. Cloning and expression of cellobiohydrolase I in *Pichia pastoris* [Clonación y expresión de una celobiohidrolasa I de *Trichoderma reesei* en *Pichia pastoris*]. Biotechnol Prog. 1999 Sep-Oct;15(5):828-33).

15 **Kanokratana y colaboradores (2008)**, **Li y colaboradores (2009)** así como el documento de patente china CN01757710 describen la expresión eficiente de unas enzimas de CBH I Cel7, sin embargo, estas proteínas carecen de los dominios de fijación de celulosa, que se requieren para un procesamiento eficiente del sustrato. (Kanokratana P, Chantasingh D, Champreda V, Tanapongpipat S, Pootanakit K, Eurwilaichitr L. Identification and expression of cellobiohydrolase (CBHI) gene from an endophytic fungus, *Fusicoccum* sp. (BCC4124) in *Pichia pastoris* [Identificación y expresión del gen de una celobiohidrolasa (CBHI) a partir de un hongo endofítico, *Fusicoccum* sp. (BCC4124), en *Pichia pastoris*]. LProtein Expr Purif. 2008 Mar;58(1):148-53. Epub 2007 Sep 19; Li YL, Li H, Li AN, Li DC. Cloning of a gene encoding thermostable cellobiohydrolase from the thermophilic fungus *Chaetomium thermophilum* and its expression in *Pichia pastoris* [Clonación de un gen que codifica una celobiohidrolasa termoestable a partir del hongo termofílico *Chaetomium thermophilum* y su expresión en *Pichia pastoris*]. J Appl Microbiol. 2009 Jun;106(6):1867-75).

20 **Voutilainen (2008)** y **Viikari (2007)** divulgan unas enzimas Cel7, que comprenden unas celobiohidrolasas termoestables, sin embargo con solamente unos niveles de expresión desde bajos hasta moderados, procedentes de *Trichoderma reesei*, (Voutilainen SP, Puranen T, Siika-Aho M, Lappalainen A, Alapuranen M, Kallio J, Hooman S, Viikari L, Vehmaanperä J, Koivula A. Cloning, expression, and characterization of novel thermostable family 7 cellobiohydrolases [Clonación, expresión y caracterización de una nueva familia termoestable de celobiohidrolasas 7]. Biotechnol Bioeng. 2008 Oct 15;101(3):515-28. PubMed PMID: 18512263; Viikari L, Alapuranen M, Puranen T, Vehmaanperä J, Siika-Aho M. Thermostable enzymes in lignocellulose hydrolysis. [Enzimas termoestables en la hidrólisis de lignocelulosas]. Adv Biochem Eng Biotechnol. 2007;108:121-45).

25 **Grassick y colaboradores (2004)** divulgan una expresión sin plegamiento de una celobiohidrolasa I de *Talaromyces emersonii* en *Escherichia coli* pero no en una levadura. (Grassick A, Murray PG, Thompson R, Collins CM, Byrnes L, Birrane G, Higgins TM, Tuohy MG. Three-dimensional structure of a thermostable native cellobiohydrolase, CBH IB, and molecular characterization of the cel7 gene from the filamentous fungus [Estructura tridimensional de una celobiohidrolasa natural termoestable, CBH IB, y caracterización molecular del gen cel7 procedente del hongo filamentoso, *Talaromyces emersonii*]. Eur J Biochem. 2004 Nov;271(22):4495-506).

30 El documento WO 2009/138877 describe un método para una expresión heteróloga de unos polipéptidos codificados por unas variantes de tipo silvestre y optimizadas con codones de CBHI y/o cbh2 a partir de los organismos fúngicos *Talaromyces emersonii* (*T. emersonii*), *Humicola grisea* (*H. grisea*), *Thermoascus aurantiacus* (*T. aurantiacus*), y *Trichoderma reesei* (*T. reesei*). Se encontró que la expresión en tales células anfitrionas de los correspondientes genes, y de las variantes y combinaciones de éstos, da como resultado una actividad específica mejorada de las celobiohidrolasas expresadas.

35 El documento WO 2009/139839 describe unos métodos y una composición para un vector y unas partículas de alfavirus que tienen una gran capacidad. En algunos aspectos se describen unos métodos para proporcionar unas partículas de alfavirus, que comprenden una proteína cápsida modificada.

40 Por lo tanto, existe una necesidad de encontrar unas enzimas celulasas con unas características mejoradas para el uso en unos procesos técnicos para la hidrólisis de celulosas. En particular, existe una necesidad de unas enzimas CBH con una actividad catalítica más alta y/o con una estabilidad más alta en las condiciones de tratamiento. Además de esto, existe una necesidad de unas enzimas CBH con una productividad más alta en la expresión en hongos y/o levaduras y unos sistemas de segregación.

Resumen del invento

El presente invento proporciona un polipéptido, que tiene una actividad de celobiohidrolasa. Este polipéptido comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 85 % con la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 ha sido modificado mediante sustitución o supresión.

Además de esto, el presente invento divulga un ácido nucleico que codifica el polipéptido del presente invento, que tiene de manera preferida una identidad de por lo menos 95 % con la SEQ ID NO: 1, un vector que comprende este ácido nucleico y un anfitrión transformado con tal vector.

La presente solicitud de patente divulga además un método para la producción de una proteína de celobiohidrolasa, que es codificada por un vector del presente invento, un método para la identificación de unos polipéptidos que tienen una actividad de celobiohidrolasa, y un método para la producción de tales polipéptidos, que tienen una actividad de celobiohidrolasa.

El presente invento proporciona también un polipéptido que tiene una actividad de celobiohidrolasa, en donde el polipéptido comprende una secuencia de aminoácidos, que tiene una identidad entre secuencias de por lo menos 85 % con la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 ha sido modificado mediante sustitución o supresión, en donde uno o más de los siguientes residuos de aminoácidos de la secuencia definida por la SEQ ID NO: 2 han sido modificados mediante sustitución o supresión: G4, A6, T15, Q28, W40, D64, E65, A72, S86, K92, V130, V152, Y155, K159, D181, E183, N194, D202, P224, T243, Y244, I277, K304, N310, S311, N318, D320, T335, T344, D346, Q349, A358, Y374, A375, T392, T393, D410, Y422, P442, N445, R446, T456, S460, P462, G463, H468 y/o V482 de los aminoácidos 1 hasta 500 de la SEQ ID NO: 2.

Más aún, la presente solicitud de patente divulga un polipéptido, que tiene una actividad de celobiohidrolasa, que es obtenible por el método de preparar un polipéptido que tiene una actividad de celobiohidrolasa de acuerdo con la presente solicitud de patente, y un polipéptido, que tiene una actividad de celobiohidrolasa, en donde el polipéptido comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 80 % con la SEQ ID NO: 5, en donde uno o más de los siguientes residuos de aminoácidos de la secuencia definida por la SEQ ID NO: 5 han sido modificados mediante sustitución o supresión: Q1, G4, A6, T15, Q28, W40, D64, E65, A72, S86, K92, V130, V152, Y155, K159, D181, E183, N194, D202, P224, T243, Y244, I277, K304, N310, S311, N318, D320, T335, T344, D346, Q349, A358, Y374, A375, T392, T393, D410 y/o Y422 de los aminoácidos 1 hasta 440 de la SEQ ID NO: 5.

La presente solicitud de patente divulga un polipéptido, que tiene una actividad de celobiohidrolasa, que comprende una secuencia de aminoácidos, que tiene una identidad entre secuencias de por lo menos 85 % con la SEQ ID NO: 12, en donde uno o más de los siguientes residuos de aminoácidos de la secuencia definida por la SEQ ID NO: 12 han sido modificados mediante sustitución o supresión: Q1, T15, Q28, W40, C72, V133, V155, Y158, T162, Y247, N307, G308, E317, S341, D345, Y370, T389, Q406, N441, R442, T452, S456, P458, G459, H464 y/o V478.

El presente invento divulga además una composición, que comprende un polipéptido del presente invento y una o más endoglucanasas y/o una o más beta-glucosidasas y/o una o más celobiohidrolasas adicionales y/o una o más xilanasas.

El presente invento proporciona además el uso de un polipéptido o de la composición del presente invento para la degradación enzimática de una biomasa lignocelulósica, y/o para el tratamiento de materiales textiles y/o como un ingrediente en detergentes y/o como un ingrediente en alimentos o composiciones alimentarias.

Breve descripción de las Figuras

Figura 1: Mapas de restricción de pV1 para la expresión constitutiva de proteínas en *Pichia pastoris*: pUC19 - ori: origen de la replicación en *E. coli*; KanR: resistencia frente a la kanamicina/G418 con las secuencias promotoras TEF1 y EMZ para la selección en *Pichia pastoris* y *E. coli*, respectivamente; 5'GAP: región promotora de la gliceraldehído-3-fosfato deshidrogenasa; 3'-GAP: región terminadora; SP MAlpha: secuencia de señal del factor de apareamiento alfa de *Saccharomyces cerevisiae*; MCS: sitio de clonación múltiple.

Figura 2: SDS-PAGE teñida con Commassie de unos materiales sobrenadantes concentrados 10 veces de unos cultivos en matraces de sacudimiento de *Pichia pastoris* CBS 7435, que contienen unos plásmidos de expresión con las secuencias codificadoras de las proteínas CBHI maduras de *Trichoderma viride* (CBH-f; pista 2), de *Humicola grisea* (CBH-d; pista 3), de *Talaromyces emersonii* (CBH-b; pista 4), de *Thermoascus aurantiacus* (CBH-e; pista 5), así como la fusión de CBHI-CBD de *Talaromyces emersonii* (CBH-a; pista 6) y la fusión de CBD de *Humicola grisea* (CBH-g; pista 7) en una fusión por el terminal de N con el péptido de señal del factor de apareamiento alfa de *Saccharomyces cerevisiae* bajo el control del promotor de la gliceraldehído-3 fosfato deshidrogenasa de *Pichia*

pastoris.

Figura 3: Mapa del plásmido de expresión pV3 para la expresión de proteínas en *Pichia pastoris*. Replicones: pUC19 - ori: origen de la replicación en *E. coli*, ZeoR: gen de resistencia frente a la zeocina con las secuencias promotoras TEF1 y EM7 para la expresión en *Pichia pastoris* y *E. coli*, respectivamente; promotor AOX I: región promotora del gen de la alcohol oxidasa I de *Pichia pastoris*; AOX 1 terminador de la transcripción: región terminadora: SP MFalpha: secuencia de señal del factor de apareamiento alfa de *Saccharomyces cerevisiae*; MCS: sitio de clonación múltiple.

Figura 4: Análisis por SDS-PAGE de unas muestras de materiales sobrenadantes del cultivo tomadas a partir de la fermentación de una cepa de *Pichia pastoris* con una integración genómica de un casete de expresión AOXI, que expresa el péptido de fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (CBH-a) en un biorreactor con una capacidad de 7 l mediando una inducción con metanol. Las muestras P1 - P7 se toman al principio de la inducción con metanol y después de 20, 45, 119,5, 142,5, 145,5 y 167 horas, respectivamente.

Figura 5: Mapa del plásmido de expresión pV4 para la expresión en *Trichoderma reesei* del péptido de fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (CBH-ah) Replicón: pUC19 para la replicación en *E. coli*. cbh1 5': región promotora 5' del gen de CBHI de *Trichoderma*; péptido de señal cbh1: secuencia codificadora para el péptido director de CBHI de *Trichoderma reesei*; CBH-a: péptido de fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei*; región codificadora para la SEQ ID NO: 18; terminador de cbh1: región terminadora en el extremo 3' del sitio de CBHI de *Trichoderma reesei*; resistencia a higromicina: región codificadora de la higromicina fosfotransferasa bajo el control de un promotor de la fosfoglicerato cinasa de *Trichoderma reesei*; cbh1 3': secuencia de homología con la región terminadora del sitio de CBHI de *Trichoderma reesei* para los sucesos de cruzamiento doble.

Figura 6: SDS-PAGE de unos materiales sobrenadantes de cultivos de *Trichoderma reesei*. La pista 1 muestra el patrón de expresión de una cepa de reemplazo que lleva una fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (CBH-ah) en lugar del gen de CBHI natural. Como comparación, la pista 2 muestra el patrón para la cepa no modificada en las mismas condiciones. M: marcador del tamaño molecular.

Figura 7: Determinación de los valores de la IT50 a partir de los diagramas de la capacidad de conversión del sustrato frente a la temperatura después de una normalización. Para la etapa de normalización, los valores de las fluorescencias máxima y mínima para la temperatura seleccionada son correlacionados con 1 o 0, respectivamente. La interpolación lineal a $F'(T) = 0,5$ entre los dos puntos de temperatura más próximos con unos valores normalizados próximos a 0,5 proporciona la temperatura IT50 definida.

Figura 8: Diagramas de la capacidad de conversión normalizada en función de la temperatura de unas fusiones de CBHI de *Talaromyces emersonii* "wt" / -CBD de *Trichoderma reesei* (CBH-ah: SEQ ID NO: 18 = SEQ ID NO: 2 + 6x marca His) y de unos mutantes basados en los resultados de la hidrólisis del 4-metilumbeliferil- β -D-lactósido, que han sido evaluados a varias temperaturas. Los valores de la fluorescencia fueron normalizados de acuerdo con la Figura 8 a lo largo del intervalo de temperaturas de 55°C a 75°C.

A...wt;
 B...G4C,A72C;
 C...G4C,A72C,Q349K;
 D...G4C,A72C,D181N,Q349K;
 E...Q1L,G4C,A72C,D181N,E183K,Q349R;
 F...QL,G4C,A72C,S86T,D181N,E183K,D320V,Q349R;
 G...G4C, A72C,E183K,D202Y,N310D,Q349R;
 H ... Q1L,G4C,A72C, A145T,H203R,Q349K,T403K;

Figure 9: Rendimientos de glucosa de la hidrólisis de una paja tratada previamente con una proteína de fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (CBH-ah), wt y mutada, después de haber llevado a cabo una hidrólisis durante 48 horas en presencia de una β -glucosidasa. Las variantes se han caracterizado para las siguientes mutaciones referidas a la SEQ ID NO: 18 y fueron expresadas a partir de *Pichia pastoris* en unos cultivos en matraces de sacudimiento, y fueron aisladas a partir del material sobrenadante por medio de una cromatografía de afinidad usando Ni-NTA.

A: wt
 B: G4C,A72C
 C: G4C,A72C,Q349K
 D: G4C,A72C, D181N,Q349K
 E: Q1L,G4C,A72C,D181N,E183K,Q349R
 F: Q1L,G4C,A72C,S86T,D181N,E183K,D320V,Q349R

G: G4C, A72C, E183K, D202Y, N310D, Q349R

Figura 10: Alineación de la SEQ ID NO: 2 con la secuencia de CBHI de *Trichoderma reesei*. La matriz de alineación blosum62mt2 con una penalización por apertura de huecos (en inglés "gap opening penalty") de 10 y una penalización por extensión de huecos de 0,1 fue utilizada para crear la alineación.

Descripción detallada del invento

El presente invento divulga un polipéptido, que tiene una actividad de celobiohidrolasa, que comprende una secuencia de aminoácidos con una identidad entre secuencias de por lo menos 85 % con la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 ha sido modificado mediante sustitución o supresión. El concepto de "celobiohidrolasa" o "CBH" se refiere a unas enzimas que disocian a una celulosa desde el extremo de la cadena de glucosa, y que producen una celobiosa como el producto principal. Unos nombres alternativos son los de 1,4-beta-D-glucano celobiohidrolasas o celulosa 1,4-beta-celobiosidasas. Las CBHS hidrolizan a los enlaces 1,4-beta-d-glucosídicos de los extremos reductores o no reductores de un polímero que contiene tales enlaces. La "celobiohidrolasa I" o "CBH I" actúa a partir del extremo reductor de la fibra de celulosa. La "celobiohidrolasa II" o "CBH II" actúa a partir del extremo no reductor de la fibra de celulosa. Las celobiohidrolasas tienen típicamente una estructura que consiste en un dominio catalítico y uno o más "dominios de fijación de celulosa" o "CBD" (acrónimo del inglés "cellulose binding domains"). Tales dominios pueden estar localizados o bien junto al extremo terminal de N o C del dominio catalítico. Los CBDs tienen una actividad de fijación de hidratos de carbono y median por la fijación de la celulosa a una celulosa cristalina, y es conocido el hecho de que la presencia o ausencia de los dominios de fijación tiene un impacto importante sobre la aptitud de de una enzima para ser procesada especialmente sobre unos sustratos poliméricos.

La secuencia parental se ha establecido en la SEQ ID NO: 2. La secuencia se deriva de la fusión junto al extremo terminal de C del dominio engarzador y del dominio de fijación de celulosa de CBHI de *Trichoderma reesei* (SEQ ID NO: 4) con el dominio catalítico de CBHI de *Talaromyces emersonii* (SEQ ID NO: 5).

En un aspecto preferido, el invento divulga unas variantes de proteínas que muestran una alta actividad a altas temperaturas durante un prolongado período de tiempo. De manera preferida, el polipéptido del presente invento mantiene un 50 % de su máxima capacidad de conversión del sustrato cuando la conversión se realiza durante 60 minutos a una temperatura de 60 °C o más alta. La respectiva temperatura se designa también como el valor IT50. En otras palabras, el valor IT50 es de manera preferida de 60 °C o más alto. La "Capacidad de Conversión del Sustrato" de una enzima se define en este contexto como el grado de conversión del sustrato, que es catalizada por una cantidad de una enzima en un determinado período de tiempo en unas condiciones definidas (la concentración del sustrato, el valor del pH, la concentración del tampón y la temperatura), tal como se pueden determinar mediante un ensayo en el punto final de la reacción enzimática en tales condiciones. La "Capacidad Máxima de Conversión del Sustrato" de una enzima se define en este caso como el valor máximo de la capacidad de conversión del sustrato que se encuentra para la enzima en un determinado número de mediciones que se efectúan tal como se ha descrito más arriba, en el que sólo se varió un parámetro, p.ej. la temperatura, dentro de un intervalo definido. De acuerdo con el presente invento, el ensayo descrito en el Ejemplo 8 se utiliza para determinar estos parámetros.

Por lo demás, los polipéptidos divulgados tienen de manera preferida un valor IT50 en el intervalo de 62 a 70 °C, de manera más preferida de 65 a 70 °C.

El polipéptido del presente invento comprende de manera preferida una secuencia de aminoácidos, que tiene identidad entre secuencias de por lo menos 90 %, de manera preferida de por lo menos 95 % y de manera más preferida de por lo menos 99 % con la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado mediante sustitución o supresión. Además, se prefiere especialmente que la secuencia de aminoácidos del polipéptido tenga la secuencia como se ha definido por la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado mediante sustitución o supresión, o una secuencia tal como se ha definido por la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado mediante sustitución o supresión, en donde han sido sustituidos, suprimidos o insertados de 1 a 75, de manera más preferida de 1 a 35 residuos de aminoácidos.

Son preferidas particularmente unas variantes de la proteína de la SEQ ID NO: 2, en las que el residuo de aminoácidos en la posición Q1 de la SEQ ID NO: 2 ha sido modificado mediante sustitución o supresión. Las "variantes de proteínas" son unos polipéptidos cuya secuencia de aminoácidos se diferencia en una o más posiciones de esta proteína parental, en donde las diferencias pueden ser un reemplazo de un aminoácido por otro, unas supresiones de uno o más aminoácidos, o una inserción de un aminoácido adicional o unos tramos de aminoácidos en la secuencia parental. Por definición, unas variantes definidas del polipéptido parental deben de ser distinguidas de otros polipéptidos mediante una comparación de la identidad entre secuencias (alineaciones) usando el algoritmo ClustalW (Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin

F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. y Higgins D.G. (2007) ClustalW and ClustalX versión 2 [ClustalW y ClustalX versión 2]. *Bioinformatics* 2007 23(21): 2947-2948). Los métodos para la generación de tales variantes de proteínas incluyen una mutagénesis aleatoria o dirigida a un sitio, una mutagénesis por saturación de un sitio, un ensamblaje de fragmentos basado en una PCR, una barajadura de los ADN, una recombinación homóloga in-vitro o in-vivo, y unos métodos de síntesis de genes.

La nomenclatura de los aminoácidos, péptidos, nucleótidos y ácidos nucleicos se realiza de acuerdo con las sugerencias de la IUPAC. Generalmente, los aminoácidos son denominados dentro de este documento de acuerdo con el código de una sola letra.

Los intercambios de unos aminoácidos individuales son descritos nombrando el código de una sola letra del aminoácido original seguido por su posición y el código de una sola letra del aminoácido que lo ha reemplazado, es decir, el intercambio de una glutamina en la posición 1 por una leucina en esta posición se describe como "Q1L". Para las supresiones de unas posiciones individuales desde la secuencia, el símbolo del aminoácido reemplazador es sustituido por la abreviación de tres letras "del", por lo tanto, la supresión de una alanina en la posición 3 sería designada como "A3del". Unos aminoácidos adicionales insertados reciben el número de la posición precedente prolongado por una letra minúscula en orden alfabético en relación con su distancia con respecto a su sitio de inserción. Por lo tanto, la inserción de dos triptófanos detrás de la posición 3 es designada como "3aW, 3bW". La introducción de unos codones no traducidos, TAA, TGA y TAG dentro de la secuencia de ácidos nucleicos se indica como "*" en la secuencia de aminoácidos, por lo tanto la introducción de un codón terminador en la posición 4 de la secuencia de aminoácidos se indica como "G4*".

Unas mutaciones múltiples son separadas por un signo "más" o por una barra o por una coma. Por ejemplo, dos mutaciones en las posiciones 20 y 21 sustituyendo una alanina y un ácido glutámico por una glicina y una serina, respectivamente, son indicadas como "A20G+E21S" o "A20G/E21S" "A20G,E21S".

Cuando un residuo de aminoácido en una posición determinada es sustituido por dos o más residuos de aminoácidos alternativos, estos residuos son separados por una coma o una barra. Por ejemplo, una sustitución de una alanina en la posición 30 por una glicina o un ácido glutámico es indicada como "A20G,E" o "A20G/E", o "A20G,A20E".

Cuando una posición apropiada para una modificación, es identificada en el presente contexto sin que se sugiera ninguna modificación específica, se ha de entender que un residuo de aminoácido arbitrario puede sustituir al residuo de aminoácido que está presente en esa posición. Así, por ejemplo, cuando se menciona, pero no se especifica, una modificación de una alanina en la posición 20, se ha de entender que la alanina puede ser suprimida o sustituida por cualquier otro residuo de aminoácido (por ejemplo, uno cualquiera de los R, N, D, C, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y y V).

Los conceptos de "mutación similar" o "sustitución similar" se refieren a una mutación de un aminoácido que un experto en la especialidad podría considerar como similar a una primera mutación. El concepto de "similar" significa dentro de este contexto que un aminoácido tiene unas características químicas similares. Si, por ejemplo, una mutación en una posición específica conduce a una sustitución de un residuo de aminoácido no alifático (p.ej. Ser) por un residuo de aminoácido alifático (p.ej. Leu), entonces una sustitución en la misma posición por un aminoácido alifático diferente (p.ej. Ile o Val) es referida como una mutación similar. Otras características más de los aminoácidos incluyen el tamaño del residuo, la hidrofobicidad, la polaridad, la carga, el valor de pK, y otras características de los aminoácidos que sean conocidas en la especialidad. De acuerdo con esto, una mutación similar puede incluir una sustitución tal como la de un residuo de carácter básico por otro residuo de carácter básico, la de un residuo de carácter ácido por otro residuo de carácter ácido, la de un residuo polar por otro residuo polar etc. Los conjuntos de aminoácidos que se derivan de esta manera probablemente deben de ser conservados por razones estructurales. Estos conjuntos pueden ser descritos en la forma del diagrama de Venn (Livingstone CD, y Barton G.J. (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" [Alineaciones de secuencias de proteínas: una estrategia para el análisis jerárquico de conservación de residuos]. *Comput. Appl Biosci.* 9: 745-756; Taylor W. R. (1986) "The classification of amino acid conservation" [La clasificación de la conservación de aminoácidos] *J.Theor.Biol.* 119: 205-218). Unas sustituciones similares se pueden realizar, por ejemplo, de acuerdo con la siguiente agrupación de aminoácidos: hidrófobos: F W Y H K M I L V A G; aromáticos: F W Y H; alifáticos: I L V; polares: W Y H K R E D C S T N; cargados eléctricamente H K R E D; cargados positivamente: H K R; cargados negativamente: E D.

Como una convención para la numeración de aminoácidos y la designación de las variantes de proteínas para la descripción de las variantes de proteínas, la primera glutamina (Q) de la secuencia de aminoácidos QQAGTA en la secuencia de la proteína parental establecida en la SEQ ID NO: 2 se indica como la posición número 1 o Q1 o glutamina 1. La numeración de todos los aminoácidos estará de acuerdo con su posición en la secuencia parental establecida en la SEQ ID NO: 2 en relación con esta posición número 1.

ES 2 549 328 T3

5 El presente invento divulga además unas variantes de los polipéptidos del presente invento con unos cambios en su secuencia en una o más de las posiciones: G4, A6, T15, Q28, W40, D64, E65, A72, S86, K92, V130, V152, Y155, K159, D181, E183, N194, D202, P224, T243, Y244, I277, K304, N310, S311, N318, D320, T335, T344, D346, Q349, A358, Y374, A375, T392, T393, D410, Y422, P442, N445, R446, T456, S460, P462, G463, H468 y/o V482 de los aminoácidos 1 hasta 500 de la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado mediante sustitución o supresión.

10 En una forma de realización preferida, la variante de los polipéptidos del presente invento comprende uno o más cambios específicos de su secuencia en las siguientes posiciones (intercambio preferido) o una mutación similar.

Posición	Intercambio preferido	Mutación similar
Q1	L	I, L, V, A, G,
G4	C	C, D, E, R, H, K
A6	G, V	I, L, V, G
T15S	S	Q, N
Q28	R	H, K, R
W40	R	H, K, R
D64	N	Q, N, S, T, E
E65	K, V	H, K, R
A72	C, V	M, I, L, V, G, C
S86	T	H, K, R, E, D, C, T, N
K92	R	H, R
V130	I	F, W, Y, H, K, M, I, L, A, G
V152	A, E	D, E, A, I, L, G, P
Y155	C	C, D, E, R, H
K159	E	D, E
D181	N	Y, H, K, R, C, S, T, N
E183	V, K	I, L, V, A, G, R, K, H
N194	C, R, Y, D, K, I, L, G, Q, S, V	F, W, C, R, Y, D, K, I, L, G, Q, S, V, H
D202	Y, N, G	A, R, N, C, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V
P224	L	I, L, V, A, G
T243	I, C, R, Y, A, F, Q, P, D, V, W, L, M	H, K, I, C, R, Y, A, F, Q, P, D, V, W, L, M
Y244	F, H	F, W, H,
I277	V	G, A
K304	R	H, R
N310	D	C, D, E, R, H
S311	G, N	I, L, V, A, G, Q, T, D,
N318	Y	F, W, Y, H
D320	V, E, N	W, Y, H, K, R, E, C, S, T, N
T335	I	F, W, Y, H, K, M, I, L, V, A, G
T344	M	F, W, Y, H, K, M, I, L, V, A, G
D346	G, A, E	I, L, V, A, G
Q349	R, K	H, K, R
A358	E	F, W, Y, H, K, M, I, L, V, G
Y374	C, P, R, H, S, A	C, P, R, H, S, A, K, E, D, C, F
A375	C, D, N, Y, R, G, L, V, E, G, T, M	H, K, C, D, N, Y, R, Q, L, V, E, G, T, M

ES 2 549 328 T3

Posición	Intercambio preferido	Mutación similar
T392	C, D, K	H, K, R, E, D, C
T393	A	I, L, V, A, G,
D410	G	F, W, Y, H, K, M, I, L, V, A, G
Y422	F	H, F, W
P442	S, del	S, del, T
N445	D	D, E, Q
R446	S, G	A, S, G, V, I, L, C, P
T456	A	S, A, G, I, L, V, C, P
S460	L, P	G, A, I, L, P, V
P462	L, del	G, A, I, L, V, del
G463	D	D, E
H468	L, Q, R	K, R, E, D, C, S, T, N
V482	A, I	G, L, A, I

De manera todavía más preferida, la variante de los polipéptidos del presente invento comprende una secuencia de aminoácidos seleccionada entre las secuencias con unas mutaciones con respecto a la SEQ ID NO: 2, en la que el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado mediante una sustitución o supresión, y opcionalmente se ha fusionado con una marcación 6x-His en el extremo terminal de C, que se enumeran en la siguiente Tabla.

5

	Mutaciones con respecto a la SEQ ID NO: 2
75	Q1L,G4C,A72C,Q349K
88	Q1L,G4C,A72C,S86T,Q349R
89	Q1L,G4C,A72C,D181N,Q349R
90	Q1L,G4C,A72C,E183K,Q349R
91	Q1L,G4C,A72C,D181N,E183K,Q349R
92	Q1L,G4C,A72C,D320V,Q349R
93	Q1L,G4C,A72C,S86T,D181N,E183K,D320V,Q349R
98	Q1L,G4C,A72C,Q349R
148	Q1L,G4C,A72C,Q349K,T392M
153	Q1L,G4C,A68T,A72C,Q349K,G439D,R453S
154	Q1L,G4C,A72C,D202N,Q349K
155	Q1L,G4C,A68T,A72C,Q349K
156	Q1L,G4C,A72C,K154R,Q349K,T3931
157	Q1L,G4C,A72C,S193P,Q349K,V482I
158	Q1L,G4C,A72C,H203R,Q349K,P442S
159	Q1L,G4C,A72C,Q349K,H468R
160	Q1L,G4C,A72C,D202N,Q349K,G486D
161	Q1L,G4C,E65K,A72C,Q349K
162	Q1L,G4C,A72C,Q349K,Y422F
163	Q1L,G4C,Q28R,A72C,Q349K,H468L
164	Q1L,G4C,A72C,D181N,D247N,Q349K
165	Q1L,G4C,A72C,D181N,Q349K,T451S
166	Q1L,G4C,Q28R,A72C,Q349K

	Mutaciones con respecto a la SEQ ID NO: 2
167	Q1L,G4C,A72C,A145T,H203R,Q349K,T403K
168	Q1L,G4C,A72C,I200F,Q349K,L500I
169	Q1L,G4G,D64N,A72C,Q349K
170	Q1L,G4C,A72C,V152A,Q349K
171	Q1L,G4C,T15S,A72C,Y244F,Q349K
172	Q1L,G4C,A6V,A72C,Q349K
173	Q1L,G4C,A72C,S311N,Q349K,G463D
174	Q1L,G4C,A72C,Y155C,Q349K
175	Q1L,G4C,A72C,S311N,Q349K
176	Q1L,G4C,A72C,D346V,Q349K
177	Q1L,G4C,A72C,Q349K,T392K
178	Q1L,G4C,A72C,S311G,Q349K
179	Q1L,G4C,A72C,S311G,Q349K,H468R

5 En otro aspecto, el presente invento divulga un ácido nucleico, que codifica el polipéptido del presente invento. El ácido nucleico es una secuencia de polinucleótidos (un ADN o ARN), que, cuando ha sido puesta bajo el control de un promotor apropiado y transferida a un adecuado anfitrión biológico o a un adecuado entorno químico, es procesada para formar el polipéptido codificado, en donde el proceso incluye también todas las etapas posteriores a la traducción y posteriores a la transcripción, que son necesarias. La secuencia codificadora puede ser adaptada fácilmente mediante una variación de unos tripletes de bases degenerados, una alteración de las secuencias de señal, o por la introducción de unos intrones, sin afectar a las propiedades moleculares de la proteína codificada. El ácido nucleico del presente invento tiene de manera preferida una identidad de por lo menos 95 %, de manera más preferida de por lo menos 97 % y de la manera más preferida de 100 % con la SEQ ID NO: 1. El presente invento proporciona también un vector que comprende este ácido nucleico y un anfitrión transformado con tal vector.

15 La presente solicitud de patente divulga también unos métodos para la producción de los polipéptidos del presente invento y unas variantes de éstos en varias células anfitrionas, que incluyen unas levaduras y unos anfitriones fúngicos. Ella divulga también el uso de las cepas resultantes para el mejoramiento de ciertas propiedades de proteínas mediante una variación de la secuencia. Además de ello, la presente solicitud de patente divulga unos métodos para la aplicación de tales polipéptidos en la hidrólisis de una celulosa.

20 Un aspecto adicional de la solicitud de patente divulga unos vectores y unos métodos para la producción de unas variantes proteínicas de la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado mediante sustitución o supresión, expresándolo en una levadura y ensayando su actividad sobre un material celulósico mediando medición de las moléculas de azúcares mono- y/u oligoméricas segregadas.

25 La presente solicitud de patente divulga además un método para la producción de una proteína de celobiohidrolasa, que comprende las etapas de:

- a. obtener una célula anfitriona, que ha sido transformada con un vector que comprende el ácido nucleico del presente invento
- b. cultivar la célula anfitriona en unas condiciones en las que se expresa la proteína de celobiohidrolasa; y
- c. recuperar la proteína de celobiohidrolasa.

35 En una forma de realización preferida, la célula anfitriona se deriva del conjunto que se compone de *Saccharomyces*, *Schizosaccharomyces*, *Kluyveromyces*, *Pichia*, *Hansenula*, *Aspergillus*, *Trichoderma*, *Penicillium*, *Candida* y *Yarrowina*. De manera preferida, la célula anfitriona es capaz de producir etanol, en donde las levaduras más preferidas incluyen a *Saccharomyces cerevisiae*, *Pichia stipitis*, *Pachysolen tannophilus*, o a una levadura metilotrófica, que se deriva de manera preferida del conjunto de células anfitrionas que comprende *Pichia methanolica*, *Pichia pastoris*, *Pichia angusta* y *Hansenula polymorpha*.

40 Sorprendentemente, se ha encontrado que el polipéptido de acuerdo con el presente invento y unas variantes de éste se puede(n) expresar a partir de una levadura a altos niveles. El concepto de "levadura" deberá de referirse en el presente caso a todos los organismos eucarióticos inferiores que muestran un estado vegetativo unicelular en su ciclo vital. Esto incluye especialmente a unos organismos de la clase *Saccharomycetes*, en particular de los géneros

Saccharomyces, Pachysolen, Pichia, Candida, Yarrowina, Debaromyces, Klyveromyces, Zygosaccharomyces.

Por lo tanto, un aspecto de la solicitud de patente divulga la expresión del polipéptido reivindicado y de unas variantes de éste en una levadura. La expresión eficiente de esta proteína de fusión (SEQ ID NO: 2) y de unas variantes de proteínas derivadas de la SEQ ID NO: 2 a partir de una levadura se puede conseguir por inserción de la molécula de ácido nucleico de la SEQ ID NO: 1 empezando por la posición de nucleótido 1 dentro de un vector de expresión que está bajo el control de por lo menos una apropiada secuencia promotora y por fusión de la molécula de nucleótido con un apropiado péptido de señal, por ejemplo con el péptido de señal del factor de apareamiento alfa de *Saccharomyces cerevisiae*.

En una forma de realización preferida, el polipéptido del presente invento y unas variantes de éste se expresan y segregan a un nivel de más que 100 mg/l, de manera más preferida de más que 200 mg/l, de manera particularmente preferida de más que 500 mg/l o de una manera sumamente preferida de más que 1 g/l dentro del material sobrenadante, después de la introducción de un ácido nucleico que codifica un polipéptido que tiene una secuencia de aminoácidos con una identidad entre secuencias de por lo menos 85 % con respecto a la SEQ ID NO: 2, dentro de una levadura. Para determinar el nivel de expresión en levaduras, la cultivación y el aislamiento del material sobrenadante se pueden llevar a cabo tal como se describe en el Ejemplo 3.

Otro aspecto adicional de la solicitud de patente divulga unos métodos para la producción de un polipéptido de acuerdo con el presente invento en un hongo filamentoso, de manera preferida en un hongo del género *Aspergillus* o *Trichoderma*, de manera más preferida en un hongo del género *Trichoderma*, de manera sumamente preferida en *Trichoderma reesei*. El concepto de "hongos filamentosos" o de "hongos" deberá de referirse en el presente caso a todos los organismos eucarióticos inferiores que muestran un crecimiento con hifas durante de por lo menos estado en su ciclo vital. Esto incluye especialmente a unos organismos de los filos *Ascomycota* y *Basidiomycota*, en particular del género *Trichoderma*, *Talaromyces*, *Aspergillus*, *Penicillium*, *Chrysosporium*, *Phanerochaete*, *Thermoascus*, *Agaricus*, *Pleutrus* o *Irpex*. El polipéptido es expresado por fusión de la región codificadora de una secuencia de señal compatible con la molécula de ácido nucleico empezando por la posición de nucleótido 52 de la SEQ ID NO: 3, tal como se hizo en la SEQ ID NO: 3 con la secuencia de señal de CBHI de *Trichoderma reesei*, y mediante la puesta del péptido de fusión bajo el control de un promotor que sea lo suficientemente fuerte, seguida por una transferencia de la construcción genética a la célula anfitriona. Unos ejemplos de tales promotores y secuencias de señal, así como unas técnicas para realizar una transferencia eficiente se han descrito en la especialidad.

En otro aspecto, la presente solicitud de patente divulga además un método para la identificación de unos polipéptidos que tienen una actividad de celobiohidrolasa, que comprende las etapas de:

- a. Generar una biblioteca de genes mutantes que codifican unas proteínas mutantes mediante mutagénesis de un ácido nucleico de acuerdo con la reivindicación 9 o de un ácido nucleico que tiene la secuencia definida por la SEQ ID NO: 6 (que codifica la SEQ ID NO: 5), que tiene de manera preferida la secuencia definida por la SEQ ID NO: 1;
- b. Insertar cada gen mutante dentro de un vector de expresión;
- c. Transformar unas células de levadura con cada vector de expresión para crear una biblioteca de transformantes de levadura;
- d. Cultivar cada transformante de levadura en unas condiciones en las que la proteína mutante es expresada y segregada;
- e. Incubar la proteína mutante expresada con un sustrato;
- f. Determinar la actividad catalítica de la proteína mutante;
- g. Seleccionar una proteína mutante de acuerdo con la actividad catalítica determinada.

Especialmente, la etapa d. se puede realizar mediante utilización de un formato de placa y de pocillos. Este formato permite de manera preferida la realización con alto rendimiento de desarrollo del método para identificar unos polipéptidos que tienen una actividad de celobiohidrolasa.

De manera preferida, las etapas e. hasta g. del método para identificar unos polipéptidos que tienen una actividad de celobiohidrolasa se realizan de la siguiente manera:

- e. Incubar la proteína mutante expresada con un material celulósico;
- f. Determinar la cantidad del azúcar liberado;
- g. Seleccionar una proteína mutante de acuerdo con la cantidad del azúcar liberado.

En otra forma de realización, el método para identificar unos polipéptidos que tienen una actividad de celobiohidrolasa comprende las etapas adicionales de:

- h. Secuenciar el gen o la proteína mutante que se ha seleccionado;

i. Identificar la(s) modificación(modificaciones) de aminoácidos por comparación de la secuencia de la proteína mutante seleccionada con la secuencia de aminoácidos de la SEQ ID NO: 2.

5 La presente solicitud de patente divulga además un método para la preparación de un polipéptido que tiene una actividad de celobiohidrolasa, que comprende las etapas de:

- a. Proporcionar un polipéptido que tiene una actividad de celobiohidrolasa, que comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 70 % con el dominio catalítico de la SEQ ID NO: 2 (SEQ ID NO: 5);
- 10 b. Identificar los aminoácidos de este polipéptido que corresponden a los aminoácidos que se han modificado con respecto a la secuencia de aminoácidos de la SEQ ID NO: 2, tal como se ha identificado en la etapa i. del método para identificar unos polipéptidos que tienen una actividad de celobiohidrolasa; y
- 15 c. Preparar un polipéptido mutante del polipéptido que se ha proporcionado en la etapa a. llevando a cabo la(s) modificación(modificaciones) de aminoácidos que se han identificado en la etapa b. por medio de una mutagénesis dirigida a un sitio.

De manera preferida, el polipéptido proporcionado en la etapa a. del método para preparar un polipéptido que tiene una actividad de celobiohidrolasa es una celobiohidrolasa de tipo silvestre que se deriva de *Trichoderma reesei*.

20 La presente solicitud de patente divulga además unos polipéptidos que tienen una actividad de celobiohidrolasa, que son obtenibles con ayuda del método para preparar un polipéptido que tiene una actividad de celobiohidrolasa de acuerdo con la presente solicitud de patente.

25 Además, el presente invento proporciona una composición que comprende un polipéptido y/o unas variantes de éste del presente invento y una o más celulasas, p.ej. una o más endoglucanasas y/o una o más beta-glucosidasas y/o una o más celobiohidrolasas adicionales y/o una o más xilanasas. Los conceptos de "celulasas" o "enzimas celulolíticas" se definen como unas enzimas, que son capaces de hidrolizar a unos substratos celulósicos o unos derivados de éstos o unas materias primas de alimentación mezcladas que comprenden unos polímeros celulósicos. A tales enzimas se hace referencia como que tienen una "actividad celulolítica", siendo ellas por lo tanto capaces de hidrolizar a unas moléculas de celulosa a partir de tal material para dar unos oligo- o monosacáridos más pequeños. Las enzimas celulolíticas incluyen unas celulasas y unas hemicelulasas, en particular ellas incluyen unas celobiohidrolasas (CBHs), unas endoglucanasas (EGs) y unas beta-glucosidasas (BGLs).

35 La presente solicitud de patente divulga además un polipéptido, que tiene una actividad de celobiohidrolasa, en donde el polipéptido comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 80 %, de manera preferida de por lo menos 95 %, de manera más preferida de por lo menos 98 %, de manera todavía más preferida de por lo menos 99 %, y de la manera más preferida de 100 % con la SEQ ID NO: 5, en donde uno o más de los siguientes residuos de aminoácidos de la secuencia definida por la SEQ ID NO: 5 son modificados por sustitución o supresión de: Q1, G4, A6, T15, Q28, W40, D64, E65, A72, S86, K92, V130, V152, Y155, K159, D181, E183, N194, D202, P224, T243, Y244, 1277, K304, N310, S311, N318, D320, T335, T344, D346, Q349, A358, Y374, A375, T392, T393, D410 y/o Y422 de los aminoácidos 1 hasta 440 de la SEQ ID NO: 5.

45 En una forma de realización preferida, el polipéptido que tiene una actividad de celobiohidrolasa con una secuencia de aminoácidos, que tiene que tiene una identidad entre secuencias de por lo menos 80 % con la SEQ ID NO: 5 comprende uno o más de los siguientes residuos de aminoácidos modificados de la secuencia definida por la SEQ ID NO: 5: Q1L, G4, A6G/V, T15S, Q28Q/R, W40R, D64N, E65K/V, A72V, S86T, K92K/R, V130I/V, V152A/E, Y155C, K159E, D181N, E183V/K, N194C/R/Y/D/K/I/UG/Q/S/V, D202Y/N/G, P224L, T243I/R/Y/A/F/Q/P/D/N/W/L/M, Y244F/H, I277V, K304R, N310D, S311G/N, N318Y, D320V/E/N, T335I, T344M, D346G/A/E/V, Q349R/K, A358E, Y374C/P/R/H/S/A, A375D/N/Y/R/Q/L/V/E/G/T/M, T392C/D/K, T393A, D410G, Y422F.

50 De manera más preferida, el polipéptido que tiene una actividad de celobiohidrolasa comprende uno o más residuos de aminoácidos modificados de la secuencia definida por la SEQ ID NO: 5 tal como se indica en la siguiente tabla:

	Mutaciones con respecto a la SEQ ID NO: 5
1	G4C,Y163C
2	E183V
3	W40R
4	S311N
5	T335I
6	T48A

ES 2 549 328 T3

	Mutaciones con respecto a la SEQ ID NO: 5
7	W40R,M234K
8	E65K
9	T344M
10	A72V
11	I277V
12	D346G
13	S418P
14	G4C,W40R,A72C
15	W40R,T344M
16	W46R,Q349K
17	W40R,T344M,Q349K
18	G4C,A72C,T344M
19	G4C,A72C,Q349K
20	G4C,A72C,T344M,Q349K
21	G4C,W40R,A72C,T344M
22	G4C,W40R,A72C,Q349K
23	G4C,W40R,A72C,T344M,Q349K
24	N54S,N194Y
25	V130I,H350Y
26	K92R
27	T325I
28	Y60H
29	T229A
30	D320E
31	E65V,Q349R
32	T392M
33	G4C,A72C,D346G
34	G4C,A72C,Q349R
35	G4C,A72C,D346G,Q349R
36	G4C,A72C,T344M,D346G,Q349R
37	G4C,A6G,A72C,Q349K
38	G4C,A72C,Q349K,A376T
39	G4C,A72C,Q349K,E65V,Q349R
40	G4C,A72C,Q349K,S24N
41	G4C,P12Q,A72C,Q349K,Q362H
42	N194C,Y374C
43	T243C,A375C,N194C,Y374C
44	G4C,A72C,Q349K,D320V
45	G4C,A72C,Q349K,S86T
46	G4C,A72C,D181N,Q349K
47	G4C,A72C,E183V,K304R,Q349K
48	Q1L,G4C,A72C,Q349K

ES 2 549 328 T3

	Mutaciones con respecto a la SEQ ID NO: 5
49	G4C,A72C,S86T,M234V,Q349K
50	G4C,A72C,E183K,Q349K
51	G4C,A72C,S311G,Q349K
52	G4C,D64N,A72C,Q349K
53	G4C,A72C,D181 N,P224L,Q349K
54	G4C,A72C,E183K,N318Y,Q349K
55	G4C,A72C,E183V,Q349K
56	G4C,A72C,N194C,Y374C
57	G4C,A72C,N194C,Q349R,Y374C
58	N194R,Y374D
59	T243R,A375D
60	N194R,T243R,Y374D,A375D
61	Q1L,G4C,A72C,S86T,Q349R
62	Q1L,G4C,A72C,D181N,Q349R
63	Q1L,G4C,A72C,E183K,Q349R
64	Q1L,G4C,A72C,D181N,E183K,Q349R
65	Q1L,G4C,A72C,D320V,Q349R
66	Q1L,G4C,A72C,S86T,D181N,E183K,D320V,Q349R
67	G4C,A72C,Q349R,V367A
68	G4C,A72C,E183K,D202Y,N310D,Q349R
69	Q1L,G4C,A72C,Q349R
70	G4C,A72C,G251R,Q349R
71	G4C,A72C,P224L,F306Y,Q349R
72	G4C,V32G,N49S,A72C,S193L,Q349R
73	G4C,A72C,D181N,Q349R
74	G4C,A72C,D202N,Q349R
75	A72V,T335I,D346A,T393A,P436S
76	A72V,T335I,D346A,T393A
77	W40R,D346A,T393A
78	A72V,D346A,T393A
79	G4C,E65V,A72C,Q349R
80	G4C,A72C,D202G,D320N,Q349R,A358E
81	G4C,A72C,D320V,Q349R
82	G4C,E65V,A72C,Y244H,Q349R
83	G4C,A72C,E183K,Q349R
84	G4C,A72C,E183K,D346E,Q349R
85	G4C,A72C,S311G,Q349R
86	G4C,A72C,P224L,Q349R
87	G4C,A72C,S236Y,Q349R
88	A72V,D320V,D346A
89	A72V,D346A
90	W40R,T335I,D346A,T393A

	Mutaciones con respecto a la SEQ ID NO: 5
91	G4C,A72C,D320V,Q349K
92	G4C,A72C,S311G,D320V,Q349K
93	Q1L,G4C,A72C,Q349K,T392M
94	D320V,Q349K
95	G4C,E172C,D202V,D320V,Q349K
96	G4C,A72C,D320V,D346V,Q349K
97	Q1L,G4C,A72C,D202N,Q349K
98	Q1L,G4C,A68T,A72C,Q349K
99	Q1L,G4C,A72C,K154R,Q349K,T393I
100	Q1L,G4C,E65K,A72C,Q349K
101	Q1L,G4C,A72C,D181N,D247N,Q349K
102	Q1L,G4C,Q28R,A72C,Q349K
103	Q1L,G4C,A72C,A145T,H203R,Q349K,T403K
104	Q1L,G4C,D64N,A72C,Q349K
105	Q1L,G4C,A72C,V152A,Q349K
106	Q1L,G4C,T15S,A72C,Y244F,Q349K
107	Q1L,G4C,A6V,A72C,Q349K
108	Q1L,G4C,A72C,Y155C,Q349K
109	C11L,G4C,A72C,S311N,Q349K
110	Q1L,G4C,A72C,D346V,Q349K
111	Q1L,G4C,A72C,Q349K,T392K
112	Q1L,G4C,A72C,S311G,Q349K
113	G4C,A72C,N194K,T243Y,Q349R,A375N
114	G4C,A72C,N194D,T243A,Q349R,Y374P,A375Y
115	G4C,A72C,N194I,T243Y,Q349R,Y374P,A375R
116	G4C,A72C,N194K,Q349R,Y374P,A375Q
117	G4C,A72C,N194L,T243Y,Q349R,Y374R,A375L
118	G4C,A72C,N194G,T243F,Q349R,Y374P,A375R
119	G4C,A72C,N194L,T243Q,Q349R,Y374P,A375V
120	G4C,A72C,N194K,T243P,Q349R,Y374H,A375E
121	G4C,A72C,N194I,T243D,Q349R,Y374P,A375Y
122	G4C,A72C,N194Q,T243V,Q349R,Y374P,A375Y
123	G4C,A72C,N194S,T243W,Q349R,Y374S,A375G
124	G4C,A72C,N194Y,T243L,Q349R,Y374S,A375R
125	G4C,A72C,N194V,T243M,Q349R,Y374A,A375T
126	G4C,A72C,T243G,Q349R,Y374P,A375M
127	G4C,A72C,T243Q,Q349R,Y374P,A375M
128	G4C,A72C,N194Y,T243V,Q349R,Y374P

Además, la presente solicitud de patente divulga un polipéptido que tiene una actividad de celobiohidrolasa, que comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 85 %, de manera preferida de por lo menos 95 %, de manera más preferida de por lo menos 98 %, de manera aún más preferida de por lo menos 99 %, y de la manera más preferida de 100 % con la SEQ ID NO:12, en donde uno o más de los siguientes residuos de aminoácidos de la secuencia definida por la SEQ ID NO:12 son modificados mediante

sustitución o supresión: Q1, T15, Q28, W40, C72, V133, V155, Y158, T162, Y247, N307, G308, E317, S341, D345, Y370, T389, Q406, N441, R442, T452, S456, P458, G459, H464 y/o V478.

- 5 En una forma de realización preferida, el polipéptido que tiene una actividad de celobiohidrolasa con una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 85 % con la SEQ ID NO: 12 comprende uno o más de los siguientes residuos de aminoácidos modificados de la secuencia definida por la SEQ ID NO: 12:

	Intercambio con respecto a la SEQ ID NO: 12
1	Q1L
2	T15S
3	Q28R
4	W40R
5	C72V
6	V133I
7	V155A,E
8	Y158C
9	T162E
10	Y247F,H
11	N307D
12	G308N
13	E317V,N
14	S341M
15	D345R,K
16	Y370P,R,H,S,A
17	T389A
18	Q406G
19	N441D
20	R442S,G
21	T452A
22	S456L,P
23	P458L,del
24	G459D
25	H464L,Q,R
26	V478A,I

- 10 Otro aspecto de la divulgación se refiere a la utilización de los polipéptidos aislados y de las variantes de éstos del presente invento para la hidrólisis completa o parcial de un material celulósico. El material celulósico puede ser de naturaleza natural, procesada o artificial. El concepto de "un material celulósico" debe de ser definido en este contexto como todos los tipos de materiales puros, no puros, mezclados, entremezclados o compuestos de otra manera, que contienen por lo menos una fracción de unos polímeros con D-glucosilo unidos por un enlace β -1-4 de por lo menos 7 subunidades consecutivas. Unos ejemplos sobresalientes de unos materiales celulósicos son todos
- 15 los tipos de materiales vegetales que contienen celulosa, tales como madera (blanda y dura), paja, granos, hierba de elefante, heno, hojas, algodón y unos materiales procesados a partir de éstos, o unas corrientes de desechos derivadas de tales procesos. A un material celulósico que se usa en una reacción enzimática se hace referencia en este contexto también como un sustrato celulósico.
- 20 La hidrólisis del material celulósico puede ser un proceso secuencial, que sigue a la producción de una celobiohidrolasa o contemporáneo con la producción en la célula de levadura (proceso biológico consolidado). La expresión de unas enzimas celulolíticas en levaduras para fermentar los azúcares (de C6 o C5) liberados para dar etanol u otros metabolitos que interesen.

Otro aspecto de la solicitud de patente se refiere por lo tanto a la aplicación de unas células enteras, que expresan el polipéptido o una variante de éste, de acuerdo con el presente invento para el tratamiento de materiales celulósicos.

5 En un aspecto particular, la presente solicitud divulga el uso de un polipéptido y de unas variantes de éste o la composición del presente invento para la degradación enzimática de un material celulósico, de manera preferida de una biomasa lignocelulósica, y/o para el tratamiento de materiales textiles y/o como un ingrediente en detergentes y/o como un ingrediente en alimentos o composiciones alimentarias.

Ejemplos

10 Ejemplo 1: Producción de un plásmido de expresión en *Pichia pastoris*

Unos plásmidos de expresión para la expresión constitutiva de una proteína a partir de unos anfitriones transformados de *Pichia pastoris* son producidos por ensamblaje de un casete de expresión, que se compone de un promotor de la gliceraldehído fosfato deshidrogenasa (GAP) de *Pichia pastoris*, un SP α (péptido de señal del factor de apareamiento alfa) de *Saccharomyces cerevisiae*, un sitio de clonación múltiple (MCS) y la secuencia de terminador 3'-GAP. Para finalidades de selección, se usa un gen de resistencia frente a la kanamicina puesto bajo el control del promotor EM7 o TEF para finalidades de selección en bacterias o levaduras, respectivamente. Los vectores plasmídicos resultantes son designados como pV1 (Figura 1) y pV2 (MCS alternativo). La transformación y la cultivación con expresión son realizadas esencialmente tal como lo describieron Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V., Cregg, J. M. (1997). Isolation of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter [Aislamiento del gen de la gliceraldehído-3-fosfato deshidrogenasa de *Pichia Pastoris* y regulación y uso de su promotor]. *Gene*, 186, 37-44, y Cregg, J.M.: *Pichia Protocols in Methods in Molecular Biology*, Second Edition, Humana Press, Totowa New Jersey 2007.

25 Ejemplo comparativo 1: Construcción de unas construcciones artificiales de expresión en *Pichia pastoris* para unas secuencias de CBHI

Los genes de CBHI de *Trichoderma viride* (CBH-f), *Humicola grisea* (CBH-d), *Thermoascus aurantiacus* (CBH-e), *Talaromyces emersonii* (CBH-b), y unas fusiones del dominio de fijación de celulosa de CBHI de *Trichoderma reesei* con la CBHI de *Talaromyces emersonii* (CBH-a) o la CBHI de *Humicola grisea* (CBH-g) se amplifican usando unos pares de oligonucleótidos y unos moldes (obtenidos por síntesis de genes) tal como se indica en la tabla. El gen de fusión que codifica la SEQ ID NO: 2 se genera mediante una PCR por extensión del solapamiento mediando utilización de los fragmentos de PCR que se han generado a partir de las SEQ ID NOs: 5 y 11. Para la PCR de amplificación PCR se utiliza la Phusion ADN polimerase (de Finnzymes).

35 Tabla 1: Cebadores y moldes para la amplificación de CBH-a, CBH-b, CBH-d, CBH-e, CBH-f y CBH-g.

	Fragmento	Cebador hacia adelante	Cebador inverso	Molde
CBH-f	CBHI de <i>Trichoderma viride</i>	GAGGCGGAAGCACCC TCTcaatctgctgaccttgca gtc	GGAGACGCAGAGCC Cttattacaggcactgcgagt agt	SEQID NO. 13
CBH-d	CBHI de <i>Humicola grisea</i>	GAGGCGGAAGCACCC TCTcagcaggctggtactatta ctgc	GGAGACGCAGAGCC Cttacacgttcacggtagaac cgattgggc	SEQID NO. 7
CBH-e	CBHI de <i>Thermoascus aurantiacus</i>	GAGGCGGAAGCACCC TCTcagcaggccggtaccgta accgc	GGAGACGCAGAGCC CTTAttagtggcgggtaag gtcgagt	SEQID NO. 9

	Fragmento	Cebador hacia adelante	Cebador inverso	Molde
CBH-b	CBHI de <i>Talaromyces emersonii</i>	GAGGCGGAAGCACCC TCTcagcaggccggcacggc gacggc	GGAGACGCAGAGCC CTTAtcacgaagcggtgaa ggtcgagt	SEQID NO. 5
CBH-a parte 1	Fragmento de fusión del dominio de fijación de CBHI de <i>Talaromyces emersonii</i> , parte1	GAGGCGGAAGCACCC TCTcagcaggccggcacggc gacggc	ATTACCTGTGCTACC gatcggaccaaacttaagtgc g	SEQID NO. 5
CBH-a parte 2	Fragmento de fusión del dominio de fijación de CBHI de <i>Trichoderma reesei</i> parte 2	AAGTTTGGTCCGATCg gtagcacaggtaatccttcagg	GGAGACGCAGAGCC CTTATTAtagacactgtga gtagtaagggt	SEQID NO.11
CBH-a	Proteína de fusión de CBHI de <i>Talaromyces emersonii</i>	GAGGCGGAAGCACCC TCTcagcaggccggcacggc gacggc	GGAGACGCAGAGCC CTTAtcattaatggtggtggt gatgatgag	5a+5b SEQID NO. 2
CBH-g part1	Fragmento de fusión de CBHI de <i>Humicola grisea</i> , parte 1	aggcgggaagcatgctgcagc aggctggtacaattactgc	ggattacctgtaagcttcaat tggtcgaatctgatgtt	SEQID NO. 19
CBH-g part2	Fragmento de fusión del dominio de fijación de CBHI de <i>Trichoderma reesei</i> , parte 2	accaattggaagcttaacaggta atccttcagggtgtaatcc	atcttgcaggctgacttatcatt aatgatgatgatggtgtgc a	SEQID NO. 11
CBH-g	Proteína de fusión de <i>Humicola grisea</i>	aggcgggaagcatgctgcagc aggctggtacaattactgc	atcttgcaggctgacttatcatt aatgatgatgatggtgtgc a	6a+6b SEQID NO. 15

5 Unos fragmentos de PCR que tienen la longitud esperada son purificados a partir de unos geles de agarosa después de una electroforesis usando el estuche Promega® SV PCR and Gel Purification kit. Las concentraciones de los fragmentos de ADN son medidas en un espectrofotómetro, y 0,2 pmoles de los fragmentos son tratados con 9 U de la ADN polimerasa de T4 en presencia de 2,5 mM de dATP durante 37,5 min a 22,5 °C, y los fragmentos tratados son reanillados con ADN plasmídico pV1 linealizados con la *SmaI* y tratados con una ADN polimerasa de T4 / dTTP, y después de esto transformados en el seno de unas células químicamente competentes de *Escherichia coli* Top10. Apartándose del procedimiento descrito, el producto generado por el par de cebadores de acuerdo con la tabla, pista 11, que codifican los fragmentos de la proteína de fusión de *Humicola grisea*, son clonados por intermedio de los sitios de *SphI* y *Sall* que han sido introducidos, para dar el pV2. Los transformantes son controlados mediante secuenciación del ADN plasmídico aislado.

15 Ejemplo comparativo 2: Expresión de los genes de CBHI en *Pichia pastoris*

Los plásmidos del Ejemplo 2 son transformados en unas células electro-competentes de *Pichia pastoris* CBS 7435, y los transformantes son utilizados para inocular a unos cultivos en un medio YPD que contiene 200 mg/l, los cuales son incubados durante 5 días a 27°C en un agitador rotatorio a 250 rpm (revoluciones por minuto). Los materiales

sobrenadantes del cultivo fueron separados por centrifugación a 5.000xg durante 30 minutos en una centrífuga Sorvall Avant. Los materiales sobrenadantes fueron concentrados en unas columnas giratorias con un tamaño de exclusión de 10 kDa. El patrón de proteínas de tales materiales sobrenadantes concentrados fue analizado mediante una SDS-PAGE (véase la cita de Laemmli y colaboradores) y unos geles se tiñeron con un colorante Commassie blue coloidal. La actividad enzimática fue determinada por incubación del material sobrenadante con unas soluciones 2 mM de p-nitrofenil-β-D-lactósido o con unas soluciones 200 μM de 4-metil-umbeliferil-β-D-lactósido en un tampón de acetato de sodio 50 mM (de pH 5) durante 1 hora. La reacción se interrumpió por adición de unos volúmenes iguales de una solución de carbonato de sodio 1 M y por determinación del p-nitrofenol liberado o de la 4-metil-umbeliferona liberada por medición de la absorbancia a 405 nm o de la fluorescencia a 360 nm/450 nm de excitación/emisión.

Ejemplo comparativo 3: Integración en el genoma y expresión de la secuencia de fusión de CBHI *Talaromyces emersonii* y de CBHII-CBD de *T. reesei* en *Pichia pastoris*

El fragmento de ADN del gen de fusión es generado mediante una PCR por extensión de solapamiento de dos etapas utilizando unos pares de oligonucleótidos y unos moldes sintéticos tal como se ha indicado en la tabla (del Ejemplo 2). Un fragmento de longitud plena tratado con la ADN polimerasa de T4 fue reanillado con el fragmento lineal del vector pV3 reduciendo lentamente la temperatura desde 75°C hasta 4°C. El plásmido pV3 contiene una fusión del péptido de señal del factor de apareamiento alfa con un sitio de clonación múltiple, situado corriente abajo del promotor AOX1 de *Pichia pastoris*. La transformación de la solución reanillada en unas células de *E. coli* químicamente competentes proporciona unos transformantes, que son seleccionados por medio de su resistencia frente a la teocina, comprobados en cuanto a si contienen el plásmido de construcción esperado mediante un análisis por restricción y un secuenciamiento. Unos preparados del plásmido pV3-CBH-a son linealizados con SmaI y aproximadamente 1 μg de unos fragmentos de ADN lineal son transformados en unas células electrocompetentes de *Pichia pastoris*. 94 transformantes procedentes de unas placas de YPD-zeocina son comprobados posteriormente en cuanto a la expresión por cultivación en unos cultivos en placas de 96 pocillos profundos con una capacidad de 500 μl en un medio BMMY que contiene 1 % de metanol y se alimentó 0,5 % de metanol cada 24 h durante 5 días (a 350 rpm/27°C; en un agitador orbital humidificado con una amplitud de 2,5 cm. Los materiales sobrenadantes son ensayados en cuanto a la actividad en 4-MUL, y los clones con los niveles más altos de expresión son seleccionados y evaluados de nuevo en las mismas condiciones.

Para la fermentación en un biorreactor Infors Multifors se selecciona la cepa que produce la concentración más alta de la enzima. Un precultivo en YPD-zeocina (100 g/l) se escoge para realizar la inoculación de un medio Mineral, que se compone de un tampón de fosfato, sulfato y cloruro de magnesio, ciertos elementos traza/biotina y glicerol, con una calibración del valor del pH mediando uso de amoníaco y ácido fosfórico. Después de realizarse el metabolismo de la tanda se mantiene una alimentación de glicerol (2 %) y de glicerol adicional durante 1 día antes de que la alimentación sea cambiada a metanol para desplazarse a unas condiciones inductivas para el promotor de AOX1. En estas condiciones se mantiene la fermentación durante 5 días. Las células son separadas con respecto del líquido de fermentación mediante una centrifugación a 5000xg durante 30 minutos. Los materiales sobrenadantes son analizados en cuanto a la cantidad total de proteínas mediando utilización del reactivo de Bradford y de unos patrones de BSA (de Biorad). Se usa una SDS-PAGE / con una tinción con Coomassie Brilliant Blue para analizar el patrón de proteínas en la SDS-PAGE.

Ejemplo 2: Construcción artificial del vector de expresión en *Trichoderma reesei*

El ADN plasmídico pSCMB100 digerido con SbfI/SwaI fue transformado en el seno de *Trichoderma reesei* SCF41 esencialmente tal como ha sido descrito por Penttilä y colaboradores, 1997. Se usaron 10 μg de un ADN lineal para la transformación de 10⁷ protoplastos. La selección de los transformantes se realizó mediante un crecimiento de los protoplastos en unas placas con un medio Andreotti de Mandel con un agar de recubrimiento, que contiene higromicina como agente de selección (100 mg/l). Los transformantes fueron purificados posteriormente por paso sobre unas placas de medio de esporulación y por reselección de las esporas sobre un medio con higromicina. El ADN genómico de los micelios que habían rebrotado fue aislado y el suceso de reemplazo se verificó mediante una PCR. Los transformantes, que se habían verificado por ser unas cepas de reemplazo genuinas, fueron ensayados posteriormente en cuanto a una segregación de la proteína recombinante.

Ejemplo 3: Expresión de la fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (CBH-ah) a partir de *Trichoderma reesei*

La expresión de unas cepas de reemplazo de CBHI recombinante de la fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* con una marcación 6x His en *Trichoderma reesei* Q6A (ATCC 13631) se llevó a cabo en unos cultivos en matraces con sacudimiento, que contenían 40 ml de un medio Mineral, el cual contenía por su parte 2 % de Avicel, en unos matraces con una capacidad de 300 ml y mediante una cultivación a 30°C/250 rpm durante 6 días. Los materiales sobrenadantes se recuperaron por centrifugación y se analizaron posteriormente mediante una SDS-PAGE y unos ensayos de Proteínas Bradford.

Ejemplo 4: Escrutinio de unas variantes con una estabilidad térmica

5 Unas bibliotecas de mutagénesis aleatoria del gen de la fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (con una marcación 6x His) se crearon mediando utilización de una PCR propensa a errores, usando unos tampones que contienen manganeso y unas concentraciones desequilibradas de dNTPs en la mezcla de reacción de la ADN polimerasa de Taq, se usaron para la amplificación por PCR, esencialmente tal como se ha descrito por Craig y Joyce (R. Craig Cadwell y G.F. Joyce, 1995. Mutagenic PCR, in PCR Primer: a laboratory manual [PCR mutagénica en cebadores de PCR un manual de laboratorio], coordinadores de edición C. W. Dieffenbach y G. S. Dveksler, Cold Spring Harbor Press, Cold Spring Harbor, ME, 583-589). Como un molde se escogió el gen de fusión de tipo silvestre (SEQ ID NO: 17) o unos mutantes de éste. Los fragmentos de PCR mutados se clonaron con el plásmido pKGM_e usando las endonucleasas SphI y HindIII y la ADN ligasa de T4.

15 Unas bibliotecas de las variantes del gen de la fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (con una marcación 6x His) fueron distribuidas en unas placas de 1.536 pocillos con un número de ocupación de los pocillos próximo a 1. La enzima fue expresada durante 7 días en un volumen de 4 µl del medio YPG-G418. Para la evaluación de las propiedades de las variantes, 2 µl de unas muestras de los materiales sobrenadantes del cultivo se transfirieron a unas placas, que contienen una suspensión de paja molida, un tampón de acetato y una beta-glucosidasa. Después de una incubación de las placas de reacción selladas durante 48 horas a unas temperaturas definidas, se determinó la concentración de glucosa usando el rojo Amplex en presencia de GOX y HRP analizando el nivel de fluorescencia. Las muestras (Hits) con los mejores resultados fueron recultivadas y evaluadas de nuevo. Los plásmidos con unas variantes confirmadas de CBH-ah fueron recuperados (con el estuche Pierce ADNzol Yeast genomic ADN Kit) y fueron secuenciados usando unos oligonucleótidos alfa-f (5' TACTATTGCCAGCATTGCTGC-3') y oli740 (5'-TCAGCTATTTACATACAAATCG-3').

25 Ejemplo 5: Determinación de la capacidad de conversión del sustrato a diferentes temperaturas para la indicación de la estabilidad térmica de las variantes de CBH-ah mediando utilización del 4-metil-umbeliferil-β-D-lactósido (4-MUL)

30 Para la comparación exacta de la estabilidad térmica, los materiales sobrenadantes del cultivo, que contenían las variantes de celobiohidrolasa segregadas, fueron diluidos diez veces en un tampón de acetato de sodio (50 mM; pH 5) y unas muestras de 10 µl se incubaron con 90 µl de 4-MUL 200 µM (en el tampón) en el gradiente de temperaturas de un Eppendorff Gradient Thermocycler. Un gradiente de temperaturas de 20°C que se extendía desde 55°C hasta 75°C se aplicó a 12 mezclas de reacción durante una hora para cada muestra. El perfil de temperaturas pudo ser registrado después de la adición de 100 µl de una solución de carbonato de sodio 1 M a cada reacción y la medición de la intensidad de fluorescencia a 360 nm/454 nm en un lector de placas Tecan Infinite M200. Para la comparación de la estabilidad térmica, los valores fueron normalizados entre 1 y 0 para los cómputos de fluorescencia máxima y mínima (Figura 7).

40 Tabla 2: Lista de los mutantes de la SEQ ID NO: 18 con unos valores IT50 mejorados

	Mutaciones con respecto a la SEQ ID NO: 18	IT50 [°C]
	Ninguna	59,8
1	N194R,Y374D	59,8
2	S467T	60,1
3	G4C,A72C,Q349K,S24N	60,1
4	T392M	60,2
5	W40R,D320V,T393A,N445D	60,2
6	W40R	60,3
7	R446G	60,6
8	D346G,R453G	60,9
9	Y496F	60,9
10	T3351,D346A,T393A,D410G	61,0
11	T2431,T325A,V482A	61,0
12	N194R,T243R,Y374D,A375D	61,1
13	E65K	61,1
14	T48A	61,2

ES 2 549 328 T3

	Mutaciones con respecto a la SEQ ID NO: 18	IT50 [°C]
15	G4C,A72C,T344M	61,2
16	T243R,A375D	61,3
17	W40R,K159E,N445D,A501S	61,3
18	T344M	61,3
19	W40R,M234K	61,3
20	Q349R,T393A,P436S,N445D	61,4
21	Q349R,A354T,D373E,N445D	61,4
22	G4C,W40R,A72C,T344M	61,4
23	G4C,A72C,N194L,T243Y,Q349R,Y374R,A375L	61,5
24	G4C,D64N,A72C,Q349K	61,5
25	W40R,T344M	61,5
26	W40R,D346A,T393A	61,6
27	N158D,G486S,Y495C	61,6
28	D320E	61,6
29	A72V	61,6
30	E183V	61,6
31	Q349R,N445D	61,6
32	W40R,C489R	61,7
33	W40R,Q349K	61,9
34	G4C,A72C,Q349K,E65V,Q349R	61,9
35	S311N	62,0
36	D320V,D346A,Q349R,T393A,N445D	62,1
37	T335I	62,2
38	A72V,D346A	62,2
39	D320V,D346A,T393A,N445D	62,2
40	G4C,A72C,N194S,T243W,Q349R,Y374S,A375G	62,3
41	E65V,Q349R	62,3
42	A72V,T335I,D346A,N445D	62,3
43	G4C,A72C,N194L,T243,Q,Q349R,Y374P,A375V	62,3
44	G4C,A72C,D346G	62,3
45	W40R,T344M,Q349K	62,3
46	G4C,Y163C	62,4
47	G4C,V32G,N49S,A72C,S193L,Q349R	62,4
48	G4C,A72C,D346G,Q349R	62,6
49	W40R,D320V,Q349K,P436S,N445D	62,7
50	G4G,A72C,D181N,Q349R	62,7
51	G4C,A72C,Q349K	62,7
52	A72V,Q349R,N445D	62,8
53	G4C,A72C,T344M,Q349K	62,8
54	G4C,A72C,D320V,Q349K	62,8
55	G4C,A72C,P224L,F306Y,Q349R	62,9
56	A72V,T335I,D346A,T393A,N445D	62,9

ES 2 549 328 T3

	Mutaciones con respecto a la SEQ ID NO: 18	IT50 [°C]
57	G4C,A72C,T344M,D346G,Q349R	63,0
58	A72V,D346A,T393A	63,0
59	G4C,A72C,Q349R,R446S,T456A	63,0
60	G4C,W40R,A72C,T344M,Q349K	63,0
61	A72V,D320V,D346A	63,1
62	G4C,A72C,N194Y,T243L,Q349R,Y374S,A375R	63,1
63	G4C,A72C,Q349K,T448A,T449A	63,2
64	G4C,E65V,A72C,Y244H,Q349R	63,3
65	G4C,A72C	63,3
66	G4C,A72C,P224L,Q349R,S409L,P429S	63,4
67	G4C,A72C,D202G,D320N,Q349R,A358E	63,4
68	G4C,A72C,D320V,Q349R,H508R	63,4
69	G4C,A72C,Q349K,S86T	63,4
70	A72V,T335I,D346A,T393A,P436S	63,4
71	G4C,A72C,E183V,K304R,Q349K	63,5
72	G4C,A72C,T243G,G3349R,Y374P,A375M	63,5
73	G4C,A72C,Q349R,T465I	63,6
74	G4C,A72C,N194V,T243M,Q349R,Y374A,A375T	63,6
75	G4C,D64N,A72C.Q349R,A358E,P464Q	63,6
76	G4C,A72C,Q349K,Q28R,S193T,Q490L	63,6
77	G4C,A72C,E183K,Q349K	63,6
78	G4C,A72C,S311N,Q349K,A455T,H509Q	63,6
79	G4C,A72C,N194K,Q349R,Y374P,A375Q	63,6
80	G4C,A72C,D181N,Q349K	63,6
81	W40R,D320V,Q349K,T393A,N445D	63,7
82	W40R,T335I,D346A,T393A	63,7
83	G4C,A72C,N194K,T243P,Q349R,Y374H,A375E	63,7
84	G4C,A72V,Q349R,P462del	63,8
85	G4C,A72C,S236Y,Q349R	63,8
86	G4C,A72C,S311G,Q349K	63,8
87	A72V,D320V,T335I,D346A,T393A,N445D	63,8
88	G4C,A72C,S86T,M234V,Q349K	63,8
89	G4C,A72C,G251R,Q349R	63,9
90	QIL,G4C,A68T,A72C,Q349K,H505N	63,9
91	A72V,T335I,D346A,T393A	63,9
92	G4C,A72C,E183K,Q349R	63,9
93	G4C,A72C,Q349R	63,9
94	Q1L,G4C,A72C,H203R,Q349K,P442S	63,9
95	G4C,A72C,Q349K,G434S,G470D	64,0
96	G4C,W40R,A72C,Q349K	64,0
97	G4C,A72C,Q349R,V367A	64,0
98	G4C,A72C,S311G,D320V,Q349K	64,0

ES 2 549 328 T3

	Mutaciones con respecto a la SEQ ID NO: 18	IT50 [°C]
99	W40R,T335I,D346A,T393A,P436S	64,0
100	A72V,D346A,T393A,N445D	64,1
101	Q1L,G4C,A72C,D202N,Q349K	64,1
102	G4C,A72C,D320V,Q349R	64,1
103	Q1L,G4C,A72C,K154R,G1349K,T393I	64,1
104	G4C,A72C,N194G,T243F,Q349R,Y374P,A375R	64,1
105	A72V,D320V,D346A,T393A,N445D	64,1
106	G4C,A72C,E183V,Q349K	64,1
107	G4C,A72C,E183K,N318Y,Q349K	64,1
108	G4C,A72C,K92R,Q349K,N493D	64,1
109	Q1L,G4C,A72C,Q349K	64,2
110	M234I,G438del	64,2
111	G4C,A72C,Q349R,G459D	64,2
112	G4C,A72C,Q349R,Y422F	64,2
113	G4C,T48A,A72C,Q349R,P480S	64,2
114	E187K,D320V,P442del	64,2
115	G4C,S24N,E65K,A72C,Q349R,I430L,G439D	64,2
116	A72V,D320V,T335I,D346A,T393A,P436S	64,3
117	Q1L,G4C,A72C,S193P,Q349K,V482I	64,3
118	G4C,A72C,D320V,Q349K,G443D,L492Q	64,3
119	G4C,A72C,Q349K,D320V	64,4
120	G4C,A72C,N194K,T243Y,Q349R,A375N	64,4
121	Q1L,G4C,Q28R,A72C,Q349K,H468L	64,5
122	G4C,E65V,A72C,Q349R	64,5
123	D320V,Q349K	64,5
124	G4C,A72C,D181N,P224L,Q349K	64,5
125	G4C,A72C,T243Q,Q349R,Y374P,A375M	64,5
126	Q1L,G4C,A72C,D320V,Q349R	64,5
127	Q1L,G4C,T15S,A72C,Y244F,Q349K	64,5
128	G4C,A72C,E183K,D346E,Q349R	64,6
129	Q1L,G4C,A72C,Q349K,T392M	64,6
130	G4C,A72C,D202N,S311N,Q349R,N493D	64,6
131	G4C,A72C,N194D,T243A,Q349R,Y374P,A375Y	64,6
132	G4C,A72C,N194Y,T243V,Q349R,Y374P	64,7
133	Q1L,G4C,A72C,Q349R	64,7
134	G4C,A72C,D202N,Q349R,H507Y	64,7
135	G4C,A72C,P224L,Q349R,A503V	64,7
136	N194C,Y374C	64,7
137	Q1L,G4C,A72C,D181N,Q349R	64,8
138	G4C,A72C,D320V,D346V,Q349K	64,8
139	Q1L,G4C,A72C,V152A,Q349K	64,8
140	G4C,A72C,E183K,D202Y,N310D,Q349R	64,8

ES 2 549 328 T3

	Mutaciones con respecto a la SEQ ID NO: 18	IT50 [°C]
141	Q1L,G4C,Q28R,A72C,Q349K	64,8
142	Q1L,G4C,A72C,Q349K,Y422F	64,8
143	G4C,A72C,D202V,D320V,Q349K,H504Y	64,8
144	Q1L,G4C,A72C,D181N,D247N,Q349K,S502T	64,8
145	Q1L,G4C,A68T,A72C,Q349K,G439D,R453S	64,9
146	Q1L,G4C,A72C,Q349K,H468R	64,9
147	Q1L,G4C,D64N,A72C,G1349K	64,9
148	G4C,A72C,E183K,Q349R,P464L	64,9
149	G4C,A72C,N1941,T243Y,Q349R,Y374P,A375R	65,0
150	Q1L,G4C,A72C,Q349K,H508L	65,0
151	Q1L,G4C,A72C,E183K,Q349R	65,0
152	G4C,A72C,S311G,Q349R	65,0
153	Q1L,G4C,A72C,S311N,Q349K,G463D	65,1
154	Q1L,G4C,A72C,S86T,Q349R	65,1
155	Q1L,G4C,E65K,A72C,Q349K	65,2
156	Q1L,G4C,A72C,D181N,Q349K,T451S	65,2
157	Q1L,G4C,A6V,A72C,Q349K	65,3
158	G4C,A72C,N1941,T243D,Q349R,Y374P,A375Y	65,3
159	G4C,A72C,N194C,Y374C	65,6
160	Q1L,G4C,A72C,A145T,H203R,Q349K,T403K	65,8
161	Q1L,G4C,A72C,S311G,Q349K	66,0
162	Q1L,G4C,A72C,D202N,Q349K,G486D	66,0
163	Q1L,G4C,A72C,I200F,Q349K,L500I	66,0
164	Q1L,G4C,A68T,A72C,Q349K	66,1
165	Q1L,G4C,A72C,D346V,Q349K	66,2
166	G4C,A72C,N194Q,T243V,Q349R,Y374P,A375Y	66,3
167	Q1L,G4C,A72C,Q349K,T392K	66,3
168	Q1L,G4C,A72C,D181N,E183K,Q349R	66,4
169	Q1L,G4C,A72C,S311N,Q349K	66,5
170	G4C,A72C,N194C,Q349R,Y374C	66,5
171	Q1L,G4C,A72C,Y155C,Q349K,A503T	67,0
172	Q1L,G4C,A72C,S86T,D181N,E183K,D320V,Q349R	67,0

Ejemplo 6: Caracterización de las variantes de la fusión de CBH1 de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (con una marcación 6x His)

5 Se concentraron 80 ml del caldo de fermentación hasta un volumen final de aprox. 1 ml. Después de la determinación de la concentración de proteínas (con el reactivo de Bradford, Biorad, Alemania, el patrón es BSA de Sigma-Aldrich, Alemania), 1,2 mg de la proteína fueron purificados con el estuche Ni-NTA Spin (de Qiagen, Alemania). La fracción purificada de CBH1 se ensayó a continuación llevando a cabo una reacción de hidrólisis con una paja de trigo que había sido tratada previamente (tratamiento previo con un ácido). 12,5 mg (masa seca) de la paja de trigo tratada previamente se mezclan con 0,0125 mg de CBH1 purificada y con 40CBU Novo188 (de Novozymes Dinamarca) por cada mg de CBH1. Se añade acetato de sodio 50 mM (de Sigma-Aldrich, Alemania) hasta llegar a un volumen de 500 µl. El ensayo se lleva a cabo a unas temperaturas que se extienden desde 50°C hasta 65°C durante 48 horas y se analiza mediante una HPLC para determinar el contenido de glucosa dependiente de la temperatura.

10

LISTA DE SECUENCIAS

<110> Süd-Chemie AG
 <120> enzimas de celulosa optimizadas
 <130> 139 202
 <160> 41

5

<170> PatentIn **versión 3.4**

<210> 1
 <211> 1509
 <212> ADN
 < 213> artificial

10

<220>
 <223> Secuencia codificadora para la fusión de CBHI de *Talaromyces emersonii* / -CBD de *Trichoderma reesei* (CBH-a madura)

15

<400> 1
 cagcaggccg gcacggcgac ggacagagaac caccgcgccc tgacatggca ggaatgcacc 60
 gcccctggga gctgcaccac ccagaacggg gcggtcgttc ttgatgcgaa ctggcgttgg 120
 gtgcacgatg tgaacggata caccaactgc tacacgggca atacctggga ccccacgtac 180
 tgccctgacg acgaaacctg cgcccagaac tgtgcgctgg acggcgcgga ttacgagggc 240
 acctacggcg tgacttcgtc gggcagctcc ttgaaactca atttcgtcac cgggtcgaac 300
 gtcggatccc gtctctacct gctgcaggac gactcgacct atcagatctt caagctcctg 360
 aaccgcgagt tcagctttga cgtcgatgtc tccaatcttc cgtgctggatt gaacggcgct 420
 ctgtactttg tcgcatgga cgccgacggc ggcgtgtcca agtaccgaa caacaaggct 480
 ggtgccaaagt acggaaccgg gtattgagac tccaatgcc cacgggacct caagttcatc 540
 gacggcgagg ccaacgtcga gggctggcag ccgtcttcga acaacgcaa caccggaatt 600
 ggcgaccacg gctctgctg tgcggagatg gatgtctggg aagcaaacag catctccaat 660
 gcggtcactc cgcacccgtg cgacacgcca ggcagacga tgtgctctgg agatgactgc 720
 ggtggcacat actctaacga tcgctacgcg ggaacctgcg atcctgacgg ctgtgacttc 780
 aacccttacc gcatgggcaa cacttctttc tacgggacct gcaagatcat cgataccacc 840
 aagcccttca ctgtcgtgac gcagttcctc actgatgatg gtacggatac tggaactctc 900
 agcgagatca agcgttctta catccagaac agcaacgtca ttccgcagcc caactcggac 960
 atcagtggcg tgaccggcaa ctcgatcacg acggagttct gcaactgctca gaagcaggcc 1020
 tttggcgaca cggacgactt ctctcagcac ggtggcctgg ccaagatggg agcggccatg 1080
 cagcagggta tggctctggt gatgagtttg tgggacgact acgcccgcga gatgctgtgg 1140
 ttggattccg actaccggac ggatgcggac cccacgacct ctggatttgc ccgtggaacg 1200
 tgtccgacgg actcgggctt cccatcggat gtcgagtcgc agagccccc aa ctctacgtg 1260
 acctactcga acattaagtt tggctcgatc ggtagcacag gtaatccttc aggtggtaat 1320
 cctccagggtg gaaacagagg aacaacgaca actagaagac cagctactac aactggttca 1380
 agtccaggtc caactcaatc aactacggg caatgtgggt gtataggta ctctggctcc 1440
 actgtttgtg cttctggtag tacttgccaa gttctgaacc cttactactc acagtgtcta 1500
 taatgataa 1509

ES 2 549 328 T3

<210> 2
 <211> 500
 <212> PRT
 <213> artificial

5 <220>
 <223> Secuencia madura de CBHI de Talaromyces emersonii / -CBD de Trichoderma reesei (CBH-a madura)

<400> 2
 Gln Gln Ala Gly Thr Ala Thr Ala Glu Asn His Pro Pro Leu Thr Trp
 1 5 10 15
 Gln Glu Cys Thr Ala Pro Gly Ser Cys Thr Thr Gln Asn Gly Ala Val
 20 25 30
 Val Leu Asp Ala Asn Trp Arg Trp Val His Asp Val Asn Gly Tyr Thr
 35 40 45
 Asn Cys Tyr Thr Gly Asn Thr Trp Asp Pro Thr Tyr Cys Pro Asp Asp
 50 55 60
 Glu Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr Glu Gly
 65 70 75 80
 Thr Tyr Gly Val Thr Ser Ser Gly Ser Ser Leu Lys Leu Asn Phe Val
 85 90 95
 Thr Gly Ser Asn Val Gly Ser Arg Leu Tyr Leu Leu Gln Asp Asp Ser
 100 105 110
 Thr Tyr Gln Ile Phe Lys Leu Leu Asn Arg Glu Phe Ser Phe Asp Val
 115 120 125
 Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala Leu Tyr Phe Val
 130 135 140
 Ala Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Asn Asn Lys Ala
 145 150 155 160
 Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys Pro Arg Asp
 165 170 175
 Leu Lys Phe Ile Asp Gly Glu Ala Asn Val Glu Gly Trp Gln Pro Ser
 180 185 190
 Ser Asn Asn Ala Asn Thr Gly Ile Gly Asp His Gly Ser Cys Cys Ala
 195 200 205

ES 2 549 328 T3

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

ES 2 549 328 T3

<210> 3
 <211> 1581
 <212> ADN
 <213> artificial

5 <220>
 <223> secuencia codificadora de la fusión de CBH-a con el péptido de señal de CBH de Trichoderma reesei

<400> 3
 atgtatcggga agttggccgt catctcgccc ttcttggcca cagctcgtgc tcagcaggcc 60
 ggcacggcga cggcagagaa ccaccgccc ctgacatggc aggaatgcac cgccccctggg 120
 agctgcacca cccagaacgg ggcggtcgtt cttgatgcga actggcgttg ggtgcacgat 180
 gtgaacggat acaccaactg ctacacgggc aatacctggg accccacgta ctgccctgac 240
 gacgaaacct ggcgccagaa ctgtgcgctg gacggcgcgg attacgaggg cacctacggc 300
 gtgacttcgt cgggcagctc cttgaaactc aatttcgtca ccgggtcga cgtcggatcc 360
 cgtctctacc tgctgcagga cgactcgacc tatcagatct tcaagctcct gaaccgcgag 420
 ttcagctttg acgtcgatgt ctccaatctt ccgtgcggat tgaacggcgc tctgtacttt 480
 gtcgccatgg acgccagcg cggcgtgtcc aagtaccga acaacaaggc tggtgccaag 540
 tacggaaccg ggtattgcga ctccaatgc ccacgggacc tcaagttcat cgacggcag 600
 gccaacgtcg agggctggca gccgtcttcg aacaacgcca acaccggaat tggcgaccac 660
 ggctcctgct gtgcggagat ggatgtctgg gaagcaaaca gcatctcaa tgcggctact 720
 ccgcaccctg gcgacacgcc agccagacg atgtgctctg gagatgactg cggtggcaca 780
 tactctaacg atcgtactcg ggaacctgc gatcctgacg gctgtgactt caacccttac 840
 cgcattggca acacttcttt ctacggcctt ggcaagatca tcgataccac caagcccttc 900
 actgtcgtga cgcagttcct cactgatgat ggtacggata ctggaactct cagcgagatc 960
 aagcgttct acatccagaa cagcaacgtc attccgcagc ccaactcgga catcagtggc 1020
 gtgaccggca actcgatcac gacggagttc tgcactgctc agaagcaggc ctttggcgac 1080
 acggacgact tctctcagca cggtgccctg gcccaagatgg gagcggccat gcagcagggt 1140
 atggtcctgg tgatgagttt gtgggacgac tacgccgcgc agatgctgtg gttggatcc 1200
 gactaccga cggatgcgga cccacgacc cctggtattg cccgtggaac gtgtccgacg 1260
 gactcgggcg tcccacgga tgtcgagtgc cagagcccca actcctacgt gacctactcg 1320
 aacattaagt ttggtccgat cggtagcaca ggtaatcctt cagggtgtaa tcctccagg 1380
 ggaaacagag gaacaacgac aactagaaga ccagctacta caactggttc aagtccagg 1440
 ccaactcaat cacactacgg tcaatgtggt ggtataggtt actctggtcc cactgtttgt 1500
 gcttctggta ctactgcca agttctgaac cttactact cacagtgtct agcttctgca 1560
 catcatcacc accaccatta a 1581

10

<210> 4
 <211> 70
 <212> PRT
 <213> artificial

15 <220>
 <223> dominio de fijación de celulosa y secuencia engarzadora de CBH1 de Trichoderma reesei

ES 2 549 328 T3

<400> 4

Gly Ser Thr Gly Asn Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Arg
1 5 10 15

Gly Thr Thr Thr Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro
20 25 30

Gly Pro Thr Gln Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser
35 40 45

Gly Pro Thr Val Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro
50 55 60

Tyr Tyr Ser Gln Cys Leu
65 70

<210> 5

<211> 437

5 <212> PRT

<213> artificial

<220>

<223> secuencia de CBHI de Talaromyces emersonii (CBH-b)

<400> 5

Gln Gln Ala Gly Thr Ala Thr Ala Glu Asn His Pro Pro Leu Thr Trp
1 5 10 15

Gln Glu Cys Thr Ala Pro Gly Ser Cys Thr Thr Gln Asn Gly Ala Val
20 25 30

Val Leu Asp Ala Asn Trp Arg Trp Val His Asp Val Asn Gly Tyr Thr
35 40 45

Asn Cys Tyr Thr Gly Asn Thr Trp Asp Pro Thr Tyr Cys Pro Asp Asp
50 55 60

Glu Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr Glu Gly
65 70 75 80

Thr Tyr Gly Val Thr Ser Ser Gly Ser Ser Leu Lys Leu Asn Phe Val
85 90 95

10 Thr Gly Ser Asn Val Gly Ser Arg Leu Tyr Leu Leu Gln Asp Asp Ser

ES 2 549 328 T3

100 105 110
 Thr Tyr Gln Ile Phe Lys Leu Leu Asn Arg Glu Phe Ser Phe Asp Val
 115 120 125
 Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala Leu Tyr Phe Val
 130 135 140
 Ala Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Asn Asn Lys Ala
 145 150 155 160
 Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys Pro Arg Asp
 165 170 175
 Leu Lys Phe Ile Asp Gly Glu Ala Asn Val Glu Gly Trp Gln Pro Ser
 180 185 190
 Ser Asn Asn Ala Asn Thr Gly Ile Gly Asp His Gly Ser Cys Cys Ala
 195 200 205
 Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser Asn Ala Val Thr Pro
 210 215 220
 His Pro Cys Asp Thr Pro Gly Gln Thr Met Cys Ser Gly Asp Asp Cys
 225 230 235 240
 Gly Gly Thr Tyr Ser Asn Asp Arg Tyr Ala Gly Thr Cys Asp Pro Asp
 245 250 255
 Gly Cys Asp Phe Asn Pro Tyr Arg Met Gly Asn Thr Ser Phe Tyr Gly
 260 265 270
 Pro Gly Lys Ile Ile Asp Thr Thr Lys Pro Phe Thr Val Val Thr Gln
 275 280 285
 Phe Leu Thr Asp Asp Gly Thr Asp Thr Gly Thr Leu Ser Glu Ile Lys
 290 295 300
 Arg Phe Tyr Ile Gln Asn Ser Asn Val Ile Pro Gln Pro Asn Ser Asp
 305 310 315 320
 Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu Phe Cys Thr Ala
 325 330 335
 Gln Lys Gln Ala Phe Gly Asp Thr Asp Asp Phe Ser Gln His Gly Gly
 340 345 350
 Leu Ala Lys Met Gly Ala Ala Met Gln Gln Gly Met Val Leu Val Met
 355 360 365
 Ser Leu Trp Asp Asp Tyr Ala Ala Gln Met Leu Trp Leu Asp Ser Asp
 370 375 380

ES 2 549 328 T3

Tyr Pro Thr Asp Ala Asp Pro Thr Thr Pro Gly Ile Ala Arg Gly Thr
385 390 400

Cys Pro Thr Asp Ser Gly Val Pro Ser Asp Val Glu Ser Gln Ser Pro
405 410 415

Asn Ser Tyr Val Thr Tyr Ser Asn Ile Lys Phe Gly Pro Ile Asn Ser
420 425 430

Thr Phe Thr Ala Ser
435

<210> 6

<211> 1590

<212> ADN

5 <213> artificial

<220>

<223> secuencia codificadora de CBHI de Talaromyces emersonii CBHI fusionada con el péptido de señal del factor alfa

<400> 6

atgagatttc	cttcaat	ttt	tactgcagtt	ttattcgcag	catcctccgc	attagctgct	60
ccagtcaaca	ctacaacaga	agatgaaacg	gcacaaattc	cggctgaagc	tgatcatcgg		120
tacttagatt	tagaagggga	tttcgatggt	gctgttttgc	cattttccaa	cagcacaaat		180
aacgggttat	tgtttataaa	tactactatt	gccagcattg	ctgctaaaga	agaaggggta		240
tctttggata	aacgtgaggc	ggaagcacc	tctcagcagg	ccggcacggc	gacggcagag		300
aaccaccgc	ccctgacatg	gcaggaatgc	accgcccctg	ggagctgcac	caccagaac		360
ggggcggtcg	ttcttgatgc	gaactggcgt	tgggtgcacg	atgtgaacgg	atacaccaac		420
tgctacacgg	gcaatacctg	ggaccccacg	tactgccctg	acgacgaaac	ctgcgccag		480
aactgtgctc	tggacggcgc	ggattacgag	ggcacctacg	gcgtgacttc	gtcgggcagc		540
tccttgaaac	tcaatttcgt	caccgggtcg	aacgtcggat	cccgtctcta	cctgctgcag		600
gacgactcga	cctatcagat	cttcaagctt	ctgaaccgcg	agttcagctt	tgacgtcgat		660
gtctccaatc	ttcctgctgg	attgaacggc	gctctgtact	ttgtcgccat	ggacgccgac		720
ggcggcgtgt	ccaagtaccc	gaacaacaag	gctggtgcca	agtacggaac	cgggtattgc		780
gactccaat	gcccacggga	cctcaagtcc	atcgacggcg	aggccaacgt	cgagggctgg		840
cagccgtctt	cgaacaacgc	caacaccgga	attggcgacc	acggctcctg	ctgtgctggag		900
atggatgtct	gggaagcaaa	cagcatctcc	aatgcggtca	ctccgcaccc	gtgcgacacg		960
ccagccaga	cgatgtgctc	tggagatgac	tgcggtggca	catactctaa	cgatcgctac		1020
gctggaaact	gcgatcctga	cggctgtgac	ttcaaccctt	accgatggg	caacacttct		1080
ttctacgggc	ctggcaagat	catcgatacc	accaagccct	tactgtcgt	gacgcagttc		1140
ctcactgatg	atggtacgga	tactggaact	ctcagcgaga	tcaagcgctt	ctacatccag		1200
10 aacagcaacg	tcattccgca	gcccactcgc	gacatcagtg	gcgtgaccgg	caactcgatc		1260

ES 2 549 328 T3

acgacggagt tctgcactgc tcagaagcag gcctttggcg acacggacga cttctctcag 1320
 cacggtggcc tggccaagat gggagcggcc atgcagcagg gtatggctct ggtgatgagt 1380
 ttgtgggacg actacgccgc gcagatgctg tggttggatt ccgactaccc gacggatgcg 1440
 gacccacga cccctggtat tgcccgtgga acgtgtccga cggactcggg cgtcccatcg 1500
 gatgtcgagt cgcagagccc caactcctac gtgacctact cgaacattaa gtttggtccg 1560
 atcaactcga ccttcaccgc ttcgtgataa 1590

<210> 7
 <211> 429
 <212> PRT
 <213> artificial

5

<220>
 <223> CBHI de Humicola grisea (CBH-d)

<400> 7
 Gln Gln Ala Gly Thr Ile Thr Ala Glu Asn His Pro Arg Met Thr Trp
 1 5 10 15
 Lys Arg Cys Ser Gly Pro Gly Asn Cys Gln Thr Val Gln Gly Glu Val
 20 25 30
 Val Ile Asp Ala Asn Trp Arg Trp Leu His Asn Asn Gly Gln Asn Cys
 35 40 45
 Tyr Glu Gly Asn Lys Trp Thr Ser Gln Cys Ser Ser Ala Thr Asp Cys
 50 55 60
 Ala Gln Arg Cys Ala Leu Asp Gly Ala Asn Tyr Gln Ser Thr Tyr Gly
 65 70 75 80
 Ala Ser Thr Ser Gly Asp Ser Leu Thr Leu Lys Phe Val Thr Lys His
 85 90 95
 Glu Tyr Gly Thr Asn Ile Gly Ser Arg Phe Tyr Leu Met Ala Asn Gln
 100 105 110
 Asn Lys Tyr Gln Met Phe Thr Leu Met Asn Asn Glu Phe Ala Phe Asp
 115 120 125
 Val Asp Leu Ser Lys Val Glu Cys Gly Ile Asn Ser Ala Leu Tyr Phe
 130 135 140
 Val Ala Met Glu Glu Asp Gly Gly Met Ala Ser Tyr Pro Ser Asn Arg
 145 150 155 160
 Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala Gln Cys Ala Arg
 165 170 175
 Asp Leu Lys Phe Ile Gly Gly Lys Ala Asn Ile Glu Gly Trp Arg Pro

10

ES 2 549 328 T3

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

<210> 8
 <211> 1563
 <212> ADN
 <213> artificial

5

<220>
 <223> secuencia codificadora de CBHI de Humicola grisea fusionada con el péptido de señal del factor alfa

ES 2 549 328 T3

```

<400> 8
atgagatttc cttcaatfff tactgcagtt ttattcgag catcctccgc attagctgct    60
ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggg    120
tacttagatt tagaagggga tttcgatggt gctgttttgc cattttccaa cagcacaaat    180
aacgggttat tgtttataaa tactactatt gccagcattg ctgctaaaga agaaggggta    240
tctttggata aacgtgaggc ggaagcaccc tctcagcagg ctggtactat tactgctgag    300
aaccacccaa gaatgacctg gaagagatgc tctggtccag gaaactgtca gactgttcag    360
ggcggaggtg tgattgacgc taattggaga tggttgcaca acaacggcca gaactgttac    420
gagggtaaca agtggacctc tcagtgttct tctgctaccg actgtgctca gagatgtgct    480
ttggacggtg ccaactacca gtctacctac ggtgcttcta cctctggtga ctctctgacc    540
ctgaagtctg ttaccaagca cgagtacgga accaacatcg gctctagatt ctacctgatg    600
gccaaccaga acaagtacca gatgttcacc ctgatgaaca acgagttcgc ctttgacggt    660
gacctgtcta aggtggagtg cggtatcaac tctgccctgt acttcggtgc tatggaagag    720
gacggtggaa tggcttctta cccatctaac agagccgggt ctaagtacgg tactggttac    780
tgtgacgccc agtgtgctag agacctgaag ttcacggtg gaaaggccaa cattgagggg    840
tggagaccat ctaccaacga cccaaacgct ggtgttggtc caatgggagc ttgtgtgccc    900
gagattgatg tgtgggagtc taacgcttac gcctacgctt ttaccccaca cgcttgcggt    960
tctaagaaca gataccacat ctgcgagacc aacaactgtg gtggaaccta ctctgacgac   1020
agattcgctg gatactgcga cgctaacggt tgtgactaca acccatacag aatgggcaac   1080
aaggacttct acggcaaggg aaagaccggt gacaccaaca gaaagttcac cgtggtgctg   1140
agattcgaga gaaacagact gtcgcagttc tttgtgcagg acggcagaaa gattgaggtc   1200
ccaccaccaa cttggccagg attgccaaac tctgccgaca ttaccccaga gttgtgtgac   1260
gctcagttca gagtgttcga cgacagaaac agatttgctg agaccggtgg ttttgacgct   1320
ttgaacgagg ctctgacct tccaatggtg ctggtgatgt ctatttggga cgaccaccac   1380
tctaacatgt tgtggctgga ctcttcttac ccaccagaga aggctggatt gccaggtggt   1440
gacagaggac catgtccaac tacttcgggt gttccagctg aggttgaggc tcagtacca   1500
gacgctcagg ttgtgtggtc gaacatcaga ttcggcccaa tcggttctac cgtgaacgtg   1560
taa                                                                    1563

```

```

<210> 9
<211> 439
<212> PRT
<213> artificial

```

5

```

<220>
<223> CBHI de Thermoascus auratiacus (CBH-e)

```

ES 2 549 328 T3

<400> 9

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

ES 2 549 328 T3

aattgtgcct tggatggagc ggattacagt ggcacctatg gtgttacgac cagtggcaac 540
 gccctgagac tgaactttgt cacccaaagc tcagggaga acattggctc gcgcctgtac 600
 ctgctgcagg acgacaccac ttatcagatc ttcaagctgc tgggtcagga gtttaccttc 660
 gatgtcgacg tctccaatct cccttgccgg ctgaacggcg ccctctactt tgtggccatg 720
 gacgccgacg gcaattttgt caaataccct ggcaacaagg caggcgctaa gtatggcact 780
 ggttactgcg actctcagtg ccctcgggat ctcaagttca tcaacggtca ggccaacgtt 840
 gaaggctggc agccgtctgc caacgaccca aatgccggcg ttggtaacca cggttcctcg 900
 tgcgctgaga tggatgtctg ggaagccaac agcatctcta ctgcggtgac gcctcaccca 960
 tgcgacacc ccggccagac catgtgccag ggagacgact gtggtggaac ctactcctcc 1020
 actcgatatg ctggtacctg cgacactgat ggctgcgact tcaatcctta ccagccaggc 1080
 aaccactcgt tctacggccc cgggaagatc gtcgacacta gtcctaaatt caccgtcgtc 1140
 acccagttca tcaccgacga cgggacacc tccggcacc tgacggagat caaacgcttc 1200
 tacgtccaga acggcaaggt gatccccag tcggagtcga cgatcagcgg cgtcaccggc 1260
 aactcaatca ccaccgagta ttgcacggcc cagaaggcag ccttcgacaa caccggcttc 1320
 ttcacgcacg ggggcttca gaagatcagt caggctctgg ctcagggcat ggtcctcgtc 1380
 atgagcctgt gggacgatca cgccgccaac atgctctggc tggacagcac ctaccggact 1440
 gatgcccacc cggacacccc tggcgtcgcg cgcggtacct gccccacgac ctccggcgtc 1500
 ccggccgacg tggagtcgca gaaccccaat tcatatgtta tctactcaa catcaaggtc 1560
 ggacctatca actcgacctt caccgccaac taa 1593

- <210> 11
- <211> 1794
- 5 <212> ADN
- <213> artificial

<220>

<223> secuencia codificadora de CBHI de Trichoderma reesei (CBH-c), incluyendo el péptido de señal del factor alfa y una marcación 6x His

10 <400> 11
 atgagatttc cttcaatttt tactgcagtt ttattcgcag catcctccgc attagctgct 60
 ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggc 120
 tacttagatt tagaagggga ttctgatgtt gctgttttgc cattttccaa cagcacaat 180
 aacgggttat tgtttataaa tactactatt gccagcattg ctgctaaaga agaaggggta 240
 tctttggata aacgtgaggc ggaagcacc tcttcagctt gtacactgca atccgagact 300
 catccacctt taacgtggca aaagtgtagt tctggcggaa cttgtactca acagactggc 360
 agtgtcgtga tagatgctaa ctggagatgg acacatgcaa cgaactcctc aactaactgc 420
 tacgatggta acacctggc ttctacattg tgcctgaca acgaaacctg cgctaagaac 480
 tgttgccttg atggagcagc ttacgcaagt acatattggtg tgactacctc tggtaacagc 540
 ctttccattg gttttgtaac ccagtcggct cagaagaatg ttgggtgctag attgtacctg 600

ES 2 549 328 T3

atggcttcag acaccacata ccaggagttt accttggtgg gaaacgagtt ctctttcgac 660
 gtagatgtgt ctgagctacc atgtggattg aatggagcct tgtactttgt ctcaatggat 720
 gcagacggag gtgtttcaaa gtaccctact aacacagctg gtgctaagta tggaactgga 780
 tactgcgatt ctcaatgccc aagagacctg aagttcatca acggacaagc taacgttgaa 840
 ggttgggaac cttctagcaa caacgcaaac actggaattg gtggtcatgg ttcttgctgt 900
 tcagagatgg acatttggga agccaactcc atcagtgaag ctttgactcc acatccatgc 960
 acaactgttg ggcaagaaat ttgcgaagg gatggttgg gtggcactta ctctgataac 1020
 agatacggcg gaacatgtga tccagatgga tgtgattgga acccatacag actgggtaac 1080
 acttcgtttt acggaccagg ttcttccttc actctagaca ctacgaagaa gttgactgtg 1140
 gtcaccaat ttgagacttc tgggtccatt aaccgatact acgtgcagaa cggagttact 1200
 ttccaacagc caaacgctga attgggtagt tactcaggca acgagcttaa cgatgactac 1260
 tgcactgctg aagaagcaga atttggtgga tcttcctttt cggataaggg tggattgacg 1320
 cagttcaaga aagctacctc tgggtggaatg gttctagtca tgagtctgtg ggacgattac 1380
 tacgctaaca tgctttggct ggactctact tacctacaa acgagacatc ttctactcct 1440
 ggtgctgtaa gaggtagctg ttctacatct tctggagttc cagcccaagt tgagagtcaa 1500
 agtccaaatg ccaaggtcac cttctccaac atcaagttcg gaccaattgg tagcacaggt 1560
 aatccttcag gtggtaatcc tccagggtgga aacagaggaa caacgacaac tagaagacca 1620
 gctactacaa ctggttcaag tccagggtcca actcaatcac actacggtca atgtggtggt 1680
 ataggttact ctggtcccac tgtttgtgct tctggtacta cttgccaagt tctgaaccct 1740
 tactactcac agtgtctagc ttctgcacac catcatcatc atcattaatg ataa 1794

<210> 12
 <211> 496
 5 <212> PRT
 <213> artificial

<220>
 <223>CBHI de Trichoderma reesei (CBH-C)

<400> 12
 Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr Trp
 1 5 10 15
 Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser Val
 20 25 30
 Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser Thr
 35 40 45
 Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp Asn
 50 55 60
 10 Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala Ser
 65 70 75 80

ES 2 549 328 T3

Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe Val
 85 90 95
 Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met Ala
 100 105 110
 Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe Ser
 115 120 125
 Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala Leu
 130 135 140
 Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Thr
 145 150 155 160
 Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys
 165 170 175
 Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly Trp
 180 185 190
 Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly Ser
 195 200 205
 Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu Ala
 210 215 220
 Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu Gly
 225 230 235 240
 Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr Cys
 245 250 255
 Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr Ser
 260 265 270
 Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys Leu
 275 280 285
 Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr Tyr
 290 295 300
 Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly Ser
 305 310 315 320
 Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu Ala
 325 330 335
 Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln Phe
 340 345 350

ES 2 549 328 T3

Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp Asp
 355 360 365
 Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr Asn
 370 375 380
 Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr Ser
 385 390 395 400
 Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys Val
 405 410 415
 Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn Pro
 420 425 430
 Ser Gly Gly Asn Pro Pro Gly Gly Asn Arg Gly Thr Thr Thr Thr Arg
 435 440 445
 Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln Ser His
 450 455 460
 Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys Ala
 465 470 475 480
 Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys Leu
 485 490 495

- <210> 13
- <211> 1767
- <212> ADN
- <213> artificial

5

- <220>
- <223> secuencia codificadora de CBHI de Trichoderma viride, incluyendo el péptido de señal del factor alfa

<400> 13
 atgagatttc cttcaat ttt tactgcagtt ttattcgcag caccctccgc attagctgct 60
 ccagtcaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggt 120
 tacttagatt tagaagggga ttctgatgtt gctgtrttgc cattttccaa cagcacaat 180
 aacgggttat tgttataaa tactactatt gccagcattg ctgctaaaga agaaggggta 240
 tctttggata aacgtgaggg ggaagcacc tctcaatctg cttgcacctt gcagtctgaa 300
 actcaccac cattgacctg gcagaagtgt tcttctggcg gtacttgta cagcagacc 360
 ggttctggtt ttatcgagc caactggaga tggactcacg ctaccaactc ttctaccaac 420
 tgctacgacg gtaacacttg gtcgtctacc ttgtgtccag acaacgagac ctgtgccaag 480
 aactgtgtt tggacggtgc tgcttacgct tctacctacg gtgttaccac ctctggtaac 540
 tcgctgtcta tcggtttcgt taccagctc gccagaaaa atgttgggtc cagactgtac 600
 ttgatggctt ctgacaacc ctaccaagag tttaccctgc tgggtaacga gttctctttc 660

ES 2 549 328 T3

gacgtggacg tttctcaact gccatgtgga ctgaacggtg cctgtactt cgtttctatg 720
gacgctgacg gtggtgtttc taagtacca accaacaccg ctggtgctaa atacggaacc 780
ggttactgcg attctcagtg cccaagagac ctgaagttca tcaacggaca ggctaacggt 840
gaaggatggg agccatcttc taacaacgcc aacaccggta ttggtgggtca cggttcttgc 900
tgttctgaga tggacatctg ggaggccaac tctatttctg aggctttgac cccacaccca 960
tgtactactg tgggtcaaga gatctgtgag ggtgatggtt gtggtgggtac ttactcggac 1020
aacagatacg gtggtacttg tgacccagac ggttgtgatt gggaccata cagactgggt 1080
aacacctctt tctacggtcc aggatcttct tttaccctgg acaccaccaa gaagttgacc 1140
gttgttacc agtttgagac ctctgggtgcc atcaacagat actacgtgca gaacgggtgtt 1200
actttccagc agccaaacgc tgaactggga tcttactctg gtaacggact gaacgacgac 1260
tactgtactg ctgaggaagc tgagttcggg ggttcttctt tctctgacaa ggttggactg 1320
accagttta agaaggctac ctctggcggg atggtgctgg ttatgtcttt gtgggacgac 1380
tactacgcta acatgctgtg gcttgactct acctaccaa ctaacgagac ctcttctacc 1440
ccaggtgctg ttgaggatc ttgctctacc tcttctgggtg ttccagctca ggttgagtct 1500
cagtcctcaa acgccaaggt gaccttctct aacatcaagt tcggtccaat cggttctact 1560
ggtgaccat ctggtggtaa cccaccaggt ggaaaccac ctggtactac cactaccaga 1620
agaccagcta ccaccactgg ttcttctcca ggtccaacc aatctacta cggtcagtgt 1680
ggtggtattg gttactctgg tccaaccgtt tgtgcttctg gaaccacctg tcaggttctg 1740
aaccatact actcgcagtg cctgtaa 1767

<210> 14
<211> 497
5 <212> PRT
<213> artificial

<220>
<223> CBHI de Trichoderma viride (CBH-f)

<400> 14
Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr Trp
1 5 10 15
Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser Val
20 25 30
Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser Thr
35 40 45
Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp Asn
50 55 60
10 Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala Ser
65 70 75 80

ES 2 549 328 T3

Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe Val
 85 90 95
 Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met Ala
 100 105 110
 Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe Ser
 115 120 125
 Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala Leu
 130 135 140
 Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Thr
 145 150 155 160
 Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys
 165 170 175
 Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly Trp
 180 185 190
 Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly Ser
 195 200 205
 Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu Ala
 210 215 220
 Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu Gly
 225 230 235 240
 Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr Cys
 245 250 255
 Asp Pro Asp Gly Cys Asp Trp Asp Pro Tyr Arg Leu Gly Asn Thr Ser
 260 265 270
 Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys Leu
 275 280 285
 Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr Tyr
 290 295 300
 Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly Ser
 305 310 315 320
 Tyr Ser Gly Asn Gly Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu Ala
 325 330 335
 Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln Phe
 340 345 350
 Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp Asp

ES 2 549 328 T3

355 360 365

Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr Asn
 370 375 380

Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr Ser
 385 390 400

Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys Val
 405 410 415

Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asp Pro
 420 425 430

Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr Thr
 435 440 445

Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln Ser
 450 455 460

His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys
 465 470 475 480

Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys
 485 490 495

Leu

- <210> 15
- <211> 1785
- <212> ADN
- <213> artificial

5

<220>

<223> secuencia codificadora de la proteína de fusión de CBHI de Humicola grisea - dominio de fijación de celulosa de CBHI de Trichoderma reesei incluyendo el péptido de señal del factor alfa y una marcación 6x His

<400> 15

atgagatttc cttcaatttt tactgcagtt ttattcgcag catcctccgc attagctgct	60
ccagtcaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgcatcgggt	120
tacttagatt tagaagggga ttccgatggt gctgttttgc cttttccaa cagcacaat	180
aacgggttat tgttataaa tactactatt gccagcattg ctgctaaaga agaaggggta	240
tctttggata aacgtgaggc ggaagcatgc tcgcagcagg ctggtacaat tactgctgag	300
aaccatccaa gaatgacgtg gaagagatgt agtgggccag gaaactgtca gactgttcag	360
ggtagggtcg tgatagatgc taactggaga tggttgcata acaacggcca gaactgctac	420
gagggtaaca agtggacctc tcagtgttct tctgctaccg actgcgctca gagatgtgct	480
cttgatggag caaactacca gagtacatat ggtgcttcta cctctggtga cagccttacc	540
ctgaagtttg taaccaagca cgagtacgga accaatatcg gttctagatt ctacctgatg	600

10

ES 2 549 328 T3

gctaaccaga acaagtacca gatgtttacc ttgatgaaca acgagttcgc cttcgacgta 660
gatctgtcta aggtggagtg tggaatcaat tctgccttgt actttgtcgc tatggaagag 720
gacggaggta tggttctta cccttctaac agagctggtg ctaagtatgg aactggatac 780
tgcgatgcc aatgcgctag agacctgaag ttcacggtg gaaaggctaa cattgaaggt 840
tggagacctt ctaccaacga cccaaacgct ggagttggtc caatgggtgc ttgctgtgcc 900
gagattgacg tgtgggaatc taacgcttac gcctacgctt ttactccaca tgcttgcggt 960
tctaagaaca gataccacat ttgcgaaacc aacaactgtg gtggcactta ctctgatgac 1020
agattcgctg gatactgtga tgctaácgga tgtgattaca acccatacag aatgggtaac 1080
aaggactttt acggaagg taagactgtt gacactaaca gaaagttcac tgtggtctcg 1140
agatttgaga gaaacagact gtcgcagttc tttgtgcagg acggaagaaa gattgaggtc 1200
ccaccaccaa cttggccagg attgccaaac tctgccgaca ttaccccaga gttgtgcgac 1260
gctcagttca gagtgtttga cgacagaaac agatttgctg agaccggtgg atttgacgct 1320
ttgaacgagg ctctgacat tccaatggtt ctagtcatga gtatttggga cgatcaccac 1380
tctaacatgc tttggctgga ctcttcttac cctccagaga aggctggatt gcctggtggt 1440
gacagaggtc catgtccaac aacttctgga gttccagccg aggttgaggc tcaataccca 1500
gacgcccagg tctgttggtc caacatcaga ttcggaccaa ttggaagctt aacaggtaat 1560
ccttcagggtg gtaatcctcc aggtggaaac agaggaacaa cgacaactag aagaccagct 1620
actacaactg gttcaagtcc aggtccaact caatcacact acggtcaatg tgggtgtata 1680
ggttactctg gtcccactgt ttgtgcttct ggtactactt gccaagttct gaacccttac 1740
tactcacagt gcttagcttc tgcacacat catcatcatc attaa 1785

<210> 16
<211> 503
5 <212> PRT
<213> artificial

<220>
<223> proteína de fusión de CBHI de Humicola grisea - dominio de fijación de celulosa de CBHI de Trichoderma reesei incluyendo una marcación 6x His (CBH-g)

10 <400> 16
Gln Gln Ala Gly Thr Ile Thr Ala Glu Asn His Pro Arg Met Thr Trp
1 5 10 15
Lys Arg Cys Ser Gly Pro Gly Asn Cys Gln Thr Val Gln Gly Glu Val
20 25 30
Val Ile Asp Ala Asn Trp Arg Trp Leu His Asn Asn Gly Gln Asn Cys
35 40 45
Tyr Glu Gly Asn Lys Trp Thr Ser Gln Cys Ser Ser Ala Thr Asp Cys
50 55 60

ES 2 549 328 T3

Ala Gln Arg Cys Ala Leu Asp Gly Ala Asn Tyr Gln Ser Thr Tyr Gly
65 70 75 80

Ala Ser Thr Ser Gly Asp Ser Leu Thr Leu Lys Phe Val Thr Lys His
85 90 95

Glu Tyr Gly Thr Asn Ile Gly Ser Arg Phe Tyr Leu Met Ala Asn Gln
100 105 110

Asn Lys Tyr Gln Met Phe Thr Leu Met Asn Asn Glu Phe Ala Phe Asp
115 120 125

Val Asp Leu Ser Lys Val Glu Cys Gly Ile Asn Ser Ala Leu Tyr Phe
130 135 140

Val Ala Met Glu Glu Asp Gly Gly Met Ala Ser Tyr Pro Ser Asn Arg
145 150 155 160

Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ala Gln Cys Ala Arg
165 170 175

Asp Leu Lys Phe Ile Gly Gly Lys Ala Asn Ile Glu Gly Trp Arg Pro
180 185 190

Ser Thr Asn Asp Pro Asn Ala Gly Val Gly Pro Met Gly Ala Cys Cys
195 200 205

Ala Glu Ile Asp Val Trp Glu Ser Asn Ala Tyr Ala Tyr Ala Phe Thr
210 215 220

Pro His Ala Cys Gly Ser Lys Asn Arg Tyr His Ile Cys Glu Thr Asn
225 230 235 240

Asn Cys Gly Gly Thr Tyr Ser Asp Asp Arg Phe Ala Gly Tyr Cys Asp
245 250 255

Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg Met Gly Asn Lys Asp Phe
260 265 270

Tyr Gly Lys Gly Lys Thr Val Asp Thr Asn Arg Lys Phe Thr Val Val
275 280 285

Ser Arg Phe Glu Arg Asn Arg Leu Ser Gln Phe Phe Val Gln Asp Gly
290 295 300

Arg Lys Ile Glu Val Pro Pro Pro Thr Trp Pro Gly Leu Pro Asn Ser
305 310 315 320

Ala Asp Ile Thr Pro Glu Leu Cys Asp Ala Gln Phe Arg Val Phe Asp
325 330 335

Asp Arg Asn Arg Phe Ala Glu Thr Gly Gly Phe Asp Ala Leu Asn Glu

ES 2 549 328 T3

340 345 350

Ala Leu Thr Ile Pro Met Val Leu Val Met Ser Ile Trp Asp Asp His
355 360 365

His Ser Asn Met Leu Trp Leu Asp Ser Ser Tyr Pro Pro Glu Lys Ala
370 375 380

Gly Leu Pro Gly Gly Asp Arg Gly Pro Cys Pro Thr Thr Ser Gly Val
385 390 395 400

Pro Ala Glu Val Glu Ala Gln Tyr Pro Asp Ala Gln Val Val Trp Ser
405 410 415

Asn Ile Arg Phe Gly Pro Ile Gly Ser Leu Thr Gly Asn Pro Ser Gly
420 425 430

Gly Asn Pro Pro Gly Gly Asn Arg Gly Thr Thr Thr Thr Arg Arg Pro
435 440 445

Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln Ser His Tyr Gly
450 455 460

Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys Ala Ser Gly
465 470 475 480

Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys Leu Ala Ser
485 490 495

Ala His His His His His His
500

- <210> 17
- <211> 1809
- <212> ADN
- <213> artificial

5

<220>
 <223> secuencia codificadora de la proteína de fusión de CBHI de Talaromyces emersonii / de -CBD de Trichoderma reesei incluyendo el péptido de señal del factor alfa y una marcación 6x His

<400> 17
 atgagatttc cttcaat ttt tactgcagtt ttattcgcag catcctccgc attagctgct 60
 ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggt 120
 tacttagatt tagaagggga tttcgatggt gctgttttgc cattttccaa cagcacaat 180
 aacgggttat tgttataaa tactactatt gccagcattg ctgctaaaga agaaggggta 240
 tctttggata aacgtgaggc ggaagcatgc tcgcagcagg ccggcacggc gacggcagag 300
 aaccaccgc ccctgacatg gcaggaatgc accgccctg ggagctgcac caccagaac 360
 ggggcggtcg ttcttgatgc gaactggcgt tgggtgcacg atgtgaacgg atacaccaac 420
 tgctacacgg gcaatacctg ggacccacg tactgccctg acgacgaaac ctgcgcccag 480

10

ES 2 549 328 T3

aactgtgctg tggacggcgc ggattacgag ggcacctacg gcgtgacttc gtcgggcagc 540
 tccttgaaac tcaatttcgt caccgggtcg aacgtcggat cccgtctcta cctgctgcag 600
 gagcactcga cctatcagat cttcaagctc ctgaaccgcg agttcagctt tgacgtcgat 660
 gtctccaatc ttccgtgctg attgaacggc gctctgtact ttgtcgccat ggacgccgac 720
 ggcggcgtgt ccaagtacc gaacaacaag gctggtgcc agtacggaac cgggtattgc 780
 gactcccaat gccacaggga cctcaagttc atcgacggcg aggccaacgt cgagggtctg 840
 cagccgtctt cgaacaacgc caacaccgga attggcgacc acggctcctg ctgtgcggag 900
 atggatgtct ggggaagcaaa cagcatctcc aatgctgtca ctccgcacc gtgctgacgc 960
 ccaggccaga cgatgtgctc tggagatgac tgcggtggca catactctaa cgatcgtac 1020
 gcgggaacct gcgatcctga cggctgtgac ttcaaccctt accgcatggg caacacttct 1080
 ttctacgggc ctggcaagat catcgatacc accaagccct tcaactgtct gacgcagttc 1140
 ctcaactgat atggtacgga tactggaact ctacagcaga tcaagcgtt ctacatccag 1200
 aacagcaacg tcattccgca gcccaactcg gacatcagtg gcgtgaccgg caactcgtc 1260
 acgacggagt tctgactcgc tcagaagcag gcctttggcg acacggacga cttctctcag 1320
 cacggtggcc tggccaagat gggagcggcc atgcagcagg gtatggctct ggtgatgagt 1380
 ttgtgggacg actacgccgc gcagatgctg tggttgatt ccgactacc gacggatgcg 1440
 gacccacga cccctggtat tgcccgtgga acgtgtccga cggactcggg cgtcccatcg 1500
 gatgtcagat cgcagagccc caactcctac gtgacctact cgaacattaa gtttggtccg 1560
 atcggtagca caggtaatcc ttcaggtggt aatcctccag gtggaacag aggaacaacg 1620
 acaactagaa gaccagctac tacaactggt tcaagtccag gtccaactca atcacactac 1680
 ggtcaatgtg gtggtatagg ttactctggt cccactgttt gtgcttctgg tactacttgc 1740
 caagttctga acccttacta ctacagtggt ctacttctg cacatcatca ccaccaccat 1800
 taatgataa 1809

<210> 18
 <211> 509
 5 <212> PRT
 <213> artificial

<220>
 <223> secuencia madura de la fusión de CBHI de Talaromyces emersonii / -CBD de Trichoderma reesei con una
 marcación 6x-His (CBH-ah)

10 <400> 18
 Gln Gln Ala Gly Thr Ala Thr Ala Glu Asn His Pro Pro Leu Thr Trp
 1 5 10 15
 Gln Glu Cys Thr Ala Pro Gly Ser Cys Thr Thr Gln Asn Gly Ala Val
 20 25 30
 Val Leu Asp Ala Asn Trp Arg Trp Val His Asp Val Asn Gly Tyr Thr
 35 40 45

ES 2 549 328 T3

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

ES 2 549 328 T3

Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu Phe Cys Thr Ala
 325 330 335

Gln Lys Gln Ala Phe Gly Asp Thr Asp Asp Phe Ser Gln His Gly Gly
 340 345 350

Leu Ala Lys Met Gly Ala Ala Met Gln Gln Gly Met Val Leu Val Met
 355 360 365

Ser Leu Trp Asp Asp Tyr Ala Ala Gln Met Leu Trp Leu Asp Ser Asp
 370 375 380

Tyr Pro Thr Asp Ala Asp Pro Thr Thr Pro Gly Ile Ala Arg Gly Thr
 385 390 395 400

Cys Pro Thr Asp Ser Gly Val Pro Ser Asp Val Glu Ser Gln Ser Pro
 405 410 415

Asn Ser Tyr Val Thr Tyr Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser
 420 425 430

Thr Gly Asn Pro Ser Gly Gly Asn Pro Pro Gly Gly Asn Arg Gly Thr
 435 440 445

Thr Thr Thr Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro
 450 455 460

Thr Gln Ser His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro
 465 470 475 480

Thr Val Cys Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr
 485 490 495

Ser Gln Cys Leu Ala Ser Ala His His His His His
 500 505

- <210> 19
- <211> 1335
- <212> ADN
- <213> artificial

5

<220>
 <223> Secuencia codificadora alternativa de CBHI de Humicola grisea con la secuencia de señal

<400> 19
 atggccagcg atctggcaca gcaggctggt acaattactg ctgagaacca tccaagaatg 60
 acgtggaaga gatgtagtgg tccaggaaac tgtcagactg ttcaggggtga ggctcgtgata 120
 gatgctaact ggagatggtt gcataacaac ggccagaact gctacgaggg taacaagtgg 180
 acctctcagt gttcttctgc taccgactgc gctcagagat gtgctcttga tggagcaaac 240
 taccagagta catatggtgc ttctacctct ggtgacagcc ttaccctgaa gtttgaacc 300

ES 2 549 328 T3

aagcacgagt acggaaccaa tateggttct agattctacc tgatggctaa ccagaacaag 360
 taccagatgt ttaccttgat gaacaacgag ttcgccttcg acgtagatct gtctaagggtg 420
 gagtgtggaa tcaattctgc cttgtacttt gtcgctatgg aagaggacgg aggtatggct 480
 tcttaccctt ctaacagagc tggtgctaag tatggaactg gatactgcga tgcccaatgc 540
 gctagagacc tgaagtcat cggtggaaag gctaacattg aaggttggag accttctacc 600
 aacgacccaa acgctggagt tggccaatg ggtgcttgct gtgccgagat tgacgtgtgg 660
 gaatctaacg cttacgccta cgcttttact ccacatgctt gcggttctaa gaacagatac 720
 cacatttgcg aaaccaacaa ctgtggtggc acttactctg atgacagatt cgctggatac 780
 tgtgatgcta acggatgtga ttacaacca tacagaatgg gtaacaagga cttttacgga 840
 aagggtgaaga ctggtgacac taacagaaag ttcactgtgg tctcgagatt tgagagaaac 900
 agactgtcgc agttctttgt gcaggacgga agaaagattg aggtcccacc accaacttgg 960
 ccaggattgc caaactctgc cgacattacc ccagagttgt gcgacgctca gttcagagtg 1020
 tttgacgaca gaaacagatt tgctgagacc ggtggatttg acgctttgaa cgaggctctg 1080
 accattccaa tggttctagt catgagtatt tgggacgatc accactctaa catgctttgg 1140
 ctggactctt cttaccctcc agagaaggct ggattgcctg gtggtgacag aggtccatgt 1200
 ccaacaactt ctggagttcc agccgagggt gaggctcaat acccagacgc ccaggctcgtg 1260
 tggccaaca tcagattcgg accaattggt agcacagtga atgtggcttc tgcacacat 1320
 catcatcatc attga 1335

<210> 20
 <211> 41
 <212> ADN
 <213> artificial

5

<220>
 <223> cebador hacia adelante

<400> 20
 gaggcggaag caccctctca atctgctgc acctgcagt c 41

10

<210> 21
 <211> 38
 <212> ADN
 <213> artificial

15

<220>
 <223> cebador inverso

<400> 21
 ggagacgcag agcccttatt acaggcactg cgagtagt 38

20

<210> 22
 <211> 41
 <212> ADN
 <213> artificial

<220>
 <223> cebador hacia adelante

25

<400> 22
 gaggcggaag caccctctca gcaggctggt actattactg c 41

<210> 23
 <211> 44
 <212> ADN
 <213> artificial

5 <220>
 <223> cebador inverso

<400> 23
 ggagacgcag agcccttaca cgttcacggt agaaccgatt gggc 44

10 <210> 24
 <211> 41
 <212> ADN
 <213> artificial

<220>
 <223> cebador hacia adelante

15 <400> 24
 gaggcggaag cacccttca cgaggccggt accgtaaccg c 41

<210> 25
 <211> 41
 <212> ADD
 20 <213> artificial

<220>
 <223> cebador inverso

<400> 25
 ggagacgcag agcccttatt agttggcggg gaaggtcgag t 41

25 <210> 26
 <211> 41
 <212> ADN
 <213> artificial

30 <220>
 <223> cebador hacia adelante

<400> 26
 gaggcggaag cacccttca gcaggccggc acggcgacgg c 41

35 <210> 27
 <211> 41
 <212> ADD
 <213> artificial

<220>
 <223> cebador inverso

40 <400> 27
 ggagacgcag agcccttacc acgaagcggg gaaggtcgag t 41

<210> 28
 <211> 41
 <212> ADN
 <213> artificial

45 <220>
 <223> cebador hacia adelante

<400> 28
 gaggcggaag cacccttca gcaggccggc acggcgacgg c 41

<210> 29
 <211> 38
 <212> ADN
 <213> artificial

5 <220>
 <223> cebador inverso

<400> 29
 attacctgtg ctaccgatcg gaccaaactt aatgttcg 38

10 <210> 30
 <211> 38
 <212> ADN
 <213> artificial

<220>
 <223> cebador hacia adelante

15 <400> 30
 aagtttggtc cgatcggtag cacaggtaat ccttcagg 38

<210> 31
 <211> 44
 <212> ADN
 20 <213> artificial

<220>
 <223> cebador inverso

<400> 31
 ggagacgcag agccctatt atagacactg tgagtagtaa ggg 44

25 <210> 32
 <211> 41
 <212> ADN
 <213> artificial

<220>
 30 <223> cebador hacia adelante

<400> 32
 gaggcggaag caccctctca gcaggccggc acggcgacgg c 41

<210> 33
 <211> 44
 35 <212> ADN
 <213> artificial

<220>
 <223> cebador inverso

<400> 33
 40 ggagacgcag agccctatc attaatggtg gtggtgatga tgag 44

<210> 34
 <211> 40
 <212> ADN
 <213> artificial

45 <220>
 <223> cebador hacia adelante

<400> 34
 aggcggaagc atgctcgag caggctggta caattactgc 40

<210> 35
 <211> 41
 <212> ADN
 <213> artificial

5 <220>
 <223> cebador inverso

<400> 35
 ggattacctg ttaagctcc aattggccg aatctgatg t 41

10 <210> 36
 <211> 42
 <212> ADN
 <213> artificial

<220>
 <223> cebador hacia adelante

15 <400> 36
 accaattgga agcttaacag gtaatcctc aggtggaat cc 42 36

<210> 37
 <211> 46
 <212> ADN
 <213> artificial

20 <220>
 <223> cebador inverso

<400> 37
 atctgcagg tcgactatc attaatgatg atgatgatgg tgtgca 46

25 <210> 38
 <211> 40
 <212> ADN
 <213> artificial

30 <220>
 <223> cebador hacia adelante

<400> 38
 aggcggaagc atgctgcag caggctgga caattactgc 40

35 <210> 39
 <211> 46
 <212> ADN
 <213> artificial

<220>
 <223> cebador inverso

40 <400> 39
 atctgcagg tcgactatc attaatgatg atgatgatgg tgtgca 46

<210> 40
 <211> 21
 <212> ADN
 <213> artificial

45 <220>
 <223> oligonucleótido alfa-f

<400> 40
 tactattgcc agcattgctg c 21

<210> 41
<211> 23
<212> ADN
<213> artificial

5 <220>
<223> oligonucleótido oli740

<400> 41
tcagctatt cacatacaaa tcg 23

10

REIVINDICACIONES

- 5 1. Un polipéptido que tiene una actividad de celobiohidrolasa, en donde el polipéptido comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 85 % con respecto a la SEQ ID NO: 2, en donde el residuo de aminoácido en la posición Q1 de la SEQ ID NO: 2 se ha modificado por sustitución o supresión.
- 10 2. El polipéptido de acuerdo con la reivindicación 1, en donde el polipéptido conserva un 50 % de su máxima capacidad de conversión del sustrato cuando la conversión se realiza durante 60 minutos a una temperatura de 60°C o más alta.
- 15 3. El polipéptido de acuerdo con la reivindicación 1 ó 2, en donde este polipéptido comprende una secuencia de aminoácidos que tiene una identidad entre secuencias de por lo menos 90 %, de manera preferida de por lo menos 95 %, de manera muy especialmente preferida de por lo menos 99 % con la SEQ ID NO: 2.
- 20 4. El polipéptido de acuerdo con una o más de las reivindicaciones precedentes, en el que la secuencia de aminoácidos del polipéptido tiene la secuencia definida por la SEQ ID NO: 2, o una secuencia de acuerdo con la SEQ ID NO: 2, en donde de 1 a 75 residuos de aminoácidos, de manera preferida más preferida de 1 a 35 residuos de aminoácidos, son sustituidos, suprimidos o insertados.
- 25 5. El polipéptido de acuerdo con la reivindicación 4, en el que adicionalmente uno o más de los siguientes residuos de aminoácidos de la secuencia definida por la SEQ ID NO: 2 son modificados por sustitución o supresión: en las posiciones G4, A6, T15, Q28, W40, D64, E65, A72, S86, K92, V130, V152, Y155, K159, D181, E183, N194, D202, P224, T243, Y244, I277, K304, N310, S311, N318, D320, T335, T344, D346, Q349, A358, Y374, A375, T392, T393, D410, Y422, P442, N445, R446, T456, S460, P462, G463, H468 y/o V482 de los aminoácidos 1 hasta 500 de la SEQ ID NO: 2.
6. El polipéptido de acuerdo con la reivindicación 5, en donde el polipéptido comprende uno o más de los siguientes intercambios preferidos con respecto a la secuencia definida por la SEQ ID NO: 2:

Posición	Intercambio preferido
Q1	L
G4	C
A6	G,V
T15	S
Q28	Q,R
W40	R
D64	N
E65	K, V
A72	C, V
S86	T
K92	K, R
V130	I, V
V152	A, E
Y155	C
K159	E
D181	N
E183	V, K
N194	C, R, Y, D, K, I, L, G, Q, S, V
D202	Y,N,G
P224	L
T243	I, C, R, Y, A, F, Q, P, D, V, W, L, M
Y244	F, H

ES 2 549 328 T3

Posición	Intercambio preferido
I277	V
K304	R
N310	D
S311	G, N
N318	Y
D320	V, E, N
T335	I
T344	M
D346	G, A, E, V
Q349	R, K
A358	E
Y374	C, P, R, H, S, A
A375	C, D, N, Y, R, Q, L, V, E, G, T, M
T392	C, D, K
T393	A
D410	G
Y422	F
P442	S, del
N445	D
R446	S, G
T456	T, A
S460	L, P
P462	L, del
G463	D
H468	L, Q, R
V482	A, I

7. El polipéptido de acuerdo con una o más de las reivindicaciones precedentes, en el que el polipéptido tiene una secuencia de aminoácidos escogida entre la lista de las siguientes mutaciones de la SEQ ID NO: 2:

75	Q1L, G4C, A72C, Q349K
88	Q1L, G4C, A72C, S86T, Q349R
89	Q1L, G4C, A72C, D181N, Q349R
90	Q1L, G4C, A72C, E183K, Q349R
91	Q1L, G4C, A72C, D181N, E183K, Q349R
92	Q1L, G4C, A72C, D320V, Q349R
93	Q1L, G4C, A72C, S86T, D181N, E183K, D320V, Q349R
98	Q1L, G4C, A72C, Q349R
148	Q1L, G4C, A72C, Q349K, T392M
153	Q1L, G4C, A68T, A72C, Q349K, G439D, R453S
154	Q1L, G4C, A72C, D202N, Q349K
155	Q1L, G4C, A68T, A72C, Q349K
156	Q1L, G4C, A72C, K154R, Q349K, T393I

157	Q1L, G4C, A72C, S193P, Q349K, V482I
158	Q1L, G4C, A72C, H203R, Q349K, P442S
159	Q1L, G4C, A72C, Q349K, H468R
160	Q1L, G4C, A72C, D202N, Q349K, G486D
161	Q1L, G4C, E65K, A72C, Q349K
162	Q1L, G4C, A72C, Q349K, Y422F
163	Q1L, G4C, Q28R, A72C, Q349K, H468L
164	Q1L, G4C, A72C, D181N, D247N, Q349K
165	Q1L, G4C, A72C, D181N, Q349K, T451S
166	Q1L, G4C, Q28R, A72C, Q349K
167	Q1L, G4C, A72C, A145T, H203R, Q349K, T403K
168	Q1L, G4C, A72C, I200F, Q349K, L500I
169	Q1L, G4C, D64N, A72C, Q349K
170	Q1L, G4C, A72C, V152A, Q349K
171	Q1L, G4C, T15S, A72C, Y244F, Q349K
172	Q1L, G4C, A6V, A72C, Q349K
173	Q1L, G4C, A72C, S311N, Q349K, G463D
174	Q1L, G4C, A72C, Y155C, Q349K
175	Q1L, G4C, A72C, S311N, Q349K
176	Q1L, G4C, A72C, D346V, Q349K
177	Q1L, G4C, A72C, Q349K, T392K
178	Q1L, G4C, A72C, S311G, Q349K
179	Q1L, G4C, A72C, S311G, Q349K, H468R

8. El polipéptido de acuerdo con una o más de las reivindicaciones precedentes, que es expresado y segregado dentro del material sobrenadante hasta un nivel de más que 100 mg/l, de manera más preferida de más que 200 mg/l, de manera especialmente preferida de más que 500 mg/l y de manera máximamente preferida de más que 1 g/l después de la introducción de un ácido nucleico que codifica un polipéptido que tiene una secuencia de aminoácidos con una identidad entre secuencias de por lo menos 85 % con la SEQ ID NO: 2, dentro de una levadura, en donde el residuo de aminoácido situado en la posición Q1 de la SEQ ID NO: 2 tiene que ser modificado por sustitución o supresión.
9. Un ácido nucleico que codifica el polipéptido de acuerdo con una o más de las reivindicaciones 1 hasta 8, que tiene de manera preferida una identidad de por lo menos 95 % con la SEQ ID NO: 1.
10. Un vector que comprende el ácido nucleico de acuerdo con la reivindicación 9.
11. Una célula anfitriona que ha sido transformada con un vector de acuerdo con la reivindicación 10.
12. La célula anfitriona de acuerdo con la reivindicación 11, en donde esta célula anfitriona se deriva del conjunto que se compone de Saccharomyces, Schizosaccharomyces, Kluyveromyces, Pichia, Pichia, Hansenula, Aspergillus, Trichoderma, Penicillium, Candida y Yarrowina.
13. Una composición que comprende el polipéptido de acuerdo con una o más de las reivindicaciones 1 hasta 9 y una o más endoglucanasas y/o una o más beta-glucosidasas y/o una o más otras celobiohidrolasas y/o una o más xilanasas.
14. Uso del polipéptido de acuerdo con una o más de las reivindicaciones 1 hasta 9 o de la composición de acuerdo con la reivindicación 13 para la degradación enzimática de una biomasa lignocelulósica, y/o para el tratamiento de materiales textiles y/o como un ingrediente en detergentes y/o como un ingrediente en alimentos o composiciones alimentarias.

Figura 1

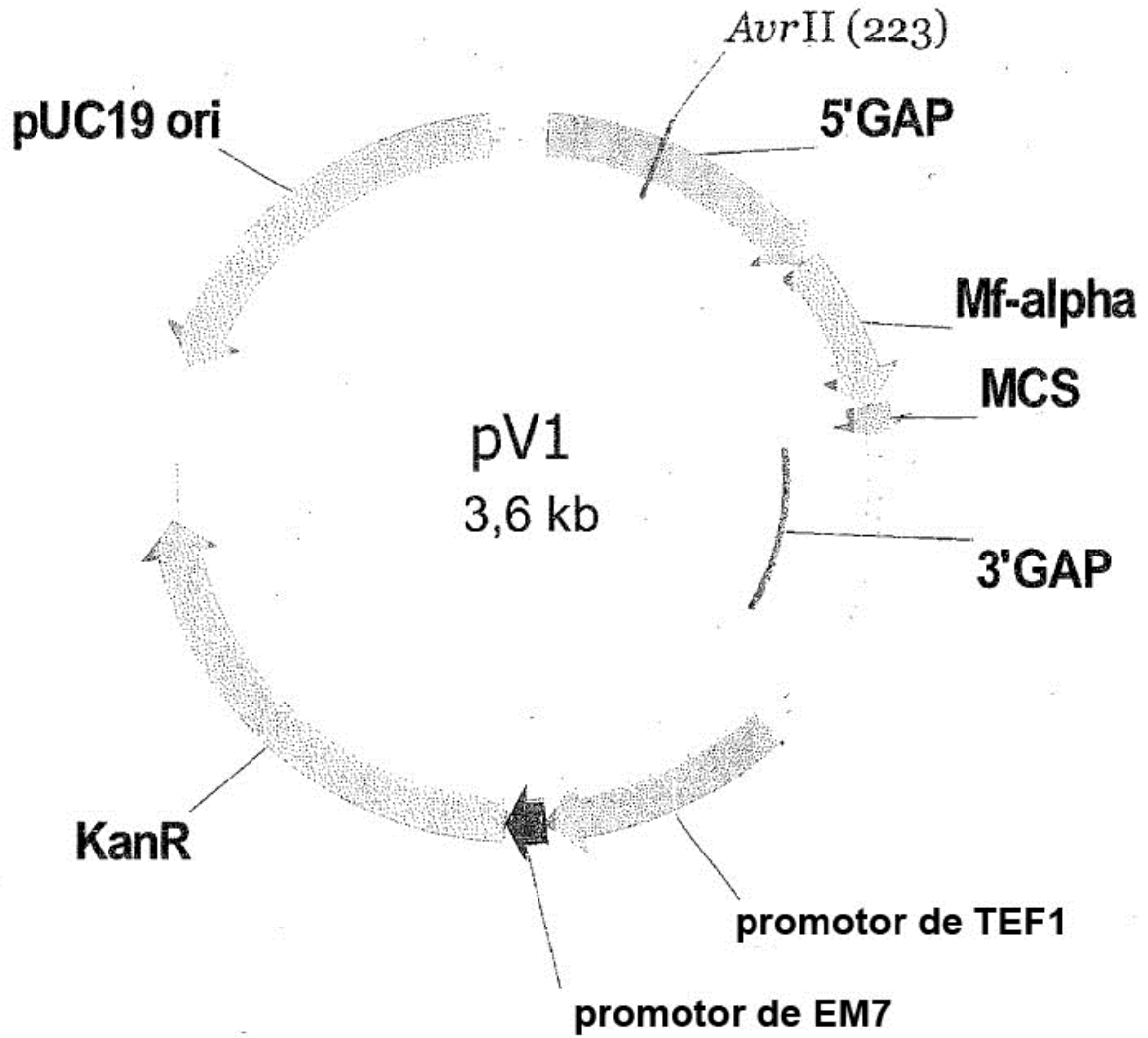
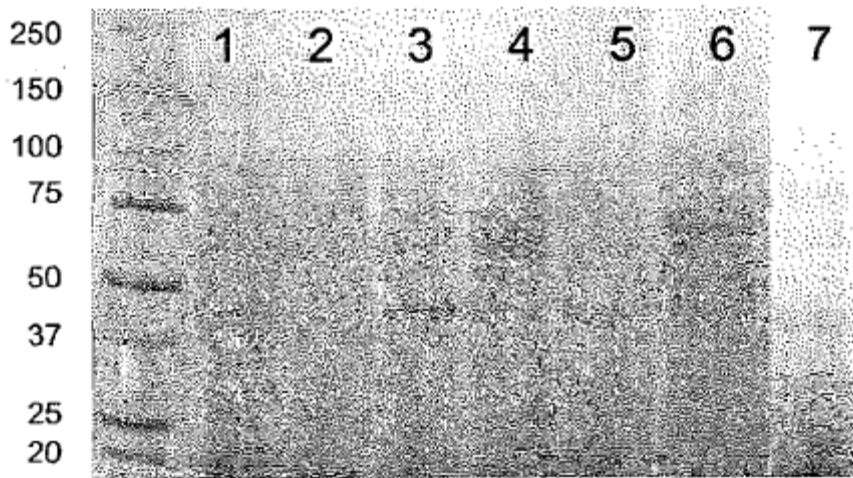


Figura 2



- 1 vector puro
- 2 CBHI de *Trichoderma viride*
- 3 CBHI de *Humicola grisea*
- 4 CBHI de *Talaromyces emersonii*
- 5 CBHI de *Thermoascus aurantiacus*
- 6 fusión de CBHI de *Talaromyces emersonii* con la CBHI CBD de *T. reesei*
- 7 fusión de CBHI de *Humicola grisea* con la CBHI CBD de *T. reesei*

Figura 3

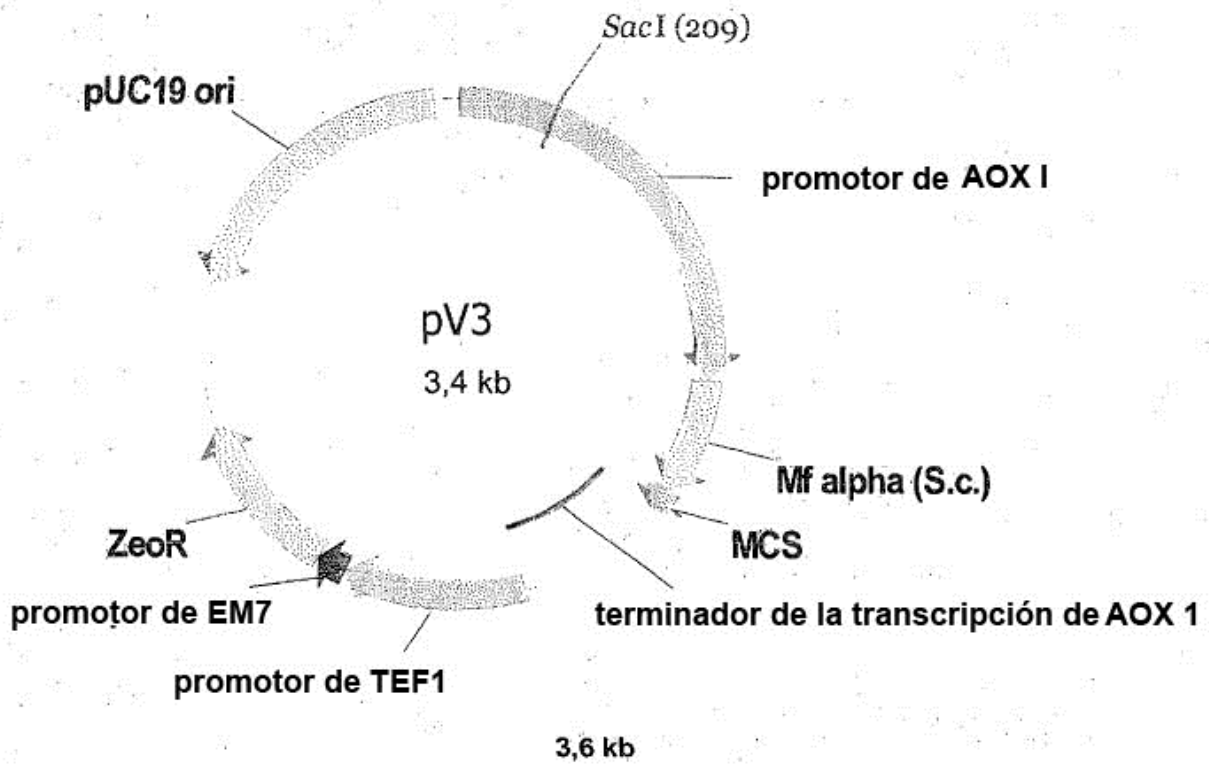


Figura 4

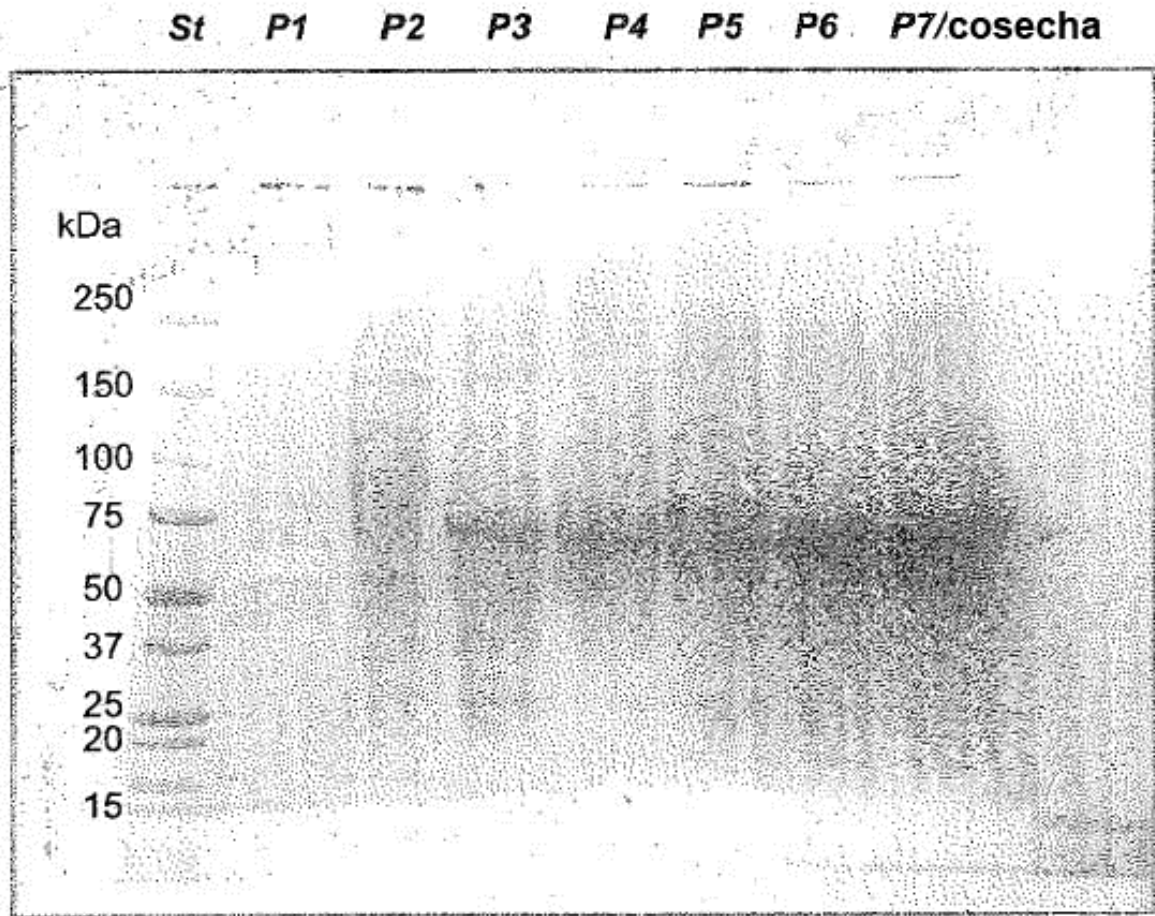


Figura 5

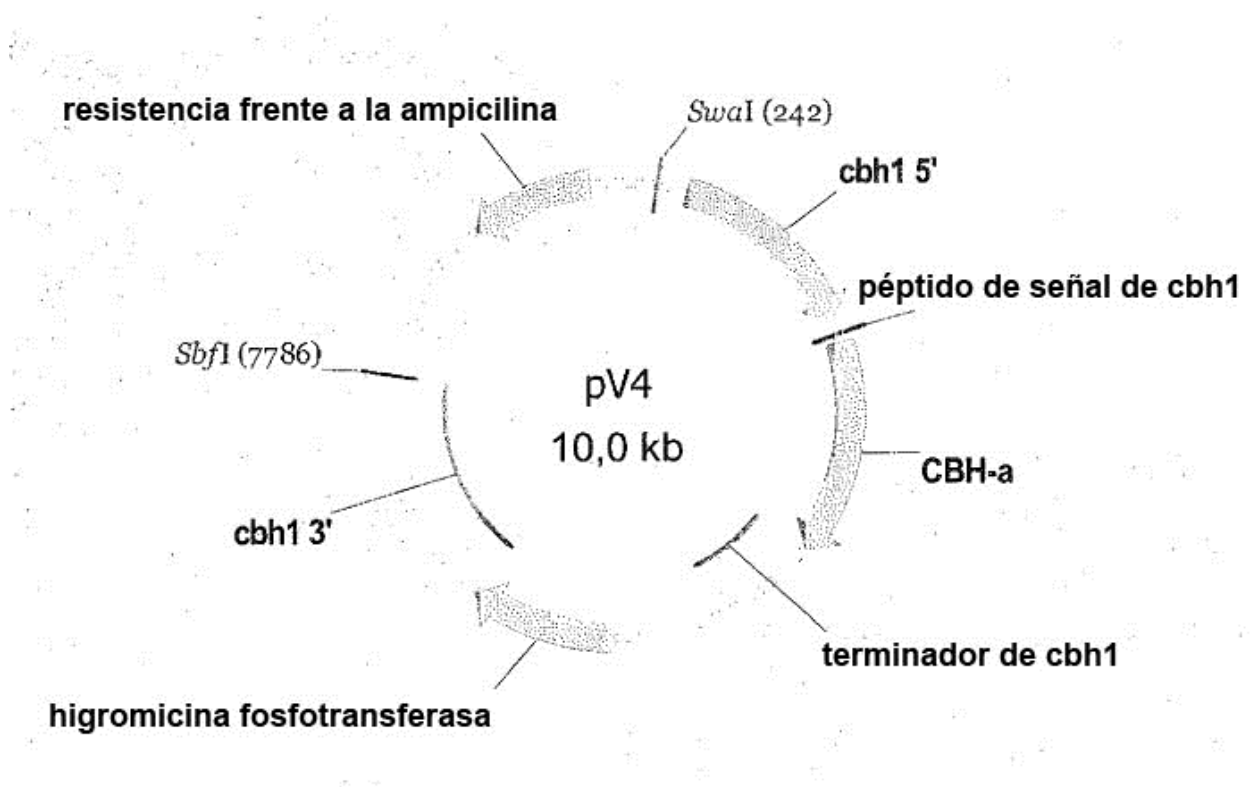


Figura 6

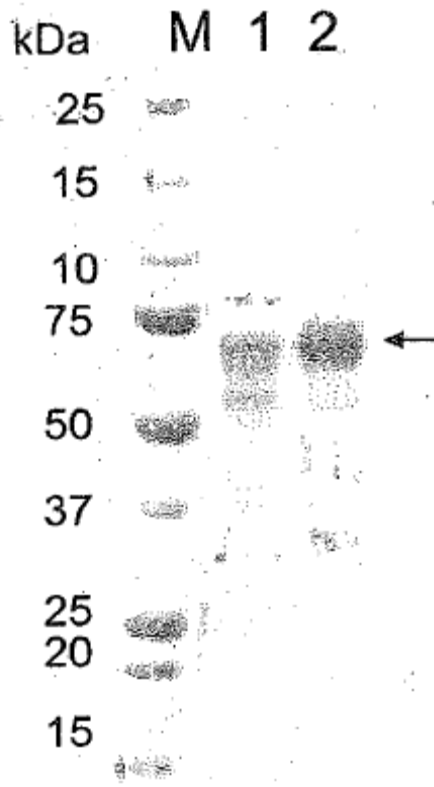


Figura 7

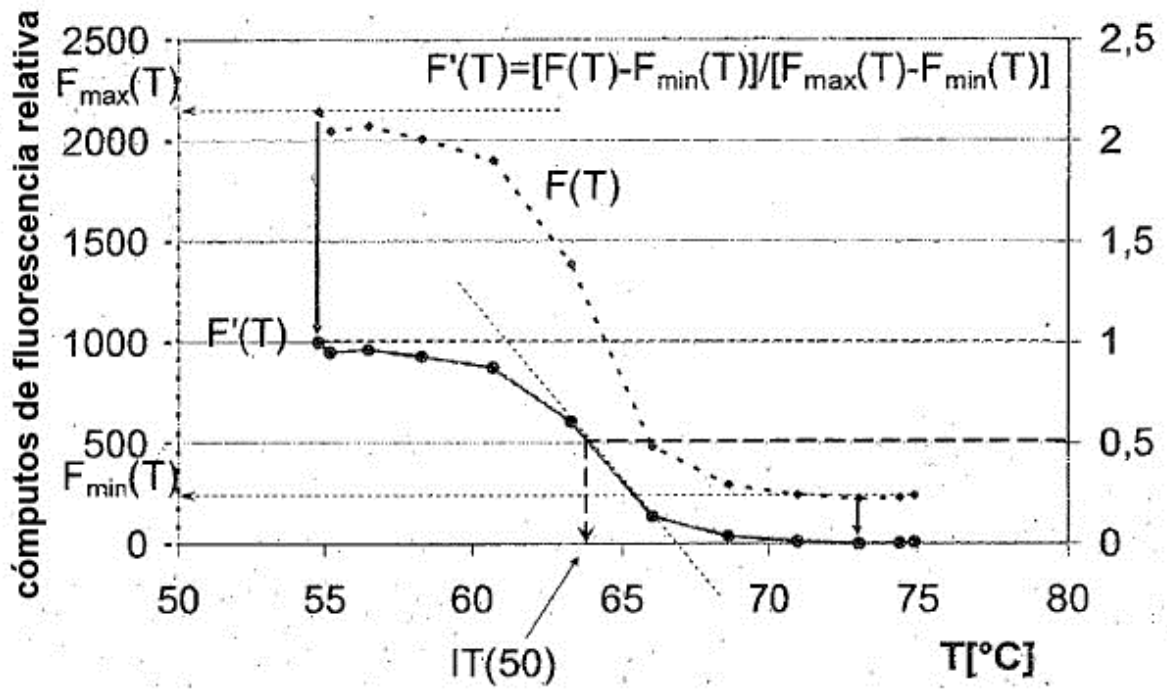


Figura 8

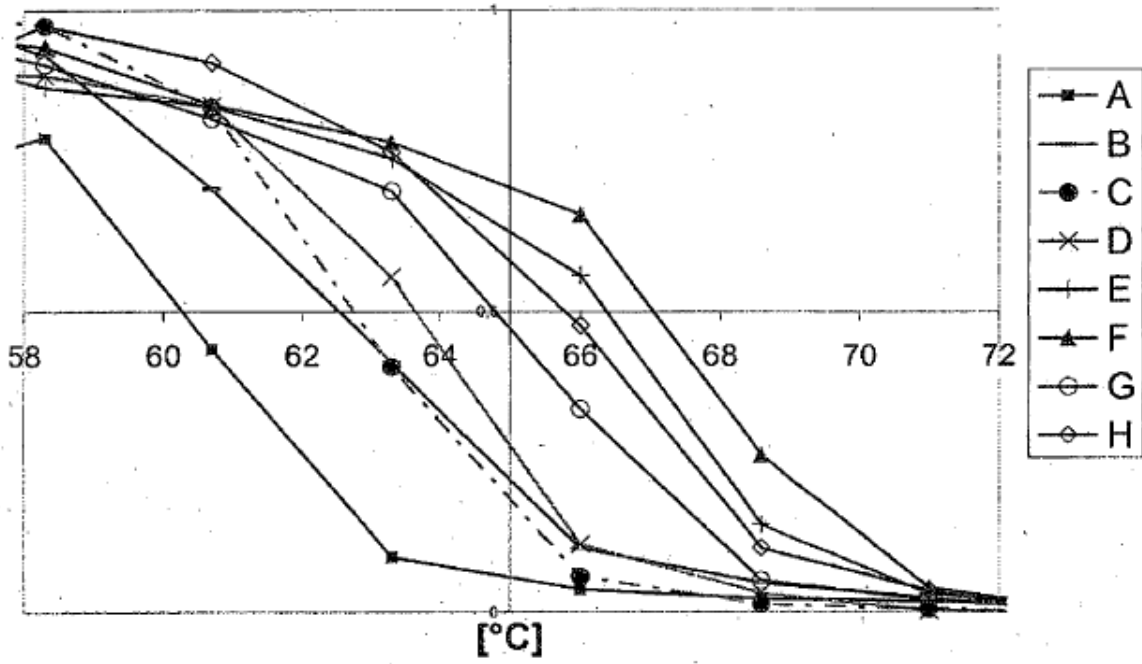


Figura 9

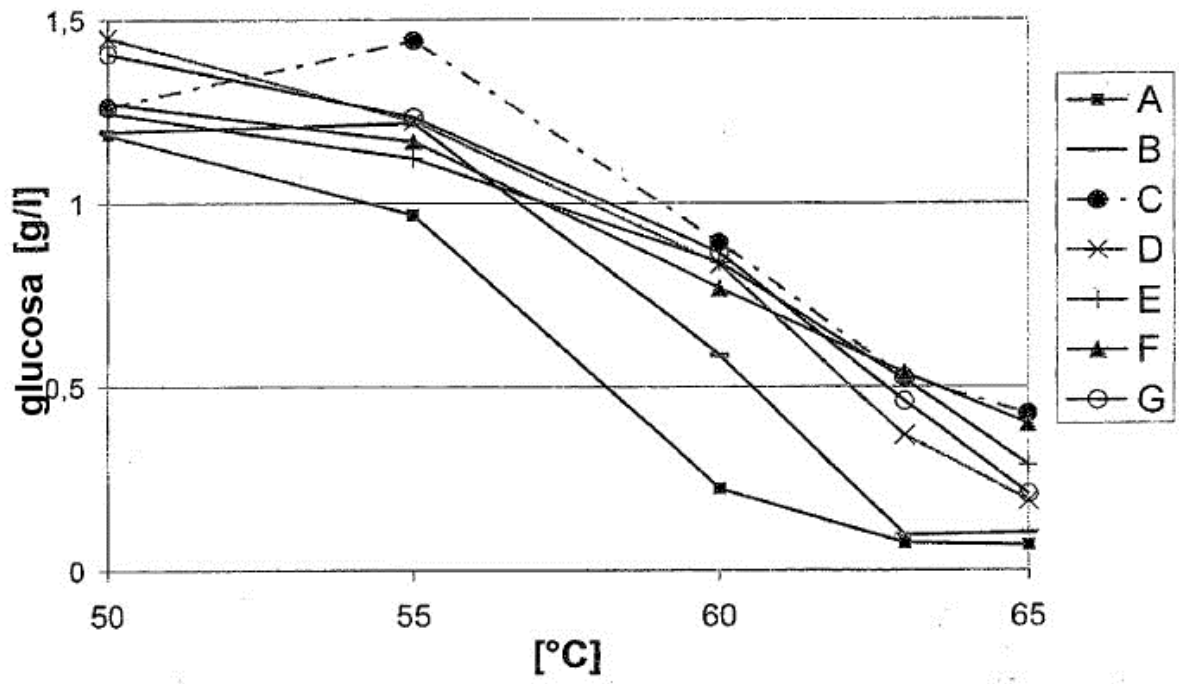


Figura 10

```

1
T._reesei_CBHI  QSACTLQSET HPPLTWQKCS SGGTCTQQTG SVVIDANWRW THATNSSTNC
SeqID_NO.2      QQAGTATAEN HPPLTWQECT  APGSCTTQNG AVVLDANWRW VHDVNGYTNC

51
T._reesei_CBHI  YDGNTWSSTL CPDNETCAKN CCLDGAAYAS TYGVTTSGNS LSIGFVTQSA
SeqID_NO.2      YTGNTWDPTY CPDDETCAQN CALDGADYEG TYGVTTSSGSS LKLNFTVG..

101
T._reesei_CBHI  QKNVGARLYL MASDTTYQEF TLLGNEFSFD VDVSQPCGL  NGALYFVSMD
SeqID_NO.2      .SNVGSRLYL LQDDSTYQIF KLLNREFSFD VDVSNLPCGL  NGALYFVAMD

151
T._reesei_CBHI  ADGGVSKYPT NTAGAKYGTG YCDSQCPRDL KPINGQANVE GWEPSSNNAN
SeqID_NO.2      ADGGVSKYPN NKAGAKYGTG YCDSQCPRDL KPIDGEANVE GWQPSNNAN

201
T._reesei_CBHI  TGIGGHGSCC SEMDIWEANS ISEALTPHPC TTVGQICEG  DGCGGTYSND
SeqID_NO.2      TGIGDHGSCC AEMDVWEANS ISNAVTPHPC DTPGQTMCSG DDCGGTYSND

251
T._reesei_CBHI  RYGGTCDPDG CDWNPYRLGN TSFYGPGSSF TLDTTKCLTV VTQFETSG..
SeqID_NO.2      RYAGTCDPDG CDFNPYRMGN TSFYGPGK.. IIDTTKPFTV VTQFLTDDGT

301
T._reesei_CBHI  .....AINR YYVQNGVTFQ QPNAELGSYS GNELNDDYCT ABEAEFGGSS
SeqID_NO.2      DTGTLSEIKR FYIQNSNVIP QPNSDISGVT GNSITTEFCT AOKQAFGDTD

351
T._reesei_CBHI  .PSDKGGLTQ FKKATSGGMV LVMSLWDDYY ANMLWLDSTY PTNETSSTPG
SeqID_NO.2      DFSQHGLAK  MGAAMQOGMV LVMSLWDDYA AQMLWLDSDY PTDADPTTPG

401
T._reesei_CBHI  AVRGCSTSS  GVPAQVESQS PNAKVTFSNI KFGPIGSTGN PSGGNPPGGN
SeqID_NO.2      IARGTCPTDS GVPSDVESQS PNSYVTYSNI KFGPIGSTGN PSGGNPPGGN

451
T._reesei_CBHI  RGTTRRPA  TTTGSSPGPT QSHYGQCGGI GYSGPTVCAS GTTCQVLNPNY
SeqID_NO.2      RGTTRRPA  TTTGSSPGPT QSHYGQCGGI GYSGPTVCAS GTTCQVLNPNY

501
T._reesei_CBHI  YSQCL
SeqID_NO.2      YSQCL

```