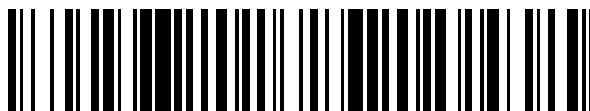


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 576 062**

51 Int. Cl.:

C07K 14/37 (2006.01)

C12N 9/24 (2006.01)

C12N 9/42 (2006.01)

C12N 15/82 (2006.01)

C12P 19/02 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **23.06.2011 E 11726850 (8)**

97 Fecha y número de publicación de la concesión europea: **16.03.2016 EP 2588492**

54 Título: **Polipéptido que tiene actividad beta-glucosidasa y usos del mismo**

30 Prioridad:

29.06.2010 US 359567 P

29.06.2010 EP 10167772

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

05.07.2016

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)

Het Overloon 1

6411 TE Heerlen, NL

72 Inventor/es:

SCHOONEVELD-BERGMANS, MARGOT

ELISABETH FRANCOISE;

HEIJNE, WILBERT HERMAN MARIE;

DAMVELD, ROBBERTUS ANTONIUS y

DE JONG, RENÉ MARCEL

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 576 062 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

Polipéptido que tiene actividad beta-glucosidasa y usos del mismo

5 Declaración de investigación y desarrollo con patrocinio federal

Esta invención se hizo con el respaldo del gobierno de los Estados Unidos conforme al n.º de subvención DE-FC36-08G018079, otorgado por el Departamento de Energía. El gobierno de los Estados Unidos puede tener ciertos derechos en esta invención.

10 Campo de la invención

La invención se refiere a secuencias que comprenden genes que codifican polipéptidos que tienen o ayudan en la actividad de degradación de material de carbohidrato. La invención destaca la secuencia codificante de longitud completa del nuevo gen, así como la secuencia de aminoácidos de la proteína funcional de longitud completa, y variantes y fragmentos del gen o la secuencia de aminoácidos. La invención también se refiere a métodos para usar estas proteínas en procesos industriales. También se incluyen en la invención células transformadas con un polinucleótido de acuerdo con la invención adecuado para producir estas proteínas. Además, la invención se refiere a la expresión satisfactoria de los genes que codifican polipéptidos que tienen actividad de degradación de material de carbohidrato en un hospedador. El hospedador puede ser cualquier hospedador adecuado, por ejemplo, *Aspergillus*, por ejemplo, *Aspergillus niger* o *Talaromyces*, por ejemplo, *Talaromyces emersonii*.

20 Antecedentes de la invención

Los carbohidratos constituyen los compuestos orgánicos más abundantes sobre la tierra. Sin embargo, gran parte de estos carbohidratos está secuestrada en polímeros complejos incluyendo el almidón (el principal carbohidrato de almacenamiento en semillas y cereales), y un conjunto de carbohidratos y lignina conocido como lignocelulosa. Los componentes principales de carbohidrato de la lignocelulosa son celulosa, hemicelulosa, y pectinas. Estos polímeros completos a menudo se mencionan de forma colectiva como lignocelulosa.

La bioconversión de biomasa lignocelulósica renovable en un azúcar fermentable que posteriormente se fermenta para producir alcohol (por ejemplo, etanol) como alternativa a los combustibles líquidos ha obtenido una atención especial de los investigadores desde la década de 1970, cuando estalló la crisis de petróleo a causa de la disminución en la producción de petróleo por la OPEC. El etanol se ha usado ampliamente como una mezcla al 10 % con gasolina en los Estados Unidos o como un combustible puro para vehículos en Brasil en las últimas dos décadas. Más recientemente, el uso de E85, una mezcla de etanol al 85 %, se ha implementado especialmente para aplicaciones de ciudades limpias. La importancia del combustible bioetanol aumentará en paralelo con los aumentos en los precios del petróleo y el agotamiento gradual de sus fuentes. Además, se están usando azúcares fermentables para producir plásticos, polímeros y otros productos de base biológica y se espera que esta industria crezca sustancialmente, aumentando, por lo tanto, la demanda de azúcares fermentables abundantes de bajo coste que pueden usarse como materia prima en lugar de materias primas basadas en petróleo.

El secuestro de dichas grandes cantidades de carbohidratos en biomasa vegetal proporciona una fuente abundante de energía potencial en forma de azúcares, que son azúcares tanto de cinco carbonos como de seis carbonos que podrían utilizarse para numerosos procesos industriales y agrícolas. Sin embargo, el enorme potencial energético de estos carbohidratos está actualmente infrautilizado porque los azúcares están encerrados en polímeros complejos, y por tanto no son fácilmente accesibles para fermentación. Métodos que generan azúcares a partir de biomasa vegetal proporcionarían materias primas abundantes, económicamente competitivas para su fermentación en agentes químicos, plásticos, tales como, por ejemplo, ácido succínico y (bio) combustibles, incluyendo etanol, metanol, butanol combustibles líquidos sintéticos y biogás.

Independientemente del tipo de materia prima celulósica, el coste y la eficacia hidrolítica de las enzimas son los factores principales que restringen la comercialización de los procesos de bioconversión de biomasa. Los costes de producción de enzimas producidas de forma microbiana están estrechamente conectados con una productividad de la cepa productora de enzimas y el rendimiento final de actividad en el caldo de fermentación.

Murray *et al.* (2004) describe la expresión y caracterización de una familia termoestable de 3 beta-glucosidasa de *Talaromyces emersonii*. Harnpicharnchai *et al.* (2009) describe una beta-glucosidasa termotolerante aislada de *Periconia* sp.

A pesar de la investigación continuada de las últimas décadas para entender la degradación enzimática de biomasa lignocelulósica y la producción de celulasa, sigue siendo deseable descubrir o diseñar por ingeniería nuevas celulasas y hemicelulasas altamente activas. También sería muy deseable construir composiciones enzimáticas altamente eficaces capaces de realizar una biodegradación rápida y eficaz de materiales lignocelulósicos, en particular dichas celulasas y hemicelulasas que tienen termoestabilidad aumentada.

Dichas enzimas pueden usarse para producir azúcares para su fermentación en agentes químicos, plásticos, tales como, por ejemplo, ácido succínico y (bio) combustibles, incluyendo etanol, metanol, butanol, combustibles líquidos sintéticos y biogás, para ensilaje, y también como enzima en otros procesos industriales, por ejemplo, en las industrias de la alimentación o los piensos, textil, de la pulpa o papel o detergentes y otras industrias.

Sumario de la invención

La presente invención proporciona polinucleótidos que codifican polipéptidos que tienen la capacidad de degradar (es decir, ayudar en la degradación de), un carbohidrato (por ejemplo, polisacáridos), en particular, lignocelulosa. Los polinucleótidos de la invención típicamente codifican un polipéptido que tiene actividad beta-glucosidasa.

La invención también proporciona polipéptidos producidos de forma natural y recombinante que tienen dicha actividad, así como líneas celulares recombinantes que producen dichas enzimas. Además, se proporcionan métodos de preparación y uso de los polinucleótidos y polipéptidos de la invención.

De acuerdo con la invención, por tanto se proporciona un polipéptido que comprende la secuencia de aminoácidos expuesta en la SEQ ID NO: 2 o una secuencia de aminoácidos codificada por la secuencia de nucleótidos de la SEQ ID NO: 1, o un polipéptido variante o polinucleótido variante de las misas, donde el polipéptido variante tiene al menos un 70 % de identidad de secuencia con la secuencia expuesta en la SEQ ID NO: 2 o el polinucleótido variante codifica un polipéptido que tiene al menos un 70 % de identidad de secuencia con la secuencia expuesta en la SEQ ID NO: 2.

Los polipéptidos de acuerdo con la invención tienen propiedades favorables, en particular una actividad degradante de celulosa. En una realización, el polipéptido de acuerdo con la invención tiene actividad beta-glucosidasa (abreviada BG).

En este documento, la beta-glucosidasa, también β -glucosidasa (EC 3.2.1.21), es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos terminales de β -D-glucosa no reductora con liberación de β -D-glucosa. Dicho polipéptido puede tener una amplia especificidad por β -D-glucósidos y también puede hidrolizar uno o más de los siguientes: un β -D-galactósido, un α -L-arabinósido, un β -D-xilósido o un β -D-fucósido. Esa enzima también puede mencionarse como amigdalasa, β -D-glucósido glucohidrolasa, celobiasa o gentiobiasa.

En una realización, el polipéptido variante tiene un resto catalítico Asp276-Glu508 (posiciones como en la SEQ ID NO: 2).

En una realización, el polipéptido variante tiene al menos un 70 % de identidad de secuencia con la secuencia expuesta en la SEQ ID NO: 2, un resto catalítico Asp276 - Glu508 y actividad BG.

Además, los polipéptidos pueden tener una alta termoestabilidad. Los polipéptidos de acuerdo con la invención pueden retener una alta actividad relativa (% de actividad inicial) como una función del tiempo de incubación (h), por ejemplo, 2 horas, 3 horas, 4 horas, cinco horas, seis horas, ocho horas, nueve horas, 10 h o más, 20 h o más, 30 h o más, en particular a altas temperaturas, por ejemplo, a 60 °C o más, a 65 °C o más, o a 70 °C o más, por ejemplo, 8 horas a 65 °C o 72 horas a 60 °C.

La invención también proporciona un polinucleótido que comprende:

- (a) la secuencia de nucleótidos expuesta en la SEQ ID NO: 1; o
- (b) una secuencia de nucleótidos que hibrida de forma selectiva con un polinucleótido que es el complemento inverso de la SEQ ID NO: 1; o
- (c) una secuencia de nucleótidos que tiene al menos aproximadamente un 75 % de identidad de secuencia con la secuencia de nucleótidos de la SEQ ID NO: 1; o
- (d) un fragmento de una secuencia de nucleótidos definida en (a), (b) o (c) que es de al menos aproximadamente 100 nucleótidos de longitud; o
- (e) una secuencia que es una degeneración, como resultado del código genético, de una secuencia definida en uno cualquiera de (a), (b), (c) o (d); o
- (f) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos definida en (a), (b), (c), (d) o (e).

También se proporciona, de acuerdo con la invención, un vector, tal como un vector de expresión, que incorpora una secuencia polinucleotídica de la invención y una célula que comprende un polipéptido, un polinucleótido o un vector de la invención.

La invención también proporciona:

- un método para la preparación de un polipéptido que tiene actividad beta-glucosidasa, comprendiendo dicho método cultivar una célula de la invención en condiciones que permiten la expresión de dicho polipéptido y, opcionalmente, recuperar el polipéptido expresado;
- un polipéptido que se puede obtener por dicho método; y

una composición que comprende: (i) un polipéptido de la invención y; (ii) una celulasa y/o una hemicelulasa y/o una pectinasa.

Los polipéptidos de la invención que tienen actividad beta-glucosidasa pueden usarse en procesos industriales. Por tanto, la invención proporciona un método para el tratamiento de un sustrato que comprende material de carbohidrato, comprendiendo dicho método poner en contacto el sustrato con un polipéptido o una composición de la invención.

En particular, la invención proporciona un método para producir un azúcar o azúcares a partir de material lignocelulósico, comprendiendo dicho método poner en contacto el material lignocelulósico con un polipéptido o una composición de la invención.

Los azúcares producidos de este modo pueden usarse en un proceso de fermentación. Por consiguiente, la invención proporciona un método para producir un producto de fermentación, comprendiendo dicho método: producir un azúcar fermentable usando los descrito anteriormente; y fermentar el azúcar fermentable resultante, para producir de ese modo un producto de fermentación.

Un polipéptido o una composición de la invención también puede usarse, por ejemplo, en la preparación de un producto alimenticio, en la preparación de un detergente, en la preparación de un pienso para animales, en el tratamiento de pulpa o en la fabricación de papel o en la preparación de una tela o tejido o en la limpieza de los mismos.

La invención también proporciona:

un material procesado que se puede obtener poniendo en contacto un material vegetal o material lignocelulósico con un polipéptido o una composición de la invención;

un alimento o pienso que comprende un polipéptido o una composición de la invención; y

una planta o una parte de la misma que comprende un polinucleótido, un polipéptido, un vector o una célula de acuerdo con la invención.

Breve descripción de los dibujos

Fig. 1: Mapa de pGBTOP para la expresión de genes en *A. niger*. Se representan el gen de interés (GOI) expresado a partir del promotor de glucoamilasa (PglA). Además, se representa el lateral de glucoamilasa (3'glA) del casete de expresión. En esta aplicación, un gen de interés es la secuencia codificante de TEMER06387 como se define posteriormente en este documento.

Breve descripción de la lista de secuencias

La SEQ ID NO: 1 presenta la secuencia codificante de TEMER06387;

La SEQ ID NO: 2 presenta la secuencia de aminoácidos de TEMER06387;

La SEQ ID NO: 3 presenta la secuencia señal de TEMER06387.

Descripción detallada de la invención

Durante toda la presente memoria descriptiva y las reivindicaciones adjuntas, las palabras "comprender" e "incluir" y variaciones tales como "comprende", "que comprende", "incluye" y "que incluye" deben interpretarse globalmente. Es decir, estas palabras pretenden transmitir la posible inclusión de otros elementos o enteros no enumerados específicamente, donde lo permita el contexto.

Los artículos "un" y "una" se usan en este documento para hacer referencia a uno o más de uno (es decir, a uno o al menos uno) de un objeto gramatical del artículo. A modo de ejemplo, "un elemento" puede significar un elemento o más de un elemento.

La presente invención proporciona polinucleótidos que codifican polipéptidos, por ejemplo, enzimas que tienen la capacidad de modificar, por ejemplo, degradar, un material de carbohidrato. Un material de carbohidrato es un material que comprende, consiste en o consiste sustancialmente en uno o más carbohidratos. Las enzimas son, en este documento, una subclase de polipéptidos.

Sustrato (también llamado materia prima) en este documento se usa para hacer referencia a una sustancia que comprende material de carbohidrato, que puede tratarse con enzimas de acuerdo con la invención, de modo que se modifique el material de carbohidrato en el mismo. Además del material de carbohidrato, el sustrato puede contener cualquier otro componente, incluyendo, aunque sin limitación material no de carbohidrato y almidón.

La presente invención proporciona polinucleótidos que codifican polipéptidos, por ejemplo, enzimas que tienen la capacidad de modificar, por ejemplo, degradar, un material de carbohidrato. Un material de carbohidrato es un material que comprende, consiste en o consiste sustancialmente en uno o más carbohidratos. Las enzimas son, en

este documento, una subclase de polipéptidos.

Típicamente, un polipéptido de la invención codifica un polipéptido que tiene al menos actividad beta-glucosidasa, provisionalmente llamada TEMER06387, que tiene una secuencia de aminoácidos de acuerdo con la SEQ ID NO: 2, o una secuencia que es una variante de la misma, típicamente equivalente funcionalmente al polipéptido que tiene la secuencia de la SEQ ID NO: 2, o una secuencia que es un fragmento de cualquiera de las mismas.

En una realización, un polipéptido de la invención puede tener una o más actividades alternativas y/o adicionales diferentes de la actividad beta-glucosidasa mencionada anteriormente, por ejemplo, una de las otras actividades degradantes de carbohidrato y/o hidrolizantes de carbohidrato mencionadas en este documento.

Carbohidrato, en este contexto, incluye todos los sacáridos, por ejemplo, polisacáridos, oligosacáridos, disacáridos o monosacáridos.

Un polipéptido de acuerdo con la invención puede modificar y/o degradar un material de carbohidrato por degradación química o degradación física de dicho material o hidrólisis del carbohidrato. La modificación física incluye, por ejemplo, interrupción de la interacción entre las microfibrillas de celulosa y/o abertura de la estructura de fibras de celulosa. La modificación química del material carbohidrato puede provocar la degradación de dicho material, por ejemplo, por hidrólisis, oxidación u otra modificación química tal como por la acción de una liasa. La modificación física puede estar acompañada o no por modificación química.

Materiales adecuados de carbohidrato

Un carbohidrato que no es almidón adecuado para modificación por un polipéptido de la invención es lignocelulosa. Los polisacáridos principales que comprenden diferentes restos lignocelulósicos, que pueden considerarse como una potencial materia prima renovable son celulosa (glucanos), hemicelulosas (xilanos, heteroxilanos y xiloglucanos). Además, puede haber algo de hemicelulosa presente como glucomananos, por ejemplo, en materias primas derivadas de la madera. La hidrólisis enzimática de estos polisacáridos en azúcares solubles, por ejemplo, glucosa, xilosa, arabinosa, galactosa, fructosa, manosa, ramnosa, ribosa, ácido D-galacturónico y otras hexosas y pentosas sucede bajo la acción de diferentes enzimas que actúan concertadamente.

Además, las pectinas y otras sustancias pécticas tales como arabinanos pueden componer una proporción considerable de la masa seca de las paredes celulares típicas de tejidos vegetales no de madera (de aproximadamente una cuarta parte de la mitad de la masa seca puede ser pectinas).

La celulosa es un polisacárido lineal compuesto de restos de glucosa unidos por enlaces β -1,4. La naturaleza lineal de las fibras de celulosa, así como la estequiometría de la glucosa β -unida (respecto a α) genera estructuras más propensas a formación de enlaces intercatenarios de hidrógeno que las estructuras α -unidas altamente ramificadas del almidón. Por tanto, los polímeros de celulosa son generalmente menos solubles, y forman fibras unidas más apretadamente que las fibras encontradas en el almidón.

La hemicelulosa es un polímero complejo, y su composición a menudo varía ampliamente de un organismo a otro, y de un tipo de tejido a otro. En general, un componente principal de la hemicelulosa es una xilosa β -1,4-unida, un azúcar de cinco carbonos. Sin embargo, esta xilosa a menudo está ramificada en O-3 y/u O-2 y puede sustituirse con enlaces a arabinosa, galactosa, manosa, ácido glucurónico, ácido galacturónico o por esterificación en ácido acético (y esterificación de ácido ferúlico en arabinosa). La hemicelulosa también puede contener glucano, que es un término general para azúcares de seis carbonos β -unidos (tales como los glucanos β -(1,3)(1,4) y los heteroglucanos mencionados previamente) y adicionalmente glucomananos (en que están presentes tanto glucosa como manosa en la estructura lineal, unidos entre sí por enlaces β).

La composición, naturaleza de sustitución, y grado de ramificación de la hemicelulosa es muy diferente en plantas dicotiledóneas (dicot., es decir, plantas cuyas semillas tienen dos cotiledones u hojas en la semilla tales como judía blanca, cacahuetes, almendras, guisantes, alubias) en comparación con las plantas monocotiledóneas (monocot.; es decir, plantas que tienen un único cotiledón u hoja en la semilla tales como maíz, trigo, arroz, gramíneas, cebada). En las dicot., la hemicelulosa está compuesta principalmente por xiloglucanos que son cadenas de glucosa 1,4- β -unidas con cadenas laterales de xilosilo 1,6- β -unidos. En monocot., incluyendo la mayoría de los cultivos de cereal, los componentes principales de la hemicelulosa son heteroxilanos. Estos están compuestos principalmente por polímeros estructurales de xilosa 1,4- β -unida con enlaces 1,3- α a arabinosa, galactosa, manosa y ácido glucurónico o ácido 4-O-metil-glucurónico, así como xilosa modificada por ácido acéticos unidos por éster. También hay β glucanos presentes compuestos por cadenas de glucosilo 1,3- y 1,4- β -unidos. En monocot., puede haber celulosa, heteroxilanos y β -glucanos presentes en cantidades casi iguales, comprendiendo cada uno aproximadamente el 15-25 % de la materia seca de las paredes celulares. Además, diferentes plantas pueden comprender diferentes cantidades de, y diferentes composiciones de, sustancias pécticas. Por ejemplo, la remolacha azucarera contiene aproximadamente el 19 % de pectina y aproximadamente el 21 % de arabinano en una base en peso seco.

Por consiguiente, una composición de la invención puede adaptarse en vista de la materia prima particular (también

llamada sustrato) que se tiene que usar. Es decir, el espectro de actividades en una composición de la invención puede variar dependiendo de la materia prima en cuestión.

Pueden administrarse combinaciones enzimáticas o tratamientos físicos de forma concomitante o secuencial. Las enzimas pueden producirse de forma exógena en microorganismos, levaduras, hongos, bacterias o plantas, después pueden aislarse y añadirse a la materia prima lignocelulósica. Como alternativa, las enzimas se producen, pero no se aíslan, y se añade caldo de fermentación de masa celular sin procesar, o material vegetal (tal como rastrojo de maíz), y similares a la materia prima. Como alternativa, la masa celular sin procesar o medio de producción enzimática o material vegetal puede tratarse para evitar el crecimiento microbiano adicional (por ejemplo, por calentamiento o adición de agentes antimicrobianos), después añadirse a la materia prima. Estas mezclas enzimáticas sin procesar pueden incluir el organismo productor de la enzima. Como alternativa, la enzima puede producirse en una fermentación que usa materia prima (tal como rastrojo de maíz) para proporcionar nutrición a un organismo que produce una o más enzimas. De este modo, las plantas que producen las enzimas pueden servir como materia prima lignocelulósica y añadirse en materia prima lignocelulósica.

Actividad enzimática

Las endo-1,4- β -glucanasas (EG) y exo-celobiohidrolasas (CBH) catalizan la hidrólisis de celulosa insoluble en celooligosacáridos (celobiosa como producto principal), mientras que las β -glucosidasas (BGL) convierten los oligosacáridos, principalmente celobiosa y celotriosa en glucosa.

Las xilanasas junto con otras enzimas accesorias, por ejemplo, α -L-arabinofuranosidasas, feruloil y acetilxilano esterases, glucuronidasas, y β -xilosidasas) catalizan la hidrólisis de parte de las hemicelulosas.

Las sustancias pécticas incluyen pectinas, arabinanos, galactanos y arabinogalactanos. Las pectinas son los polisacáridos más complejos en la pared celular de las plantas. Se construyen alrededor de una cadena central de unidades de ácido D-galacturónico α (1,4)-unido entremezcladas en algún grado con L-ramnosa. En una parece celular cualquier hay varias unidades estructurales que se ajusta a esta descripción y generalmente se ha considerado que, en una única molécula péctica, las cadenas centrales de diferentes unidades estructurales son continuas con las otras.

Las pectinasas incluyen, por ejemplo, una endo-poligalacturonasa, una pectin metil esterasa, una endo-galactanasa, una beta galactosidasa, una pectin acetil esterasa, una endo-pectin liasa, pectato liasa, alfa ramnosidasa, una exo-galacturonasa, una expoligalacturonato liasa, una ramnogalacturonano hidrolasa, una ramnogalacturonano liasa, una ramnogalacturonano acetil esterasa, una ramnogalacturonano galacturonohidrolasa, una xilogalacturonasa, una α -arabinofuranosidasa.

Los tipos principales de unidad estructural son: galacturonano (homogalacturonano), que puede estar sustituido con metanol en el grupo carboxilo y acetato en O-2 y O-3; ramnogalacturonano I (RGI), en que las unidades de ácido galacturónico alternan con unidades de ramnosa que portan cadenas laterales de galactano (1,4)-unido y arabinano (1,5)-unido. Las cadenas laterales de arabinano pueden unirse directamente a ramnosa o indirectamente a través de las cadenas de galactano; xilogalacturonano, con unidades únicas de xilosilo en O-3 de ácido galacturónico (muy relacionado con RGI); y ramnogalacturonano II (RGII), una unidad minoritaria particularmente compleja que contiene azúcares inusuales, por ejemplo, apiosa. Una unidad RGII puede contener dos restos de apiosilo que, en condiciones iónicas adecuadas, pueden formar de forma reversible ésteres con borato.

Como se ha expuesto anteriormente, un polipéptido de la invención típicamente tendrá actividad beta-glucosidasa. Sin embargo, un polipéptido de la invención puede tener una o más de las actividades expuestas anteriormente además de o como alternativa a esa actividad. Además, una composición de la invención como se describe en este documento puede tener una o más de las actividades mencionadas anteriormente además de la proporcionada por un polipéptido de la invención que tiene actividad beta-glucosidasa.

Secuencia polinucleotídica

La invención proporciona secuencias polinucleotídicas genómicas que comprenden el gen que codifica TEMER06387, así como su secuencia codificante. Por consiguiente, la invención se refiere a un polinucleótido aislado que comprende la secuencia genómica de nucleótidos de acuerdo con la secuencia codificante de nucleótidos de acuerdo con la SEQ ID NO: 1 y a variantes, tales como equivalentes funcionales, de cualquiera de las mismas.

En particular, la invención se refiere a un polinucleótido aislado que es capaz de hibridar selectivamente, por ejemplo, en condiciones rigurosas, preferiblemente en condiciones altamente rigurosas, con el complemento inverso de un polinucleótido que comprende la secuencia expuesta en la SEQ ID NO: 1.

Más específicamente, la invención se refiere a un polinucleótido que comprende o consiste esencialmente en una secuencia de nucleótidos de acuerdo con la SEQ ID NO: 1.

La invención también se refiere a un polinucleótido aislado que comprende o consiste esencialmente en una secuencia que codifica al menos un dominio funcional de un polipéptido de acuerdo con la SEQ ID NO: 2 o una variante de la misma, tal como un equivalente funcional, o un fragmento de cualquiera de las mismas.

Como se usa en este documento, las expresiones "gen" y "gen recombinante" se refieren a moléculas de ácido nucleico que pueden aislarse de ADN cromosómico, que incluyen una fase de lectura abierta que codifica una proteína, por ejemplo, la proteína beta-glucosidasa de acuerdo con la presente invención.

Un gen puede incluir secuencias codificantes, secuencias no codificantes, intrones y/o secuencias reguladoras. Además, el término "gen" puede referirse a una molécula aislada de ácido nucleico como se define en este documento.

Una molécula de ácido nucleico de la presente invención, tal como una molécula de ácido nucleico que tiene la secuencia de nucleótidos de la SEQ ID NO: 1 o una variante de la misma, tal como un equivalente funcional, puede aislarse usando técnicas convencionales de biología molecular y la información de secuencia proporcionada en este documento. Por ejemplo, usando todo o una parte de la secuencia de ácido nucleico de la SEQ ID NO: 1 como sonda de hibridación, pueden aislarse moléculas de ácido nucleico de acuerdo con la invención usando técnicas convencionales de hibridación y clonación (por ejemplo, como se describe en Sambrook, J., Fritsh, E. F., y Maniatis, T. Molecular Cloning: A Laboratory Manual. 2ª ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Además, puede aislarse una molécula de ácido nucleico que abarca todo o una parte de la SEQ ID NO: 1 por la reacción en cadena de la polimerasa (PCR) usando cebadores oligonucleotídicos sintéticos diseñados en base a la información de secuencia contenida en la SEQ ID NO: 1.

Un ácido nucleico de la invención puede amplificarse usando ADNc, ARNm o, como alternativa, ADN genómico, como molde y cebadores oligonucleotídicos apropiados de acuerdo con técnicas convencionales de amplificación por PCR. El ácido nucleico así amplificado puede clonarse en un vector apropiado y caracterizarse por análisis de secuencia de ADN.

Además, pueden prepararse oligonucleótidos correspondientes a o que pueden hibridar con una secuencia de nucleótidos de acuerdo con la invención por técnicas sintéticas convencionales, por ejemplo, usando un sintetizador automatizado de ADN.

En una realización preferida, una molécula aislada de ácido nucleico de la invención comprende la secuencia de nucleótidos mostrada en la SEQ ID NO: 1.

En otra realización preferida, una molécula aislada de ácido nucleico de la invención comprende una molécula de ácido nucleico que es el complemento inverso de la secuencia de nucleótidos mostrada en la SEQ ID NO: 1 o una variante, tal como un equivalente funcional, de cualquiera de dichas secuencias de nucleótidos.

Una molécula de ácido nucleico que es complementaria a otra secuencia de nucleótidos es una que es suficientemente complementaria a la otra secuencia de nucleótidos de modo que pueda hibridar con la otra secuencia de nucleótidos formando de ese modo un dúplex estable.

Un aspecto de la invención se refiere a moléculas de ácido nucleico que codifican un polipéptido de la invención o una variante, tal como un equivalente funcional del mismo, por ejemplo, un fragmento o dominio biológicamente activo, así como moléculas de ácido nucleico suficientes para su uso como sondas de hibridación para identificar moléculas de ácido nucleico que codifican un polipéptido de la invención y fragmentos de dichas moléculas de ácido nucleico adecuados para su uso como cebadores de PCR para la amplificación o mutación de moléculas de ácido nucleico.

Un polinucleótido de acuerdo con la invención puede estar "aislado". En el contexto de esta invención, un "polinucleótido aislado" o "ácido nucleico aislado" es un ADN o ARN que no está inmediatamente contiguo con una o las dos secuencias codificantes con las que está inmediatamente contiguo (una en el extremo 5' y una en el extremo 3') en el genoma de origen natural del organismo del cual se obtiene. Por tanto, en una realización, un ácido nucleico aislado incluye algunas o todas las secuencias 5' no codificantes (por ejemplo, promotoras) que están inmediatamente contiguas a la secuencia codificante. La expresión, por lo tanto, incluye, por ejemplo, un ADN recombinante que se incorpora en un vector, en un plásmido de replicación autónoma o virus, o en el ADN genómico de un procarionte o eucarionte, o que existe como una molécula separada (por ejemplo, un ADNc o un fragmento de ADN genómico producido por PCR o tratamiento endonucleasa de restricción) independiente de otras secuencias. También incluye un ADN recombinante que es parte de un gen híbrido que codifica un polipéptido adicional que está sustancialmente libre de material celular, material vírico, o medio de cultivo (cuando se produce por técnicas de ADN recombinante), o precursores químicos u otros agentes químicos (cuando se sintetiza químicamente). Además, un "fragmento de ácido nucleico aislado" es un fragmento de ácido nucleico que no existe de forma natural como un

fragmento y no se encontraría en estado natural.

Como se usa en este documento, las expresiones "polinucleótido" o "molécula de ácido nucleico" pretenden incluir moléculas de ADN (por ejemplo, ADNc o ADN genómico) y moléculas de ARN (por ejemplo, ARNm) y análogos del ADN o ARN generado usando análogos nucleotídicos. La molécula de ácido nucleico puede ser monocatenaria o bicatenaria, pero preferiblemente es ADN bicatenario. El ácido nucleico puede sintetizarse usando análogos o derivados oligonucleotídicos (por ejemplo, nucleótidos de inosina o fosforotioato). Dichos oligonucleótidos pueden usarse, por ejemplo, para preparar ácidos nucleicos que tienen capacidades alteradas de apareamiento de bases o resistencia aumentada a nucleasas.

Otra realización de la invención proporciona una molécula aislada de ácido nucleico que es antisentido a una molécula de ácido nucleico de TEMER06387, por ejemplo, la hebra codificante de una molécula de ácido nucleico de TEMER06387. También se incluyen dentro del alcance de la invención las hebras complementarias de las moléculas de ácido nucleico descritas en este documento.

Salvo que se indique de otro modo, todas las secuencias de nucleótidos determinadas por secuenciación de una molécula de ADN en este documento se determinaron usando un secuenciador automatizado de ADN y todas las secuencias de aminoácidos de polipéptidos codificadas por moléculas de ADN determinadas en este documento se predijeron por traducción de una secuencia de ADN determinada como anteriormente. Por lo tanto, como se sabe en la técnica para cualquier secuencia de ADN determinada por este enfoque automatizado, cualquier secuencia de nucleótidos determinada en este documento puede contener algunos errores. Las secuencias de nucleótidos determinadas por automatización son típicamente al menos aproximadamente un 90 % idénticas, más típicamente al menos aproximadamente un 95 % a al menos aproximadamente un 99,9 % idénticas a la secuencia real de nucleótidos de la molécula de ADN secuenciada.

La secuencia real puede determinarse de forma más precisa por otros enfoques incluyendo métodos de secuenciación manual de ADN bien conocidos en la técnica. Como también se sabe en la técnica, una única inserción o delección en una secuencia de nucleótidos determinada en comparación con la secuencia real causará un desplazamiento de fase en la traducción de la secuencia de nucleótidos de modo que la secuencia de aminoácidos predicha codificada por una secuencia de nucleótidos determinada será completamente diferente de la secuencia de aminoácidos realmente codificada por la molécula de ADN secuenciada, empezando en el punto de dicha inserción o delección.

Los expertos en la materia son capaces de identificar dichas bases identificadas de forma errónea y conocen el modo de corregir dichos errores.

Una molécula de ácido nucleico de acuerdo con la invención puede comprender solamente una parte o un fragmento de la secuencia de ácido nucleico mostrada en la SEQ ID NO: 1 (o de una variante de cualquiera de las mismas), por ejemplo, un fragmento que puede usarse como sonda o cebador o un fragmento que codifica una parte de un polipéptido TEMER06387.

La secuencia de nucleótidos determinada a partir de la clonación del gen y el ADNc de TEMER06387 permite la generación de sondas y cebadores diseñados para su uso en la identificación y/o clonación de otros miembros de la familia de TEMER06387, así como homólogos de TEMER06387 de otras especies.

La sonda/cebador típicamente comprende un oligonucleótido sustancialmente purificado que típicamente comprende una región de secuencia de nucleótidos que hibrida preferiblemente en condiciones altamente rigurosas con de al menos aproximadamente 12 a aproximadamente 15, preferiblemente de aproximadamente 18 a aproximadamente 20, preferiblemente de aproximadamente 22 a aproximadamente 25, más preferiblemente aproximadamente 30, aproximadamente 35, aproximadamente 40, aproximadamente 45, aproximadamente 50, aproximadamente 55, aproximadamente 60, aproximadamente 65, o aproximadamente 75 o más nucleótidos consecutivos de una secuencia de nucleótidos mostrada en la SEQ ID NO: 1 o de una variante, tal como un equivalente funcional, de cualquiera de las mismas.

Pueden usarse sondas basadas en las secuencias de nucleótidos de TEMER06387 para detectar transcritos o secuencias genómicas de TEMER06387 que codifican los mismos polipéptidos u homólogos, por ejemplo, en otros organismos. En realizaciones preferidas, la sonda comprende adicionalmente un grupo marcador unido a la misma, por ejemplo, el grupo marcador puede ser un radioisótopo, un compuesto fluorescente, una enzima, o un cofactor enzimático. Dichas sondas también pueden usarse como parte de un kit de ensayo de diagnóstico para identificar células que expresan un polipéptido TEMER06387.

Los polinucleótidos de este documento pueden ser polinucleótidos sintéticos. Los polinucleótidos sintéticos puede optimizarse en el uso de codones, preferiblemente de acuerdo con los métodos descritos en el documento WO2006/077258 y/o el documento PCT/EP2007/055943. El documento PCT/EP2007/055943 aborda la optimización de pares de codones. La optimización de pares de codones es un método donde las secuencias de nucleótidos que codifican un polipéptido se han modificado con respecto a su uso de codones, en particular los pares de codones

que se usan para obtener expresión mejorada de la secuencia de nucleótidos que codifica el polipéptido y/o producción mejorada del polipéptido codificado. Los pares de codones se definen como un conjunto de dos tripletes (codones) posteriores en una secuencia codificante.

La invención se refiere adicionalmente a una construcción de ácido nucleico que comprende el polinucleótido que se ha descrito anteriormente. "Construcción de ácido nucleico" se define en este documento como una molécula de ácido nucleico, mono o bicatenaria, que se aísla de un gen de origen natural o que se ha modificado para que contenga segmentos de ácido nucleico que se combinan y yuxtaponen de un modo que no existiría de otro modo en la naturaleza. La expresión construcción de ácido nucleico es sinónima de la expresión "casete de expresión" cuando la construcción de ácido nucleico contiene todas las secuencias de control necesarias para la expresión de una secuencia codificante. La expresión "secuencia codificante" como se define en este documento es una secuencia, que se transcribe en ARNm y se traduce en un activador transcripcional de un promotor de proteasa de la invención. Los límites de la secuencia codificante generalmente se determinan por el codón de inicio ATG en el extremo 5' del ARNm y una secuencia de codón de parada de la traducción que termina la fase de lectura abierta en el extremo 3' del ARNm. Una secuencia codificante puede incluir, pero sin limitación, ADN, ADNc, y secuencias recombinantes de ácido nucleico. Preferiblemente, el ácido nucleico tiene alto contenido en GC. El contenido en GC en este documento indica la cantidad de nucleótidos G y C en la construcción, dividida por la cantidad total de nucleótidos, expresada en %. El contenido en GC es preferiblemente del 56 % o más, del 57 % o más, del 58 % o más, del 59 % o más, del 60 % o más, o en el intervalo del 56-70 % o el intervalo del 58-65 %.

Preferiblemente, la construcción de ADN comprende una secuencia de ADN promotora, una secuencia codificante en asociación funcional con dicha secuencia de ADN promotora y secuencias de control tales como:

- una secuencia de terminación de la traducción orientada en dirección 5' hacia 3' seleccionada de la siguiente lista de secuencias: TAAG, TAGA y TAAA, preferiblemente TAAA, y/o
- una secuencia codificante iniciadora de la traducción orientada en la dirección 5' hacia 3' seleccionada de la siguiente lista de secuencias: GCTACCCCC; GCTACCTCC; GCTACCCTC; GCTACCTTC; GCTCCCCCC; GCTCCCTCC; GCTCCCCTC; GCTCCCTTC; GCTGCCCCC; GCTGCCTCC; GCTGCCCTC; GCTGCCTTC; GCTTCCCCC; GTTTCTCC; GTTTCCCTC; y GCTTCCTTC, preferiblemente GCT TCC TTC, y/o
- una secuencia iniciadora de la traducción seleccionada de la siguiente lista de secuencias: 5'-mwChkyCAAA-3'; 5'-mwChkyCACA-3' o 5'-mwChkyCAAG-3', que usan códigos de ambigüedad para los nucleótidos: m (A/C); w (A/T); y (C/T); k (G/T); h (A/C/T), preferiblemente 5'-CACCGTCAAA-3' o 5'-CGCAGTCAAG-3'.

En el contexto de esta invención, la expresión "secuencia codificante iniciadora de la traducción" se define como los nueve nucleótidos inmediatamente cadena abajo del codón iniciador o de inicio de la fase de lectura abierta de una secuencia codificante de ADN. El codón iniciador o de inicio codifica la metionina AA. El codón iniciador es típicamente ATG, pero también puede ser cualquier codón de inicio funcional tal como GTG.

En el contexto de esta invención, la expresión "secuencia de terminación de la traducción" se define como los cuatro nucleótidos partiendo del codón de parada de la traducción en el extremo 3' de la fase de lectura abierta o secuencia codificante de nucleótidos y orientada en la dirección 5' hacia 3'.

En el contexto de esta invención, la expresión "secuencia iniciadora de la traducción" se define como los diez nucleótidos inmediatamente cadena arriba del codón iniciador o de inicio de la fase de lectura abierta de una secuencia de ADN que codifica un polipéptido. El codón iniciador o de inicio codifica la metionina AA. El codón iniciador es típicamente ATG, pero también puede ser cualquier codón de inicio funcional tal como GTG. Es bien sabido en la técnica que el uracilo, U, reemplaza el desoxinucleótido timina, T, en el ARN.

Homología e identidad

Se dice que las secuencias de aminoácidos o nucleótidos son homólogas cuando muestran un cierto nivel de similitud. Dos secuencias que son homólogas indican un origen evolutivo común. Que dos secuencias homólogas estén estrechamente relacionadas o relacionadas de forma más distante se indica por el "porcentaje de identidad" o "porcentaje de similitud", que es alta o baja respectivamente. Aunque está cuestionado, indicar "porcentaje de identidad" o "porcentaje de similitud", "nivel de homología" o "porcentaje de homología" frecuentemente se usan de forma intercambiable.

Las expresiones "homología", "porcentaje de homología", "porcentaje de identidad" o "porcentaje de similitud" se usan de forma intercambiable en este documento. Para el propósito de esta invención, se define aquí que para determinar el porcentaje de identidad de dos secuencias de aminoácidos o de dos secuencias de ácido nucleico, las secuencias completas se alinean con propósitos de comparación óptima. Para optimizar la alineación entre las dos secuencias pueden introducirse huecos en cualquiera de las dos secuencias que se comparan. Dicha alineación se realiza sobre la longitud completa de las secuencias que se están comparando. Como alternativa, la alineación puede realizarse sobre una longitud más corta, por ejemplo, sobre aproximadamente 20, aproximadamente 50, aproximadamente 100 o más ácidos nucleicos/bases o aminoácidos. La identidad es el porcentaje de coincidencias idénticas entre las dos secuencias sobre la región alineada presentada.

Una comparación de secuencias y determinación del porcentaje de identidad entre dos secuencias puede realizarse usando un algoritmo matemático. Los expertos en la materia serán conscientes del hecho de que están disponibles varios programas informáticos diferentes para alinear dos secuencias y determinar la homología entre dos secuencias (Kruskal, J. B. (1983) An overview of sequence comparison En D. Sankoff y J. B. Kruskal, (ed.), Time warps, string edits and macromolecules: the theory and practice of sequence comparison, pág. 1-44 Addison Wesley). El porcentaje de identidad entre dos secuencias de aminoácidos puede determinarse usando el algoritmo de Needleman y Wunsch para la alineación de dos secuencias. (Needleman, S. B. y Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). El algoritmo alinea secuencias de aminoácidos, así como secuencias de nucleótidos. El algoritmo de Needleman-Wunsch se ha implementado en el programa informático NEEDLE. Para el propósito de esta invención, se usó el programa NEEDLE del paquete EMBOSS (versión 2.8.0 o superior, *EMBOSS: The European Molecular Biology Open Software Suite* (2000) Rice, P. Longden, I. y Bleasby, A. Trends in Genetics 16, (6) pág. 276-277, <http://emboss.bioinformatics.nl/>). Para secuencias polipeptídicas, se usa BLOSUM62 para la matriz de sustitución. Para secuencias de nucleótidos, se usa EDNAFULL. Pueden especificarse otras matrices. Los parámetros opcionales usados para la alineación de secuencias de aminoácidos son una penalización por apertura de hueco de 10 y una penalización por extensión de hueco de 0,5. Los expertos en la materia apreciarán que todos estos diferentes parámetros producirán resultados ligeramente diferentes pero que el porcentaje global de identidad de dos secuencias no se altera significativamente cuando se usan diferentes algoritmos.

Definición de homología global

La homología o identidad es el porcentaje de coincidencias idénticas entre las dos secuencias completas sobre la región alineada total incluyendo cualquier hueco o extensión. La homología o identidad entre las dos secuencias alineadas se calcula del siguiente modo: Cantidad de posiciones correspondientes en la alineación que muestran un aminoácido idéntico en ambas secuencias dividida por la longitud total de la alineación incluyendo los huecos. La identidad definida en este documento puede obtenerse a partir de NEEDLE y se etiqueta en la salida del programa como "IDENTIDAD".

Definición de identidad más larga

La homología o identidad entre las dos secuencias alineadas se calcula del siguiente modo: Cantidad de posiciones correspondientes en la alineación que muestran un aminoácido idéntico en ambas secuencias dividida por la longitud total de la alineación después de restar la cantidad total de huecos en la alineación. La identidad definida en este documento puede obtenerse a partir de NEEDLE usando la opción NOBRIEF y se etiqueta en la salida del programa como "identidad más larga". Para los propósitos de la invención el nivel de identidad (homología) entre dos secuencias (de aminoácidos o nucleótidos) se calcula de acuerdo con la definición de "identidad más larga" que puede realizarse usando el programa NEEDLE.

Las secuencias polipeptídicas de la presente invención pueden usarse adicionalmente como una "secuencia de consulta" para realizar una búsqueda frente a bases de datos de secuencias, por ejemplo, para identificar otros miembros de la familia o secuencias relacionadas. Dichas búsquedas pueden realizarse usando los programas BLAST. El software para realizar análisis BLAST está disponible al público a través del National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov>). BLASTP se usa para secuencias de aminoácidos y BLASTN para secuencias de nucleótidos. El programa BLAST usa como defectos:

- Valor de apertura de hueco: defecto = 5 para nucleótidos/11 para polipéptidos
- Valor de hueco extendido: defecto = 2 para nucleótidos/1 para polipéptidos
- Penalización por desapareamiento de nucleótidos: defecto = -3
- Recompensa por apareamiento de nucleótidos: defecto = 1
- Valor esperado: defecto = 10
- Tamaño de palabra: defecto = 11 para nucleótidos/28 para megablast/3 para polipéptidos

Además, el grado de identidad local (homología) entre la secuencia de aminoácidos de consulta o la secuencia de ácido nucleico de consulta y las secuencias homólogas recuperadas se determina por el programa BLAST. Sin embargo, se comparan solamente aquellos segmentos de secuencia que dan una coincidencia por encima de un cierto umbral. Por consiguiente, el programa calcula la identidad solamente para estos segmentos coincidentes. Por lo tanto, la identidad calculada de este modo se menciona como identidad local.

Hibridación

Como se usa en este documento, la expresión "que hibrida selectivamente", "hibrida selectivamente" y expresiones similares pretenden describir condiciones para hibridación y lavado en que las secuencias de nucleótidos al menos aproximadamente un 50 %, al menos aproximadamente un 55 %, al menos aproximadamente un 60 %, al menos aproximadamente un 65 %, al menos aproximadamente un 70 %, al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, más preferiblemente al menos aproximadamente un 85 %, incluso más preferiblemente al menos aproximadamente un 90 %, preferiblemente al menos un 95 %, más preferiblemente al menos aproximadamente un 98 % o más preferiblemente al menos aproximadamente un 99 % homólogas entre sí

permanecen típicamente hibridadas entre sí. Es decir, dichas secuencias de hibridación pueden compartir al menos aproximadamente un 50 %, al menos aproximadamente un 55 %, al menos aproximadamente un 60 %, al menos aproximadamente un 65 %, al menos aproximadamente un 70 %, al menos aproximadamente un 75 %, al menos aproximadamente un 80 %, más preferiblemente al menos aproximadamente un 85 %, incluso más preferiblemente al menos aproximadamente un 90 %, más preferiblemente al menos un 95 %, más preferiblemente al menos un 98 % o más preferiblemente al menos aproximadamente un 99 % de identidad de secuencia.

Un ejemplo preferido no limitante de dichas condiciones de hibridación es la hibridación en cloruro sódico/citrato sódico (SSC) 6 X a aproximadamente 45 °C, seguido de uno o más lavados en SSC 1 X, SDS al 0,1 % a aproximadamente 50 °C, preferiblemente a aproximadamente 55 °C, preferiblemente a aproximadamente 60 °C e incluso más preferiblemente a aproximadamente 65 °C.

Las condiciones altamente rigurosas incluyen, por ejemplo, hibridación a aproximadamente 68 °C en SSC 5 x/solución de Denhardt 5 x/SDS al 1,0 % y lavado en SSC 0,2 x/SDS al 0,1 % a temperatura ambiente. Como alternativa, el lavado puede realizarse a 42 °C.

Los expertos en la materia conocerán las condiciones a aplicar para condiciones rigurosas y altamente rigurosas de hibridación. Están fácilmente disponibles en la técnica directrices adicionales respecto a dichas condiciones, por ejemplo, en Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, N.Y.; y Ausubel et al. (eds.), 1995, *Current Protocols in Molecular Biology*, (John Wiley & Sons, N.Y.).

Por supuesto, un polinucleótido que hibrida solamente con una secuencia poli A (tal como el tramo poli(A) 3' terminal de ARNm), o con un tramo complementario de restos T (o U), no se incluiría en un polinucleótido de la invención usado para hibridar específicamente con una parte de un ácido nucleico de la invención, ya que dicho polinucleótido hibridaría con cualquier molécula de ácido nucleico que contenga un tramo poli (A) o el complemento del mismo (por ejemplo, prácticamente cualquier clon de ADNc bicatenario).

En un enfoque típico, pueden explorarse bibliotecas de ADNc construidas a partir de otros organismos, por ejemplo, un hongo filamentoso, en particular de la familia de microorganismos *Trichomaceae*, por ejemplo, del género *Penicillium* tal como *Penicillium decumbens*.

Por ejemplo, pueden explorarse cepas de *Penicillium* para polinucleótidos TEMER06387 homólogos por análisis de transferencia de Northern. Tras la detección de transcritos homólogos a polinucleótidos de acuerdo con la invención, pueden construirse bibliotecas de ADNc a partir del ARN aislado de la cepa apropiada, utilizando técnicas convencionales bien conocidas para los expertos en la materia. Como alternativa, puede explorarse una biblioteca de ADN genómico total usando una sonda capaz de hibridar con un polinucleótido TEMER06387 de acuerdo con la invención.

Pueden aislarse secuencias génicas homólogas, por ejemplo, realizando PCR usando dos combinaciones de cebadores oligonucleotídicos degenerados diseñados en base a las secuencias de nucleótidos mostradas en este documento.

El molde para la reacción puede ser ADNc obtenido por transcripción inversa de ARNm preparado a partir de cepas conocidas por o sospechosas de expresar un polinucleótido de acuerdo con la invención. El producto de PCR puede subclonarse y secuenciarse para asegurar que las secuencias amplificadas representan las secuencias de una nueva secuencia de ácido nucleico de TEMER06387, o un equivalente funcional de la misma.

El fragmento de PCR después puede usarse para aislar un clon de ADNc de longitud completa por una diversidad de métodos conocidos. Por ejemplo, el fragmento amplificado puede marcarse y usarse para explorar una biblioteca de ADNc de bacteriófagos o cósmidos. Como alternativa, el fragmento marcado puede usarse para explorar una biblioteca genómica.

También puede usarse la tecnología PCR para aislar secuencias de ADNc de longitud completa de otros organismos. Por ejemplo, puede aislarse ARN, siguiendo procedimientos convencionales, de una fuente celular o tisular apropiada. Puede realizarse una reacción de transcripción inversa sobre el ARN usando un cebador oligonucleotídico específico para el extremo más 5' del fragmento amplificado para el cebador de la síntesis de la primera hebra.

Al híbrido ARN/ADN resultante después se le puede "añadir una cola" (por ejemplo, con guaninas) usando una reacción convencional de transferasa terminal, puede digerirse el híbrido con RNasa H, y puede después cebarse la síntesis de la segunda hebra (por ejemplo, con un cebador poli-C). por tanto, pueden aislarse fácilmente las secuencias de ADNc cadena arriba del fragmento amplificado. Para una revisión de estrategias útiles de clonación, véase, por ejemplo, Sambrook et al., supra; y Ausubel et al., supra.

Vectores

Otro aspecto de la invención se refiere a vectores, incluyendo vectores de clonación y expresión, que comprenden un polinucleótido de la invención que codifica un polipéptido TEMER06387 o un equivalente funcional del mismo y métodos para cultivar, transformar o transfectar dichos vectores en una célula hospedadora adecuada, por ejemplo, en condiciones en que sucede la expresión de un polipéptido de la invención. Como se usa en este documento, el término "vector" se refiere a una molécula de ácido nucleico capaz de transportar otro ácido nucleico al cual se ha unido.

Los polinucleótidos de la invención pueden incorporarse en un vector replicable recombinante, por ejemplo, un vector de clonación o expresión. El vector puede usarse para replicar el ácido nucleico en una célula hospedadora compatible. Por tanto, en una realización adicional, la invención proporciona un método para preparar polinucleótidos de la invención introduciendo un polinucleótido de la invención en un vector replicable, introduciendo el vector en una célula hospedadora compatible, y cultivando la célula hospedadora en condiciones que provoquen la replicación del vector. El vector puede recuperarse de la célula hospedadora. Se describen a continuación células hospedadoras adecuadas.

El vector en que se inserta el casete de expresión o polinucleótido de la invención puede ser cualquier vector que pueda someterse convenientemente a procedimientos de ADN recombinante, y la elección del vector a menudo dependerá de la célula hospedadora en que tiene que introducirse.

A vector de acuerdo con la invención puede ser un vector de replicación autónoma, es decir, un vector que existe como entidad extracromosómica, cuya replicación es independiente de la replicación cromosómica, por ejemplo, un plásmido. Como alternativa, el vector puede ser uno que, cuando se introduce en una célula hospedadora, se integra en el genoma de la célula hospedadora y se replica junto con el cromosoma o cromosomas en que se ha integrado.

Un tipo de vector es un "plásmido", que se refiere a un lazo circular de ADN bicatenario en que pueden ligarse segmentos adicionales de ADN. Otro tipo de vector es un vector vírico, donde pueden ligarse segmentos adicionales de ADN en el genoma vírico. Ciertos vectores tienen capacidad de replicación autónoma en una célula hospedadora en que se introducen (por ejemplo, vectores bacterianos que tienen un origen bacteriano de replicación y vectores episómicos de mamífero). Otros vectores (por ejemplo, vectores no episómicos de mamífero) se integran en el genoma de una célula hospedadora tras la introducción en la célula hospedadora, y de ese modo se replican junto con el genoma del hospedador. Además, ciertos vectores son capaces de dirigir la expresión de genes a los cuales están unidos de forma funcional. Dichos vectores se mencionan en este documento como "vectores de expresión". En general, los vectores de expresión de utilidad en técnicas de ADN recombinante a menudo están en forma de plásmidos. Las expresiones "plásmido" y "vector" pueden usarse de forma intercambiable en este documento ya que el plásmido es la forma más habitualmente usada de vector. Sin embargo, la invención pretende incluir otras formas de vectores de expresión, tales como cósmidos, vectores víricos (por ejemplo, retrovirus defectuosos en la replicación, adenovirus y virus adeno-asociados) y vectores fágicos que logran funciones equivalentes.

Para la mayoría de hongos filamentosos y levaduras, el vector o construcción de expresión preferiblemente se integra en el genoma de la célula hospedadora para obtener transformantes estables. Sin embargo, para ciertas levaduras también están disponibles vectores episómicos adecuados en que la construcción de expresión puede incorporarse para expresión estable y de alto nivel, ejemplos de los mismos incluyen vectores derivados de los plásmidos 2 μ y pKD1 de *Saccharomyces* y *Kluyveromyces*, respectivamente, o vectores que contienen una secuencia AMA (por ejemplo, AMA1 de *Aspergillus*). En caso de que las construcciones de expresión se integren en el genoma de las células hospedadoras, las construcciones se integran en loci aleatorios en el genoma, o en loci diana predeterminados usando recombinación homóloga, en cuyo caso los loci diana comprenden preferiblemente un gen altamente expresado.

Por consiguiente, los vectores de expresión útiles en la presente invención incluyen vectores derivados de cromosomas, episomas y virus, por ejemplo, vectores derivados de plásmidos bacterianos, bacteriófagos, episomas de levadura, elementos cromosómicos de levadura, virus tales como baculovirus, papovavirus, virus vaccinia, adenovirus, virus de la viruela aviar, virus de la pseudorrabia y retrovirus, y vectores derivados de combinaciones de los mismos, tales como los derivados de elementos genéticos de plásmidos y bacteriófagos, tales como cósmidos y fagómidos.

Los vectores de acuerdo con la invención pueden usarse *en vitro*, por ejemplo, para la producción de ARN o usarse para transfectar o transformar una célula hospedadora. El vector de expresión recombinante puede transcribirse y traducirse *en vitro*, por ejemplo, usando secuencias reguladoras del promotor T7 y la polimerasa T7.

Un vector de la invención puede comprender dos o más, por ejemplo, tres, cuatro o cinco, polinucleótidos de la invención, por ejemplo, para sobre-expresión.

Los vectores de expresión recombinante de la invención comprenden un ácido nucleico de la invención en una forma adecuada para la expresión del ácido nucleico en una célula hospedadora, que significa que el vector de expresión recombinante incluye una o más secuencias reguladoras, seleccionadas en base a las células hospedadoras a usarse para la expresión, que están unidas de forma funcional a la secuencia de ácido nucleico a expresarse.

Dentro de un vector, tal como un vector de expresión, "unido de forma funcional" pretende indicar que la secuencia de nucleótidos de interés está unida a la secuencia o secuencias reguladoras de un modo que permite la expresión de la secuencia de nucleótidos (por ejemplo, en un sistema de transcripción/traducción *en vitro* o en una célula hospedadora cuando se introduce el vector en la célula hospedadora), es decir la expresión "unido de forma funcional" se refiere a una yuxtaposición donde los componentes descritos están en una relación que les permite funcionar en su modo pretendido. Una secuencia reguladora tal como un promotor, potenciador u otra señal de regulación de la expresión "unida de forma funcional" a una secuencia codificante se posiciona de tal modo que se consiga la expresión de la secuencia codificante en condiciones compatibles con las secuencias de control o las secuencias se disponen de modo que funcionen concertadamente para su propósito pretendido, por ejemplo, la transcripción se inicia en un promotor y prosigue a través de la secuencia de ADN que codifica el polipéptido.

Un vector o construcción de expresión para una célula hospedadora dada puede, por tanto, comprender los siguientes elementos unidos de forma funcional entre sí en un orden consecutivo desde el extremo 5' hasta el extremo 3' respecto a la hebra codificante de la secuencia que codifica el polipéptido de la primera invención: (1) una secuencia promotora capaz de dirigir la transcripción de la secuencia de nucleótidos que codifica el polipéptido en la célula hospedadora dada; (2) opcionalmente, una secuencia señal capaz de dirigir la secreción del polipéptido a partir de una célula hospedadora dada en un medio de cultivo; (3) una secuencia de ADN de la invención que codifica una forma madura y preferiblemente activa de un polipéptido que tiene actividad beta-glucosidasa; y preferiblemente también (4) una región de terminación de la transcripción (terminador) capaz de terminar la transcripción cadena abajo de la secuencia de nucleótidos que codifica el polipéptido.

Cadena abajo de la secuencia de nucleótidos de acuerdo con la invención puede haber una región no traducida 3' que contiene uno o más sitios de terminación de la transcripción (por ejemplo, un terminador). El origen del terminador es menos crítico. El terminador puede ser, por ejemplo, nativo para la secuencia de ADN que codifica el polipéptido. Sin embargo, se usa preferiblemente un terminador de levadura en células hospedadoras de levadura y se usa un terminador de hongo filamentosos en células hospedadoras de hongos filamentosos. Más preferiblemente, el terminador es endógeno para la célula hospedadora (en que la secuencia de nucleótidos que codifica el polipéptido tiene que expresarse). En la región transcrita, puede estar presente un sitio de unión al ribosoma para la traducción. La parte codificante de los transcritos maduros expresada por las construcciones incluirá un AUG de unión de la traducción en el principio y un codón de terminación posicionado apropiadamente al final del polipéptido a traducirse.

También puede conseguirse expresión potenciada del polinucleótido de la invención mediante la selección de regiones reguladoras heterólogas, por ejemplo, regiones promotoras, líderes de secreción y/o terminadoras, que pueden servir para aumentar la expresión y, si se desea, los niveles de secreción del polipéptido de interés desde el hospedador de expresión y/o para proporcionar el control inducible de la expresión de un polipéptido de la invención.

Se apreciará por los expertos en la materia que el diseño del vector de expresión puede depender de factores tales como la elección de la célula hospedadora a transformarse, el nivel de expresión del polipéptido deseado, etc. Los vectores, tales como vectores de expresión, de la invención pueden introducirse en células hospedadoras para producir de ese modo proteínas o péptidos, codificados por ácidos nucleicos como se describe en este documento (por ejemplo, polipéptidos TEMER06387, formas mutantes de polipéptidos TEMER06387, fragmentos, variantes o equivalentes funcionales de los mismos). Los vectores, tales como vectores de expresión recombinantes, de la invención pueden diseñarse para la expresión de polipéptidos TEMER06387 en células procariotas o eucariotas.

Por ejemplo, pueden expresarse polipéptidos TEMER06387 en células bacterianas tales como *E. coli*, células de insecto (usando vectores de expresión de baculovirus), hongos filamentosos, células de levadura o células de mamífero. Se analizan adicionalmente células hospedadoras adecuadas en Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Posteriormente en este documento se describen ejemplos representativos de hospedadores apropiados.

Se conocen en la técnica medios y condiciones de cultivo apropiados para las células hospedadoras descritas anteriormente.

Se define en este documento que la expresión "secuencias de control" o "secuencias reguladoras" incluyen al menos cualquier componente que pueda ser necesario y/o ventajoso para la expresión de un polipéptido. Cualquier secuencia de control puede ser nativa o foránea para la secuencia de ácido nucleico de la invención que codifica un polipéptido. Dichas secuencias de control pueden incluir, aunque sin limitación, un promotor, un líder, secuencias óptimas de inicio de la traducción (como se describe en Kozak, 1991, J. Biol. Chem. 266:19867-19870), una secuencia señal de secreción, una secuencia de pro-péptido, una secuencia de poliadenilación, un terminador de la transcripción. Como mínimo, las secuencias de control típicamente incluyen un promotor, y señales de parada de la transcripción y la traducción. Como se ha expuesto anteriormente, la expresión "unido de forma funcional" se define en este documento como una configuración en que una secuencia de control está situada apropiadamente en una posición respecto a la secuencia codificante de la secuencia de ADN de modo que la secuencia de control dirige la producción de un polipéptido.

Las secuencias de control pueden proporcionarse con enlazadores con el propósito de introducir sitios específicos de restricción que faciliten el ligamiento de las secuencias de control con la región codificante de la secuencia de ácido nucleico que codifica un polipéptido. La expresión "unido de forma funcional" se define en este documento como una configuración en que una secuencia de control está situada apropiadamente en una posición respecto a la secuencia codificante de la secuencia de ADN de modo que la secuencia de control dirija la producción de un polipéptido.

La secuencia de control puede ser una secuencia promotora apropiada, una secuencia de ácido nucleico, que se reconoce por una célula hospedadora para la expresión de la secuencia de ácido nucleico. La secuencia promotora contiene secuencias de control de la transcripción, que median la expresión del polipéptido. El promotor puede ser cualquier secuencia de ácido nucleico, que muestra actividad transcripcional en la célula, incluyendo promotores mutantes, truncados, e híbridos, y pueden obtenerse de genes que codifican polipéptidos extracelulares o intracelulares homólogos o heterólogos para la célula.

La secuencia de control también puede ser una secuencia adecuada de terminación de la transcripción, una secuencia reconocida por una célula de hongo filamentoso para terminar la transcripción. La secuencia terminadora está unida de forma funcional al extremo 3' de la secuencia de ácido nucleico que codifica el polipéptido. Cualquier terminador, que sea funcional en la célula, puede usarse en la presente invención.

La secuencia de control también puede ser un terminador. Se obtienen terminadores preferidos para células de hongos filamentosos de los genes que codifican TAKA amilasa de *A. oryzae*, glucoamilasa (glaA) de *A. niger*, antranilato sintasa de *A. nidulans*, alfa-glucosidasa de *A. niger*, el gen trpC y la proteasa tipo tripsina de *Fusarium oxisporum*.

La secuencia de control también puede ser una secuencia líder adecuada, una región no traducida de un ARNm que es importante para la traducción por la célula de hongo filamentoso. La secuencia líder está unida de forma funcional al extremo 5' de la secuencia de ácido nucleico que codifica el polipéptido. Puede usarse cualquier secuencia líder, que sea funcional en la célula, en la presente invención. Se obtienen líderes preferidos para células de hongos filamentosos de los genes que codifican TAKA amilasa de *A. oryzae* y triosa fosfato isomerasa de *A. nidulans* y glaA de *A. niger*.

Pueden aislarse otras secuencias de control del gen IPNS de *Penicillium*, o el gen pcbC, el gen de la beta tubulina. Las secuencias de control se han citado en el documento WO 01/21779.

La secuencia de control también puede ser una secuencia de poliadenilación, una secuencia que está unida de forma funcional extremo 3' de la secuencia de ácido nucleico y que, cuando se transcribe, se reconoce por la célula de hongo filamentosos como una señal para añadir restos de poliadenosina al ARNm transcrito. Puede usarse cualquier secuencia de poliadenilación, que sea funcional en la célula, en la presente invención. Se obtienen secuencias preferidas de poliadenilación para células de hongos filamentosos de los genes que codifican TAKA amilasa de *A. oryzae*, glucoamilasa de *A. niger*, antranilato sintasa de *A. nidulans*, proteasa tipo tripsina de *Fusarium oxisporum* y alfa-glucosidasa de *A. niger*.

Cuando el polipéptido de acuerdo con la invención tiene que secretarse desde la célula hospedadora en el medio de cultivo, puede añadirse una secuencia señal apropiada al polipéptido para dirigir el polipéptido sintetizado *de novo* a la ruta de secreción de la célula hospedadora. Los expertos en la materia saben seleccionar una secuencia señal apropiada para un hospedador específico. La secuencia señal puede ser nativa para la célula hospedadora, o puede ser foránea para la célula hospedadora. Como ejemplo, puede usarse una secuencia señal de un polipéptido nativo para la célula hospedadora. Preferiblemente, dicho polipéptido nativo es un polipéptido altamente secretado, es decir, un polipéptido que se secreta en cantidades mayores del 10 % de la cantidad total de polipéptido que se está secretando. Las secuencias señal preferiblemente usadas de acuerdo con la invención son, por ejemplo: pmeA.

Como alternativa para una secuencia señal, el polipéptido de la invención puede fusionarse a un polipéptido vehículo secretado, o parte del mismo. Dicha construcción quimérica se dirige a la ruta de secreción mediante la secuencia señal del polipéptido vehículo, o parte del mismo. Además, el polipéptido vehículo proporcionará un efecto estabilizador al polipéptido de acuerdo con la invención y/o puede potenciar la solubilidad. Dicho polipéptido vehículo puede ser cualquier polipéptido. Preferiblemente, se usa un polipéptido altamente secretado como polipéptido vehículo. El polipéptido vehículo puede ser nativo o foráneo para el polipéptido de acuerdo con la invención. El polipéptido vehículo puede ser nativo o puede ser foráneo para la célula hospedadora. Ejemplos de dichos polipéptidos vehículo son glucoamilasa, prepro secuencia del factor alfa-acoplamiento, el dominio de unión a celulosa de *Clostridium cellulovorans* proteína A de unión a celulosa, glutatión S-transferasa, el dominio de unión a quitina de *Bacillus circulans* quitinasa A1, el dominio de unión a maltosa codificado por el gen malE de *E. coli* K12, beta-galactosidasa, y fosfatasa alcalina. Un polipéptido vehículo preferido para la expresión de dicha construcción quimérica en células de *Aspergillus* es glucoamilasa. La proteína y polipéptido vehículo pueden contener un motivo específico de aminoácidos para facilitar el aislamiento del polipéptido; el polipéptido de acuerdo con la invención puede liberarse por un agente de liberación especial. El agente de liberación puede ser una enzima proteolítica o un

agente químico. Un ejemplo de dicho motivo de aminoácidos es el sitio de escisión por proteasa KEX, que es bien conocido para los expertos en la materia.

Puede usarse una secuencia señal para facilitar la secreción y aislamiento de una proteína o polipéptido de la invención. Las secuencias señal típicamente se caracterizan por un núcleo de aminoácidos hidrófobos, que generalmente se escinden de la proteína madura durante la secreción en uno o más eventos de escisión. Dichos péptidos señal contienen sitios de procesamiento que permiten la escisión de la secuencia señal de las proteínas maduras según pasan a través de la ruta secretora. La secuencia señal dirige la secreción de la proteína, tal como desde un hospedador eucariota en que se introduce el vector de expresión por transformación, y la secuencia señal se escinde posterior o concurrentemente. La proteína después puede purificarse fácilmente del medio extracelular por métodos conocidos. Como alternativa, la secuencia señal puede unirse a la proteína de interés usando una secuencia, que facilita la purificación, tal como con un dominio GST. Por tanto, por ejemplo, la secuencia que codifica el polipéptido puede fusionarse a una secuencia marcadora, tal como una secuencia que codifica un péptido, que facilita la purificación del polipéptido fusionado. En ciertas realizaciones preferidas de este aspecto de la invención, la secuencia marcadora es un péptido de hexa-histidina, tal como la marca proporcionada en un vector pQE (Qiagen, Inc.), entre otras, muchas de las cuales están disponibles en el mercado. Como se describe en Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), por ejemplo, la hexa-histidina proporciona purificación conveniente de la proteína de fusión. La marca HA es otro péptido útil para la purificación, que corresponde a un epítipo derivado de la proteína hemaglutinina de la influenza, que se ha descrito por Wilson et al., Cell 37:767 (1984), por ejemplo.

Preferiblemente, se produce una proteína de fusión TEMER06387 de la invención por técnicas convencionales de ADN recombinante. Por ejemplo, se ligan juntos fragmentos de ADN que codifican las diferentes secuencias polipeptídicas, en fase, de acuerdo con técnicas convencionales, por ejemplo, empleando extremos de acabado romo o acabado escalonado para el ligamiento, digestión con enzimas de restricción para proporcionar extremos apropiados, rellenado de extremos cohesivos según lo apropiado, tratamiento con fosfatasa alcalina para evitar uniones indeseables, y ligamiento enzimático. En otra realización, puede sintetizarse el gen de fusión por técnicas convencionales incluyendo sintetizadores automatizados de ADN. Como alternativa, puede realizarse amplificación por PCR de fragmentos génicos usando cebadores de anclaje, que dan lugar a salientes complementarios entre dos fragmentos génicos consecutivos que posteriormente pueden hibridarse y reamplificarse para generar una secuencia génica quimérica (véase, por ejemplo, Current Protocols in Molecular Biology, eds. Ausubel et al. John Wiley & Sons: 1992). Además, están disponibles en el mercado muchos vectores de expresión que ya codifican un resto de fusión (por ejemplo, un polipéptido GST). Puede clonarse un ácido nucleico que codifica TEMER06387 en dicho vector de expresión de modo que el resto de fusión se una en fase a la proteína TEMER06387.

Preferiblemente, la eficacia de integración dirigida en el genoma de la célula hospedadora, es decir, la integración en un locus diana predeterminado, se aumenta mediante capacidades incrementadas de recombinación homóloga de la célula hospedadora. Dicho fenotipo de la célula implica preferiblemente un gen deficiente hdfA o hdfB como se describe en el documento WO2005/095624. El documento WO2005/095624 describe un método preferido para obtener una célula de hongo filamentoso que comprende eficacia aumentada de integración dirigida.

Opcionalmente, la célula hospedadora comprende una elevada respuesta de proteína no plegada (UPR) en comparación con la célula de tipo silvestre para potenciar las capacidades de producción de un polipéptido de interés. Puede aumentarse la UPR por técnicas descritas en el documento US2004/0186070A1 y/o el documento US2001/0034045A1 y/o el documento WO01/72783A2 y/o el documento WO2005/123763. Más específicamente, se ha modulado el nivel de proteína de HAC1 y/o IRE1 y/o PTC2, y/o la proteína SEC61 se ha modificado por ingeniería para obtener una célula hospedadora que tiene una UPR elevada.

Como alternativa, o en combinación con una UPR elevada, la célula hospedadora se modifica genéticamente para obtener un fenotipo que presente inferior expresión de proteasa y/o secreción de proteasa en comparación con la célula de tipo silvestre para potenciar las capacidades de producción de un polipéptido de interés. Dicho fenotipo puede obtenerse por delección y/o modificación y/o inactivación de un regulador transcripcional de la expresión de proteasas. Dicho regulador transcripcional es, por ejemplo, prtT. La disminución de la expresión de proteasas mediante la modulación de prtT puede realizarse por técnicas descritas en el documento US2004/0191864A1.

Como alternativa, o en combinación con una UPR elevada y/o un fenotipo que presenta inferior expresión de proteasa y/o secreción de proteasa, la célula hospedadora presente un fenotipo deficiente en oxalato para potenciar el rendimiento de producción de un polipéptido de interés. Un fenotipo deficiente en oxalato puede obtenerse por técnicas descritas en el documento WO2004/070022A2.

Como alternativa, o en combinación con una UPR elevada y/o un fenotipo que presenta inferior expresión de proteasa y/o secreción de proteasa y/o deficiencia de oxalato, la célula hospedadora presenta una combinación de diferencias fenotípicas en comparación con la célula de tipo silvestre para potenciar el rendimiento de producción del polipéptido de interés. Estas diferencias pueden incluir, aunque sin limitación, expresión disminuida de glucoamilasa y/o alfa-amilasa A neutra y/o alfa-amilasa B neutra, proteasa, y ácido oxálico hidrolasa. Dichas diferencias fenotípicas presentadas por la célula hospedadora pueden obtenerse por modificación genética de acuerdo con las

técnicas descritas en el documento US2004/0191864A1.

Como alternativa, o en combinación con una UPR elevada y/o un fenotipo que presenta inferior expresión de proteasa y/o secreción de proteasa y/o deficiencia de oxalato y una combinación de diferencias fenotípicas en comparación con la célula de tipo silvestre para potenciar el rendimiento de producción del polipéptido de interés, la célula hospedadora presenta una deficiencia en genes de toxina, que inutiliza la capacidad de la célula hospedadora de hongo filamentosos de expresar toxinas. Dichas toxinas incluyen, aunque sin limitación, ocratoxinas, fumonisinas, ácido ciclapiazónico, ácido 3-nitropropiónico, emodina, malformina, aflatoxinas y ácidos secalónicos. Dicha deficiencia es preferiblemente tal como la descrita en el documento WO2000/039322.

(Sobre)expresión

En una realización preferida, los polinucleótidos de la presente invención como se describe en este documento pueden sobre-expresarse en una cepa microbiana de la invención en comparación con la cepa microbiana precursora en que dicho gen no se sobre-expresa. La sobre-expresión de una secuencia polinucleotídica se define en este documento como la expresión de dicha secuencia génica que provoca la actividad de la enzima codificada por dicha secuencia en una cepa microbiana que es al menos aproximadamente 1,5 veces la actividad de la enzima en la cepa microbiana precursora; preferiblemente la actividad de dicha enzima es al menos aproximadamente 2 veces, más preferiblemente al menos aproximadamente 3 veces, más preferiblemente al menos aproximadamente 4 veces, más preferiblemente al menos aproximadamente 5 veces, incluso más preferiblemente al menos aproximadamente 10 veces y mucho más preferiblemente al menos aproximadamente 20 veces la actividad de la enzima en la cepa microbiana precursora.

El vector puede incluir adicionalmente secuencias que flanquean el polinucleótido dando lugar a ARN que comprende secuencias homólogas a secuencias genómicas eucariotas o secuencias genómicas víricas. Esto permitirá la introducción de los polinucleótidos de la invención en el genoma de una célula hospedadora.

Un vector de clonación integrador puede integrarse en un locus aleatorio o en un locus diana predeterminado en el cromosoma o cromosomas de la célula hospedadora en que tiene que integrarse. En una realización preferida de la invención, un vector de clonación integrador puede comprender un fragmento de ADN que es homólogo a una secuencia de ADN en un locus diana predeterminado en el genoma de la célula hospedadora para dirigir la integración del vector de clonación a este locus predeterminado. Para promover la integración dirigida, el vector de clonación preferiblemente puede linealizarse para la transformación de la célula hospedadora. La linealización puede realizarse preferiblemente de modo que al menos uno, pero preferiblemente cualquier extremo del vector de clonación esté flanqueado por secuencias homólogas al locus diana. La longitud de las secuencias homólogas que flanquean el locus diana es preferiblemente al menos aproximadamente 0,1 kb, tal como aproximadamente al menos 0,2 kb, más preferiblemente al menos aproximadamente 0,5 kb, incluso más preferiblemente al menos aproximadamente 1 kb, mucho más preferiblemente al menos aproximadamente 2 kb. Preferiblemente, las cepas hospedadoras precursoras pueden modificarse para una frecuencia mejorada de integración dirigida de ADN como se describe en el documento WO05/095624 y/o el documento WO2007/115886.

El ejemplo de delección proporcionado en la presente invención, usa el promotor del gen como flanco 5' y el gen como flanco 3' para insertar un marcador de selección entre el promotor y el gen, alterando de ese modo (es decir, inactivando funcionalmente) la transcripción génica. Las secuencias génicas dadas anteriormente pueden usarse para preparar genes funcionalmente inactivados similares. Los genes pueden dividirse en dos, produciendo un flanco 5' y un flanco 3', pero el gen también puede usarse para clonar un trazo más grande de ADN genómico que contiene las regiones promotora y terminadora del gen, que entonces pueden funcionar como flanco 5' y flanco 3'.

El sistema de vector puede ser un único vector, tal como un único plásmido, o dos o más vectores, tales como dos o más plásmidos, que juntos contienen el ADN total a introducirse en el genoma de la célula hospedadora.

El vector puede contener un polinucleótido de la invención orientado en una dirección antisentido para proporcionar la producción de ARN antisentido.

El ADN del vector puede introducirse en células procariotas o eucariotas mediante técnicas convencionales de transformación o transfección. Como se usa en este documento, los términos "transformación" y "transfección" pretenden referirse a una diversidad de técnicas reconocidas en la técnica para introducir ácido nucleico foráneo (por ejemplo, ADN) en una célula hospedadora, incluyendo co-precipitación con fosfato cálcico o cloruro cálcico, transfección mediada por DEAE-dextrano, transducción, infección, lipofección, transfección mediada por lípidos catiónicos o electroporación. Los métodos adecuados para transformar o transfectar células hospedadoras pueden encontrarse en Sambrook, et al. (Molecular Cloning: A Laboratory Manual, 2ª, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., Basic Methods in Molecular Biology (1986) y otros manuales de laboratorio.

Para transfección estable de células de mamífero, se sabe que, dependiendo del vector de expresión y la técnica de transfección usados, solamente una pequeña fracción de células puede integrar el ADN foráneo en su genoma. Para

identificar y seleccionar estos integrantes, generalmente se introduce un gen que codifica un marcador de selección (por ejemplo, resistencia a antibióticos) en las células hospedadoras junto con el gene de interés. Los marcadores de selección preferidos incluyen, aunque sin limitación, aquellos que confieren resistencia a fármacos o que complementan un defecto en la célula hospedadora. Incluyen, por ejemplo, genes marcadores versátiles que pueden usarse para la transformación de la mayoría de hongos filamentosos y levaduras tales como genes o ADNc de acetamidasa (los genes o ADNc de *amdS*, *niaD*, *facA* de *A. nidulans*, *A. oryzae* o *A. niger*), o genes que proporcionan resistencia a antibióticos como resistencia a G418, higromicina, bleomicina, kanamicina, metotrexato, fleomicina o benomil (*benA*). Como alternativa, pueden usarse marcadores de selección específicos tales como marcadores auxotróficos que requieren cepas hospedadoras mutantes correspondientes: por ejemplo, *URA3* (de *S. cerevisiae* o genes análogos de otras levaduras), *pyrG* o *pyrA* (de *A. nidulans* o *A. niger*), *argB* (de *A. nidulans* o *A. niger*) o *trpC*. En una realización preferida, el marcador de selección se deleciona de la célula hospedadora transformada después de la introducción de la construcción de expresión para obtener células hospedadoras transformadas capaces de producir el polipéptido que están libres de genes marcadores de selección.

Otros marcadores incluyen ATP sintetasa, subunidad 9 (*oliC*), orotidina-5'-fosfatodescarboxilasa (*pvrA*), el gen de resistencia a G418 bacteriano (éste también puede usarse en levaduras, pero no en hongos), el gen de resistencia a ampicilina (*E. coli*), el gen de resistencia a neomicina (*Bacillus*), el gen de resistencia a nourseotricina *nat1* de *Streptomyces nursei*, el gen de resistencia a piriamina *ptrA* de *Aspergillus oryzae* y el gen *uidA* de *E. coli*, que codifica la β -glucuronidasa (GUS). Pueden usarse vectores *in vitro*, por ejemplo, para la producción de ARN o usarse para transfectar o transformar una célula hospedadora.

La expresión de proteínas en procariotas a menudo se realiza en *E. coli* con vectores que contienen promotores constitutivos o inducibles que dirigen la expresión de proteínas de fusión o no de fusión. Los vectores de fusión añaden varios aminoácidos a una proteína codificada en los mismos, por ejemplo, al extremo amino de la proteína recombinante. Dichos vectores de fusión típicamente logran tres propósitos: 1) aumentar la expresión de proteína recombinante; 2) aumentar la solubilidad de la proteína recombinante; y 3) ayudar en la purificación de la proteína recombinante actuando como ligando en purificación por afinidad. A menudo, en vectores de expresión de fusión, se introduce un sitio de escisión proteolítica en la unión del resto de fusión y la proteína recombinante para posibilitar la separación de la proteína recombinante del resto de fusión posterior a purificación de la proteína de fusión.

Como se indica, los vectores de expresión contendrán preferiblemente marcadores de selección. Dichos marcadores incluyen resistencia a la dihidrofolato reductasa o neomicina para cultivo de células eucariotas y resistencia a tetraciclina o ampicilina para cultivo en *E. coli* y otras bacterias.

Se describen vectores preferidos para su uso en bacterias, por ejemplo, en el documento WO-A1-2004/074468, que se adjunta en la presente por referencia. Otros vectores adecuados serán fácilmente evidentes para los expertos en la materia.

Para la secreción de la proteína traducida en el lumen del retículo endoplasmático, en el espacio periplásmico o en el entorno extracelular, puede incorporarse una señal apropiada de secreción en el polipéptido expresado. Las señales pueden ser endógenas para el polipéptido o pueden ser señales heterólogas.

El polipéptido TEMER06387 puede expresarse en una forma modificada, tal como una proteína de fusión, y puede incluir no solamente señales de secreción, sino también regiones funcionales heterólogas adicionales. Por tanto, por ejemplo, puede añadirse una región de aminoácidos adicionales, particularmente aminoácidos cargados, al extremo N-terminal del polipéptido para mejorar la estabilidad y persistencia en la célula hospedadora, durante la purificación o durante la posterior manipulación y almacenamiento. Además, pueden añadirse restos peptídicos al polipéptido para facilitar la purificación.

La invención proporciona un polipéptido aislado que tiene la secuencia de aminoácidos de acuerdo con la SEQ ID NO: 2, y una secuencia de aminoácidos que se puede obtener expresando el polinucleótido de la SEQ ID NO: 1 en un hospedador apropiado. Además, un péptido o polipéptido que comprende una variante de los polipéptidos anteriores, tales como un equivalente funcional, está comprendido dentro de la presente invención. Los polipéptidos anteriores están colectivamente comprendidos en la expresión "polipéptidos de acuerdo con la invención".

La expresión "péptido variante" o "polipéptido variante" se define en este documento como un péptido o polipéptido, respectivamente, que comprende una o más alteraciones, tales como sustituciones, inserciones, delecciones y/o truncamientos de uno o más restos de aminoácido específicos en una o más posiciones específicas en el péptido o polipéptido, respectivamente. Por consiguiente, un péptido señal variante es un péptido señal que comprende una o más alteraciones, tales como sustituciones, inserciones, delecciones y/o truncamientos de uno o más restos de aminoácido específicos en una o más posiciones específicas en el péptido señal.

El término "polinucleótido" es idéntico a la expresión "molécula de ácido nucleico" y en este documento pueden leerse de forma intercambiable. El término se refiere a una molécula polinucleotídica, que es una molécula de ácido ribonucleico (ARN) o ácido desoxirribonucleico (ADN), monocatenario o bicatenario. Un polinucleótido puede estar presente en forma aislada, o puede estar comprendido en moléculas de ácido nucleico recombinantes o vectores, o

puede estar comprendido en una célula hospedadora.

La expresión "polinucleótido variante" se define en este documento como un polinucleótido que comprende una o más alteraciones, tales como sustituciones, inserciones, deleciones y/o truncamientos de uno o más nucleótidos en una o más posiciones específicas en el polinucleótido.

Los términos "péptido" y "oligopéptido" se consideran sinónimos (como está reconocido habitualmente) y cada término puede usarse de forma intercambiable, según lo requiera el contexto, para indicar una cadena de al menos dos aminoácidos acoplados por enlaces peptídico. La palabra "polipéptido" se usa en este documento para cadenas que contienen más de siete restos de aminoácido. Todas las fórmulas o secuencias de oligopéptidos y polipéptidos de este documento se escriben de izquierda a derecha y en la dirección desde el extremo amino hasta el extremo carboxi. El código de una letra de aminoácidos usado en este documento es habitualmente conocido en la técnica y puede encontrarse en Sambrook, et al. (Molecular Cloning: A Laboratory Manual, 2ª, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Por polipéptido o proteína "aislada" se entiende un polipéptido o proteína retirada de su entorno nativo. Por ejemplo, polipéptidos y proteínas producidos de forma recombinante expresados en células hospedadoras se consideran aislados para el propósito de la invención ya que son polipéptidos nativos o recombinantes que se han purificado sustancialmente por cualquier técnica adecuada tal como, por ejemplo, el método de purificación de una única etapa descrito en Smith y Johnson, Gene 67:31-40 (1988).

El polipéptido de beta-glucosidasa TEMER06387 de acuerdo con la invención puede recuperarse y purificarse de cultivos de células recombinantes por métodos conocidos en la técnica. Más preferiblemente, se emplea cromatografía líquida de alto rendimiento ("HPLC") para la purificación.

Los polipéptidos de la presente invención incluyen productos purificados de forma natural, productos de procedimientos sintéticos químicos, y productos producidos por técnicas recombinantes de un hospedador procariota o eucariota incluyendo, por ejemplo, células bacterianas, de levadura, de plantas superiores, de insecto y de mamífero. Dependiendo del hospedador empleado en un procedimiento de producción recombinante, los polipéptidos de la presente invención pueden estar glucosilados o pueden estar no glucosilados. Además, los polipéptidos de la invención también pueden incluir un resto de metionina modificado inicial, en algunos casos como resultado de procesos mediados por el hospedador.

La invención también destaca fragmentos biológicamente activos de los polipéptidos de acuerdo con la invención.

Los fragmentos biológicamente activos de un polipéptido de la invención incluyen polipéptidos que comprende secuencias de aminoácidos suficientemente idénticas a o derivadas de la secuencia de aminoácidos del polipéptido TEMER06387 (por ejemplo, la secuencia de aminoácidos de la SEQ ID NO: 2), que incluyen menos aminoácidos que el polipéptido de longitud completa pero que muestran al menos una actividad biológica del correspondiente polipéptido de longitud completa. Típicamente, los fragmentos biológicamente activos comprenden un dominio o motivo con al menos una actividad del polipéptido TEMER06387.

Un fragmento biológicamente activo de un polipéptido de la invención puede ser un polipéptido que es, por ejemplo, de aproximadamente 10, aproximadamente 25, aproximadamente 50, aproximadamente 100 o más aminoácidos de longitud o de al menos aproximadamente 100 aminoácidos, al menos 150, 200, 250, 300, 350, 400 aminoácidos de longitud, o de una longitud hasta la cantidad total de aminoácidos del polipéptido de la invención.

Además, pueden prepararse otras partes biológicamente activas, en que están delecionadas otras regiones del polipéptido, por técnicas recombinantes y pueden evaluarse para una o más de las actividades biológicas de la forma nativa de un polipéptido de la invención.

La invención también destaca fragmentos de ácido nucleico que codifican los anteriores fragmentos biológicamente activos del polipéptido TEMER06387.

Polipéptidos

En otro aspecto de la invención, se proporciona polipéptidos TEMER06387 mejorados. Los polipéptidos TEMER06387 mejorados son polipéptidos donde al menos una actividad biológica está mejorada. Dichos polipéptidos pueden obtenerse introduciendo aleatoriamente mutaciones a lo largo de todo o parte de la secuencia codificante de TEMER06387, tal como por mutagénesis de saturación, y los mutantes resultantes pueden expresarse de forma recombinante y seleccionarse por actividad biológica.

Pueden obtenerse variantes mejoradas de las secuencias de aminoácidos de la presente invención que conducen a una función beta-glucosidasa mejorada por los correspondientes genes de la presente invención. Entre dichas modificaciones se incluyen:

1. PCR propensa a errores para introducir mutaciones aleatorias, seguida de una exploración de variantes obtenidas y aislamiento de las variantes con propiedades cinéticas mejoradas
2. Reorganización de familias de variantes relacionadas de los genes que codifican la enzima beta-glucosidasa, seguida de una exploración de las variantes obtenidas y aislamiento de variantes con propiedades cinéticas mejoradas.

Pueden obtenerse variantes de los genes de la presente invención que conducen a un nivel aumentado de ARNm y/o polipéptido, que provoca más actividad beta-glucosidasa por las secuencias polinucleotídicas de dichos genes. Entre dichas modificaciones se incluyen:

1. Mejora del uso de codones de tal modo que los codones estén adaptados (de forma óptima) al hospedador microbiano precursor.
2. Mejora del uso de pares de codones de tal modo que los codones estén adaptados (de forma óptima) al hospedador microbiano precursor.
3. Adición de secuencias estabilizadoras a la información genómica que codifica el polipéptido beta-glucosidasa que produce moléculas de ARNm con una semi-vida aumentada.

Se describen métodos preferidos para aislar variantes con propiedades catalíticas mejoradas o niveles aumentados de ARNm o polipéptido en el documento WO03/010183 y el documento WO03/01311. Se describen métodos preferidos para optimizar el uso de codones en cepas microbianas precursoras en el documento PCT/EP2007/05594. Se describen métodos preferidos para la adición de elementos estabilizadores a los genes que codifican el polipéptido beta-glucosidasa de la invención en el documento WO2005/059149.

En una realización preferida, el polipéptido TEMER06387 tiene una secuencia de aminoácidos de acuerdo con la SEQ ID NO: 2. En otra realización, el polipéptido TEMER06387 es sustancialmente homólogo a la secuencia de aminoácidos de acuerdo con la SEQ ID NO: 2 y retiene al menos una actividad biológica de un polipéptido de acuerdo con la SEQ ID NO: 2, aunque difiere en secuencia de aminoácidos debido a la variación o mutagénesis natural como se ha descrito.

En una realización preferida adicional, el polipéptido TEMER06387 tiene una secuencia de aminoácidos codificada por un fragmento de ácido nucleico aislado capaz de hibridar con un ácido nucleico de acuerdo con la SEQ ID NO: 1, preferiblemente en condiciones de hibridación altamente rigurosas.

Por consiguiente, el polipéptido TEMER06387 es preferiblemente un polipéptido que comprende una secuencia de aminoácidos al menos un 70 %, al menos un 71 %, al menos un 72 %, al menos un 73 %, al menos un 74 %, al menos un 75 %, al menos un 76 %, al menos un 77 %, al menos un 78 %, al menos un 79 %, al menos un 80 %, al menos un 81 %, al menos un 82 %, al menos un 83 %, al menos un 84 %, al menos un 85 %, al menos un 86 %, al menos un 87 %, al menos un 88 %, al menos un 89 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 % o más homóloga a la secuencia de aminoácidos mostrada en la SEQ ID NO: 2 y, típicamente, retiene al menos una actividad funcional del polipéptido de acuerdo con la SEQ ID NO: 2.

También pueden identificarse equivalentes funcionales de un polipéptido de acuerdo con la invención, por ejemplo, por exploración de bibliotecas combinatorias de mutantes, por ejemplo, mutantes de truncamiento, del polipéptido de la invención para actividad beta-glucosidasa. En una realización, se genera una biblioteca diversificada de variantes por mutagénesis combinatoria a nivel de ácido nucleico. Una biblioteca diversificada de variantes puede producirse por, por ejemplo, ligamiento enzimático de una mezcla de oligonucleótidos sintéticos en secuencias génicas de modo que se pueda expresar un conjunto degenerado de secuencias polipeptídicas potenciales como polipéptidos individuales, o como alternativa, como un conjunto de proteínas de fusión más grandes (por ejemplo, para presentación en fagos). Existe una diversidad de métodos que pueden usarse para producir bibliotecas de variantes potenciales de los polipéptidos de la invención a partir de una secuencia oligonucleotídica degenerada. Se conocen en la técnica métodos para sintetizar oligonucleótidos degenerados (véase, por ejemplo, Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

Además, pueden usarse bibliotecas de fragmentos de la secuencia codificante de un polipéptido de la invención para generar una población diversificada de polipéptidos para explorar una posterior selección de variantes. Por ejemplo, puede generarse una biblioteca de fragmentos de secuencia codificante tratando un fragmento de PCR bicatenario de la secuencia codificante de interés con una nucleasa en condiciones donde sucede formación de mella sólo aproximadamente una vez por molécula, desnaturalizando el ADN bicatenario, renaturalizando el ADN para formar ADN bicatenario que puede incluir pares con sentido/antisentido de diferentes productos con mella, retirando las partes monocatenarias de los dúplex vueltos a formar por tratamiento con nucleasa S1, y ligando la biblioteca resultante de fragmentos en un vector de expresión. Por este método, puede obtenerse una biblioteca de expresión que codifica fragmentos N-terminales e internos de diversos tamaños de la proteína de interés.

Se conocen en la técnica varias técnicas para explorar productos génicos de bibliotecas combinatorias preparadas

por mutaciones puntuales de truncamiento, y para explorar bibliotecas de ADNc para productos génicos que tienen una propiedad seleccionada. Las técnicas más ampliamente usadas, que son susceptibles a análisis de alto rendimiento, para explorar bibliotecas génicas grandes típicamente incluyen clonación de la biblioteca génica en vectores de expresión replicables, transformación de células apropiadas con la biblioteca resultante de vectores, y expresión de los genes combinatorios en condiciones en que la detección de una actividad deseada facilita el aislamiento del vector que codifica el gen cuyo producto se detectó. Puede usarse mutagénesis de ensamblaje reiterado (REM), una técnica que potencia la frecuencia de mutantes funcionales en las bibliotecas, en combinación con los ensayos de exploración para identificar variantes de un polipéptido de la invención (Arkin y Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3): 327-331).

Además de la secuencia génica de TEMER06387 mostrada en la SEQ ID NO: 1, será evidente para los expertos en la materia que pueden existir polimorfismos de secuencia de ADN dentro de una población dada, que pueden conducir a cambios en la secuencia de aminoácidos del polipéptido TEMER06387. Dichos polimorfismos genéticos pueden existir en células de diferentes poblaciones o dentro de una población debido a variación alélica natural. Las variantes alélicas también incluyen equivalentes funcionales.

Fragmentos de un polinucleótido de acuerdo con la invención también pueden comprender polinucleótidos que no codifican polipéptidos funcionales. Dichos polinucleótidos pueden funcionar como sondas o cebadores para una reacción de PCR.

Pueden usarse ácidos nucleicos de acuerdo con la invención independientemente de si codifican polipéptidos funcionales o no funcionales como sondas de hibridación o cebadores de reacción en cadena de la polimerasa (PCR). Usos de las moléculas de ácido nucleico de la presente invención que no codifican un polipéptido que tiene una actividad TEMER06387 incluyen, entre otros, (1) aislamiento del gen que codifica el polipéptido TEMER06387, o variantes alélicas del mismo procedentes de una biblioteca de ADNc, por ejemplo, de microorganismos adecuados; (2) hibridación *in situ* (por ejemplo, FISH) a extensiones cromosómicas en metafase para proporcionar localización cromosómica precisa del gen TEMER06387 como se describe en Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, Nueva York (1988); (3) análisis de transferencia de Northern para detectar la expresión de ARNm de TEMER06387 en tejidos y/o células específicas y 4) sondas y cebadores que pueden usarse como herramienta de diagnóstico para analizar la presencia de un ácido nucleico que puede hibridar con la sonda de TEMER06387 en una muestra biológica (por ejemplo, tejido) dada.

También está englobado por la invención un método de obtención de un equivalente funcional de un gen de TEMER06387. Dicho método implica obtener una sonda marcada que incluye un ácido nucleico aislado que codifica todo o una parte de la secuencia polipeptídica de acuerdo con la SEQ ID NO: 2 o una variante de la misma; explorar una biblioteca de fragmentos de ácido nucleico con la sonda marcada en condiciones que permitan la hibridación de la sonda a fragmentos de ácido nucleico en la biblioteca, formando de ese modo dúplex de ácido nucleico, y preparar una secuencia génica de longitud completa a partir de los fragmentos de ácido nucleico en cualquier dúplex marcado para obtener un gen relacionado con el gen de TEMER06387.

En una realización, un ácido nucleico de TEMER06387 de la invención es al menos un 70 %, al menos un 71 %, al menos un 72 %, al menos un 73 %, al menos un 74 %, al menos un 75 %, al menos un 76 %, al menos un 77 %, al menos un 78 %, al menos un 79 %, al menos un 80 %, al menos un 81 %, al menos un 82 %, al menos un 83 %, al menos un 84 %, al menos un 85 %, al menos un 86 %, al menos un 87 %, al menos un 88 %, al menos un 89 %, al menos un 90 %, al menos un 91 %, al menos un 92 %, al menos un 93 %, al menos un 94 %, al menos un 95 %, al menos un 96 %, al menos un 97 %, al menos un 98 %, al menos un 99 %, o más homólogo a una secuencia de ácido nucleico mostrada en la SEQ ID NO: 1 o el complemento de la misma.

Se proporcionan adicionalmente: células hospedadoras que comprenden un polinucleótido o vector de la invención. El polinucleótido puede ser heterólogo para el genoma de la célula hospedadora. El término "heterólogo", habitualmente con respecto a la célula hospedadora, significa que el polinucleótido no existe de forma natural en el genoma de la célula hospedadora o que el polipéptido no se produce de forma natural por esa célula.

En otra realización, la invención destaca células, por ejemplo, células hospedadoras transformadas o células hospedadoras recombinantes que contienen un ácido nucleico englobado por la invención. Una "célula transformada" o "célula recombinante" es una célula en que (o en un progenitor de la cual) se ha introducido, mediante técnicas de ADN recombinante, un ácido nucleico de acuerdo con la invención. Se incluyen células tanto procariontas como eucariotas, por ejemplo, bacterias, hongos, levaduras, y similares, siendo especialmente preferidas células de hongos filamentosos, tales como *Aspergillus niger* o *Talaromyces emersonii*.

Puede elegirse una célula hospedadora que module la expresión de las secuencias insertadas, o modifique y procese en producto génico de un modo específico, deseado. Dichas modificaciones (por ejemplo, glucosilación) y procesamiento (por ejemplo, escisión) de productos proteicos pueden facilitar el funcionamiento óptimo de la proteína.

Diversas células hospedadoras tienen características y mecanismos específicos para el procesamiento post-

traduccional y modificación de proteínas y productos génicos. Pueden elegirse líneas celulares o sistemas hospedadores apropiados familiares para los expertos en la materia de biología molecular y/o microbiología para asegurar la modificación y procesamiento deseado y correcto de la proteína foránea expresada. Para este fin, pueden usarse células hospedadoras eucariotas que poseen la maquinaria celular para el apropiado procesamiento del transcrito primario, la glucosilación, y fosforilación del producto génico. Dichas células hospedadoras son bien conocidas en la técnica.

Si se desea, puede usarse una célula como se ha descrito anteriormente en la preparación de un polipéptido de acuerdo con la invención. Dicho método típicamente comprende cultivar una célula hospedadora (por ejemplo, transformada o transfectada con un vector de expresión como se ha descrito anteriormente) en condiciones para proporcionar la expresión (por el vector) de una secuencia codificante que codifica el polipéptido, y opcionalmente recuperar el polipéptido expresado. Pueden incorporarse polinucleótidos de la invención en un vector replicable recombinante, por ejemplo, un vector de expresión. El vector puede usarse para replicar el ácido nucleico en una célula hospedadora compatible. Por tanto, en una realización adicional, la invención proporciona un método de preparación de un polinucleótido de la invención introduciendo un polinucleótido de la invención en un vector replicable, introduciendo el vector en una célula hospedadora compatible, y cultivando la célula hospedadora en condiciones que consiguen la replicación del vector. El vector puede recuperarse de la célula hospedadora.

Preferiblemente, el polipéptido se produce como una proteína secretada en cuyo caso la secuencia de nucleótidos que codifica una forma madura del polipéptido en la construcción de expresión está unida de forma funcional a una secuencia de nucleótidos que codifica una secuencia señal. Preferiblemente, la secuencia señal es nativa (homóloga) para la secuencia de nucleótidos que codifica el polipéptido. Como alternativa, la secuencia señal es foránea (heteróloga) a la secuencia de nucleótidos que codifica el polipéptido, en cuyo caso la secuencia señal es preferiblemente endógena para la célula hospedadora en que se expresa la secuencia de nucleótidos de acuerdo con la invención. Ejemplos de secuencias señal adecuadas para células hospedadoras de levadura son las secuencias señal derivadas de genes de factor- α de levaduras. Asimismo, una secuencia señal adecuada para células hospedadoras fúngicas es, por ejemplo, una secuencia señal derivada de un gen de amiloglucosidasa (AG) de hongo filamentoso, por ejemplo, el gen *glaA* de *A. niger*. Éste puede usarse en combinación con el propio promotor de amiloglucosidasa (también llamada (gluco) amilasa), así como en combinación con otros promotores. También pueden usarse secuencias señal híbridas con el contexto de la presente invención.

Secuencias líder de secreción heterólogas preferidas son aquellas que se originan a partir del gen de la amiloglucosidasa (AG) fúngica (*glaA*-ambas versiones de 18 y 24 aminoácidos, por ejemplo, de *Aspergillus*), el gen de factor- α (levaduras, por ejemplo, *Saccharomyces* y *Kluyveromyces*) o el gen de la α -amilasa (*Bacillus*).

Los vectores pueden introducirse por transformación o transfección en una célula hospedadora adecuada como se ha descrito anteriormente para proporcionar la expresión de un polipéptido de la invención. Este proceso puede comprender el cultivo de una célula hospedadora transformada con un vector de expresión como se ha descrito anteriormente en condiciones para proporcionar la expresión por el vector de una secuencia codificante que codifica el polipéptido.

Células hospedadoras

La invención, por tanto, proporciona células hospedadoras transformadas o transfectadas con o que comprenden un polinucleótido o vector de la invención. Preferiblemente, el polinucleótido se transporta en un vector para la replicación y expresión del polinucleótido. Las células se elegirán para que sean compatibles con dicho vector y pueden ser, por ejemplo, células procariotas (por ejemplo, bacterianas), fúngicas, de levadura o vegetales.

También puede elegirse un hospedador heterólogo donde el polipéptido de la invención se produce en una forma que está sustancialmente libre de otras enzimas degradantes de celulosa o degradantes de hemicelulosa. Esto puede conseguirse eligiendo un hospedador que normalmente no produce dichas enzimas.

La invención abarca procesos para la producción del polipéptido de la invención mediante expresión recombinante de una secuencia de ADN que codifica el polipéptido. Para este propósito, la secuencia de ADN de la invención puede usarse para amplificación génica y/o intercambio de señales de expresión, tales como promotores, secuencias señal de secreción, para permitir la producción económica del polipéptido en una célula hospedadora homóloga o heteróloga adecuada. Una célula hospedadora homóloga es una célula hospedadora que es de la misma especie o que es una variante dentro de la misma especie que la especie de la cual se obtiene la secuencia de ADN.

Células hospedadoras adecuadas son preferiblemente microorganismos procariotas tales como bacterias, o más preferiblemente organismos eucariotas, por ejemplo, hongos, tales como levaduras u hongos filamentosos, o células vegetales. En general, se prefieren células de levadura sobre las células fúngicas porque son más fáciles de manipular. Sin embargo, algunas proteínas se secretan mal desde las levaduras, o en algunos casos no se procesan apropiadamente (por ejemplo, hiperglucosilación en levaduras). En estos casos, debe seleccionarse un organismo hospedador fúngico.

La célula hospedadora sobre-expresa el polipéptido, y se conocen bien técnicas para diseñar por ingeniería la sobre-expresión. El hospedador, por tanto, puede tener dos o más copias del polinucleótido codificante (y el vector puede, por tanto, tener dos o más copias en consecuencia).

Bacterias del género *Bacillus* son muy adecuadas como hospedadores heterólogos a causa de su capacidad para secretar proteínas en el medio de cultivo. Otras bacterias adecuadas como hospedadores son aquellas de los géneros *Streptomyces* y *Pseudomonas*. Una célula hospedadora de levadura preferida para la expresión de la secuencia de ADN que codifica el polipéptido es de los géneros *Saccharomyces*, *Kluyveromyces*, *Hansenula*, *Pichia*, *Yarrowia*, y *Schizosaccharomyces*.

Más preferiblemente, se selecciona una célula hospedadora de levadura del grupo que consiste en las especies *Saccharomyces cerevisiae*, *Kluyveromyces lactis* (también conocida como *Kluyveromyces marxianus* var. *lactis*), *Hansenula polymorpha*, *Pichia pastoris*, *Yarrowia lipolytica* y *Schizosaccharomyces pombe*.

Las más preferidas son, sin embargo, células hospedadoras fúngicas (por ejemplo, filamentosas). Las células hospedadoras de hongos filamentosos preferidas se seleccionan del grupo que consiste en los géneros *Aspergillus*, *Trichoderma/Hypocrea*, *Fusarium*, *Disporotrichum*, *Penicillium*, *Acremonium*, *Neurospora*, *Thermoascus*, *Myceliophthora*, *Sporotrichum*, *Thielavia*, *Chrysosporium*, *Fusarium*, *Humicola*, *Neurospora* y *Talaromyces*.

Más preferiblemente, una célula hospedadora de hongos filamentosos es de las especies que incluyen, aunque sin limitación, *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus tubingensis*, *Aspergillus aculeatus*, *Aspergillus foetidus*, *Aspergillus nidulans*, *Aspergillus japonicus*, *Aspergillus oryzae* y *Aspergillus ficuum*, *Trichoderma reesei*, *Hypocrea jecorina*, *Fusarium graminearum*, *Talaromyces emersonii*, *Penicillium decumbens*, *Acremonium alabamense*, *Neurospora crassa*, *Myceliophthora thermophilurri*, *Sporotrichum cellulophilum*, *Disporotrichum dimorfosporum*, *Talaromyces emersonii*, *Talaromyces stipitatus* y *Thielavia terrestris*.

Las células hospedadoras de acuerdo con la invención incluyen células vegetales, y la invención por lo tanto se extiende hasta organismos transgénicos, tales como plantas y partes de las mismas, que contienen una o más células de la invención. Las células pueden expresar de forma heteróloga el polipéptido de la invención o pueden contener de forma heteróloga uno o más de los polinucleótidos de la invención. La planta transgénica (o modificada genéticamente) puede, por lo tanto, haber insertado (por ejemplo, de forma estable) en su genoma una secuencia que codifica uno o más de los polipéptidos de la invención. La transformación de células vegetales puede realizarse usando técnicas conocidas, por ejemplo, usando un plásmido Ti o Ri de *Agrobacterium tumefaciens*. El plásmido (o vector) puede contener, por tanto, secuencias necesarias para infectar una planta, y pueden emplearse derivados de los plásmidos Ti y/o Ri.

Como alternativa, puede lograrse la infección directa de una parte de una planta, tal como una hoja, raíz o tallo. En esta técnica, la planta a infectarse puede lacerarse, por ejemplo, cortando la planta con una navaja o perforando la planta con una aguja o frotando la planta con un abrasivo. La laceración después se inocula con el *Agrobacterium*. La planta o parte de la planta después se cultiva en un medio de cultivo adecuado y se deja desarrollar en una planta madura. La regeneración de las células transformadas en plantas modificadas genéticamente puede conseguirse usando técnicas conocidas, por ejemplo, por selección de brotes transformados usando un antibiótico y por sub-cultivos de los brotes en un medio que contiene los nutrientes apropiados, hormonas vegetales y similares.

La invención también incluye células que se han modificado para que expresen el polipéptido de beta-glucosidasa de la invención o una variante del mismo. Dichas células incluyen líneas celulares transitorias, o preferiblemente estable de eucariotas superiores, tales como células de mamífero o células de insecto, células eucariotas inferiores, tales como células de levadura y fúngicas (por ejemplo, filamentosas) o células procariotas tales como células bacterianas.

También es posible que los polipéptidos de la invención se expresen de forma transitoria en una línea celular o sobre una membrana, tal como, por ejemplo, en un sistema de expresión de baculovirus. Dichos sistemas, que están adaptados para expresar los polipéptidos de acuerdo con la invención, también se incluyen dentro del alcance de la presente invención.

De acuerdo con la presente invención, la producción del polipéptido de la invención puede lograrse mediante el cultivo de hospedadores de expresión microbianos, que se han transformado con uno o más polinucleótidos de la presente invención, en un medio convencional de fermentación con nutrientes.

Producción de polipéptido/enzima

Las células hospedadoras recombinantes de acuerdo con la invención pueden cultivarse usando procedimientos conocidos en la técnica. Para cada combinación de un promotor y una célula hospedadora, están disponibles condiciones de cultivo que conducen a la expresión de la secuencia de ADN que codifica el polipéptido. Después de alcanzar la densidad celular o título deseado del polipéptido, se detiene el cultivo y se recupera el polipéptido usando

procedimientos conocidos.

El medio de fermentación puede comprender un medio de cultivo conocido que contiene una fuente de carbono (por ejemplo, glucosa, maltosa, melazas, almidón, celulosa, xilano, pectina, hidrolizado de biomasa lignocelulolítica, etc.), una fuente de nitrógeno (por ejemplo, sulfato de amonio, nitrato de amonio, cloruro de amonio, etc.), una fuente de nitrógeno orgánico (por ejemplo, extracto de levadura, extracto de malta, peptona, etc.) y fuentes de nutrientes inorgánicos (por ejemplo, fosfato, magnesio, potasio, zinc, hierro, etc.). Opcionalmente, puede incluirse un inductor (por ejemplo, celulosa, pectina, xilano, maltosa, maltodextrina o xilogalacturonano).

La selección del medio apropiado puede basarse en la elección del hospedador de expresión y/o puede basarse en las necesidades reguladoras de la construcción de expresión. Dichos medios son conocidos para los expertos en la materia. El medio puede contener, si se desea, componentes adicionales que favorecen a los hospedadores de expresión transformados sobre otros microorganismos potencialmente contaminantes.

La producción de polipéptido por el hospedador transformado (fermentación) puede realizarse de acuerdo con cualquier procedimiento conocido. El tiempo de producción puede extenderse durante un periodo de aproximadamente 0,5 a aproximadamente 30 días. Puede ser un proceso discontinuo, continuo o semi-continuo, adecuadamente a una temperatura en el intervalo de 0-100 °C o 0-80 °C, por ejemplo, de aproximadamente 0 a aproximadamente 60 °C y/o a un pH, por ejemplo, de aproximadamente 2 a aproximadamente 10, o de aproximadamente 3 a aproximadamente 9. Son condiciones de fermentación preferidas, una temperatura en el intervalo de aproximadamente 20 a aproximadamente 55 °C y/o a un pH de aproximadamente 3 a aproximadamente 5. Las condiciones apropiadas se seleccionan habitualmente en base a la elección del hospedador de expresión y el polipéptido a expresarse.

Después de la fermentación, si fuera necesario, las células pueden retirarse del caldo de fermentación mediante centrifugación o filtración. Después de haber detenido la fermentación o después de la retirada de las células, el polipéptido de la invención entonces puede recuperarse y, si se desea, purificarse y aislarse por medios convencionales.

Composiciones de polipéptido/enzima

La invención proporciona una composición que comprende un polipéptido de la invención y una celulasa y/o una hemicelulasa y/o una pectinasa.

Cuando el polipéptido de la invención es una celulasa, una composición de la invención típicamente comprenderá una hemicelulasa y/o una pectinasa además del polipéptido de la invención.

Cuando el polipéptido de la invención es una hemicelulasa, una composición de la invención típicamente comprenderá una celulasa y/o una pectinasa además del polipéptido de la invención.

Cuando el polipéptido de la invención es una pectinasa, una composición de la invención típicamente comprenderá una celulasa y/o una hemicelulasa además del polipéptido de la invención.

Una composición de la invención puede comprender una, dos o tres o más clases de celulasa, por ejemplo, una, dos o todas de una endo-1,4-β-glucanasa (EG), una exocelobiohidrolasa (CBH) y una β-glucosidasa (BG).

Una composición de la invención puede comprender un polipéptido que tiene la misma actividad enzimática, por ejemplo, el mismo tipo de actividad celulasa, hemicelulasa y/o pectinasa que la proporcionada por un polipéptido de la invención.

Una composición de la invención puede comprender un polipéptido que tiene un tipo diferente de actividad celulasa y/o actividad hemicelulasa y/o actividad pectinasa que la proporcionada por un polipéptido de la invención. Por ejemplo, una composición de la invención puede comprender un tipo de celulasa y/o actividad hemicelulasa y/o actividad pectinasa proporcionada por un polipéptido de la invención y un segundo tipo de celulasa y/o actividad hemicelulasa y/o actividad pectinasa proporcionada por una hemicelulasa/pectinasa adicional.

En este documento, una celulasa es cualquier polipéptido que sea capaz de degradar celulosa. Un polipéptido que es capaz de degradar celulosa es uno que es capaz de catalizar el proceso de descomposición de la celulosa en unidades más pequeñas, de forma parcial, por ejemplo, en celodextrinas, o de forma completa en monómeros de glucosa. Una celulasa de acuerdo con la invención puede dar lugar a una población mixta de celodextrinas y monómeros de glucosa cuando se pone en contacto con la celulosa. Dicha degradación típicamente tendrá lugar mediante una reacción de hidrólisis.

En este documento, una hemicelulasa es cualquier polipéptido que sea capaz de degradar hemicelulosa. Es decir, una hemicelulasa puede ser capaz de degradar uno o más de xilano, glucuronoxilano, arabinoxilano, glucomanano y xiloglucano. Un polipéptido que es capaz de degradar una hemicelulosa es uno que es capaz de catalizar el proceso

de descomposición de la hemicelulosa en polisacáridos más pequeños, de forma parcial, por ejemplo, en oligosacáridos, o de forma completa en monómeros de azúcar, por ejemplo, monómeros de azúcar de hexosa o pentosa. Una hemicelulasa de acuerdo con la invención puede dar lugar a una población mixta de oligosacáridos y monómeros de azúcar cuando se pone en contacto con la hemicelulasa. Dicha degradación típicamente tendrá lugar mediante una reacción de hidrólisis.

En este documento, una pectinasa es cualquier polipéptido que sea capaz de degradar pectina. Un polipéptido que es capaz de degradar pectina es uno que es capaz de catalizar el proceso de descomposición de la pectina en unidades más pequeñas, de forma parcial, por ejemplo, en oligosacáridos, o de forma completa en monómeros de azúcar. Una pectinasa de acuerdo con la invención puede dar lugar a una población mixta de oligosacáridos y monómeros de azúcar cuando se pone en contacto con la pectinasa. Dicha degradación típicamente tendrá lugar mediante una reacción de hidrólisis.

Por consiguiente, una composición de la invención puede comprender cualquier celulasa, por ejemplo, una celobiohidrolasa, una endo- β -1,4-glucanasa, una β -glucosidasa o una β -(1,3)(1,4)-glucanasa.

En este documento, una celobiohidrolasa es cualquier polipéptido que sea capaz de catalizar la hidrólisis de enlaces 1,4- β -D-glucosídicos en celulosa o celotetraosa, liberando celobiosa desde los extremos de las cadenas. Esta enzima también puede mencionarse como celulasa, 1,4- β -celobiosidasa, 1,4- β -celobiohidrolasa, 1,4- β -D-glucano celobiohidrolasa, avicelasa, exo-1,4- β -D-glucanasa, exo-glucanasa o exoglucanasa. Puede tener el código EC, EC 3.2.1.91. Las celobiohidrolasas pueden subdividirse en celobiohidrolasa I (CBH I) y celobiohidrolasa II (CBH II). CBH I se define como celobiohidrolasas que hidrolizan celulosa predominantemente desde los extremos reductores, separando celobiosa. CBH II se define como celobiohidrolasas que hidrolizan celulosa predominantemente desde los extremos no reductores, separando celobiosa.

En este documento, una endo- β -1,4-glucanasa (EC 3.2.1.4) es cualquier polipéptido que sea capaz de catalizar la endohidrólisis de enlaces 1,4- β -D-glucosídicos en celulosa, liquenina o β -D-glucanos de cereal. Dicho polipéptido puede ser capaz también de hidrolizar enlaces 1,4 en β -D-glucanos que también contienen enlaces 1,3. Esta enzima también puede mencionarse como celulasa, avicelasa, β -1,4-endoglucano hidrolasa, β -1,4-glucanasa, carboximetil celulasa, celudextrinasa, endo-1,4- β -D-glucanasa, endo-1,4- β -D-glucanohidrolasa, endo-1,4- β -glucanasa o endoglucanasa. CEA es en este documento una endoglucanasa EC 3.2.1.4 que en base a su estructura 3D se clasifica en la familia de Glucosil Hidrolasa 5 (GH5). Las actividades conocidas para miembros de la familia GH5 incluyen quitosanasa (EC 3.2.1.132); β -manosidasa (EC 3.2.1.25); celulasa (EC 3.2.1.4); glucano 1,3- β -glucosidasa (EC 3.2.1.58); liqueninasa (EC 3.2.1.73); glucano endo-1,6- β -glucosidasa (EC 3.2.1.75); manano endo- β -1,4-manosidasa (EC 3.2.1.78); endo- β -1,4-xilanasa (EC 3.2.1.8); celulosa β -1,4-celobiosidasa (EC 3.2.1.91); endo- β -1,6-galactanasa (EC 3.2.1.-); β -1,3-mananasa (EC 3.2.1.-); endo- β -1,4-glucanasa específica de xiloglucano (EC 3.2.1.151); manano transglucosilasa (EC 2.4.1.-). CEB en este documento es una endoglucanasa EC 3.2.1.4 que en base a su estructura 3D se clasifica en la familia de Glucosil Hidrolasa 7 (GH7). Las actividades conocidas para miembros de la familia GH7 incluyen endo- β -1,4-glucanasa (EC 3.2.1.4); [acción en extremos reductores] celobiohidrolasa (EC 3.2.1.-); quitosanasa (EC 3.2.1.132); endo- β -1,3-1,4-glucanasa (EC 3.2.1.73).

En este documento, una β -glucosidasa (abreviada BG) (EC 3.2.1.21) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos terminales de β -D-glucosa no reductores con liberación de β -D-glucosa. Dicho polipéptido puede tener una amplia especificidad por β -D-glucósidos y también puede hidrolizar uno o más de los siguientes: un β -D-galactósido, un α -L-arabinósido, un β -D-xilósido o un β -D-fucósido. Esta enzima también puede mencionarse como amigdalasa, β -D-glucósido glucohidrolasa, celobiasa o gentiobiasa.

En este documento, una β -(1,3)(1,4)-glucanasa (EC 3.2.1.73) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de enlaces 1,4- β -D-glucosídicos en β -D-glucanos que contienen enlaces 1,3 y 1,4. Dicho polipéptido puede actuar sobre liquenina y β -D-glucanos de cereal, pero no sobre β -D-glucanos que contienen solamente enlaces 1,3 o 1,4. Esta enzima también puede mencionarse como liqueninasa, 1,3-1,4- β -D-glucano 4-glucanohidrolasa, β -glucanasa, endo- β -1,3-1,4 glucanasa, liquenasa o β -glucanasa de enlaces mixtos. Una alternativa para este tipo de enzima es EC 3.2.1.6, que se describe como endo-1,3(4)-beta-glucanasa. Este tipo de enzima hidroliza enlaces 1,3 o 1,4 en beta-D-glucanos cuando el resto de glucosa cuyo grupo reductor está implicado en el enlace a hidrolizarse está en sí mismo sustituido en C-3. Nombres alternativos incluyen endo-1,3-beta-glucanasa, laminarinasa, 1,3-(1,3;1,4)-beta-D-glucano 3 (4) glucanohidrolasa; sustratos incluyen laminarina, liquenina y beta-D-glucanos de cereal.

Una composición de acuerdo con la invención puede comprender una enzima de la familia GH61 o cualquier otra enzima que tenga actividad potenciadora de celulasa.

La actividad potenciadora de celulasa en este documento se define como la potenciación de la actividad de al menos una celulasa. Cuando la proteína beta-glucosidasa de la invención está presente en una mezcla con una o más celulasas, por ejemplo, en una mezcla con celobiohidrolasa (CBH) y beta-glucosidasa (BG), potenciará la actividad de estas celulasas, que provocará una mayor actividad de la mezcla para degradar la celulosa. Las enzimas en la familia GH61 originalmente se clasificaron como una familia de glucósido hidrolasa en base a la medición de una

actividad endo-1,4-b-D-glucanasa muy débil en un miembro de la familia (endoglucanasa (EC 3.2.1.4)). La estructura y modo de acción de estas enzimas son ciertamente no canónicos y no pueden considerarse como glucosidasas auténticas. Sin embargo, se mantienen en la clasificación CAZy en base a su capacidad de potenciar la descomposición de lignocelulosa cuando se usa junto con una celulasa o una mezcla de celulasas. Se da una visión general de los miembros conocidos de la familia GH61 en la figura 5 de Harris, PV et al., Biochemistry 2010, **49**, 3305-3316.

Una composición de la invención puede comprender cualquier hemicelulasa, por ejemplo, una endoxilanasas, una β -xilosidasa, una α -L-arabinofuranosidasa, una α -D-glucuronidasa, una acetil xilano esterasa, una feruloil esterasa, una cumaroil esterasa, una α -galactosidasa, una β -galactosidasa, una β -mananasa o una β -manosidasa.

En este documento, una endoxilanasas (EC 3.2.1.8) es cualquier polipéptido que sea capaz de catalizar la endohidrólisis de enlaces 1,4- β -D-xilosídico en xilanos. Esta enzima también puede mencionarse como endo-1,4- β -xilanasas o 1,4- β -D-xilano xilanolhidrolasa. Una alternativa es EC 3.2.1.136, una glucuronoarabinoxilano endoxilanasas, una enzima que es capaz de hidrolizar enlaces 1,4-xilosídicos en glucuronoarabinoxilanos.

En este documento, una β -xilosidasa (EC 3.2.1.37) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de 1,4- β -D-xilanos, para retirar restos sucesivos de D-xilosa de los extremos no reductores. Dichas enzimas también pueden hidrolizar xilobiosa. Esta enzima también puede mencionarse como xilano 1,4- β -xilosidasa, 1,4- β -D-xilano xilohidrolasa, exo-1,4- β -xilosidasa o xilobiasa.

En este documento, una α -L-arabinofuranosidasa (EC 3.2.1.55) es cualquier polipéptido que sea capaz de actuar sobre α -L-arabinofuranósidos, α -L-arabinanos que contienen enlaces (1,2) y/o (1,3) y/o (1,5), arabinoxilanos y arabinogalactanos. Esta enzima también puede mencionarse como α -N-arabinofuranosidasa, arabinofuranosidasa o arabinosidasa.

En este documento, una α -D-glucuronidasa (EC 3.2.1.139) es cualquier polipéptido que sea capaz de catalizar una reacción de la siguiente forma: alfa-D-glucuronósido + H(2)O = un alcohol + D-glucuronato. Esta enzima también puede mencionarse como alfa-glucuronidasa o alfa-glucosiduronasa. Estas enzimas también pueden hidrolizar ácido glucurónico 4-O-metilado, que también puede estar presente como sustituyente en xilanos. Una alternativa es EC 3.2.1.131: xilano alfa-1,2-glucuronosidasa, que cataliza la hidrólisis de enlaces alfa-1,2-(4-O-metil)glucuronosilo.

En este documento, una acetil xilano esterasa (EC 3.1.1.72) es cualquier polipéptido que sea capaz de catalizar la desacetilación de xilanos y xilo-oligosacáridos. Dicho polipéptido puede catalizar la hidrólisis de grupos acetilo desde xilano polimérico, xilosa acetilada, glucosa acetilada, acetato de alfa-naptilo o acetato de p-nitrofenilo pero, típicamente, no desde triacilglicerol. Dicho polipéptido típicamente no actúa sobre manano acetilado o pectina.

En este documento, una feruloil esterasa (EC 3.1.1.73) es cualquier polipéptido que sea capaz de catalizar una reacción de la forma: feruloil-sacárido + H(2)O = ferulato + sacárido. El sacárido puede ser, por ejemplo, un oligosacárido o un polisacárido. Puede catalizar típicamente la hidrólisis del grupo 4-hidroxi-3-metoxicinnamoilo (feruloilo) desde un azúcar esterificado, que es habitualmente arabinosa en sustratos 'naturales'. El acetato de p-nitrofenol y ferulato de metilo son típicamente peores sustratos. Esta enzima también puede mencionarse como cinamoil éster hidrolasa, ácido ferúlico esterasa o hidroxycinamoil esterasa. También puede mencionarse como una enzima accesoria de hemicelulasa, ya que puede ayudar a las xilanasas y pectinasas a descomponer la hemicelulosa y pectina de la pared celular de las plantas.

En este documento, una cumaroil esterasa (EC 3.1.1.73) es cualquier polipéptido que sea capaz de catalizar una reacción de la forma: cumaroil-sacárido + H(2)O = cumarato + sacárido. El sacárido puede ser, por ejemplo, un oligosacárido o un polisacárido. Esta enzima también puede mencionarse como trans-4-cumaroil esterasa, trans-p-cumaroil esterasa, p-cumaroil esterasa o ácido p-cumárico esterasa. Esta enzima también está dentro de EC 3.1.1.73 de modo que también puede mencionarse como feruloil esterasa.

En este documento, una α -galactosidasa (EC 3.2.1.22) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos terminales de α -D-galactosa no reductores en α -D-galactósidos, incluyendo oligosacáridos de galactosa, galactomananos, galactanos y arabinogalactanos. Dicho polipéptido también puede ser capaz de hidrolizar α -D-fucósidos. Esta enzima también puede mencionarse como melibiasa.

En este documento, una β -galactosidasa (EC 3.2.1.23) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos terminales de β -D-galactosa no reductores en β -D-galactósidos. Dicho polipéptido también puede ser capaz de hidrolizar α -L-arabinósidos. Esta enzima también puede mencionarse como exo-(1- $\rightarrow$ 4)- β -D-galactanasas o lactasa.

En este documento, una β -mananasa (EC 3.2.1.78) es cualquier polipéptido que sea capaz de catalizar la hidrólisis aleatoria de enlaces 1,4- β -D-manosídicos en mananos, galactomananos y glucomanos. Esta enzima también puede mencionarse como manano endo-1,4- β -manosidasa o endo-1,4-mananasa.

En este documento, una β -manosidasa (EC 3.2.1.25) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos terminales de β -D-manosa no reductores en β -D-manósidos. Esta enzima también puede mencionarse como mananasa o manasa.

Una composición de la invención puede comprender cualquier pectinasa, por ejemplo, una endo poligalacturonasa, una pectin metil esterasa, una endo-galactanasa, una beta galactosidasa, una pectin acetil esterasa, una endo-pectin liasa, pectato liasa, alfa ramnosidasa, una exo-galacturonasa, una exo-poligalacturonato liasa, una ramnogalacturonano hidrolasa, una ramnogalacturonano liasa, una ramnogalacturonano acetil esterasa, una ramnogalacturonano galacturonohidrolasa o una xilogalacturonasa.

En este documento, una endo-poligalacturonasa (EC 3.2.1.15) es cualquier polipéptido que sea capaz de catalizar la hidrólisis aleatoria de enlaces 1,4- α -D-galactosidurónicos en pectato y otros galacturonanos. Esta enzima también puede mencionarse como poligalacturonasa pectin despolimerasa, pectinasa, endopoligalacturonasa, pectolasa, pectin hidrolasa, pectin poligalacturonasa, poli- α -1,4-galacturónido glucanohidrolasa, endogalacturonasa; endo-D-galacturonasa o poli(1,4- α -D-galacturónido) glucanohidrolasa.

En este documento, una pectin metil esterasa (EC 3.1.1.11) es cualquier enzima que sea capaz de catalizar la reacción: pectina + n H₂O = n metanol + pectato. Esta enzima también puede conocerse como pectinesterasa, pectin desmetoxilasa, pectin metoxilasa, pectin metilesterasa, pectasa, pectinoesterasa o pectin pectilhidrolasa.

En este documento, una endo-galactanasa (EC 3.2.1.89) es cualquier enzima capaz de catalizar la endohidrólisis de enlaces 1,4- β -D-galactosídicos en arabinogalactanos. La enzima también puede conocerse como arabinogalactano endo-1,4- β -galactosidasa, endo-1,4- β -galactanasa, galactanasa, arabinogalactanasa o arabinogalactano 4- β -D-galactanohidrolasa.

En este documento, una pectin acetil esterasa se define en este documento como cualquier enzima que tiene una actividad acetil esterasa que cataliza la desacetilación de los grupos acetilo en los grupos hidroxilo de restos GalUA de pectina.

En este documento, una endo-pectin liasa (EC 4.2.2.10) es cualquier enzima capaz de catalizar la escisión eliminadora de (1 $\rightarrow$ 4)- α -D-galacturonano metil éster para dar oligosacáridos con grupos 4-desoxi-6-O-metil- α -D-galact-4-enuronosilo en sus extremos no reductores. La enzima también puede conocerse como pectin liasa, pectin trans-eliminasa; endo-pectin liasa, polimetilgalacturónico transeliminasa, pectin metiltranseliminasa, pectoliasa, PL, PNL o PMGL o (1 $\rightarrow$ 4)-6-O-metil- α -D-galacturonano liasa.

En este documento, una pectato liasa (EC 4.2.2.2) es cualquier enzima capaz de catalizar la escisión eliminadora de (1 $\rightarrow$ 4)- α -D-galacturonano para dar oligosacáridos con grupos 4-desoxi- α -D-galact-4-enuronosilo en sus extremos no reductores. La enzima también puede conocerse como poligalacturónico transeliminasa, ácido péctico transeliminasa, poligalacturonato liasa, endopectin metiltranseliminasa, pectato transeliminasa, endogalacturonato transeliminasa, ácido péctico liasa, péctico liasa, ácido α -1,4-D-endopoligalacturónico liasa, PGA liasa, PPasa-N, ácido endo- α -1,4-poligalacturónico liasa, ácido poligalacturónico liasa, pectin trans-eliminasa, ácido poligalacturónico trans-eliminasa o (1 $\rightarrow$ 4)- α -D-galacturonano liasa.

En este documento, una alfa ramnosidasa (EC 3.2.1.40) es cualquier polipéptido que sea capaz de catalizar la hidrólisis de restos terminales de α -L-ramnosa no reductores en α -L-ramnósidos o como alternativa en ramnogalacturonano. Esta enzima también se conoce como α -L-ramnosidasa T, α -L-ramnosidasa N o α -L-ramnósido ramnoidrolasa.

En este documento, exo-galacturonasa (EC 3.2.1.82) es cualquier polipéptido con capacidad de hidrólisis de ácido péctico desde el extremo no reductor, liberando digalacturonato. La enzima también puede conocerse como exo-poli- α -galacturonosidasa, exopoligalacturonosidasa o exopoligalacturanosidasa.

En este documento, exo-galacturonasa (EC 3.2.1.67) es cualquier polipéptido capaz de catalizar: (1,4- α -D-galacturónido)_n + H₂O = (1,4- α -D-galacturónido)_{n-1} + D-galacturonato. La enzima también puede conocerse como galacturano 1,4- α -galacturonidasa, exopoligalacturonasa, poli(galacturonato) hidrolasa, exo-D-galacturonasa, exo-D-galacturonanasa, exopoli-D-galacturonasa o poli(1,4- α -D-galacturónido) galacturonohidrolasa.

En este documento, exopoligalacturonato liasa (EC 4.2.2.9) es cualquier polipéptido capaz de catalizar la escisión eliminadora de 4-(4-desoxi- α -D-galact-4-enuronosil)-D-galacturonato desde el extremo reductor de pectato, es decir, pectina des-esterificada. Esta enzima puede conocerse como pectato disacárido-liasa, pectato exo-liasa, ácido exopéctico transeliminasa, exopectato liasa, ácido exopoligalacturónico-trans-eliminasa, PATE, exo-PATE, exo-PGL o extremo reductor de (1 $\rightarrow$ 4)- α -D-galacturonano-disacárido-liasa.

En este documento, ramnogalacturonano hidrolasa es cualquier polipéptido que sea capaz de hidrolizar el enlace entre ácido galactosilurónico y ramnopiranosilo en un modo endo en estructuras de ramnogalacturonano estrictamente alternantes, que consisten en el disacárido [ácido (1,2-alfa-L-ramnoil-(1,4)-alfa-galactosilurónico)].

En este documento, ramnogalacturonano liasa es cualquier polipéptido que es cualquier polipéptido que sea capaz de escindir enlaces α -L-Rap-(1 $\rightarrow$ 4)- α -D-GalpA en un modo endo en ramnogalacturonano por beta-eliminación.

- 5 En este documento, ramnogalacturonano acetil esterasa es cualquier polipéptido que cataliza la desacetilación de la estructura de restos alternantes de ramnosa y ácido galacturónico en ramnogalacturonano.

En este documento, ramnogalacturonano galacturonohidrolasa es cualquier polipéptido que sea capaz de hidrolizar ácido galacturónico desde el extremo no reductor de estructuras de ramnogalacturonano estrictamente alternantes en un modo exo.

En este documento, xilogalacturonasa es cualquier polipéptido que actúe sobre xilogalacturonano escindiendo la estructura de ácido galacturónico sustituido con β -xilosa en un modo endo. Esta enzima también puede conocerse como xilogalacturonano hidrolasa.

15 En este documento, una α -L-arabinofuranosidasa (EC 3.2.1.55) es cualquier polipéptido que sea capaz de actuar sobre α -L-arabinofuranósidos, α -L-arabinanos que contienen enlaces (1,2) y/o (1,3) y/o (1,5), arabinoxilanos y arabinogalactanos. Esta enzima también puede mencionarse como α -N-arabinofuranosidasa, arabinofuranosidasa o arabinosidasa.

20 En este documento, endo-arabinanasa (EC 3.2.1.99) es cualquier polipéptido que sea capaz de catalizar la endohidrólisis de enlaces 1,5- α -arabinofuranosídicos en 1,5-arabinanos. La enzima también puede conocerse como endo-arabinasa, arabinano endo-1,5- α -L-arabinosidasa, endo-1,5- α -L-arabinanasa, endo- α -1,5-arabanasa; endo-arabanasa o 1,5- α -L-arabinano 1,5- α -L-arabinanohidrolasa.

25 Una composición de la invención típicamente comprenderá al menos una celulasa y/o al menos una hemicelulasa y/o al menos una pectinasa (una de las cuales es un polipéptido de acuerdo con la invención). Una composición de la invención puede comprender una celobiohidrolasa, una endoglucanasa y/o una β -glucosidasa. Dicha composición también puede comprender una o más hemicelulasas y/o una o más pectinasas.

30 Una o más (por ejemplo, dos, tres, cuatro o todas) de una amilasa, una proteasa, una lipasa, una ligninasa, una hexosiltransferasa o una glucuronidasa pueden estar presentes en una composición de la invención.

35 "Proteasa" incluye enzimas que hidrolizan enlaces peptídicos (peptidasas), así como enzimas que hidrolizan enlaces entre péptidos y otros restos, tales como azúcares (glucopeptidasas). Muchas proteasas se caracterizan en EC 3.4, y son adecuadas para su uso en la invención. Algunos tipos específicos de proteasas incluyen, cisteína proteasas incluyendo pepsina, papaína y serina proteasas incluyendo quimotripsinas, carboxipeptidasas y metaloendopeptidasas.

40 "Lipasa" incluye enzimas que hidrolizan lípidos, ácidos grasos, y acilglicéridos, incluyendo fosfoglicéridos, lipoproteínas, diacilgliceroles, y similares. En plantas, se usan lípidos como componentes estructurales para limitar la pérdida de agua y la infección por patógenos. Estos lípidos incluyen ceras derivadas de ácidos grasos, así como cutina y suberina.

45 "Ligninasa" incluye enzimas que pueden hidrolizar o descomponer la estructura de polímeros de lignina. Enzimas que pueden descomponer la lignina incluyen lignina peroxidasas, manganeso peroxidasas, lacasas y feruloil esterases, y otras enzimas descritas en la técnica conocidas por despolimerizar o descomponer de otro modo polímeros de lignina. También se incluyen enzimas capaces de hidrolizar enlaces formados entre azúcares hemicelulósicos (notablemente arabinosa) y lignina. Las ligninasas incluyen, aunque sin limitación, el siguiente grupo de enzimas: lignina peroxidasas (EC 1.11.14), manganeso peroxidasas (EC 1.11.1.13), lacasas (EC 1.10.3.2) y feruloil esterases (EC 3.1.1.73).

50 "Hexosiltransferasa" (2.4.1-) incluye enzimas que son capaces de transferir grupos glicosilo, más específicamente grupos hexosilo. Además de transferir un grupo glicosilo de un donante que contiene glicosilo a otro compuesto que contiene glicosilo, el aceptor, las enzimas también pueden transferir el grupo glicosilo a agua como aceptor. Esta reacción también se conoce como reacción de hidrólisis, en lugar de una reacción de transferencia. Un ejemplo de una hexosiltransferasa que puede usarse en la invención es una β -glucanosiltransferasa. Dicha enzima puede ser capaz de catalizar la degradación de (1,3)(1,4)glucano y/o celulosa y/o un producto de degradación de celulosa.

55 "Glucuronidasa" incluye enzimas que catalizan la hidrólisis de un glucuronósido, por ejemplo, β -glucuronósido para producir un alcohol. Se han caracterizado, muchas glucuronidasas y pueden ser adecuadas para su uso en la invención, por ejemplo, β -glucuronidasa (EC 3.2.1.31), hialurono-glucuronidasa (EC 3.2.1.36), glucuronosil-disulfoglucosamina glucuronidasa (3.2.1.56), glicirrizinato β -glucuronidasa (3.2.1.128) o α -D-glucuronidasa (EC 3.2.1.139).

60 Una composición de la invención puede comprender una expansina o proteína tipo expansina, tal como una

swollenina (véase, Salheimo et al., Eur. J. Biochem. 269, 4202-4211, 2002) o una proteína tipo swollenina.

Las expansinas están implicadas en la distensión de la estructura de pared celular durante el crecimiento de las células vegetales. Se ha propuesto que las expansinas alteran los enlaces de hidrógeno entre la celulosa y otros polisacáridos de la pared celular sin tener actividad hidrolítica. De este modo, se cree que permiten el deslizamiento de las fibras de celulosa y el agrandamiento de la pared celular. La swollenina, una proteína tipo expansina, contiene un dominio N-terminal de la familia 1 de módulo de unión a carbohidrato (CBD) y un dominio C-terminal tipo expansina. Para los propósitos de esta invención, una proteína tipo expansina o proteína tipo swollenina puede comprender uno o ambos de dichos dominios y/o puede alterar la estructura de las paredes celulares (tal como alterar la estructura de la celulosa), opcionalmente sin producir cantidades detectables de azúcares reductores.

Una composición de la invención puede comprender el producto polipeptídico de una proteína integradora de celulosa, scaffoldina una proteína tipo scaffoldina, por ejemplo, CipA o CipC de *Clostridium thermocellum* o *Clostridium cellulolyticum*, respectivamente.

Las scaffoldinas y proteínas integradoras de celulosa son subunidades integradoras multi-funcionales que puede organizar subunidades celolíticas en un complejo multi-enzimático. Esto se consigue mediante la interacción de dos clases complementarias de dominio, es decir, un dominio de cohesión en scaffoldina y un dominio dockerina en cada unidad enzimática. La subunidad scaffoldina también alberga un módulo de unión a celulosa (CBM) que media la unión del celulosoma a su sustrato. Una scaffoldina o proteína integradora de celulosa para los propósitos de esta invención puede comprender uno o ambos de dichos dominios.

Una composición de la invención puede comprender una proteína inducida por o proteína moduladora de celulosa, por ejemplo, codificada por el gen cip1 o cip2 o genes similares de *Trichoderma reesei* / *Hypocrea jecorina* (véase, Foreman et al., J. Biol. Chem. 278(34), 31988-31997, 2003). El producto polipeptídico de estos genes son proteínas bimodulares, que contienen un módulo de unión a celulosa y un dominio cuya función o actividad no puede relacionarse con familias conocidas de glicosil hidrolasa. Además, la presencia de un módulo de unión a celulosa y la co-regulación de la expresión de estos genes con componentes celulasas indica actividades previamente no reconocidas con papel potencial en degradación de biomasa.

Una composición de la invención puede estar compuesta por un miembro de cada una de las clases de los polipéptidos mencionados anteriormente, varios miembros de una clase de polipéptidos, o cualquier combinación de estas clases de polipéptidos.

Una composición de la invención puede estar compuesto por polipéptidos, por ejemplo, enzimas, de (1) proveedores comerciales; (2) genes clonados que expresan polipéptidos, por ejemplo, enzimas; (3) caldo complejo (tal como el resultante del cultivo de una cepa microbiana en medios, donde las cepas secretan proteínas y enzimas en el medio; (4) lisados celulares de cepas cultivadas como en (3); y/o (5) material vegetal que expresa polipéptidos, por ejemplo, enzimas. Pueden obtenerse diferentes polipéptidos, por ejemplo, enzimas, en una composición de la invención a partir de diferentes fuentes.

Uso de los polipéptidos

Los polipéptidos y composiciones polipeptídicas de acuerdo con la invención pueden usarse en muchas aplicaciones diferentes. Por ejemplo, pueden usarse para producir azúcares fermentables. Los azúcares fermentables pueden convertirse después, como parte de un proceso de biocombustible, en biogás o etanol, butanol, isobutanol, 2 butanol u otras sustancias adecuadas. Como alternativa, los polipéptidos y sus composiciones pueden usarse como enzima, por ejemplo, en la producción de productos alimenticios, en composiciones detergentes, en la industria del papel y la pulpa y en formulaciones antibacterianas, en productos farmacéuticos tales como pastillas para la garganta, dentífricos, y enjuagues bucales. Algunos de los usos se ilustrarán en más detalle a continuación.

En los usos y métodos descritos a continuación, los componentes de las composiciones descritas anteriormente pueden proporcionarse de forma concomitante (es decir, como una única composición *per se*) o por separado o secuencialmente.

La invención también se refiere al uso del polipéptido de beta-glucosidasa de acuerdo con la invención y composiciones que comprenden dicha enzima en procesos industriales.

A pesar de la experiencia a largo plazo con estos procesos, el polipéptido de beta-glucosidasa de acuerdo con la invención puede mostrar varias ventajas significativas sobre enzimas actualmente usadas. Dependiendo de la aplicación específica, estas ventajas pueden incluir aspectos tales como inferiores costes de producción, mayor especificidad hacia el sustrato, antigenicidad reducida, menores actividades secundarias indeseables, mayores rendimientos cuando se producen en un microorganismo adecuado, intervalos más adecuados de pH y temperatura, ausencia de inhibición por productos hidrófobos derivados de lignina o menos inhibición del producto o, en el caso de la industria de la alimentación, un mejor sabor o textura de un producto final así como grado de calidad de los alimentos y aspectos kosher.

En principio, un polipéptido de beta-glucosidasa o composición de la invención puede usarse en cualquier proceso que requiera el tratamiento de un material que comprende polisacárido. Por tanto, un polipéptido o composición de la invención puede usarse en el tratamiento de material de polisacárido. En este documento, material de polisacárido es un material que comprende o consiste esencialmente en uno o, más típicamente, más de un polisacárido.

Típicamente, las plantas y material derivado de las mismas comprenden cantidades significativas de material de polisacárido que no es almidón. Por consiguiente, un polipéptido de la invención puede usarse en el tratamiento de un material vegetal o fúngico o un material derivado del mismo.

Lignocelulosa

Un componente importante del material vegetal de polisacárido que no es almidón es lignocelulosa (también mencionada en este documento como biomasa lignocelulolítica). La lignocelulosa es material vegetal que comprende celulosa y hemicelulosa y lignina. Los polímeros de carbohidrato (celulosa y hemicelulosas) están unidos estrechamente a la lignina por enlaces de hidrógeno y covalentes. Por consiguiente, un polipéptido de la invención puede usarse en el tratamiento de material lignocelulolítico. En este documento, material lignocelulolítico es un material que comprende o consiste esencialmente en lignocelulosa. Por tanto, en un método de la invención para el tratamiento de un polisacárido que no es almidón, el polisacárido que no es almidón puede ser un material/biomasa lignocelulósica.

Por consiguiente, la invención proporciona un método para tratar un sustrato que comprende polisacárido que no es almidón en que el tratamiento comprende la degradación y/o hidrólisis y/o modificación de celulosa y/o hemicelulosa y/o una sustancia péctica.

Degradación, en este contexto, indica que el tratamiento provoca la generación de productos de hidrólisis de celulosa y/o hemicelulosa y/o una sustancia péctica, es decir, hay sacáridos de longitud más corta presentes como resultado del tratamiento, que los presentes en un polisacárido que no es almidón no tratado similar. Por tanto, la degradación, en este contexto, puede provocar la liberación de oligosacáridos y/o monómeros de azúcar.

Todas las plantas contienen polisacárido que no es almidón como casi todos los materiales de polisacárido derivados de plantas. Por consiguiente, en un método de la invención para el tratamiento de sustrato que comprende un polisacárido que no es almidón, dicho sustrato puede proporcionarse en forma de una planta o material derivado de planta o un material que comprende una planta o material derivado de planta, por ejemplo, una pulpa vegetal, un extracto vegetal, un producto alimenticio o ingrediente para el mismo, una tela, un tejido o una prenda de vestir.

La biomasa lignocelulolítica adecuada para su uso en la invención incluye biomasa y puede incluir biomasa virgen y/o biomasa no virgen tal como biomasa agrícola, compuestos orgánicos comerciales, desechos de construcción y demolición, desechos sólidos municipales, desperdicios de papel y residuos de jardinería. Las formas comunes de biomasa incluyen árboles, matorrales y gramíneas, trigo, paja de trigo, bagazo de caña de azúcar, maíz, espigas del maíz, mazorcas del maíz, granos de maíz incluyendo fibra de los granos, productos y sub-productos de la molienda de granos tales como maíz, trigo y cebada (incluyendo molienda en húmedo y molienda en seco) a menudo llamado "salvado o fibra" así como desechos sólidos municipales, desperdicios de papel y residuos de jardinería. La biomasa también puede ser, aunque sin limitación, material herbáceo, residuos agrícolas, residuos forestales, desechos sólidos municipales, desperdicios de papel, y residuos de fabricación de pulpa y papel. "Biomasa agrícola" incluye ramas, setos, cañas, maíz y espigas del maíz, cultivos energéticos, bosques, frutos, flores, granos, gramíneas, cultivos herbáceos, hojas, corteza, agujas, leños, raíces, árboles jóvenes, cultivos leñosos de rotación corta, matorrales, pastos varilla, árboles, vegetales, mondas, vides, pulpa de remolacha azucarera, moyuelo de trigo, cáscaras de avena, y maderas duras y blandas (sin incluir maderas con materiales nocivos). Además, la biomasa agrícola incluye materiales orgánicos de desecho generados por procesos agrícolas incluyendo actividades de agricultura y silvicultura, específicamente incluyendo residuos forestales de madera. La biomasa agrícola puede ser cualquiera de las mencionadas anteriormente de forma individual o en cualquier combinación o mezcla de las mismas. Ejemplos adicionales de biomasa adecuada son preparaciones de huerto, chaparral, residuos de fábricas, residuos urbanos de madera, residuos municipales, residuos del madereo, rastrojos forestales, cultivos leñosos de rotación corta, residuos industriales, paja de trigo, paja de avena, paja de arroz, paja de cebada, paja de centeno, cáñamo de lino, cáscaras de soja, cáscaras de arroz, paja de arroz, forraje con gluten de maíz, cáscaras de avena, caña de azúcar, rastrojo de maíz, tallos de maíz, mazorcas del maíz, espigas del maíz, césped de pradera, maicillo, cola de zorra; pulpa de remolacha azucarera, pulpa de frutos cítricos, cáscaras de semillas, residuos animales celulósicos, recortes de césped, algodón, algas marinas, árboles, matorrales, gramíneas, trigo, paja de trigo, bagazo de caña de azúcar, maíz, espigas del maíz, mazorcas de maíz, granos de maíz, fibra de granos, productos y sub-productos de molienda en húmedo o seco de granos, desechos sólidos municipales, desperdicios de papel, residuos de jardinería, material herbáceo, residuos agrícolas, residuos forestales, desechos sólidos municipales, desperdicios de papel, pulpa, residuos de fabricación del papel, ramas, setos, canes, maíz, espigas del maíz, un cultivo energético, bosque, una fruta, una flor, una grana, una gramínea, un cultivo herbáceo, una hoja, corteza, una aguja, un leño, una raíz, un árbol joven, un arbusto, pasto varilla, un árbol, una hortaliza, monda, una vid, pulpa de remolacha azucarera, moyuelo de trigo, cáscaras de avena, madera dura o blanda, material orgánico de desecho

generado de un proceso agrícola, residuos forestales de madera, o una combinación de dos cualesquiera o más de los mismos.

- 5 Aparte de la biomasa virgen o las materias primas ya procesadas en las industrias de los alimentos y piensos o el papel y la pulpa, la biomasa/materia prima puede pretratarse adicionalmente con calor, modificación mecánica y/o química o cualquier combinación de dichos métodos para potenciar la degradación enzimática.

Pretratamiento

- 10 Antes del tratamiento enzimático, la materia prima puede pretratarse opcionalmente con calor, modificación mecánica y/o química o cualquier combinación de dichos métodos para potenciar la accesibilidad del sustrato a hidrólisis enzimática y/o hidrolizar la hemicelulosa y/o solubilizar la hemicelulosa y/o celulosa y/o lignina, de cualquier modo conocido en la técnica. El pretratamiento puede comprender la exposición del material lignocelulósico a agua (caliente), vapor (explosión de vapor), un ácido, una base, un disolvente, calor, un peróxido, ozono, desfibración mecánica, trituración, molienda o despresurización rápida, o una combinación de dos cualesquiera o más de los mismos. A menudo se combina un pretratamiento químico con pretratamiento por calor, por ejemplo, entre 150-220 °C durante 1 a 30 minutos.

Presacarificación

- 20 Después de la etapa de pretratamiento, puede utilizarse una etapa de licuefacción/hidrólisis o presacarificación que implica incubación con una enzima o mezcla de enzimas. La etapa de presacarificación puede realizarse a muchas temperaturas diferentes, pero se prefiere que la etapa de presacarificación sucede a la mejor temperatura adecuada para la mezcla de enzimas que se está aplicando, o el óptimo enzimático predicho de las enzimas a aplicar. La temperatura de la etapa de presacarificación puede variar de aproximadamente 10 °C a aproximadamente 95 °C, de aproximadamente 20 °C a aproximadamente 85 °C, de aproximadamente 30 °C a aproximadamente 70 °C, de aproximadamente 40 °C a aproximadamente 60 °C, de aproximadamente 37 °C a aproximadamente 50 °C, preferiblemente de aproximadamente 37 °C a aproximadamente 80 °C, más preferiblemente a aproximadamente 60-70 °C, incluso más preferiblemente a aproximadamente 65 °C. El pH de la mezcla de presacarificación puede variar de aproximadamente 2,0 a aproximadamente 10,0, pero es preferiblemente de aproximadamente 3,0 a aproximadamente 7,0, más preferiblemente de aproximadamente 4,0 a aproximadamente 6,0, incluso más preferiblemente de aproximadamente 4,0 a aproximadamente 5,0. De nuevo, el pH puede ajustarse para maximizar la actividad enzimática y puede ajustarse con la adición de la enzima.
- 35 La reacción de la etapa de licuefacción/hidrólisis o presacarificación puede suceder desde varios minutos hasta varias horas, tal como de aproximadamente 1 hora a aproximadamente 120 horas, preferiblemente de aproximadamente 2 horas a aproximadamente 48 horas, más preferiblemente de aproximadamente 2 a aproximadamente 24 horas, mucho más preferiblemente durante de aproximadamente 2 a aproximadamente 6 horas. El tratamiento con celulasa puede suceder desde varios minutos hasta varias horas, tal como de aproximadamente 6 horas a aproximadamente 120 horas, preferiblemente de aproximadamente 12 horas a aproximadamente 72 horas, más preferiblemente de aproximadamente 24 a 48 horas.

Sacarificación

- 45 La invención proporciona un método para producir un azúcar a partir de un material lignocelulósico, comprendiendo dicho método poner en contacto un polipéptido de la invención o una composición de la invención con el material lignocelulósico.

- 50 Dicho método permite la generación de azúcares libres (monómeros) y/u oligosacáridos a partir de biomasa lignocelulósica. Estos métodos implican la conversión de biomasa lignocelulósica en azúcares libres y pequeños oligosacáridos con un polipéptido o composición de la invención.

- 55 El proceso de conversión de un carbohidrato complejo tal como lignocelulosa en azúcares preferiblemente permite la conversión en azúcares fermentables. Dicho proceso puede mencionarse como "sacarificación". Por consiguiente, un método de la invención puede provocar la liberación de uno o más azúcares de hexosa y/o pentosa, tales como uno o más de glucosa, xilosa, arabinosa, galactosa, ácido galacturónico, ácido glucurónico, manosa, ramnosa, ribosa y fructosa.

- 60 Por consiguiente, otro aspecto de la invención incluye métodos que utilizan el polipéptido o composición de la invención descrita anteriormente junto con enzimas adicionales o tratamiento físicos tales como temperatura y pH para convertir la biomasa vegetal lignocelulósica en azúcares y oligosacáridos.

- 65 Aunque la composición se ha analizado como una única mezcla, se reconoce que las enzimas pueden añadirse secuencialmente, donde la temperatura, pH, y otras condiciones pueden alterarse para aumentar la actividad de cada enzima individual. Como alternativa, puede determinarse un pH y temperatura óptimos para la mezcla de enzimas.

Las enzimas se hacen reaccionar con el sustrato en cualquier condición apropiada. Por ejemplo, las enzimas pueden incubarse a aproximadamente 25 °C, aproximadamente 30 °C, aproximadamente 35 °C, aproximadamente 37 °C, aproximadamente 40 °C, aproximadamente 45 °C, aproximadamente 50 °C, aproximadamente 55 °C, aproximadamente 60 °C, aproximadamente 65 °C, aproximadamente 70 °C, aproximadamente 75 °C, aproximadamente 80 °C, aproximadamente 85 °C, aproximadamente 90 °C o mayor. Es decir, pueden incubarse a una temperatura de aproximadamente 20 °C a aproximadamente 95 °C, por ejemplo, en tampones de baja a media fuerza iónica y/o de pH bajo a neutro. Por "fuerza iónica media" se entiende que el tampón tiene una concentración de iones de aproximadamente 200 milimolar (mM) o menos para cualquier componente iónico individual. El pH puede variar de aproximadamente pH 2,5, aproximadamente pH 3,0, aproximadamente pH 3,5, aproximadamente pH 4,0, aproximadamente pH 4,5, aproximadamente pH 5, aproximadamente pH 5,5, aproximadamente pH 6, aproximadamente pH 6,5, aproximadamente pH 7, aproximadamente pH 7,5, aproximadamente pH 8,0, a aproximadamente pH 8,5. Generalmente, el intervalo de pH será de aproximadamente pH 3,0 a aproximadamente pH 7. Para la producción de etanol se prefiere un medio ácido, por ejemplo, pH = 4, mientras que para la producción de biogás es deseable pH neutro, por ejemplo, pH = 7. La incubación de combinaciones de enzimas en estas condiciones provoca la emisión o liberación de cantidades sustanciales del azúcar desde la lignocelulosa. Por cantidad sustancial se entiende al menos el 20 %, 30 %, 40 %, 50 %, 60 %, 70 %, 80 %, 90 %, 95 % o más del azúcar disponible.

Los polipéptidos, tales como enzimas, pueden producirse de forma exógena en microorganismos, levaduras, hongos, bacterias o plantas, después pueden aislarse y añadirse, por ejemplo, a materia prima lignocelulósica. Como alternativa, las enzimas se producen, pero no se aíslan, y puede añadirse caldo de fermentación de masa celular sin procesar, o material vegetal (tal como rastrojo de maíz), y similares a, por ejemplo, la materia prima. Como alternativa, la masa celular sin procesar o el medio de producción enzimática o el material vegetal puede tratarse para evitar el crecimiento microbiano adicional (por ejemplo, por calentamiento o adición de agentes antimicrobianos), después se añade a, por ejemplo, una materia prima. Estas mezclas de enzimas sin procesar pueden incluir el organismo que produce la enzima. Como alternativa, la enzima puede producirse en una fermentación que usa materia prima (tal como rastrojo de maíz) para proporcionar nutrición a un organismo que produce una o más enzimas. De este modo, las plantas que producen las enzimas pueden servir, en sí mismas, como materia prima lignocelulósica y pueden añadirse a materia prima lignocelulósica.

Fermentación de azúcares

Los azúcares fermentables pueden convertirse en productos útiles de fermentación de valor añadido, cuyos ejemplos no limitantes incluyen aminoácidos, vitaminas, agentes farmacéuticos, suplementos de piensos animales, agentes químicos especializados, materias primas químicas, plásticos, disolventes, combustibles, u otros polímeros orgánicos, ácido láctico, y etanol, incluyendo etanol combustible. En particular, los azúcares pueden usarse como materias primas para fermentación en agente químicos, plásticos, tales como, por ejemplo, ácido succínico y (bio) combustibles, incluyendo etanol, metanol, butanol, combustibles líquidos sintéticos y biogás.

Por ejemplo, en el método de la invención, una enzima o combinación de enzimas actúa sobre un sustrato lignocelulósico o biomasa vegetal, sirviendo como materia prima, para convertir este sustrato complejo en azúcares simples y oligosacáridos para la producción de etanol u otros productos útiles de fermentación.

Los azúcares liberados de biomasa pueden convertirse en productos útiles de fermentación, incluyendo algunos de ellos, aunque sin limitación, aminoácidos, vitaminas, agentes farmacéuticos, suplementos de piensos animales, agentes químicos especializados, materias primas químicas, plásticos, y etanol, incluyendo etanol combustible.

Por consiguiente, la invención proporciona un método para la preparación de un producto de fermentación, comprendiendo dicho método:

- a. degradación de lignocelulosa usando un método como se describe en este documento; y
- b. fermentación del material resultante,

para preparar de ese modo un producto de fermentación.

La fermentación puede realizarse en condiciones aeróbicas o anaeróbicas. Preferiblemente, el proceso se realiza en condiciones micro-aerófilas o de oxígeno limitado.

Un proceso anaeróbico de fermentación en este documento se define como un proceso de fermentación ejecutado en ausencia de oxígeno o en que no se consumen sustancialmente nada de oxígeno, preferiblemente aproximadamente 5 o menos, aproximadamente 2,5 o menos o aproximadamente 1 mmol/l/h o menos, y donde las moléculas orgánicas sirven como donantes de electrones y también como aceptores de electrones.

Un proceso de fermentación de oxígeno limitado es un proceso en que el consumo de oxígeno está limitado por la transferencia de oxígeno desde el gas hasta el líquido. El grado de limitación de oxígeno se determina por la

cantidad y composición del flujo de gas existente, así como la mezcla real/propiedades de transferencia de masas del equipo de fermentación usado. Preferiblemente, en un proceso en condiciones de oxígeno limitado, la tasa de consumo de oxígeno es de al menos aproximadamente 5,5, más preferiblemente de al menos aproximadamente 6 e incluso más preferiblemente de al menos aproximadamente 7 mmol/l/h.

Un método para la preparación de un producto de fermentación puede comprender opcionalmente la recuperación del producto de fermentación.

SSF

La fermentación y sacarificación también pueden ejecutarse en modo simultáneo de sacarificación y fermentación (SSF). Una de las ventajas de este modo es la reducción de inhibición de azúcar sobre hidrólisis enzimática (la inhibición de azúcar sobre celulasas se describe por Caminal B&B Vol XXVII pág. 1282-1290).

Productos de fermentación

Los productos de fermentación que pueden producirse de acuerdo con la invención incluyen aminoácidos, vitaminas, agentes farmacéuticos, suplementos de piensos animales, agentes químicos especializados, materias primas químicas, plásticos, disolventes, combustibles, u otros polímeros orgánicos, ácido láctico, y etanol, incluyendo etanol combustible (entendiéndose que el término "etanol" incluye alcohol etílico o mezclas de alcohol etílico y agua).

Productos específicos de valor añadido que pueden producirse por los métodos de la invención incluyen, aunque sin limitación, biocombustibles (incluyendo etanol y butanol y un biogás); ácido láctico; un plástico; un agente químico especializado; un ácido orgánico, incluyendo ácido cítrico, ácido succínico, ácido fumárico, ácido itacónico y ácido maleico; ácido 3-hidroxi-propiónico, ácido acrílico; ácido acético; 1,3-propano-diol; etileno, glicerol; un disolvente; un suplemento de pienso animal; una gente farmacéutico, tal como un antibiótico de β -lactama o una cefalosporina; vitaminas; un aminoácido, tal como lisina, metionina, triptófano, treonina, y ácido aspártico; una enzima industrial, tal como una proteasa, una celulasa, una amilasa, una glucanasa, una lactasa, una lipasa, una liasa, una oxidorreductasa, una transferasa o una xilanasas; y una materia prima química.

Biogás

La invención también proporciona el uso de un polipéptido o composición como se describe en este documento en un método para la preparación de biogás. Biogás típicamente se refiere a un gas producido por la descomposición biológica de materia orgánica, por ejemplo, material que contiene carbohidrato que no es almidón, en ausencia de oxígeno. El biogás se origina a partir de material biogénico y es un tipo de biocombustible. Un tipo de biogás se produce por la digestión anaeróbica o fermentación de materiales biodegradables tales como biomasa, estiércol o aguas negras, residuos municipales, y cultivos energéticos. Este tipo de biogás está compuesto principalmente de metano y dióxido de carbono. El gas metano puede quemarse u oxidarse con oxígeno. El aire contiene un 21 % de oxígeno. Esta liberación de energía permite el uso del biogás como combustible. El biogás puede usarse como combustible de bajo coste en cualquier país para cualquier propósito de calentamiento, tal como cocinar. También puede utilizarse en instalaciones modernas de tratamiento de residuos donde puede usarse para hacer funcionar cualquier tipo de motor térmico, para generar energía mecánica o eléctrica.

La primera etapa en la producción de biogás microbiano consiste en la degradación enzimática de polímeros y sustratos complejos (por ejemplo, carbohidrato que no es almidón). Por consiguiente, la invención proporciona un método para la preparación de un biogás, en que un sustrato que comprende carbohidrato que no es almidón se pone en contacto con un polipéptido o composición de la invención, para producir de ese modo material fermentable que puede convertirse en un biogás por un organismo, tal como un microorganismo. En dicho método, puede proporcionarse un polipéptido de la invención mediante un organismo, por ejemplo, un microorganismo que expresa dicho polipéptido.

Uso de enzimas en productos alimenticios

Los polipéptidos y composiciones de la invención pueden usarse en un método de procesamiento de material vegetal para degradar o modificar los constituyentes de celulosa o hemicelulosa o sustancia pécica de las paredes celulares del material vegetal o fúngico. Dichos métodos pueden ser útiles en la preparación de productos alimenticios. Por consiguiente, la invención proporciona un método para preparar un producto alimenticio, comprendiendo dicho método incorporar un polipéptido o composición de la invención durante la preparación del producto alimenticio.

La invención también proporciona un método de procesamiento de un material vegetal, comprendiendo dicho método poner en contacto el material vegetal con un polipéptido o composición de la invención para degradar o modificar la celulosa en el material (vegetal). Preferiblemente, el material vegetal es una pulpa vegetal o extracto vegetal, tal como zumos.

Los materiales vegetales y que contienen celulosa/hemicelulosa/sustancia péctica incluyen pulpa vegetal, partes de plantas y extractos vegetales. En el contexto de esta invención, un extracto de un material vegetal es cualquier sustancia que puede obtenerse de material vegetal por extracción (mecánica y/o química), procesamiento o por otras técnicas de separación. El extracto puede ser zumo, néctar, base, o concentrados hechos de los mismos. El material vegetal puede comprender u obtenerse de hortalizas, por ejemplo, zanahorias, apio, cebollas, legumbres o plantas leguminosas (soja, judías de soja, guisantes) o fruta, por ejemplo, fruta de hueso o de pepitas (manzanas, peras, membrillo, etc.), uvas, tomates, cítricos (naranja, limón, lima, mandarina), melones, ciruelas, cerezas, grosellas negras, grosellas rojas, frambuesas, fresas, arándanos, piña y otras frutas tropicales, árboles y partes de los mismos (por ejemplo, polen, de pinos), o cereales (avena, cebada, trigo, maíz, arroz). El material (a hidrolizarse) puede ser también residuos agrícolas, tales como pulpa de remolacha azucarera, mazorcas de maíz, paja de trigo, cáscaras de nuez (molidas), o materiales reciclables, por ejemplo, papel (residual).

Los polipéptidos de la invención, por tanto, pueden usarse para tratar material vegetal, incluyendo pulpa vegetal y extractos vegetales. También pueden usarse para tratar productos alimenticios líquidos o sólidos o ingredientes comestibles de productos alimenticios, o pueden usarse en la extracción de aceites vegetales, almidón o como espesante en alimentos.

Típicamente, los polipéptidos de la invención se usan como una composición/preparación de enzimas como se ha descrito anteriormente. La composición generalmente se añade a pulpa vegetal que se puede obtener por, por ejemplo, procesamiento mecánico, tal como trituración o molienda de material vegetal. La incubación de la composición con la planta típicamente se realizará durante un tiempo de 10 minutos a 5 horas, tal como de 30 minutos a 2 horas, preferiblemente durante aproximadamente 1 hora. La temperatura de procesamiento es preferiblemente de aproximadamente 10 °C a aproximadamente 55 °C, por ejemplo, de aproximadamente 15 °C a aproximadamente 25 °C, óptimamente a aproximadamente 20 °C y se pueden usarse de aproximadamente 10 g a aproximadamente 300 g, preferiblemente de aproximadamente 30 g a aproximadamente 70 g, óptimamente aproximadamente 50 g de enzima por tonelada de material a tratarse.

Toda la enzima o enzimas o sus composiciones usadas puede añadirse secuencialmente o al mismo tiempo a la pulpa vegetal. Dependiendo de la composición de la preparación de enzimas, el material vegetal primero puede macerarse (por ejemplo, hasta un puro) o licuarse. Usando los polipéptidos de la invención pueden mejorarse parámetros de procesamiento tales como el rendimiento de la extracción, la viscosidad del extracto y/o la calidad del extracto.

Como alternativa, o además de lo anterior, puede añadirse un polipéptido de la invención al zumo sin procesar obtenido del prensado o licuado de la pulpa vegetal. El tratamiento del zumo sin procesar se realizará de una manera similar a la pulpa vegetal respecto a la dosificación, temperatura y tiempo de espera. De nuevo, pueden incluirse otras enzimas tales como las analizadas previamente. Las condiciones típicas de incubación son como las descritas en el párrafo previo.

Una vez se ha incubado el zumo sin procesar con los polipéptidos de la invención, entonces se centrifuga el zumo o (ultra) filtra para producir el producto final.

Después del tratamiento con el polipéptido de la invención, el producto (final) puede tratarse con calor, por ejemplo, a aproximadamente 100 °C durante un tiempo de aproximadamente 1 minuto a aproximadamente 1 hora, en condiciones de inactivación parcial o completa del polipéptido o polipéptidos de la invención.

Una composición que contiene un polipéptido de la invención también puede usarse durante la preparación de purés de frutas u hortalizas.

En repostería el polipéptido puede mejorar la estructura de la masa, modificar su pegajosidad o flexibilidad, mejorar el volumen de la barra y/o estructura de la miga o conferir mejores características texturales tales como el partido, el desmenuzado o la calidad de la miga.

La presente invención, por tanto, se refiere a métodos para preparar una masa o un producto alimenticio basado en cereal, que comprende incorporar en la masa un polipéptido o composición de la presente invención. Esto puede mejorar una o más propiedades de la masa o el producto alimenticio basado en cereal obtenidas de la masa respecto a una masa o un producto alimenticio basado en cereal en que no está incorporado el polipéptido.

La preparación del producto alimenticio basado en cereal de acuerdo con la invención puede comprender adicionalmente etapas conocidas en la técnica tales como hervido, secado, fritura, cocción al vapor u horneado de la masa obtenida.

Los productos que se preparan a partir de una masa que se hierve son, por ejemplo, fideos hervidos, budines de carne, los productos que se preparan a partir de masa frita son, por ejemplo, rosquillas, buñuelos, fideos fritos, los productos que se preparan a partir de masa cocida al vapor son, por ejemplo, panecillos al vapor y fideos al vapor, ejemplos de productos preparados a partir de masa seca son pasta y fideos secos y ejemplos de productos

preparados a partir de masa horneada son pan, galletas, tarta.

La expresión "propiedad mejorada" se define en este documento como cualquier propiedad de una masa y/o un producto obtenido a partir de la masa, particularmente un producto alimenticio basado en cereal, que se mejora por la acción del polipéptido de acuerdo con la invención respecto a una masa o producto en que no está incorporado el polipéptido de acuerdo con la invención. La propiedad mejorada puede incluir, aunque sin limitación, resistencia aumentada de la masa, elasticidad aumentada de la masa, estabilidad aumentada de la masa, maquinabilidad mejorada de la masa, resistencia mejorada al levado de la masa, pegajosidad reducida de la masa, extensibilidad mejorada de la masa, volumen aumentado del producto alimenticio basado en cereal, ampollamiento reducido del producto alimenticio basado en cereal, estructura mejorada de la miga del producto horneado, suavidad mejorada del producto alimenticio basado en cereal, aroma mejorado del producto alimenticio basado en cereal, anti-enranciado mejorado del producto alimenticio basado en cereal. Propiedades mejoradas relacionadas con productos basados en cereal de tipo pasta y fideos son, por ejemplo, firmeza mejorada, pegajosidad reducida, cohesión mejorada y pérdida reducida en la cocción.

La propiedad mejorada puede determinarse por comparación de una masa y/o un producto alimenticio basado en cereal preparado con y sin adición de un polipéptido de la presente invención. Las cualidades organolépticas pueden evaluarse usando procedimientos bien establecidos en la industria panadera, y pueden incluir, por ejemplo, el uso de un panel de evaluadores cualificados del sabor.

El término "masa" se define en este documento como una mezcla de harina de cereal y otros ingredientes, suficientemente firme para amasarla o enrollarla. Ejemplos de cereales son trigo, centeno, abatí, maíz, cebada, arroz, avena mondada, trigo sarraceno y avena. Trigo aquí y a partir de ahora en este documento pretende abarcar todas las especies conocidas del género *Triticum*, por ejemplo, *aestivum*, *durum* y/o *spelta*. Ejemplos de otros ingredientes adecuados son: el polipéptido de beta-glucosidasa de acuerdo con la presente invención, enzimas adicionales, aditivos químicos y/o auxiliares de procesamiento. La masa puede fresca, congelada, pre-cortada, o pre-horneada. La preparación de una masa a partir de los ingredientes descritos anteriormente es bien conocida en la técnica y comprende mezcla dichos ingredientes y auxiliares de procesamiento y una o más etapas de moldeo y opcionalmente fermentación. La preparación de masa congelada se describe por Kulp y Lorenz en *Frozen and Refrigerated Doughs and Batters*.

La expresión "producto alimenticio basado en cereal" se define en este documento como cualquier producto preparado a partir de una masa, de carácter blando o crujiente. Ejemplos de productos alimenticios basados en cereal, sean de tipo blanco, ligero u oscuro, que pueden producirse ventajosamente por la presente invención son pan (en particular pan blanco, integral o de centeno), típicamente en forma de barras o panecillos, pan tipo baguette francesa, pasta, fideos, rosquillas, roscos, tarta, pan pita, tortillas, tacos, tortas, panqueque, panecillos, galletas, costras de pastel, pan al vapor, y pan crujiente, y similares.

El término "producto horneado" se define en este documento como cualquier producto alimenticio basado en cereal preparado por horneado de la masa.

Los polisacáridos que no son almidón (NSP) pueden aumentar la viscosidad del alimento digerido que puede, a su vez, disminuir la disponibilidad de nutrientes y el desempeño animal. El uso del polipéptido de beta-glucosidasa de la presente invención puede mejorar la utilización de fósforo, así como los minerales catiónicos y polipéptidos durante la digestión del animal.

Añadir nutrientes específicos al pienso mejora la digestión del animal y de ese modo reduce los costes de alimentación. Actualmente están en uso muchos aditivos de piensos y se están desarrollando continuamente nuevos conceptos. El uso de enzimas específicas como enzimas degradantes de carbohidrato que no es almidón podría descomponer la fibra liberando energía, así como aumentando la digestibilidad de las proteínas debido a una mejor accesibilidad de la proteína cuando la fibra queda descompuesta. De este modo, el coste de alimentación podría descenderse, así como los niveles de proteínas en el pienso también podrían reducirse.

Los polisacáridos que no son almidón (NSP) también están presentes en casi todos los ingredientes de piensos de origen vegetal. Los NSP se utilizan mal y pueden, cuando se solubilizan, ejercer efectos adversos sobre la digestión. Enzimas exógenas pueden contribuir a una mejor utilización de estos NSP y como consecuencia pueden reducir cualquier efecto anti-nutricional. Pueden usarse enzimas degradantes de carbohidrato que no es almidón de la presente invención para este propósito en dietas basadas en cereales para aves de corral y, a un mejor grado, para cerdos y otras especies.

Puede usarse un polipéptido/enzima degradante de carbohidrato que no es almidón de la invención (o una composición que comprende el polipéptido/enzima de la invención) en la industria de detergentes, por ejemplo, para la eliminación en lavandería de manchas basadas en carbohidrato. Una composición detergente puede comprender un polipéptido/enzima de la invención y, además, una o más de una celulosa, una hemicelulasa, una pectinasa, una proteasa, una lipasa, una cutinasa, una amilasa o una carbohidrasa.

Uso de enzimas en composiciones detergentes

Una composición detergente que comprende un polipéptido o composición de la invención puede estar en cualquier forma conveniente, por ejemplo, una pasta, un gel, un polvo o un líquido. Un detergente líquido puede ser acuoso, 5 conteniendo típicamente hasta aproximadamente un 70 % de agua y de aproximadamente un 0 a aproximadamente un 30 % de disolvente orgánico o material no acuoso.

Dicha composición detergente puede formularse, por ejemplo, como una composición detergente de lavado a mano o a máquina incluyendo una composición aditiva de lavado adecuada para el pre-tratamiento de telas manchadas y 10 una composición suavizante de telas añadida en aclarado, o puede formularse como una composición detergente para su uso en tareas generales de limpieza de superficies duras domésticas, o puede formularse para tareas de lavado de platos a mano o a máquina.

En general, las propiedades de la enzima deben ser compatibles con el detergente seleccionado (por ejemplo, pH 15 óptimo, compatibilidad con otros ingredientes enzimáticos y/o no enzimáticos, etc.) y la enzima o enzimas deben estar presentes en una cantidad eficaz.

Una composición detergente puede comprender un tensioactivo, por ejemplo, un tensioactivo aniónico o no iónico, un formador de detergente o agente de formación de complejos, uno o más polímeros, un sistema de blanqueado 20 (por ejemplo, una fuente de H₂O₂) o un estabilizador enzimático. Una composición detergente también puede comprender otro ingrediente detergente convencional tal como, por ejemplo, un suavizante incluyendo una arcilla, un realzador de espuma, un supresor de agua de lavado, una gente anticorrosión, un agente que suspende la suciedad, un agente que redeposita la suciedad, un colorante, un bactericida, un abrillantador óptico, un hidrótrofo, un inhibidor de deslustre o un perfume.

Uso de enzimas en procesamiento de papel y pulpa

Un polipéptido o composición de la presente invención puede usarse en la industria del papel y la pulpa, entre otros, en el proceso de blanqueado para potenciar el brillo de las pulpas blanqueadas, mediante lo cual puede reducirse la 30 cantidad de cloro usada en las etapas de blanqueado, y para aumentar la liberación de pulpas en el proceso de papel reciclado (Eriksson, K. E. L., Wood Science and Technology 24 (1990):79-101; Paice, et al., Biotechnol. and Bioeng. 32 (1988):235-239 y Pommier et al., Tappi Journal (1989):187-191). Además, un polipéptido o composición de la invención puede usarse para el tratamiento de pulpa lignocelulósica para mejorar la capacidad de blanqueado de la misma. De ese modo puede reducirse la cantidad de cloro necesaria para obtener un blanqueado satisfactorio 35 de la pulpa.

Un polipéptido o composición de la invención puede usarse en un método para reducir la tasa a la cual telas que contienen celulosa quedan ásperas o para reducir la aspereza de telas que contienen celulosa, comprendiendo el 40 método tratar telas que contienen celulosa con un polipéptido o composición como se ha descrito anteriormente. La presente invención se refiere adicionalmente a un método que proporciona aclarado del color de telas coloreadas que contienen celulosa, comprendiendo el método tratar telas coloreadas que contienen celulosa con un polipéptido o composición como se ha descrito anteriormente, y un método para proporcionar una variación localizada en el color de telas coloreadas que contienen celulosa, comprendiendo el método tratar telas coloreadas que contienen 45 celulosa con un polipéptido o composición como se ha descrito anteriormente. Los métodos de la invención pueden realizarse tratando telas que contienen celulosa durante el lavado. Sin embargo, si se desea, el tratamiento de las telas también puede realizarse durante el remojo o aclarado o simplemente añadiendo el polipéptido o composición como se ha descrito anteriormente al agua en que las telas están sumergidas o se sumergirán.

Otros usos de las enzimas

Además, un polipéptido o composición de la presente invención también puede usarse en una formulación antibacteriana, así como en productos farmacéuticos tales como pastillas para la garganta, dentífricos, y enjuagues bucales.

Los siguientes Ejemplos ilustran la invención:

Ejemplos**Materiales y métodos*****Procedimientos de ADN***

Se realizaron procedimientos convencionales de ADN como se describe en otra parte (Sambrook et al., 1989, Molecular cloning: a laboratory manual, 2ª Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York) salvo que se indique de otro modo. Se amplificó el ADN usando la enzima con corrección de errores 65 polimerasa Phusion (Finnzymes). Las enzimas de restricción fueron de Invitrogen o New England Biolabs.

Preparación de muestras de celulasa

Se expresaron celulasas originarias de *Talaromyces emersonii* en *Aspergillus niger*. Se produjeron filtrados concentrados de las enzimas como se describe en el documento WO2004/030468. Después de cultivar *Aspergillus niger* que contenía los plásmidos de expresión apropiados, se prepararon sobrenadantes sin células por centrifugación del caldo de fermentación a 5000 x g durante 30 minutos a 4 °C. Opcionalmente el sobrenadante puede ajustarse a pH = 5 con KOH 4 N y filtrarse a esterilidad sobre un filtro de 2 µm (tapa de botella) con succión para retirar cualquier material fúngico. Además, los sobrenadantes pueden filtrarse adicionalmente sobre un filtro de microfibras de vidrio GF/A Whatmann (150 mm Ø) para retirar cualquier sólido. Los sobrenadantes se ultrafiltraron, concentraron, y almacenaron hasta su uso a 4 °C o congelados a -20 °C.

Método para la determinación de proteína total

El método fue una combinación de precipitación de proteína usando ácido tricloroacético (TCA) para retirar sustancias alterantes y permitir la determinación de la concentración de proteína con la reacción colorimétrica de Biuret. En la reacción de Biuret, se reduce un ión de cobre (II) en cobre (I), que forma un complejo con los nitrógenos y carbonos de los enlaces peptídicos en una solución alcalina. Un color violeta indica la presencia de proteínas. La intensidad del color, y por tanto la absorción a 546 nm, es directamente proporcional a la concentración de proteína, de acuerdo con la ley de Beer-Lambert. La normalización se realizó usando BSA (albúmina sérica bovina) y el contenido de proteína se expresó en g de proteína como equivalente de BSA/l o mg de proteína como equivalente de BSA/ml. El contenido de proteína se calculó usando protocolos convencionales de cálculo conocidos en la técnica, representando la DO_{546} frente a la concentración de muestras con concentración conocida, seguido del cálculo de la concentración de las muestras desconocidas usando la ecuación generada a partir de la línea de calibración.

Métodos para la determinación de actividad beta-glucosidasa

Se determinó la actividad beta-glucosidasa usando para-nitrofenil-beta-glucósido a concentración 1,5 mM en tampón acetato sódico 50 mM a pH 4,5. Se realizó incubación a 40 °C durante 30 min. Se detuvo la reacción usando bicarbonato sódico 1 M. La actividad se calculó a partir de la extinción determinada a 405 nm, y el coeficiente de extinción molar de para-nitrofenol en condiciones alcalinas, como saben los expertos en la materia. La actividad se expresa como µmol de pNP liberados por ml por minuto.

Un método alternativo para la determinación de la actividad beta-glucosidasa es la incubación de la enzima con celobiosa 10 mM en tampón acetato sódico 50 mM. Después de 30 minutos a 40 °C la reacción se detiene con hidróxido sódico. Se ultrafiltraron y analizaron las muestras usando cromatografía de intercambio aniónico de alto rendimiento con detección amperométrica por pulsos (HPAEC-PAD). Se aplicó una columna CarboPac PA-20 termo-regulada a 30 °C, y NaOH 200 mM a un caudal de 0,5 ml/min para separar la glucosa y celobiosa. La actividad beta-glucosidasa se expresó en µmol de glucosa liberados por ml por minuto.

Preparación de sustrato de paja de trigo pretratado lavado

Puede obtenerse paja de trigo pretratada con ácido diluido como se describe en Linde, M. et al, Biomass and Bioenergy 32 (2008), 326-332 y puede usarse el equipo descrito en Schell, D.J., Applied Biochemistry and Biotechnology (2003), vol. 105-108, pág. 69-85. La paja de trigo pretratada se lavó con agua hasta que la solución con paja de trigo estuvo a pH 6,0 o superior. El agua de lavado se retiró por filtración.

Hidrólisis prolongada de celulosa

Se realizaron incubaciones de combinaciones de enzimas en solución de sustrato de materia seca (dm) de paja de trigo pretratada (PWS) con ácido lavada al 2 % en tampón acetato sódico 50 mM pH 4,5, a escala de 10 ml. Se añadieron combinaciones de enzimas a dosis fija de proteína por gramo de materia seca de sustrato. Se recogieron muestras en el tiempo, hasta 72 horas de incubación a 65 °C. Las reacciones se terminaron en el momento dado, por centrifugación del residuo, pipeteo del sobrenadante y congelación de las muestras hasta el análisis.

Se realizó análisis de la cantidad de glucosa liberada usando RMN de flujo. Se registraron los espectros de RMN de 1H en un sistema de RMN Bruker AVANCE II BEST que funcionaba a frecuencia de protones de 500 MHz y temperatura de sonda de 27 °C.

Ejemplo 11.1. Construcción de plásmidos de expresión

Se clonó la secuencia que tiene la SEQ ID NO: 1 en el vector pGBTOP (Fig. 1) usando sitios EcoRI y SnaBI, que comprenden la secuencia promotora y terminado de la glucoamilasa. La parte *E. coli* se retiró por digestión con NotI

antes de la transformación de *A. niger* CBS 513.88.

1.2. Transformación de *A. niger*

- 5 *A. niger* WT-1: Esta cepa de *A. niger* es CBS513.88 que comprende delecciones de los genes que codifican la glucoamilasa (*glaA*), amilasa fúngica y amilasa ácida. *A. niger* WT 1 se construye usando el enfoque "LIBRE DE GEN MARCADOR" como se describe en el documento EP 0 635 574 B1.

- 10 Las construcciones de expresión se introdujeron por co-transformación en la cepa *A. niger* WT-1 de acuerdo con el método descrito por Tilburn, J. et al. (1983) Gene 26, 205-221 y Kelly, J. y Hynes, M. (1985) EMBO J., 4, 475-479 con las siguientes modificaciones:

- 15 - Se germinan esporas y se cultivan durante 16 horas a 30 grados Celsius en un frasco de agitación colocado en un agitador rotario a 300 rpm en medio mínimo de *Aspergillus* (100 ml). El medio mínimo de *Aspergillus* contiene por litro: 6 g de NaNO₃, 0,52 g de KCl, 1,52 g de KH₂PO₄, 1,12 ml de KOH 4 M, 0,52 g de MgSO₄·7H₂O, 10 g de glucosa, 1 g de casaminoácidos, 22 mg de ZnSO₄·7H₂O, 11 mg de H₃BO₃, 5 mg de FeSO₄·7H₂O, 1,7 mg de CoCl₂·6H₂O, 1,6 mg de CuSO₄·5H₂O, 5 mg de MnCl₂·2H₂O, 1,5 mg de Na₂MoO₄·2H₂O, 50 mg de EDTA, 2 mg de riboflavina, 2 mg de tiamina-HCl, 2 mg de nicotinamida, 1 mg de piridoxina-HCl, 0,2 mg de ácido pantoténico, 4 g de biotina, 10 ml de solución de penicilina (5000 UI/ml) y estreptomycin (5000 UG/ml) (Gibco).

- 20 - Se usa Novozym 234[™] (Novo Industries) en lugar de helicasa para la preparación de protoplastos;
 - Después de la formación de protoplastos (60-90 minutos), se añade tampón KC (KCl 0,8 M, ácido cítrico 9,5 mM, pH 6,2) a un volumen final de 45 ml, se centrifuga la suspensión de protoplastos durante 10 minutos a 3000 rpm a 4 grados Celsius en un rotor de cubo oscilatorio. Los protoplastos se resuspenden en 20 ml de tampón KC y posteriormente se añaden 25 ml de tampón STC (sorbitol 1,2 M, Tris-HCl 10 mM pH 7,5, CaCl₂ 50 mM). La
 25 suspensión de protoplastos se centrifuga durante 10 minutos a 3000 rpm a 4 grados Celsius en un rotor de cubo oscilatorio, se lava en tampón en STC y se resuspende en tampón STC a una concentración de 10E8 protoplastos/ml;

- A 200 microlitros de la suspensión de protoplastos, se añade el fragmento de ADN, disuelto en 10 microlitros de tampón TE (Tris-HCl 10 mM pH 7,5, EDTA 0,1 mM) y 100 microlitros de solución de PEG (PEG 4000 (Merck) al 20
 30 %, sorbitol 0,8 M, Tris-HCl 10 mM pH 7,5, CaCl₂ 50 mM);

- Después de incubación de la suspensión de ADN-protoplastos durante 10 minutos a temperatura ambiente, se añaden lentamente 1,5 ml de solución de PEG (PEG 4000 (Merck) al 60 %, Tris-HCl 10 mM pH 7,5, CaCl₂ 50 mM), con mezcla repetida de los tubos. Después de incubación durante 20 minutos a temperatura ambiente, se diluyen las suspensiones con 5 ml de sorbitol 1,2 M, se mezclan por inversión y se centrifugan durante 10 minutos a 4000 rpm a
 35 temperatura ambiente. Los protoplastos se resuspenden suavemente en 1 ml de sorbitol 1,2 M y se siembran en medio de regeneración selectivo sólido que consiste en medio mínimo de *Aspergillus* sin riboflavina, tiamina-HCl, nicotinamida, piridoxina, ácido pantoténico, biotina, casaminoácidos y glucosa. En caso de selección con acetamida, el medio contiene acetamida 10 mM como única fuente de nitrógeno y sacarosa 1 M como agente osmótico y fuente de C. Como alternativa, los protoplastos se siembran en PDA (agar de dextrosa de patata, Oxoid) suplementado con
 40 1-50 microgramos/ml de fleomicina y sacarosa 1 M como agente osmótico. Las placas de regeneración se solidifican usando agar al 2 % (agar n.º 1, Oxoid L11). Después de incubación durante 6-10 días a 30 grados Celsius, se transfieren las conidiosporas de los transformantes a placas que consisten en medio selectivo de *Aspergillus* (medio mínimo que contiene acetamida como única fuente de nitrógeno en el caso de selección con acetamida o PDA suplementado con 1-50 microgramos/ml de fleomicina en el caso de selección con fleomicina) con glucosa al 2 % y
 45 agarosa (Invitrogen) al 1,5 % y se incuban durante 5-10 días a 30 grados Celsius. Se aíslan transformantes individuales y esta etapa de purificación selectiva se repite una vez tras lo cual se almacenan los transformantes purificados.

- Después de la transformación, se seleccionaron los transformantes en medios que comprenden acetamida como
 50 única fuente de nitrógeno y se purificaron las colonias. Se estimaron las cantidades de copias por PCR cuantitativa y se seleccionaron transformantes con baja y alta cantidad de copias. Los transformantes con alta cantidad de copias se cultivaron en matraces de agitación en 100 ml de medio CSM-MES como se describe en el documento EP 635 574 a 34 °C a 170 rpm en un agitador de incubación usando un matraz de agitación de 500 ml con deflectores. Después de 3 y 4 días de fermentación, se recogieron muestras de sobrenadante para determinar la expresión por
 55 SDS-PAGE.

1.3 Contenido de proteína

- 60 Los sobrenadantes ultrafiltrados y concentrados de las fermentaciones en matraz de agitación de los transformantes que expresaban beta-glucosidasa (TEMER06387) se analizaron para el contenido de proteína. El contenido de proteína se determinó en 4 mg de proteína como equivalente de BSA/ml.

Ejemplo 2

- 65 2.1 Preparación de muestras de celulasa

Una proteína potenciadora de celulasa, TEMER07589 (como se describe en la solicitud de patente en trámite junto con la presente DSM Caso 27666, presentada el mismo día que esta solicitud), y dos exoglucanasas, que son CBHI como se describe en la solicitud de patente EP09158739.4 y CBHII (como se describe en la solicitud de patente en trámite junto con la presente DSM Caso 27829, presentada el mismo que esta solicitud), todas de *Talaromyces emersonii*, se prepararon como se ha descrito en el Ejemplo 1 para TEMER06387. Se determinaron los contenidos de proteína de estas muestras y variaban de 20 a 60 mg de proteína como equivalente de BSA/ml. Además, también se preparó uno de los homólogos más cercanos de la beta-glucosidasa TEMER06387, que es una beta-glucosidasa de *Sclerotinia sclerotiorum* (XP_001591700; NCBI) como se ha descrito anteriormente. Esta beta-glucosidasa se indicará como BG-SS.

2.2 Hidrólisis prolongada de mezclas de 4 enzimas

La beta-glucosidasa TEMER06387 se mezcló con las otras 3 proteínas celulolíticas, en cantidades relativas del 9 % de TEMER06387, el 37 % de TEMER07589, el 30 % de CBHI y el 24 % de CBHII, de la proteína total en la mezcla. También se preparó una mezcla similar con la BG-SS, en que se reemplazó TEMER06387 por BG-SS. Estas mezclas se aplicaron en hidrólisis prolongada a escala de 10 ml con un 2 % de paja de trigo pretratada con DM, a pH 4,5 y 65 °C. La dosis de proteína total en cada una de las incubaciones fue de 15 mg de proteína como equivalentes de BSA por gramo de materia seca de paja de trigo pretratada. Las incubaciones duraron 72 horas y se tomaron muestras a varios intervalos temporales. Se centrifugaron las muestras, y se congeló el sobrenadante hasta el análisis por RMN. Se muestra la liberación de glucosa y celobiosa en el tiempo en la Tabla 1.

Tabla 1. Liberación de glucosa y celobiosa, expresada como mmol/l, durante incubación de paja de trigo pretratada lavada con DM al 2 %, a pH 4,5 y 65 °C, por una mezcla de 4 enzimas que contiene 1 BG (TEMER06387 o BG-SS), 1 CBHI, 1 CBHII y 1 proteína potenciadora de celulasa, en cantidades relativas del 9 %, 30 %, 24 % y 37 % de proteína total de la mezcla.

	Glc (mmol/l)				Celobiosa (mmol/l)			
	3 h	21 h	45 h	72 h	3 h	21 h	45 h	72 h
Mezcla TEMER06387	13,5	26,5	32,7	37,2	0,1	0,7	0,9	0,6

A partir de este experimento, está claro que TEMER06387 es muy capaz de hidrolizar celulosa en glucosa en una mezcla con otras 3 proteínas celulolíticas.

LISTA DE SECUENCIAS

<110> DSM IP Assets B.V.

<120> Polipéptido que tiene actividad beta-glucosidasa y usos del mismo

<130> 27667-WO-PCT

<160> 3

<170> PatentIn versión 3.5

<210> 1

<211> 2541

<212> ADN

<213> *T. emersonii*

<400> 1

atgcagttcc tctgtcctct cctgctgttt ctctcactct gcttaggccca acaaaatgaa	60
ggatctgtat acaccccccg gataagctca ggctatggca attggtctga agcatacaga	120
aaagcccgcg aattcgtctg tcagttaacg ttgaccgaaa aagtcaacct cacaacgcaa	180
gtcggcacgg gccagaccca aggctacgac attggcatca tccccgggt gaactttcga	240
ggatttgagg ctgatgactc gccctcggc gtgcgaggaa ccgactatc ctccgcctt	300
goggggggac ttaacctggc catgtcctgg gatagagagc taatotatgc tcaatcatac	360
gccaacggag cggagcacaa gggcaaaggc gtcaatatcg tctatgccc cgtggccggt	420
ccgctggggc gagcgccaac cggggggcgc attttcgaga gctactctcc agatccgtat	480
ctctcagga tcgccttcgg gacatcgatt gcaggaatgc aggtgccgg agtgatcgcc	540
gtcgggaaac acttcatcct ctatgagcag gagcatttcc gccaagtcac cgaatggaac	600
tctacggggc tgggcccga tatcacggag ccgtactcgt cgaactgtga tgaccggact	660
ttgcacgagc tctatctgtt cccgtggtac gacgctgtat acagcggcat ggccggggtg	720
atgtgttctt acaaccagat caacaatacc caggcgtgcc agaacaatca catcctgaat	780
gacatcctca agggcgaact cggcttcag gggtttgtga ccagcgacga cgggtcccag	840
cacgacggag tcctcaccgc gctggcagga atggacatga ctattcctgg agagacctct	900
ccgaacggta tcgccgggct agggatctca ttctgggggc ccaatctcac catggccgtg	960
ctcaatggca gcgtgccga gtggcggtt gacgacatgg ccaccaggat tatggctgct	1020
tactactacc tcggccagga ccaggatttc ccaccactga gctttgcagc agcagattac	1080
tctacgtacg gatacgagta tccgcaagcg caggctgact acacgcagat caactggcac	1140
gtcgatgtgc gtcagaatca ctctcagatc atccgtgcgg tgggcgcaa ctogatcgtg	1200
ctgctgaaga ataacaacgg cacgcttctc ctcaggaaac ccagacagat tgccgtcatc	1260
ggcgaggatg ctggtccttc gccctacggt cccaacggct gcactgatcg gggctgcgtt	1320
aatggcacgc ttgcgatggg ctgggggagt ggtagcacia atttccgta tctggtcgac	1380
cccttgtctg ccatccagca gggggcgtt aaagacggta ccgaagtga atatgtcttg	1440

```

aataatttcg ccactgcgca gatcgactct gtagcgctgc aagccaccgt gtgtctcgtg 1500
ttcatcaatg ccgatagcgg agagggatat atcgaggtgg gcggcaacta cggcgacagg 1560
aacaatctca cggcatggct gggcggagat gctctcgtcc aacgggttgc tgcgaactgc 1620
agtcagacgg cggcatcatc ccattgcgca gggccattc tgatggagga ctggatcgac 1680
caccgcaatg tgacgggggt gctctttgcc gggctcccgg gacaggaaag cgggaatttg 1740
ctggtcgacg tgctgtacgg cgacgtcaac ccattctgga aacttcctg gaccgtgggc 1800
aaggctcgcg aggactacgg gactgatgtg ttgtatgaac ccaaagggcc agtgccgcag 1860
atcgacttca cagagggcct gttcatcgac tatcgctggt tcgacaagca gggcatcgat 1920
cctcgctttg aattcggctt tgggctgagc tacacgacgt tcgactatca gaacctcagc 1980
atcgagacct taaaagcctc tccagcgaca gcaacaacca catcgccagt agcttcaacc 2040
accgtcaatc cttctaggtg cccgaccacc agcctcaatc cgtccgacta caogtttcca 2100
tcgacgatca gctcgtgac gaattacatc taccgtaca tcgctccaa tgcgacaatc 2160
tctaccgtg ccgccggtac accgacgccc actatcatcc cggaggcagc aggtcctcag 2220
ggaggaaacc cggcgttgta cgatgtgctg taccgggtga cattgacgct gaagaacaac 2280
ggaaccgttg cggggcaaga ggtggcccag ctgtatctgg acttggacaa tggggagccc 2340
agtcgccagc ttccgggggt caacaagacg ttctcgagg ccggggagac caaggacgtc 2400
acgtttgatc tgcggaggag agatgtggcg atttgggatg ttgtgaagca ggagtgggtc 2460
gaggtcaaga aactggggac aaagattggg gtgtatgttg gggccagctc gagggatttg 2520
cgtctgacgg gggagatctg a 2541

```

<210> 2
 <211> 846
 <212> PRT
 <213> *T. emersonii*

<400> 2

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

Gly Phe Ala Ala Asp Asp Ser Pro Ser Gly Val Arg Gly Thr Asp Tyr
85 90 95

Ser Ser Ala Phe Ala Ala Gly Leu Asn Leu Ala Met Ser Trp Asp Arg
100 105 110

Glu Leu Ile Tyr Ala Gln Ser Tyr Ala Asn Gly Ala Glu His Lys Gly
115 120 125

Lys Gly Val Asn Ile Val Tyr Ala Pro Val Ala Gly Pro Leu Gly Arg
130 135 140

Ala Pro Thr Gly Gly Arg Ile Phe Glu Ser Tyr Ser Pro Asp Pro Tyr
145 150 155 160

Leu Ser Gly Ile Ala Phe Gly Thr Ser Ile Ala Gly Met Gln Ala Ala
165 170 175

Gly Val Ile Ala Val Gly Lys His Phe Ile Leu Tyr Glu Gln Glu His
180 185 190

Phe Arg Gln Val Ile Glu Trp Asn Ser Tyr Gly Leu Gly Pro Asn Ile
195 200 205

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

Tyr Leu Phe Pro Trp Tyr Asp Ala Val Tyr Ser Gly Met Ala Gly Val
225 230 235 240

Met Cys Ser Tyr Asn Gln Ile Asn Asn Thr Gln Ala Cys Gln Asn Asn
245 250 255

His Ile Leu Asn Asp Ile Leu Lys Gly Glu Leu Gly Phe Gln Gly Phe
260 265 270

Val Thr Ser Asp Asp Gly Ser Gln His Asp Gly Val Leu Thr Ala Leu
275 280 285

Ala Gly Met Asp Met Thr Ile Pro Gly Glu Thr Ser Pro Asn Gly Ile
290 295 300

Ala Gly Leu Gly Ile Ser Phe Trp Gly Pro Asn Leu Thr Met Ala Val
305 310 315 320

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

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

Gly Lys Leu Pro Trp Thr Val Gly Lys Ala Arg Glu Asp Tyr Gly Thr
595 600 605

Asp Val Leu Tyr Glu Pro Lys Gly Pro Val Pro Gln Ile Asp Phe Thr
610 615 620

Glu Gly Leu Phe Ile Asp Tyr Arg Trp Phe Asp Lys Gln Gly Ile Asp
625 630 635 640

Pro Arg Phe Glu Phe Gly Phe Gly Leu Ser Tyr Thr Thr Phe Asp Tyr
645 650 655

Gln Asn Leu Ser Ile Glu Thr Leu Lys Ala Ser Pro Ala Thr Ala Thr
660 665 670

Thr Thr Ser Pro Val Ala Ser Thr Thr Val Asn Pro Ser Arg Cys Pro
675 680 685

Thr Thr Ser Leu Asn Pro Ser Asp Tyr Thr Phe Pro Ser Thr Ile Ser
690 695 700

Ser Leu Thr Asn Tyr Ile Tyr Pro Tyr Ile Ala Pro Asn Ala Thr Ile
705 710 715 720

Ser Thr Ala Ala Ala Gly Thr Pro Thr Pro Thr Ile Ile Pro Glu Ala
725 730 735

Ala Gly Pro Gln Gly Gly Asn Pro Ala Leu Tyr Asp Val Leu Tyr Arg
740 745 750

Val Thr Leu Thr Leu Lys Asn Asn Gly Thr Val Ala Gly Gln Glu Val
755 760 765

Ala Gln Leu Tyr Leu Asp Leu Asp Asn Gly Glu Pro Ser Arg Gln Leu
770 775 780

Arg Gly Phe Asn Lys Thr Phe Leu Gln Ala Gly Glu Thr Lys Asp Val
785 790 795 800

Thr Phe Asp Leu Arg Arg Arg Asp Val Ala Ile Trp Asp Val Val Lys
805 810 815

Gln Glu Trp Val Glu Val Lys Lys Leu Gly Thr Lys Ile Gly Val Tyr
820 825 830

Val Gly Ala Ser Ser Arg Asp Leu Arg Leu Thr Gly Glu Ile
835 840 845

<210> 3

<211> 16

<212> PRT

5 <213> *T. emersonii*

<400> 3

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

REIVINDICACIONES

1. Un polipéptido que comprende la secuencia de aminoácidos expuesta en la SEQ ID NO: 2 o una secuencia de aminoácidos codificada por la secuencia de nucleótidos de la SEQ ID NO: 1, o un polipéptido variante o polinucleótido variante de la misma, en el que el polipéptido variante tiene al menos un 70 % de identidad de secuencia con la secuencia expuesta en la SEQ ID NO: 2 o el polinucleótido variante codifica un polipéptido que tiene al menos un 70 % de identidad de secuencia con la secuencia expuesta en la SEQ ID NO: 2 y en el que el polipéptido tiene actividad beta-glucosidasa.
2. Un polinucleótido que comprende:
 - (a) la secuencia de nucleótidos expuesta en la SEQ ID NO: 1; o
 - (b) una secuencia de nucleótidos que hibrida selectivamente con un polinucleótido que es el complemento inverso de la SEQ ID NO: 1; o
 - (c) una secuencia de nucleótidos que tiene al menos un 75 % de identidad de secuencia con la secuencia de nucleótidos de la SEQ ID NO: 1; o
 - (d) un fragmento que es de al menos aproximadamente 100 nucleótidos de longitud de una secuencia de nucleótidos definida en (a), (b) o (c); o
 - (e) una secuencia que está degenerada como resultado del código genético de una secuencia definida en uno cualquiera de (a), (b), (c) o (d); o
 - (f) una secuencia de nucleótidos que es el complemento inverso de una secuencia de nucleótidos definida en (a), (b), (c), (d) o (e), en el que el polinucleótido codifica un polipéptido de acuerdo con la reivindicación 1.
3. Un polinucleótido de acuerdo con la reivindicación 2 que es una secuencia de ADN.
4. Una construcción de ácido nucleico que comprende el polinucleótido de acuerdo con las reivindicaciones 2 o 3.
5. Una construcción de ácido nucleico de acuerdo con la reivindicación 4, en la que el contenido en GC es del 56 % o más, el 58 % o más, o del 58-65 %.
6. Un vector que incorpora una secuencia polinucleotídica de acuerdo con la reivindicación 2 o 3.
7. Una célula hospedadora que está transformada con un polinucleótido de acuerdo con la reivindicación 2 o 3, una construcción de ácido nucleico de acuerdo con la reivindicación 4 o 5 o un vector de acuerdo con la reivindicación 6, siendo dicha célula capaz de sobre-expresar el polipéptido.
8. Una célula de acuerdo con la reivindicación 7, en la que la célula es una célula eucariota.
9. Una célula de acuerdo con la reivindicación 8, en la que la célula es una célula fúngica.
10. Una célula de acuerdo con la reivindicación 9, en la que la célula fúngica se selecciona del grupo que consiste en los géneros *Aspergillus*, *Trichoderma/Hypocrea*, *Fusarium*, *Disporotrichum*, *Penicillium*, *Acremonium*, *Neurospora*, *Thermoascus*, *Myceliophthora*, *Sporotrichum*, *Thielavia*, *Chrysosporium*, *Fusarium*, *Humicola*, *Neurospora* y *Talaromyces*.
11. Una célula de acuerdo con la reivindicación 10, en la que la célula fúngica es de la especie *Aspergillus oryzae*, *Aspergillus sojae*, *Aspergillus nidulans*, o *Aspergillus niger*.
12. Una célula de acuerdo con cualquiera de las reivindicaciones 7 - 11, en la que uno o más genes están deletados, eliminados o alterados por completo o en parte, en la que opcionalmente el gen codifica una proteasa.
13. Una planta transgénica o parte de la misma que comprende un polinucleótido de acuerdo con la reivindicación 2 o 3, una construcción de ácido nucleico de acuerdo con la reivindicación 4 o 5, un vector de acuerdo con la reivindicación 6 o una célula de acuerdo con cualquiera de las reivindicaciones 7 - 12, en la que opcionalmente la planta es una planta de maíz, una planta de sorgo, una planta de tomate, una planta de trigo, una planta de semilla oleaginosa, una planta de colza, una planta de soja, una planta de cebada, una gramínea, o una planta de tabaco; o una planta de los géneros *Anacardium*, *Arachis*, *ragus*, *Atropa*, *Avena*, *Brassica*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *ferae*, *Cucumis*, *Cucurbita*, *Daucus*, *Elaeis*, *Fragaria*, *Glycine*, *Gossypium*, *Helianthus*, *Ocallis*, *Hordeum*, *Hyoscyamus*, *Lactuca*, *Linum*, *Lolium*, *Lupinus*, *Lycopersicon*, *Malus*, *Majorana*, *Medicago*, *Nicotiana*, *Olea*, *Oryza*, *Panieum*, *Pannisetum*, *Persea*, *eolus*, *Pistachio*, *Pisum*, *Pirus*, *Prunus*, *Raphanus*, *Ricinus*, *Secale*, *Senecio*, *Sinapis*, *Sorghum*, *Theobromus*, *Trigonella*, *Triticum*, *Vicia*, *Vitis*, *Vigna* o *Zea*.
14. Un método para la preparación de un polipéptido de acuerdo con la reivindicación 1, que tiene actividad degradante de material de carbohidrato y/o hidrolizante de carbohidrato, comprendiendo dicho método cultivar una célula de acuerdo con cualquiera de las reivindicaciones 7 a 12 en condiciones que permiten la expresión de dicho

polipéptido y, opcionalmente, recuperar el polipéptido expresado.

- 5 15. Una composición que comprende: (i) un polipéptido de acuerdo con la reivindicación 1 y; (ii) una celulasa y/o una hemicelulasa y/o una pectinasa y/o un polipéptido que pertenece a la familia GH61.
16. Una composición de acuerdo con la reivindicación 15, en la que la celulasa es una celobiohidrolasa, celobiohidrolasa I, celobiohidrolasa II, una endo- β -1,4-glucanasa, una β -glucosidasa o una β -(1,3)(1,4)-glucanasa.
- 10 17. Una composición de acuerdo con la reivindicación 15 o 16, en la que la hemicelulasa es una endoxilanas, una β -xilosidasa, una α -L-arabinofuranosidasa, una α -D-glucuronidasa, una acetil xilano esterasa, una feruloil esterasa, una cumaroil esterasa, una α -galactosidasa, una β -galactosidasa, una β -mananasa o una β -manosidasa.
- 15 18. Una composición de acuerdo con una cualquiera de las reivindicaciones 15 a 17, en la que la pectinasa es una endo poligalacturonasa, una pectin metil esterasa, una endo-galactanasa, una beta galactosidasa, una pectin acetil esterasa, una endo-pectin liasa, pectato liasa, alfa ramnosidasa, una exo-galacturonasa, una expoligalacturonato liasa, una ramnogalacturonano hidrolasa, una ramnogalacturonano liasa, una ramnogalacturonano acetil esterasa, una ramnogalacturonano galacturonohidrolasa, una xilogalacturonasa, una alfa-arabinofuranosidasa o una endo-arabinanasa.
- 20 19. Una composición de acuerdo con una cualquiera de las reivindicaciones 15 a 18 que comprende una ligninasa, una expansina, un polipéptido tipo expansina, una swollenina o el producto polipeptídico de un gen que codifica una proteína integradora de celulosa o una proteína inducida por celulosa.
- 25 20. Un método para el tratamiento de un sustrato, que comprende material de carbohidrato, opcionalmente un material vegetal, comprendiendo dicho método poner en contacto el sustrato con un polipéptido de acuerdo con la reivindicación 1 y/o una composición de acuerdo con una cualquiera de las reivindicaciones 15 a 19.
- 30 21. Un método de acuerdo con la reivindicación 20, en el que el sustrato es un material vegetal y el material vegetal se proporciona en forma de una planta, una pulpa vegetal, un extracto vegetal, un producto alimenticio o ingrediente derivado del mismo o una tela, tejido o prenda de vestir que comprende un material vegetal.
- 35 22. Un método de acuerdo con la reivindicación 20 y/o reivindicación 21, en el que el tratamiento comprende la degradación y/o modificación de celulosa y/o hemicelulosa y/o una sustancia péctica.
- 40 23. Un método para producir un producto de fermentación, comprendiendo dicho método producir un azúcar fermentable poniendo en contacto material lignocelulósico con un polipéptido de acuerdo con la reivindicación 1 y/o una composición de acuerdo con una cualquiera de las reivindicaciones 15 a 19, y fermentando el azúcar fermentable resultante para producir un producto de fermentación.
24. Uso de un polipéptido de acuerdo con la reivindicación 1 y/o una composición de acuerdo con cualquiera de las reivindicaciones 15 a 19 para producir azúcar a partir de un material lignocelulósico.

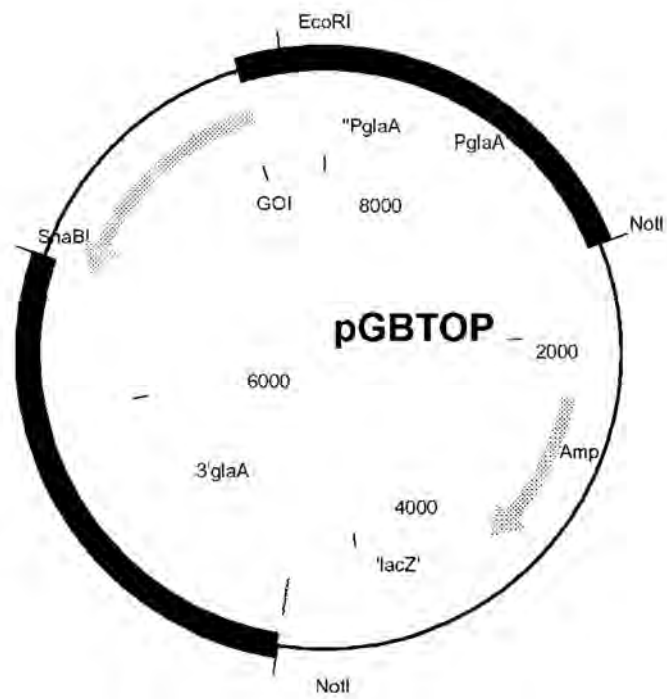


Fig. 1