

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 438 146**

51 Int. Cl.:

**C12N 9/00** (2006.01)

**C12N 9/14** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.06.2002 E 10179578 (9)**

97 Fecha y número de publicación de la concesión europea: **18.09.2013 EP 2277997**

54 Título: **Polipéptidos que tienen actividad de celobiohrolasa I y polinucleótidos que codifican los mismos**

30 Prioridad:

**26.06.2001 DK 200101000**

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

**16.01.2014**

73 Titular/es:

**NOVOZYMES A/S (100.0%)  
Krogshøjvej 36  
2880 Bagsvaerd, DK**

72 Inventor/es:

**LANGE, LENE;  
WU, WENPING;  
AUBERT, DOMINIQUE;  
LANDVIK, SARA;  
SCHNORR, KIRK MATTHEW y  
CLAUSEN, IB GROTH**

74 Agente/Representante:

**TOMAS GIL, Tesifonte Enrique**

**ES 2 438 146 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**

Polipéptidos que tienen actividad de celobiohidrolasa I y polinucleótidos que codifican los mismos

5 **Campo de la invención**

[0001] La presente invención se refiere a polipéptidos con actividad de celobiohidrolasa I (también llamada CBH I o CBH 1) y polinucleótidos con una secuencia de nucleótidos que codifica los polipéptidos. La invención también se refiere a constructos de ácidos nucleicos, vectores, y células huésped que comprenden los constructos de ácidos nucleicos al igual que métodos para producir y utilizar los polipéptidos.

**Antecedentes de la invención**

15 [0002] La celulosa es una materia prima industrial importante y una fuente de energía renovable. La estructura física y la morfología de la celulosa nativa son complejas y los detalles finos de su estructura han sido difíciles de determinar experimentalmente. No obstante, la composición química de la celulosa es simple, consiste en residuos de D-glucosa enlazados por enlaces beta-1,4-glicosídicos para formar polímeros lineales con longitud de cadenas de más de 10 000 residuos glicosídicos.

20 [0003] Para ser eficaz, la digestión de la celulosa requiere diferentes tipos de enzimas que actúan cooperativamente. Al menos tres categorías de enzimas son necesarias para convertir celulosa en glucosa: endo (1,4)-beta-D-glucanasas (EC 3.2.1.4) que cortan al azar las cadenas de celulosa; celobiohidrolasas (EC 3.2.1.91) que disocian unidades de celobiosil de los extremos de la cadena de celulosa y beta-glucosidasas (EC 3.2.1.21) que convierten celobiosa y celodextrinas solubles en glucosa. Entre estas tres categorías de enzimas implicadas en la biodegradación de celulosa, las celobiohidrolasas son las enzimas clave para la degradación de celulosa cristalina nativa.

[0004] Exo-celobiohidrolasas (celobiohidrolasa I o CBH I) se refieren a las celobiohidrolasas que degradan celulosa mediante hidrólisis de la celobiosa del extremo no reducido de las cadenas de polímero de celulosa.

30 [0005] Es un objeto de la presente invención proporcionar polipéptidos mejorados con actividad de celobiohidrolasa I y polinucleótidos que codifiquen los polipéptidos. Los polipéptidos mejorados pueden tener actividad específica mejorada y/o estabilidad mejorada, en particular termoestabilidad mejorada. Los polipéptidos también pueden tener una capacidad mejorada para resistir a la inhibición por celobiosa.

35 **Resumen de la invención**

[0006] En un primer aspecto, la presente invención se refiere a un polipéptido con actividad de celobiohidrolasa I, seleccionado del grupo que consiste en:

40 (a) un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 90% de identidad con los aminoácidos 1 a 457 de SEC ID n.º: 8,

(b) un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 90% de identidad con el polipéptido codificado por los nucleótidos 1 a 1371 de SEC ID n.º: 7.

45 [0007] En un segundo aspecto, la presente invención se refiere a un polinucleótido con una secuencia de nucleótidos que codifica el polipéptido de la invención.

50 [0008] En un tercer aspecto, la presente invención se refiere a un constructo de ácidos nucleicos que comprende la secuencia de nucleótidos, que codifica el polipéptido de la invención, operativamente enlazado a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.

[0009] En un cuarto aspecto, la presente invención se refiere a un vector de expresión recombinante que comprende el constructo de ácidos nucleicos de la invención.

55 [0010] En un quinto aspecto, la presente invención se refiere a una célula huésped recombinante que comprende el constructo de ácidos nucleicos de la invención.

60 [0011] En un sexto aspecto, la presente invención se refiere a un método para producir un polipéptido de la invención, el método comprende:

(a) cultivo de una célula huésped recombinante de la invención bajo condiciones propicias para la producción del polipéptido; y

(b) recuperación del polipéptido.

[0012] En un séptimo aspecto, la presente invención se refiere a un método para la producción in situ de un polipéptido de la invención, el método comprende:

(a) cultivo una célula huésped recombinante de la invención bajo condiciones propicias para la producción del polipéptido; y

(b) puesta en contacto del polipéptido con un sustrato deseado sin recuperación previa del polipéptido.

[0013] Otros aspectos de la presente invención serán aparentes de la descripción que se realizará más adelante y de las reivindicaciones anexas

### Definiciones

[0014] Antes de discutir la presente invención en detalles adicionales, los siguientes términos y convenios serán definidos primeramente:

*Polipéptido sustancialmente puro*: en el presente contexto, el término "polipéptido sustancialmente puro" significa una preparación de polipéptido que contiene como mucho un 10% en peso de otro material polipeptídico con el cual es originalmente asociado (porcentajes inferiores de otro material polipeptídico son preferidos, por ejemplo como mucho un 8% en peso, como mucho un 6% en peso, como mucho un 5% en peso, como mucho un 4% como mucho 3% en peso, como mucho un 2% en peso, como mucho un 1% en peso, y como mucho un  $\frac{1}{2}$  % en peso). Así, se prefiere que el polipéptido sustancialmente puro sea al menos un 92% puro, es decir que el polipéptido constituya al menos un 92% en peso del material polipeptídico total presente en la preparación, y porcentajes más altos son preferidos tales como al menos un 94% puro, al menos un 95% puro, al menos un 96% puro, al menos un 96% puro, al menos un 97% puro, al menos un 98% puro, al menos un 99%, y como mucho un 99.5% puro. Los polipéptidos descritos aquí están preferiblemente en forma sustancialmente pura. En particular, se prefiere que los polipéptidos descritos aquí estén en "forma esencialmente pura", es decir que la preparación de polipéptido esté esencialmente libre de otro material polipeptídico con el cual es originalmente asociado. Este puede ser realizado, por ejemplo, por preparación del polipéptido mediante métodos recombinantes bien conocidos. Aquí, el término "polipéptido sustancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada".

*Actividad de Celobiohidrolasa I*: El término "actividad de celobiohidrolasa I" es definido aquí como una actividad de celulosa 1,4-beta-celobiosidasa (también llamada exo-glucanasa, exo-celobiohidrolasa o 1,4-beta-celobiohidrolasa), como se define en la clase enzimática EC 3.2.1.91, que cataliza la hidrólisis de enlaces de 1,4-beta-D-glucosídico en celulosa y celotetraosa, liberando celobiosa de los extremos no reductores de las cadenas.

[0015] Para fines de la presente invención, la actividad de celobiohidrolasa I puede ser determinada según el procedimiento descrito en ejemplo 2.

[0016] En una forma de realización, la actividad de celobiohidrolasa I puede ser determinada según el procedimiento descrito en Deshpande MV et al., Methods in Enzymology, pp. 126-130 (1988): "Selective Assay for Exo-1,4-Beta-Glucanases". Según este procedimiento, una unidad de actividad de celobiohidrolasa I (actividad de escisión de enlace aglucónico) es definida como 1,0  $\mu$ mol de p-nitrofenol producido por minuto a 50°C, pH 5,0.

[0017] Los polipéptidos de la presente invención deberían preferiblemente tener como mínimo un 20% de la actividad de celobiohidrolasa I de un polipéptido que en una secuencia de aminoácidos de SEC ID n.º:8. En una forma de realización preferida particular, los polipéptidos deberían tener al menos un 40%, tal como al menos un 50%, preferiblemente al menos un 60%, tal como al menos un 70%, más preferiblemente al menos un 80%, tal como al menos un 90%, de la forma más preferible al menos un 95%, tal como acerca de o al menos un 100% de la actividad de celobiohidrolasa I del polipéptido que consiste en los aminoácidos 1 a 457 de SEC ID n.º:8.

[0018] *Identidad*: en el presente contexto, la homología entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos es descrita por el parámetro "identidad".

[0019] Para fines de la presente invención, el grado de identidad entre dos secuencias de aminoácidos es determinado usando el programa FASTA incluido en la versión 2.0x del paquete FASTA (véase W. R. Pearson and D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448 ; and W. R. Pearson (1990) "Rapid and Sensitive Sequence Comparison with FASTP and FASTA", Methods in Enzymology 183:63-98). La matriz de puntuación usada fue

BLOSUM50, penalización del espacio fue -12, y penalización de extensión del espacio fue -2.

5 [0020] El grado de identidad entre dos secuencias de nucleótidos es determinado usando el mismo algoritmo y paquete de software, como se ha descrito anteriormente. La matriz de puntuación usada fue la matriz de identidad, penalización del espacio fue -16, y penalización de extensión del espacio fue -4.

10 [0021] Fragmento: cuando se usa aquí, un "fragmento" de una secuencia seleccionada del grupo que consiste en SEC ID n.º:8 es un polipéptido con uno o más aminoácidos eliminados del extremo amino y/o carboxilo terminal de esta secuencia de aminoácidos. Preferiblemente, un fragmento es un polipéptido con la secuencia de aminoácidos eliminada correspondiente al "dominio de enlace de celulosa" y/o el "dominio enlazador" de Celobiohidrolasa I de *Trichoderma reesei* como está descrito en el número de acceso de SWISS-PROT P00725. Más preferiblemente, un fragmento comprende la secuencia de aminoácidos correspondiente al "dominio catalítico" de celobiohidrolasa I de *Trichoderma reesei* como se describe en el número de acceso de SWISS-PROT P00725. De la forma más preferible, un fragmento contiene al menos 434 residuos de aminoácidos, por ejemplo, los residuos de aminoácidos 1 a 434 de SEC ID n.º:8. En particular, un fragmento contiene al menos 215 residuos de aminoácidos, por ejemplo, los residuos de aminoácidos 200 a 434 de SEC ID n.º:8.

20 [0022] Variante alélica: en el presente contexto, el término "variante alélica" denota cualquiera de dos o más formas alternativas de un gen que ocupa el mismo locus cromosómico. La variación alélica surge naturalmente a través de mutaciones, y puede tener como resultado el polimorfismo dentro de poblaciones. Las mutaciones de genes pueden ser silenciosas (ningún cambio en el polipéptido codificado) o puede codificar polipéptidos con secuencias de aminoácidos alteradas. Una variante alélica de un polipéptido es un polipéptido codificado por una variante alélica de un gen.

25 [0023] Polinucleótido sustancialmente puro: el término "polinucleótido sustancialmente puro" como se utiliza en este caso se refiere a una preparación polinucleótida, donde el polinucleótido ha sido retirado de su ambiente genético natural, y está así libre de otras secuencias de codificación indeseadas o extrañas y está en una forma adecuada para el uso dentro de sistemas de producción de proteínas creada genéticamente. Así, un polinucleótido sustancialmente puro contiene como mucho un 10% en peso de otro material polinucleótido con el cual está originalmente asociado (porcentajes inferiores de otro material polinucleótido son preferidos, por ejemplo como mucho un 8% en peso, como mucho un 6% en peso, como mucho un 5% en peso, como mucho un 4% como mucho un 3% en peso, como mucho un 2% en peso, como mucho un 1% en peso, y como mucho %% en peso). Un polinucleótido sustancialmente puro puede, no obstante, incluir regiones no traducidas 5' y 3' existentes de forma natural, tales como promotores y terminadores. Se prefiere que el polinucleótido sustancialmente puro sea al menos un 92% puro, es decir que el polinucleótido constituya al menos un 92% en peso del material de polinucleótido total presente en la preparación, y porcentajes más altos son preferidos tal como al menos un 94% puro, al menos un 95% puro, al menos un 96% puro, al menos un 96% puro, al menos un 97% puro, al menos un 98% puro, al menos un 99%, y como mucho un 99,5% puro. Los polinucleótidos descritos aquí están preferiblemente en una forma sustancialmente pura. En particular, se prefiere que los polinucleótidos descritos aquí estén en "forma esencialmente pura", es decir que la preparación polinucleótida esté esencialmente libre de otro material polinucleótido con el cual está originalmente asociado. Aquí, el término "polinucleótido sustancialmente puro" es sinónimo de los términos "polinucleótido aislado" y "polinucleótido en forma aislada".

45 [0024] Modificación(es): en el contexto de la presente invención el término "modificación(es)" significa cualquier modificación química de un polipéptido que consista en una secuencia de aminoácidos seleccionado de SEQ ID n.º: 8, al igual que manipulación genética del ADN que codifica ese polipéptido. La(s) modificación(es) puede(n) ser sustitución(es) de la(s) cadena(s) lateral(es) de aminoácido, sustitución(es), delección(es) y/o inserciones(s) en o al(a los) aminoácido(s) de interés.

50 [0025] Variante artificial: cuando se usa aquí, el término "variante artificial" significa un polipéptido con actividad de celobiohidrolasa I, que ha sido producido por un organismo que está expresando un gen modificado en comparación con la SEQ ID n.º : 7. El gen modificado, donde dicha variante se produce cuando se expresa en un huésped adecuado, se obtiene a través de la intervención humana por modificación de una secuencia de nucleótidos que consiste en SEC ID n.º: 7.

55 [0026] ADNc: el término "ADNc" cuando se usa en el presente contexto, se destina a cubrir una molécula de ADN que se puede preparar por transcripción inversa a partir de una molécula de ARN madura, empalmada, derivada de una célula eucariota. Las secuencias de ADNc carecen de las secuencias de intrón que están normalmente presentes en el ADN genómico correspondiente. La transcripción de ARN inicial primaria es un precursor para ARNm y va a través de una serie de eventos de tratamiento antes de aparecer como ARNm empalmado maduro. Estos eventos incluyen la eliminación de secuencias de intrón por un proceso llamado empalme. Cuando el ADNc es derivado de ARNm carece por lo tanto de secuencias de intrón.

60 [0027] Constructo de ácidos nucleicos: cuando se usa aquí, el término "constructo de ácidos nucleicos" significa una molécula de ácido nucleico, mono o bicatenario, que se aísla de un gen de origen natural o que ha sido modificada para

contener segmentos de ácidos nucleicos de manera que de otra forma no existirían en la naturaleza. El término constructo de ácidos nucleicos es sinónimo del término "cassette de expresión" cuando el constructo de ácidos nucleicos contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención.

5 [0028] Secuencia de control: el término "secuencias de control" es definido aquí para incluir todos los componentes que son necesarios o ventajosos para la expresión de un polipéptido de la presente invención. Cada secuencia de control puede ser nativa o extraña a la secuencia de nucleótidos que codifica el polipéptido. Tales secuencias de control incluyen, pero de forma no limitativa, una secuencia líder, secuencia de poliadenilación, secuencia de propéptido, promotor, secuencia de péptido señal, y terminador de transcripción. A un mínimo, las secuencias de control incluyen un promotor, y señales de parada traduccional y transcripcional. Las secuencias de control se pueden proporcionar con enlaces para introducir sitios de restricción específica facilitando la ligadura de las secuencias de control con la región de codificación de la secuencia de nucleótidos que codifica un polipéptido.

10 [0029] Operativamente enlazado: el término "operativamente enlazado" es definido aquí como una configuración en la que una secuencia de control está apropiadamente colocada en una posición relativa a la secuencia de codificación de la secuencia de ADN tal que la secuencia de control dirige la expresión de un polipéptido.

15 [0030] Secuencia codificante: cuando se usa aquí el término "secuencia codificante" se destina a cubrir una secuencia de nucleótidos que especifica directamente la secuencia de aminoácidos de su producto proteínico. Los límites de la secuencia codificante están generalmente determinados por un marco de lectura abierto, que comienza normalmente con el codón de inicio ATG. La secuencia codificante típicamente incluye ADN, ADNc y secuencias de nucleótidos recombinantes.

20 [0031] Expresión: en el presente contexto, el término "expresión" incluye cualquier paso implicado en la producción del polipéptido incluyendo, pero no limitado a, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.

25 [0032] Vector de expresión: en el presente contexto, el término "vector de expresión" cubre una molécula de ADN, lineal o circular, que incluye un segmento que codifica un polipéptido de la invención y que está operativamente enlazado a segmentos adicionales que proveen para su transcripción.

30 [0033] Célula huésped: el término "célula huésped", como se utiliza en este caso, incluye cualquier tipo celular que sea susceptible de transformación con un constructo de ácidos nucleicos.

35 [0034] Los términos, "sonda polinucleótida", "hibridación" al igual que las varias condiciones de astringencia se definen en la sección titulada "Polipéptidos que tienen actividad de celobiohidrolasa I".

[0035] Termoestabilidad: el término "thermostability", como se utiliza en este caso, es medido como se describe en el ejemplo 2.

#### 40 Descripción detallada de la invención

##### Polipéptidos con actividad de celobiohidrolasa I

45 [0036] En una primera forma de realización, la presente invención se refiere a polipéptidos con actividad de celobiohidrolasa I y donde los polipéptidos comprenden, preferiblemente consisten en, una secuencia de aminoácidos que tiene un grado de identidad para una secuencia de aminoácidos de SEC ID n.º: 8, (es decir, el polipéptido maduro) de al menos un 90%, preferiblemente al menos un 95%, por ejemplo al menos un 96%, tal como al menos un 97%, e incluso de la forma más preferible al menos un 98%, tal como al menos un 99% (de ahora en adelante "polipéptidos homólogos"). En una forma de realización interesante, la secuencia de aminoácidos difiere en como mucho diez aminoácidos (p. ej. por diez aminoácidos), en particular por como mucho cinco aminoácidos (p. ej. en cinco aminoácidos), tal como en como mucho cuatro aminoácidos (p. ej. en cuatro aminoácidos), por ejemplo en como mucho tres aminoácidos (p. ej. en tres aminoácidos) de una secuencia de aminoácidos de SEC ID n.º: 8. En una forma de realización interesante particular, la secuencia de aminoácidos difiere en como mucho dos aminoácidos (p. ej. en dos aminoácidos), tal como por un aminoácido de una secuencia de aminoácidos de SEC ID n.º: 8.

50 [0037] Preferiblemente, los polipéptidos de la presente invención comprenden una secuencia de aminoácidos de SEC ID n.º:8. En otra forma de realización preferida, el polipéptido de la presente invención consiste en SEC ID n.º: 8.

55 [0038] El polipéptido de la invención puede ser una celobiohidrolasa I de tipo salvaje identificada y aislada de una fuente natural. Tales polipéptidos de tipo salvaje pueden ser específicamente seleccionados por técnicas estándar conocidas en la técnica, tal como selección molecular como está descrito en el ejemplo 1. Además, el polipéptido de la invención se puede

preparar por la técnica de redistribución de ADN, tal como se describe en J.E. Ness et al. Nature Biotechnology 17, 893-896 (1999). Por otra parte, el polipéptido de la descripción puede ser una variante artificial que comprende, preferiblemente consiste en, una secuencia de aminoácidos que tiene al menos una sustitución, eliminación y/o inserción de un aminoácido en comparación con una secuencia de aminoácidos de SEC ID n.º: 8. Tales variantes artificiales se pueden construir por técnicas estándar conocidas en la técnica, tal como por mutagénesis dirigida al sitio/aleatoria del polipéptido que incluye una secuencia de aminoácidos de SEC ID n.º: 8. En una forma de realización de la descripción, los cambios de aminoácidos (en la variante artificial al igual que en polipéptidos tipo salvaje) son de naturaleza menor, esto es, sustituciones de aminoácidos conservadoras que significativamente no afectan el pliegue y/o actividad de la proteína; deleciones pequeñas, típicamente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones en el amino o carboxilo terminal, tal como un residuo de metionina aminoterminal; un péptido enlazador pequeño de aproximadamente 20-25 residuos; o una extensión pequeña que facilita la purificación por cambio de la carga neta u otra función, tal como un tracto de polihistidina, un epítipo antigénico o un dominio de unión.

[0039] Ejemplos de sustituciones conservadoras son en el grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina y asparagina), aminoácidos hidrofóbicos (leucina, isoleucina, valina y metionina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina), y aminoácidos pequeños (glicina, alanina, serina y treonina). Sustituciones de aminoácidos que generalmente no alteran la actividad específica se conocen en la técnica y están descritas, por ejemplo, por H. Neurath and R.L. Hill, 1979, In, The Proteins, Academic Press, New York. Los intercambios que más frecuentemente ocurren son Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, y Asp/Gly igual que estos a la inversa.

[0040] En una forma de realización interesante de la descripción, los cambios de aminoácidos son de tal naturaleza que las propiedades físico químicas de los polipéptidos son alteradas. Por ejemplo, cambios de aminoácidos pueden ser realizados, que mejoran la termoestabilidad del polipéptido, alteran la especificidad de sustrato, cambian el pH óptimo y similares.

[0041] Preferiblemente, el número de tales sustituciones, deleciones y/o inserciones comparadas con una secuencia de aminoácidos de SEC ID n.º: 8 es como máximo 10, tal como máximo 9, por ejemplo como máximo 8, más preferiblemente como máximo 7, por ejemplo como máximo 6, tal como máximo 5, más preferiblemente como máximo 4, por ejemplo como máximo 3, tal como máximo 2, en particular como máximo 1.

[0042] Los presentes inventores han aislado polipéptidos de codificación de secuencias de nucleótidos con actividad de celobiohidrolasa I del microorganismo *Thermoascus aurantiacus*. Así, en una segunda forma de realización, la presente invención se refiere a polipéptidos que incluyen una secuencia de aminoácidos que tiene al menos un 90% identidad con el polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado de *Thermoascus aurantiacus*. En una forma de realización interesante de la invención, el polipéptido comprende una secuencia de aminoácidos que tiene al menos un 90%, preferiblemente al menos un 95%, por ejemplo al menos un 96%, tal como al menos un 97%, e incluso de la forma más preferible al menos un 98%, tal como al menos un 99% de identidad con el polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado de *Thermoascus aurantiacus* (de ahora en adelante "polipéptidos homólogos"). En una forma de realización interesante, la secuencia de aminoácidos difiere en como mucho diez aminoácidos (p. ej. en diez aminoácidos), en particular en como mucho cinco aminoácidos (p. ej. por cinco aminoácidos), tal como en como mucho cuatro aminoácidos (p. ej. en cuatro aminoácidos), por ejemplo en como mucho tres aminoácidos (p. ej. en tres aminoácidos) del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado de *Thermoascus aurantiacus*. En una forma de realización interesante particular, la secuencia de aminoácidos difiere en como mucho dos aminoácidos (p. ej. por dos aminoácidos), tal como en un aminoácido del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado de *Thermoascus aurantiacus*.

[0043] Preferiblemente, los polipéptidos de la presente invención comprenden la secuencia de aminoácidos del polipéptido codificados por la parte codificante de celobiohidrolasa I de la secuencia de nucleótidos insertados en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n.º 0582. En otra forma de realización preferida, el polipéptido de la presente invención consiste en la secuencia de aminoácidos del polipéptido codificado por la parte codificante de celobiohidrolasa I de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n.º 0582.

[0044] De una manera similar como se ha descrito anteriormente, el polipéptido de la descripción puede ser una variante artificial que comprende, preferiblemente consiste en, una secuencia de aminoácidos que tiene al menos una sustitución, eliminación y/o inserción de un aminoácido en comparación con la secuencia de aminoácidos codificada por la parte codificante de celobiohidrolasa I de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n.º 0582.

[0045] En otra forma de realización, la presente divulgación se refiere a polipéptidos con actividad de celobiohidrolasa I que se codifican por secuencias de nucleótidos que hibridan bajo condiciones de astringencia muy bajas, preferiblemente bajo condiciones de astringencia bajas, más preferiblemente bajo condiciones de astringencia media, más preferiblemente bajo condiciones de astringencia medio altas, incluso más preferiblemente bajo condiciones de astringencia altas, y de la forma más preferible bajo condiciones de astringencia altísimas con una sonda polinucleótida seleccionada del grupo que consiste en: (i) la cadena complementaria de nucleótidos 1 a 1371 de SEC ID n.º: 7;

(ii) la cadena complementaria de nucleótidos 1 a 500 de SEC ID n.º: 7 y

(iii) la cadena complementaria de nucleótidos 1 a 200 de SEC ID n.º: 7

(J. Sambrook, E.F. Fritsch, and T. Maniatus, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York).

[0046] La presente divulgación se refiere a polipéptidos con actividad de celobiohidrolasa I que se codifican por la parte codificante de celobiohidrolasa I de la secuencia de nucleótidos presentes en un microorganismo del género *Thermoascus* (bajo *Ascomycota*, *Eurotiomycetes*, *Eurotiales*, *Trichocomoaceae*) por ejemplo *Thermoascus aurantiacus*.

[0047] Una secuencia de nucleótidos seleccionada de SEQ ID n.º 7 o una subsecuencia de la misma, al igual que una secuencia de aminoácidos seleccionada de SEQ ID n.º 8 o un fragmento de la misma se puede utilizar para diseñar una sonda polinucleótida para identificar y clonar polipéptidos de codificación de ADN con actividad de celobiohidrolasa I de cepas de diferentes géneros o especies, según métodos bien conocidos en la técnica. En particular, tales sondas se pueden usar para hibridación con el ADNc o genómico del género o especie de interés, siguiendo procedimientos de transferencia de Southern estándar, para identificar y aislar el gen correspondiente en este. Tales sondas pueden ser considerablemente más cortas que la secuencia entera, pero deberían tener al menos 15, preferiblemente al menos 25, más preferiblemente al menos 35 nucleótidos de longitud, tal como al menos 70 nucleótidos de longitud. Es, no obstante, preferido que la sonda polinucleótida tenga al menos 100 nucleótidos de longitud. Por ejemplo, la sonda polinucleótida puede tener al menos 200 nucleótidos de longitud, al menos 300 nucleótidos de longitud, al menos 400 nucleótidos de longitud o al menos 500 nucleótidos de longitud. Sondas más largas incluso pueden ser utilizadas, por ejemplo, sondas polinucleótidas que tienen al menos 600 nucleótidos de longitud, al menos 700 nucleótidos de longitud, al menos 800 nucleótidos de longitud, o al menos 900 nucleótidos de longitud. Tanto sondas de ADN como de ARN pueden ser usadas. Las sondas están típicamente marcadas para la detección del gen correspondiente (por ejemplo, con <sup>32</sup>P, <sup>3</sup>H, <sup>35</sup>S, biotina, o avidina).

[0048] Así, un ADN genómico o genoteca de ADNc obtenida a partir de tales otros organismos se pueden seleccionar para ADN que hibridiza con las sondas anteriormente descritas y que codifica un polipéptido con actividad de celobiohidrolasa I.

[0049] Para fines de la presente invención, la hibridación indica que la secuencia de nucleótidos hibridiza a una sonda de polinucleótido etiquetada que hibridiza una sonda de polinucleótidos marcada bajo condiciones de astringencia de muy baja a altísima. Moléculas a las que la sonda polinucleótida hibridiza bajo estas condiciones pueden ser detectadas usando una película radiográfica o por cualquier otro método conocido en la técnica. Siempre que el término "sonda polinucleótida" se use en el presente contexto, debe entenderse que tal sonda contiene al menos 15 nucleótidos.

[0050] En una forma de realización interesante, la sonda polinucleótida es la cadena complementaria de nucleótidos 1 a 1371 de SEC ID n.º: 7; o la cadena complementaria de nucleótidos 1 a 500 de SEC ID n.º: 7; o la cadena complementaria de nucleótidos 1 a 200 de SEC ID n.º: 7.

[0051] La sonda polinucleótida puede ser la cadena complementaria de la secuencia de nucleótidos que codifica un polipéptido de SEC ID n.º:8. La sonda polinucleótida puede ser la cadena complementaria de una secuencia de nucleótidos de SEC ID n.º:7. La sonda polinucleótida puede ser la cadena complementaria de la secuencia de nucleótidos contenida en un plásmido que se contiene en un microorganismo depositado seleccionado de CGMCC n.º 0582.

[0052] Para sondas largas de al menos 100 nucleótidos de longitud, condiciones de astringencia de muy baja a altísima son definidas como prehibridación e hibridación a 42 °C en SSPE 5X, SDS 1,0%, solución de Denhardt 5X, 100 µg/ml ADN de esperma de salmón desnaturalizado y cortado, siguiendo procedimientos de transferencia de Southern estándar. Preferiblemente, las sondas largas de al menos 100 nucleótidos no contienen más de 1000 nucleótidos. Para sondas largas de al menos 100 nucleótidos de longitud, el material portador es finalmente lavado tres veces cada uno durante 15 minutos usando SSC 2 x, SDS 0,1% a 42 °C (astringencia muy baja), preferiblemente lavado tres veces cada uno durante 15 minutos usando SSC 0,5 x, SDS 0,1% a 42 °C (astringencia baja), más preferiblemente lavado tres veces cada uno durante 15 minutos usando SSC 0,2 x, SDS 0,1% a 42 °C (astringencia media), incluso más preferiblemente lavado tres veces cada

uno durante 15 minutos usando SSC 0,2 x, SDS 0,1% a 55 °C (astringencia media alta), de la forma más preferible lavado tres veces cada uno durante 15 minutos usando SSC 0,1 x, SDS 0,1% a 60 °C (astringencia alta), en particular lavado tres veces cada uno durante 15 minutos usando SSC 0,1 x, SDS 0,1% a 68 °C (astringencia altísima).

5 [0053] Aunque no es particularmente preferido, se contempla que sondas más cortas, por ejemplo sondas que tienen de aproximadamente 15 a 99 nucleótidos de longitud, tal como de aproximadamente 15 a aproximadamente 70 nucleótidos de longitud, pueden también ser usadas. Para este tipo de sondas cortas, condiciones de astringencia son definidas como prehibridación, hibridación, y post-hibridación de lavado a 5 °C a 10 °C por debajo de la  $T_m$  calculada utilizando el cálculo según Bolton y McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) en NaCl 0,9 M, Tris-HCl 10  
10 0,09 M pH 7,6, EDTA 6 mM, NP-40 0,5%, solución de Denhardt 1x, pirofosfato de sodio 1 mM, fosfato monobásico de sodio 1 mM, ATP 0,1 mM, y 0,2 mg de ARN de levadura por ml siguiendo los procedimientos de transferencia de Southern estándar.

15 [0054] Para sondas cortas que tienen aproximadamente 15 nucleótidos a 99 nucleótidos de longitud, el material portador se lava una vez en SCC 6X más SDS 0,1% durante 15 minutos y dos veces cada uno durante 15 minutos usando SSC 6X a 5 °C hasta 10 °C por debajo de la  $T_m$  calculada.

#### Fuentes para Polipéptidos con Actividad de Celobiohidrolasa I

20 [0055] Un polipéptido de la presente invención puede ser un polipéptido de *Thermoascus*.

[0056] En una forma de realización preferida, el polipéptido es un polipéptido de *Thermoascus aurantiacus*.

25 [0057] En una forma de realización más preferida, el polipéptido es un polipéptido de *Thermoascus aurantiacus*, por ejemplo, el polipéptido consistiendo en una secuencia de aminoácidos de SEC ID n.º: 8.

[0058] Será entendido que para las especies ya mencionadas, la invención abarca los estados imperfectos y perfectos, y otros equivalentes taxonómicos, por ejemplo, anamorfos, independientemente del nombre de la especie por el que son conocidos. Expertos en la técnica reconocerán la identidad de los equivalentes apropiados.

30 [0059] Cepas de estas especies están fácilmente accesibles al público en un número de colecciones de cultivo, como the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), China General Microbiological Culture Collection Center (CGMCC), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

35 [0060] Además, tales polipéptidos se pueden identificar y obtener de otras fuentes que incluyen microorganismos aislados de la naturaleza (p. ej., tierra, agua, plantas, animales, etc.) usando las sondas arriba mencionadas. Técnicas para aislar microorganismos de hábitats naturales se conocen en la técnica. La secuencia de nucleótidos se puede luego derivar por selección de forma similar de una genoteca de ADNc o genómica de otro microorganismo. Una vez que una secuencia de nucleótidos que codifica un polipéptido ha sido detectada con la(s) sonda(s), la secuencia se puede aislar o clonar por utilización de técnicas que se conocen por aquellos de habilidad común en la técnica (véase, por ejemplo, Sambrook et al., 1989, *supra*).

40 [0061] Polipéptidos codificados por secuencias de nucleótidos de la presente invención también incluyen polipéptidos fusionados o polipéptidos divisibles de fusión en el que otro polipéptido se fusiona al extremo N-terminal o el C-terminal del polipéptido o fragmento del mismo. Un polipéptido fusionado se produce por fusión de una secuencia de nucleótidos (o una parte de la misma) que codifica otro polipéptido a una secuencia de nucleótidos (o una parte de la misma) de la presente invención. Técnicas para producir polipéptidos de fusión se conocen en la técnica e incluyen ligar las secuencias de codificación que codifican los polipéptidos de modo que éstas están en bastidor y que la expresión del polipéptido fusionado está bajo control del(de los) mismo(s) promotor(es) y terminador.

#### Polinucleótidos y Secuencias de Nucleótidos

45 [0062] La presente invención también se refiere a polinucleótidos con una secuencia de nucleótidos que codifica un polipéptido de la invención. En particular, la presente invención se refiere a polinucleótidos que consisten en una secuencia de nucleótidos que codifica un polipéptido de la invención. En una forma de realización preferida, la secuencia de nucleótidos es SEC ID n.º: 7. En una forma de realización más preferida, la secuencia de nucleótidos es el polipéptido maduro que codifica la región contenida en un plásmido que está contenido en un microorganismo depositado seleccionado de CGMCC n.º 0582. La presente invención también abarca polinucleótidos que comprenden, preferiblemente que consisten en, secuencias de nucleótidos que codifican un polipéptido que consiste en una secuencia de aminoácidos seleccionada de SEC ID n.º: 8, que difiere de una secuencia de nucleótidos seleccionada de SEC ID n.º: 7 en virtud de la degeneración del

código genético.

5 [0063] La presente divulgación también se refiere a polinucleótidos que comprenden, preferiblemente consisten en, una subsecuencia de una secuencia de nucleótidos seleccionada de SEC ID n.º: 7 que codifica fragmentos de una secuencia de aminoácidos seleccionada de SEC ID n.º: 8 que tiene actividad de celobiohidrolasa I. Una subsecuencia de una secuencia de nucleótidos seleccionada de SEC ID n.º: 7 es una secuencia de nucleótidos encerrada por una secuencia seleccionada de SEC ID n.º: 7 excepto que uno o más nucleótidos del extremo 5' y/o 3' han sido eliminados.

10 [0064] La presente divulgación también se refiere a polinucleótidos con, preferiblemente consistiendo en, una secuencia de nucleótidos modificada que comprende al menos una modificación en la secuencia codificante de polipéptido maduro seleccionada de SEC ID n.º: 7, y donde la secuencia de nucleótidos modificada codifica un polipéptido que consiste de una secuencia de aminoácidos seleccionada de SEC ID n.º: 8.

15 [0065] Las técnicas usadas para aislar o clonar una secuencia de nucleótidos que codifica un polipéptido se conocen en la técnica e incluyen el aislamiento de ADN genómico, preparación de ADNc, o una combinación de las mismas. La clonación de las secuencias de nucleótidos de la presente invención de tal ADN genómico puede ser efectuada, por ejemplo, usando la bien conocida reacción en cadena de polimerasa (PCR) o selección de anticuerpos de genotecas de expresión para detectar fragmentos de ADN clonados con características estructurales compartidas. Véase, por ejemplo, Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Otros procedimientos de amplificación tales como  
20 reacción en cadena de ligasa (LCR), transcripción activada ligada (LAT) y amplificación basada en secuencias de nucleótidos (NASBA) pueden ser utilizados. La secuencia de nucleótidos se puede clonar de una cepa seleccionada del grupo que consiste en *Acremonium*, *Scytalidium*, *Teramoascus*, *Thielavia*, *Verticillium*, *Neotermes*, *Melanocarpus*, *Poitrasia*, *Coprinus*, *Trichotecium*, *Humicola*, *Cladorrhinum*, *Diplodia*, *Myceliophthora*, *Rhizomucor*, *Meripilus*, *Exidia*, *Xylaria*, *Trichophaea*, *Chaetomium*, *Chaetomidium*, *Sporotrichum*, *Thielavia*, *Aspergillus*, *Scopulariopsis*, *Fusarium*, *Pseudoplectania* y *Phytophthora* u otro o relacionado organismo y así, por ejemplo, puede ser una variante alélica o de especies del  
25 polipéptido que codifica la región de la secuencia de nucleótidos.

[0066] La secuencia de nucleótidos se puede obtener por procedimientos de clonación estándar usados en ingeniería genética para recolocar la secuencia de nucleótidos desde su ubicación natural a un sitio diferente donde será reproducida. Los procedimientos de clonación pueden implicar escisión y aislamiento de un fragmento deseado que comprende la secuencia de nucleótidos que codifica el polipéptido, inserción del fragmento en una molécula de vector, e incorporación del vector recombinante en una célula huésped donde copias múltiples o clones de la secuencia de nucleótidos serán replicadas. La secuencia de nucleótidos puede ser de origen genómico, ADNc, ARN, origen semisintético, sintético, o cualquier combinación de los mismos.  
35

[0067] La presente invención también se refiere a un polinucleótido que comprende, preferiblemente consiste en, una secuencia de nucleótidos que tiene un grado de identidad con los nucleótidos 1 a 1371 de SEC ID n.º: 7 de al menos un 90% de identidad, más preferiblemente al menos un 95% de identidad, tal como al menos un 96% de identidad, por ejemplo al menos un 97% de identidad, incluso más preferiblemente al menos un 98% de identidad, tal como al menos un 99%. El grado de identidad entre dos secuencias de nucleótidos se determina como se describe previamente (véase la sección titulada "Definiciones").  
40

[0068] En una forma de realización preferida, el grado de identidad con la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n.º 0582 es al menos un 90%, más preferiblemente al menos un 95%, tal como al menos un 96%, por ejemplo al menos un 97%, incluso más preferiblemente al menos un 98%, tal como al menos un 99%. Preferiblemente, la secuencia de nucleótidos comprende la celobiohidrolasa I que codifica la parte de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n.º 0582. En una forma de realización aún más preferida, la secuencia de nucleótidos consiste en la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n.º 0582.  
50

[0069] La modificación de una secuencia de nucleótidos que codifica un polipéptido de la presente invención puede ser necesaria para la síntesis de un polipéptido, que comprende una secuencia de aminoácidos que tiene al menos una sustitución, delección y/o inserción al compararla con una secuencia de aminoácidos de SEC ID n.º: 8. Estas variantes artificiales pueden diferir en alguna forma creada genéticamente del polipéptido aislado de su fuente nativa, por ejemplo, variantes que difieren en la actividad específica, termoestabilidad, pH óptimo o similares.  
55

[0070] Será aparente para los expertos en la técnica que tales modificaciones pueden ser hechas fuera de las regiones críticas a la función de la molécula y todavía suponer un polipéptido activo. Residuos de aminoácidos esenciales para la actividad del polipéptido codificado por la secuencia de nucleótidos de la invención, y por lo tanto preferiblemente no sujetos a modificación, tal como sustitución, se pueden identificar según procedimientos conocidos en la técnica, tal como  
60

mutagénesis dirigida o mutagénesis de escaneado de alanina (véase, por ejemplo, Cunningham and Wells, 1989, Science 244: 1081-1085). En la anterior técnica, mutaciones se introducen en cada residuo positivamente cargado de la molécula, y las moléculas mutantes resultantes se evalúan para actividad de celobiohidrolasa I para identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Sitios de interacción de enzima-sustrato pueden también ser determinados por análisis de la estructura tridimensional como determinado por tales técnicas como análisis de resonancia magnética nuclear, cristalografía o etiquetado de fotoafinidad (véase, por ejemplo, de Vos et al., 1992, Science 255: 306-312 ; Smith et al., 1992, Journal of Molecular Biology 224: 899-904; Wlodaver et al., 1992, FEBS Letters 309: 59-64).

[0071] Por otra parte, una secuencia de nucleótidos que codifica un polipéptido de la presente invención se puede modificar por introducción de sustituciones de nucleótidos que no dan lugar a otra secuencia de aminoácidos del polipéptido codificado por la secuencia de nucleótidos, pero que corresponden al uso de codón del organismo huésped destinado a la producción de la enzima.

[0072] La introducción de una mutación en la secuencia de nucleótidos para intercambiar un nucleótido por otro nucleótido se puede realizar por mutagénesis dirigida que usa cualquiera de los métodos conocidos en la técnica. Particularmente útil es el procedimiento que utiliza un vector de ADN bicatenario superenrollado con un inserto de interés y dos cebadores sintéticos que contienen la mutación deseada. Los cebadores oligonucleótidos, cada uno complementario de las cepas opuestas del vector, se extienden durante la variación cíclica de la temperatura mediante *Pfu* ADN polimerasa. En la incorporación de los cebadores, un plásmido mutado con cortes en bisel es generado. Después del ciclo de temperatura, el producto se trata con *DpnI* que es específico para ADN hemimetilado y metilado para digerir el molde de ADN progenitor y para seleccionar para ADN sintetizado con mutación. Otros procedimientos conocidos en la técnica pueden también ser usados. Para una descripción general de sustitución de nucleótidos, véase, por ejemplo, Ford et al., 1991, Protein Expression and Purification 2: 95-107.

[0073] La presente divulgación también se refiere a un polinucleótido que comprende, preferiblemente que consiste en, una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I, y que hibridiza bajo condiciones de astringencia muy bajas, preferiblemente bajo condiciones de astringencia bajas, más preferiblemente bajo condiciones de astringencia media, más preferiblemente bajo condiciones de astringencia medias altas, incluso más preferiblemente bajo condiciones de astringencia altas, y de la forma más preferible bajo condiciones de astringencia altísimas con una sonda polinucleótida seleccionada del grupo consistiendo en

1. (i) la cadena complementaria de los nucleótidos seleccionados de:

nucleótidos 1 a 1371 de SEC ID n.º: 7;

2. (ii) la cadena complementaria de los nucleótidos seleccionados de:

nucleótidos 1 a 500 de SEC ID n.º: 77 y

3. (iii) la cadena complementaria de los nucleótidos seleccionados de:

nucleótidos 1 a 200 de SEC ID n.º: 7.

[0074] Como será entendido, detalles y particulares en lo que se refiere a hibridación de las secuencias de nucleótidos serán iguales o análogos a los aspectos de hibridación discutidos en la sección titulada "Polipéptidos con actividad de celobiohidrolasa I" aquí expuesta.

### **Constructos de Acidos Nucleicos**

[0075] La presente divulgación también se refiere a constructos de ácidos nucleicos que comprenden una secuencia de nucleótidos de la presente invención operativamente enlazados a una o más secuencias de control que dirigen la expresión de la secuencia codificante en una célula huésped adecuada bajo condiciones compatibles con las secuencias de control.

[0076] Una secuencia de nucleótidos que codifica un polipéptido de la presente invención se puede manipular en una variedad de vías para proveer a expresión del polipéptido. La manipulación de la secuencia de nucleótidos antes de la inserción su en un vector puede ser deseable o necesaria dependiendo del vector de expresión. Las técnicas para modificar secuencias de nucleótidos utilizando métodos de ADN recombinante se conocen bien en la técnica.

[0077] La secuencia de control puede ser una secuencia del promotor apropiada, una secuencia de nucleótidos que se reconoce por una célula huésped para la expresión de la secuencia de nucleótidos. La secuencia promotora contiene secuencias de control transcripcionales, que median la expresión del polipéptido. El promotor puede ser cualquier secuencia

de nucleótidos que muestre actividad transcripcional en la célula huésped de elección, incluidos promotores mutantes truncados e híbridos, y se pueden obtener de genes que codifican polipéptidos intracelulares o extracelulares ya sean heterólogos u homólogos de la célula huésped.

- 5 [0078] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácidos nucleicos de la presente invención, especialmente en una célula huésped bacteriana, son los promotores obtenidos del operón *lac* de *E. coli*, gen de agarasa de *Streptomyces coelicolor* (*dagA*), gen de levansucrasa de *Bacillus subtilis* (*sacB*), gen de alfa-amilasa de *Bacillus licheniformis* (*amyL*), gen de amilasa maltogénica de *Bacillus stearothermophilus* (*amyM*), gen de alfa-amilasa de *Bacillus amyloliquefaciens* (*amyQ*), gen de penicilinas de *Bacillus licheniformis* (*penP*), genes *xylA* y *xylB* de *Bacillus subtilis*, y gen de beta-lactamasa procariótica (Villa-Kamaroff et al., 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), así como el promotor *tac* (DeBoer et al., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Más promotores son descritos en "Useful proteins from recombinant bacteria" en Scientific American, 1980, 242: 74-94; y en Sambrook et al., 1989, *supra*.
- 10
- 15 [0079] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de ácidos nucleicos de la presente invención en una célula huésped filamentosa fúngica son promotores obtenidos de los genes para TAKA amilasa de *Aspergillus oryzae*, proteinasa aspártica de *Rhizomucor miehei*, alfa-amilasa neutra de *Aspergillus niger*, alfa-amilasa estable en medio ácido de *Aspergillus niger*, glucoamilasa de *Aspergillus niger* o de *Aspergillus awamori* (*glaA*), lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, triosa fosfato isomerasa de *Aspergillus oryzae*, acetamidasa de *Aspergillus nidulans* y proteasa de tipo tripsina de *Fusarium oxysporum* (WO 96/00787), al igual que el promotor NA2-tpi (un híbrido de los promotores de los genes para alfa-amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae*) y promotores mutantes, truncados e híbridos de los mismos.
- 20
- 25 [0080] En un huésped de levadura, promotores útiles se obtienen de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactocinasa de *Saccharomyces cerevisiae* (GAL1), alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH2/GAP) y 3-fosfoglicerato-cinasa de *Saccharomyces cerevisiae*. Otros promotores útiles para células huésped de levadura están descritos por Romanos et al., 1992, Yeast 8: 423-488.
- 30 [0081] La secuencia de control también puede ser una secuencia terminadora de transcripción adecuada, una secuencia reconocida por una célula huésped para terminar la transcripción. La secuencia terminadora está operativamente enlazada al extremo 3' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier terminador que sea funcional en la célula huésped de elección se puede utilizar en la presente invención.
- 35 [0082] Terminadores preferidos para células huésped fúngicas filamentosas se obtienen de los genes para amilasa TAKA de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, sintasa de antranilato de *Aspergillus nidulans*, alfa-glucosidasa de *Aspergillus niger* y proteasa de tipo tripsina de *Fusarium oxysporum*.
- 40 [0083] Terminadores preferidos para células huésped de levadura se obtienen de los genes para enolasa de *Saccharomyces cerevisiae*, citocromo C de *Saccharomyces cerevisiae* (CYC1), y de gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae*. Otros terminadores útiles para células huésped de levadura son descritas por Romanos et al., 1992, *supra*.
- 45 [0084] La secuencia de control puede también ser una secuencia líder adecuada, una región no traducida de un ARNm que es importante para traducción por la célula huésped. La secuencia líder está operativamente enlazada al extremo 5' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier secuencia líder que es funcional en la célula huésped de elección se puede utilizar en la presente invención.
- 50 [0085] Las secuencias líder preferidas para células huésped fúngicas filamentosas son obtenidas de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.
- 55 [0086] Secuencias líderes adecuadas para células huésped de levadura son obtenidas de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), quinasa de 3-fosfoglicerato de *Saccharomyces cerevisiae*, factor-alfa de *Saccharomyces cerevisiae*, y alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa de *Saccharomices cerevisiae* (ADH2/GAP).
- 60 [0087] La secuencia de control puede también ser una secuencia de poliadenilación, una secuencia operativamente enlazada al extremo 3' de la secuencia de nucleótidos y que, cuando se transcribe, es reconocida por la célula huésped como una señal para añadir residuos de poliadenosina al ARNm transcrito. Cualquier secuencia de poliadenilación que sea funcional en la célula huésped de elección se puede utilizar en la presente invención.
- [0088] Secuencias de poliadenilación preferidas para células huésped filamentosas fúngicas se obtienen de los genes para

TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, sintasa de antranilato de *Aspergillus nidulans*, proteasa de tipo tripsina de *Fusarium oxysporum*, y alfa-glicosidasa de *Aspergillus niger*.

5 [0089] Secuencias de poliadenilación útiles para células huésped de levadura están descritas por Guo and Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.

10 [0090] La secuencia de control puede también ser una región de codificación de péptido de señal que codifica para una secuencia de aminoácidos enlazada al amino terminal de un polipéptido y que dirige el polipéptido codificado en la vía secretora de la célula. El extremo 5' de la secuencia codificante de la secuencia de nucleótidos puede intrínsecamente contener una región de codificación de péptido señal naturalmente enlazado en el marco de lectura de traducción con el segmento de la región de codificación que codifica el polipéptido segregado. Alternativamente, el extremo 5' de la secuencia codificante puede contener una región de codificación de péptido señal que es extraña a la secuencia codificante. La región de codificación de péptido señal extraña puede ser requerida donde la secuencia codificante no contenga naturalmente una región de codificación de péptido señal. Alternativamente, la región de codificación del péptido señal foráneo puede simplemente reemplazar la región de codificación de péptido señal natural para mejorar la secreción del polipéptido. No obstante, cualquier región de codificación de péptido señal que dirige el polipéptido expresado en la vía secretora de una célula huésped de elección se puede utilizar en la presente invención.

20 [0091] Regiones de codificación de péptido señal eficaces para células huésped bacterianas son las regiones de codificación de péptido señal obtenidas de los genes para amilasa maltogénica de *Bacillus* NCIB 11837, alfa-amilasa de *Bacillus stearothermophilus*, subtilisina de *Bacillus licheniformis*, beta-lactamasa de *Bacillus licheniformis*, proteasas neutras de *Bacillus stearothermophilus* (nprT, nprS, nprM), y *Bacillus subtilis prsA*. Más péptidos señal están descritos por Simonen and Palva, 1993, Microbiological Reviews 57: 109-137.

25 [0092] Regiones de codificación de péptido señal eficaces para células huésped filamentosas fúngicas son las regiones de codificación de péptido señal obtenidas de los genes para TAKA amilasa de *Aspergillus oryzae*, amilasa neutra de *Aspergillus niger*, glucoamilasa de *Aspergillus niger*, proteinasa aspártica de *Rhizomucor miehei*, celulasa de *Humicola insolens* y lipasa de *Humicola lanuginosa*.

30 [0093] Péptidos señal útiles para células huésped de levadura se obtienen de los genes para factor alfa de *Saccharomyces cerevisiae* e invertasa de *Saccharomyces cerevisiae*. Otras regiones de codificación de péptido señal útil son descritas por Romanos et al., 1992, *supra*

35 [0094] La secuencia de control también puede ser un propéptido que codifica una región que codifica la secuencia de aminoácidos situada en el extremo amino terminal de un polipéptido. El polipéptido resultante es conocido como una proenzima o propolipéptido (o un zimógeno en algunos casos). Un propolipéptido está generalmente inactivo y se puede convertir a un polipéptido maduro activo por escisión autocatalítica o catalítica del propéptido del propolipéptido. La región codificante de propéptido se puede obtener de los genes para proteasa alcalina de *Bacillus subtilis* (*aprE*), proteasa neutra de *Bacillus subtilis* (*nprT*), factor-alfa de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei*, y lacasa de *Myceliophthora thermophila* (WO 95/33836).

40 [0095] Donde ambos péptidos señal y regiones de propéptido están presentes en el extremo amino-terminal de un polipéptido, la región del propéptido está situada junto al amino terminal de un polipéptido y la región de péptido señal está situada junto al amino terminal de la región de propéptido.

45 [0096] También puede ser deseable añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido en relación al crecimiento de la célula huésped. Ejemplos de sistemas reguladores son los que causan que la expresión del gen sea activada o desactivada en respuesta a un estímulo físico o químico, incluida la presencia de un compuesto regulador. Sistemas reguladores en sistemas procarióticos incluyen sistemas de operador *lac*, *tac* y *trp*. En la levadura, el sistema ADH2 o el sistema GAL1 pueden ser utilizados. En hongos filamentosos, el promotor de alfa-amilasa TAKA, promotor de glucoamilasa de *Aspergillus niger* y promotor de glucoamilasa de *Aspergillus oryzae* se pueden utilizar como secuencias reguladoras. Otros ejemplos de secuencias reguladoras son los que permiten la amplificación génica. En sistemas eucarióticos, estos incluyen el gen de dihidrofolato-reductasa que se amplifica en presencia de metotrexato y los genes de metalotioneína que se amplifican con metales pesados. En estos casos, la secuencia de nucleótidos que codifica el polipéptido estaría operativamente enlazada con la secuencia regulador.

### Vectores de Expresión

60 [0097] La presente invención también se refiere a vectores de expresión recombinantes que incluyen el constructo de ácidos nucleicos de la invención. Las varias secuencias de nucleótidos y de control anteriormente descritas pueden ser juntadas para producir un vector de expresión recombinante que puede incluir uno o más sitios de restricción convenientes para

5 permitir la inserción o sustitución de la secuencia de nucleótidos que codifica el polipéptido en tales sitios. Alternativamente, la secuencia de nucleótidos de la presente invención se puede expresar por inserción de la secuencia de nucleótidos o un constructo de ácidos nucleicos que comprende la secuencia en un vector apropiado para la expresión. En la creación del vector de expresión, la secuencia codificante se localiza en el vector de modo que la secuencia codificante está operativamente enlazada con las secuencias de control apropiadas para la expresión.

10 [0098] El vector de expresión recombinante puede ser cualquier vector (p. ej., un plásmido o virus) que pueda ser convenientemente sometido a procedimientos de ADN recombinante y pueda provocar la expresión de la secuencia de nucleótidos. La elección del vector típicamente dependerá de la compatibilidad del vector con la célula huésped en la que el vector va a ser introducido. Los vectores pueden ser plásmidos lineales o circulares cerrados.

15 [0099] El vector puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, la replicación del cual es independiente de la replicación cromosómica, por ejemplo, un plásmido, un elemento extracromosómico, un minicromosoma, o un cromosoma artificial.

20 [0100] El vector puede contener cualquier medio para asegurar la autorreplicación. Alternativamente, el vector puede ser uno que, cuando es introducido en la célula huésped, se integra en el genoma y es replicado con el(los) cromosoma(s) en que ha sido integrado. Además, un único vector o plásmido o dos o más vectores o plásmidos que juntos contienen el ADN total que debe ser introducido en el genoma de la célula huésped o un transposón pueden ser utilizados.

[0101] Los vectores de la presente invención preferiblemente contienen uno o más marcadores seleccionables que permiten la selección fácil de células transformadas. Un marcador seleccionable es un gen cuyo producto proporciona resistencia biocida o vírica, resistencia a metales pesados, prototrofia para auxótrofos y similares.

25 [0102] Ejemplos de marcadores bacterianos seleccionables son los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia antibiótica tales como resistencia a ampicilina, canamicina, cloranfenicol o de tetraciclina. Marcadores adecuados para células huésped de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1, y URA3. Marcadores seleccionables para uso en una célula huésped fúngica filamentosa incluyen, pero de forma no limitativa, *amdS* (acetamidasa), *argB* (ornitina-carbamoiltransferasa), *bar* (fosfonitricina acetiltransferasa), *hygB* (higromicina fosfotransferasa), *niaD* (nitrato-reductasa), *pyrG* (orotidina-5'-fosfato-d Descarboxilasa), *sC* (adeniltransferasa de sulfato), *trpC* (antranilato sintasa), al igual que equivalentes han de lo mismo.

35 [0103] Preferidos para uso en una célula de *Aspergillus* son los genes de *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen *bar* de *Streptomyces hygrosopicus*.

[0104] Los vectores de la presente invención preferiblemente contienen un(os) elemento(s) que permite(n) integración estable del vector en el genoma de la célula huésped o replicación autónoma del vector en la célula independiente del genoma.

40 [0105] Para la integración en el genoma de la célula huésped, el vector puede contar con la secuencia de nucleótidos que codifica el polipéptido o cualquier otro elemento del vector para integración estable del vector en el genoma por recombinación homóloga o no homóloga. Alternativamente, el vector puede contener secuencias de nucleótidos adicionales para dirigir la integración por recombinación homóloga en el genoma de la célula huésped. Las secuencias de nucleótidos adicionales permiten que el vector sea integrado en el genoma de la célula huésped en una(s) ubicación(es) precisa del cromosoma(s). Para aumentar la probabilidad de integración en una ubicación precisa, los elementos integracionales deberían preferiblemente contener un número suficiente de nucleótidos, tal como 100 a 1500 pares de bases, preferiblemente 400 a 1500 pares de bases y, de la forma más preferible 800 a 1500 pares de bases, que son altamente homólogos a la secuencia objetivo correspondiente para mejorar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga a la secuencia objetivo del genoma de la célula huésped. Además, los elementos integracionales pueden ser secuencias de nucleótidos codificantes o no codificantes. Por otro lado, el vector se puede integrar en el genoma de la célula huésped por recombinación no homóloga.

55 [0106] Para replicación autónoma, el vector puede comprender además un origen de replicación permitiendo al vector replicarse de manera autónoma en la célula huésped en cuestión. Ejemplos de orígenes bacterianos de replicación son los orígenes de replicación de plásmidos pBR322; pUC19; pACYC177 y pACYC184 que permiten replicación en *E. coli*, y pUB110; pE194; pTA1060, y pAMB1 que permiten replicación en *Bacillus*. Ejemplos de orígenes de replicación para uso en una célula huésped de levadura son el origen de 2 micras de replicación, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6. El origen de replicación puede ser uno que tiene una mutación que hace sensible su temperatura de funcionamiento en la célula huésped (véase, por ejemplo, Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75: 1433).

[0107] Más de una copia de una secuencia de nucleótidos de la presente invención se puede insertar en la célula huésped para aumentar la producción del producto génico. Un aumento en el número de copias de la secuencia de nucleótidos se puede obtener por integración de al menos una copia adicional de la secuencia en el genoma de la célula huésped o incluyendo un gen marcador seleccionable amplificable con la secuencia de nucleótidos donde células que contienen copias amplificadas del gen marcador seleccionable, y así copias adicionales de la secuencia de nucleótidos se puede seleccionar para por cultivo de las células en presencia del agente apropiado seleccionable.

[0108] Los procedimientos usados para enlazar los elementos anteriormente descritos para construir los vectores de expresión recombinantes de la presente invención son bien conocidos por un experto en la técnica (véase, por ejemplo, Sambrook et al., 1989, *supra*).

### Células Huésped

[0109] La presente invención también se refiere a recombinar una célula huésped que incluye el constructo de ácidos nucleicos de la descripción, que es ventajosamente usado en la producción recombinante de los polipéptidos. Un vector que comprende una secuencia de nucleótidos de la presente invención se introduce en una célula huésped de modo que el vector es mantenido como un integrante cromosómico o como un vector extracromosómico que se autoreplica como se describe anteriormente.

[0110] La célula huésped puede ser un microorganismo unicelular, por ejemplo, un microorganismo procarionta o uno no unicelular, por ejemplo, un eucariota.

[0111] Células útiles unicelulares son células bacterianas tales como bacterias gram positivas incluidas, pero no limitadas a, una célula de *Bacillus*, por ejemplo, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, y *Bacillus thuringiensis*; o una célula de *Streptomyces*, por ejemplo, *Streptomyces lividans* o *Streptomyces murinus* o bacterias gram negativas tales como *E. coli* y *Pseudomonas* sp. En una forma de realización preferida, la célula huésped bacteriana es una célula de *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus*, o *Bacillus subtilis*. En otra forma de realización preferida, la célula de *Bacillus* es un *Bacillus* alcalofílico.

[0112] La introducción de un vector en una célula huésped bacteriana puede, por ejemplo, ser efectuada por transformación de protoplasto (ver, por ejemplo, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), usando células competentes (ver, por ejemplo, Young and Spizizin, 1961, *Journal of Bacteriology* 81: 823-829, o Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), electroporación (ver, por ejemplo, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), o conjugación (ver, por ejemplo, Koehler and Thome, 1987, *Journal of Bacteriology* 169: 5771-5278).

[0113] La célula huésped puede ser una eucariota, tal como un mamífero, insecto, planta, o célula fúngica.

[0114] En una forma de realización preferida, la célula huésped es una célula fúngica. "Fúngica", como se utiliza en este caso, incluye filo Ascomycota, Basidiomycota, Chytridiomycota y Zygomycota (como definido por Hawksworth et al., In, Ainsworth and Bisby's Dictionary of The Fungi, 8ª edición, 1995, CAB International, University Press, Cambridge, UK) así como la Oomycota (como citado en Hawksworth et al., 1995, *supra*, página 171) y todos los hongos mitospóricos (Hawksworth et al., 1995, *supra*).

[0115] En una forma de realización más preferida, la célula huésped fúngica es una célula de levadura. "Levadura" como se utiliza en este caso incluye levadura ascoesporógena (Endomycetales), levadura basidioesporogénea, y levadura de los hongos imperfectos (Blastomycetos). Dado que la clasificación de levadura puede cambiar en el futuro, para los fines de esta invención, la levadura debe ser definida como se describe en *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

[0116] En una forma de realización aún más preferida, la célula huésped de levadura es una célula de *Candida*, *Aschbyii*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*.

[0117] En una forma de realización más preferida, la célula huésped de levadura es una célula de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* o *Saccharomyces oviformis*. En otra forma de realización más preferida, la célula huésped de levadura es una célula de *Kluyveromyces lactis*. En otra forma de realización más preferida, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.

[0118] En otra forma de realización más preferida, la célula huésped fúngica es una célula micótica filamentosa. "Hongos

filamentosos" incluyen todas las formas filamentosas de la subdivisión Eumycota y Oomycota (como definido por Hawksworth et al., 1995, *supra*). Los hongos filamentosos se caracterizan por una pared micelial compuesta por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. Crecimiento vegetativo es por alargamiento hifal y catabolismo de carbono es estrictamente aeróbico. En cambio, el crecimiento vegetativo por levaduras tal como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y catabolismo de carbono pueden ser fermentativos.

[0119] En una forma de realización aún más preferida, la célula huésped filamentosa fúngica es una célula de una especie de, pero no limitado a, *Acremonium*, *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolipocladium* o *Trichoderma*.

[0120] En una forma de realización más preferida, la célula huésped filamentosa fúngica es una célula de *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otra forma de realización más preferida, la célula huésped filamentosa fúngica es una célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, o *Fusarium venenatum*. En una forma de realización incluso más preferida, la célula madre filamentosa fúngica es una célula de *Fusarium venenatum* (Nirenberg sp. nov.). En otra forma de realización más preferida, la célula huésped filamentosa fúngica es una célula de *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

[0121] Células fúngicas se pueden transformar por un proceso que implica la formación de protoplasto, transformación de los protoplastos y regeneración de la pared celular de una manera conocida *per se*. Procedimientos adecuados para transformación de células huésped *Aspergillus* son descritas en EP 238 023 y Proceedings of the National Academy of Sciences USA 81: 1470-1474. Métodos adecuados para transformar especies de *Fusarium* son descritas por Malardier et al., 1989, Gene 78: 147-156 and WO 96/00787. La levadura se puede transformar utilizando los procedimientos descritos por Becker and Guarente, In Abelson, J.N. and Simon, M.I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volume 194, pp 182-187, Academic Press, Inc., New York ; Ito et al., 1983, Journal of Bacteriology 153: 163 ; and Hinnen et al., 1978, Proceedings of the National Academy of Sciences USA 75: 1920.

### Métodos de Producción

[0122] La presente divulgación también se refiere a métodos para producir un polipéptido de la presente invención que comprende (a) cultivo una cepa, que en su forma tipo salvaje es capaz de producir el polipéptido; y (b) recuperación del polipéptido. Preferiblemente, la cepa es seleccionada del grupo consistiendo en *Acremonium*, *Scytalidium*, *Thermoascus*, *Thielavia*, *Verticillium*, *Neotermes*, *Melanocarpus*, *Poitrasia*, *Coprinus*, *Trichothecium*, *Humicola*, *Cladorrhinum*, *Diplodia*, *Myceliophthora*, *Rhizomucor*, *Meripilus*, *Exidia*, *Xylaria*, *Trichophaea*, *Chaetomium*, *Chaetomidium*, *Sporotrichum*, *Thielavia*, *Aspergillus*, *Scopulariopsis*, *Fusarium*, *Pseudoplectania*, y *Phytophthora*; más preferiblemente del grupo que consiste en *Acremonium thermophilum*, *Chaetomium thermophilum*, *Scytalidium thermophilum*, *Thermoascus aurantiacus*, *Thielavia australiensis*, *Verticillium tenerum*, *Neotermes castaneus*, *Melanocarpus albomyces*, *Poitrasia circinans*, *Coprinus cinereus*, *Trichothecium roseum*, *Humicola nigrescens*, *Cladorrhinum foecundissimum*, *Diplodia gossypina*, *Myceliophthora thermophila*, *Rhizomucor pusillus*, *Meripilus giganteus*, *Exidia glandulosa*, *Xylaria hypoxylon*, *Trichophaea saccata*, *Chaetomidium pingtungium*, *Myceliophthora thermophila*, *Myceliophthora hinnulea*, *Sporotrichum prunosum*, *Thielavia cf. microspora*, *Pseudoplectania nigrella*, y *Phytophthora infestans*.

[0123] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que comprende (a) cultivo una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

[0124] La presente invención también se refiere a métodos para la producción in situ de un polipéptido de la presente invención que comprende (a) cultivo una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) puesta en contacto del polipéptido con un sustrato deseado, tal como un sustrato de celulosa, sin recuperación previa del polipéptido. El término "producción in situ" significa que el polipéptido se produce directamente en la localización en la que está destinado a ser usado, tal como en un proceso de fermentación para la producción de etanol.

[0125] En los métodos de producción de la presente invención, las células se cultivan en un medio nutritivo adecuado para la producción de los polipéptidos usando métodos conocidos en la técnica. Por ejemplo, la célula se puede cultivar por cultivo en matraz vibrante, fermentación a pequeña o gran escala (incluida fermentación continua, de lote, de lote alimentado o de estado sólido) en el laboratorio o fermentadores industriales realizados en un medio adecuado y bajo condiciones que permiten al polipéptido ser expresado y/o aislado. El cultivo se desarrolla en un medio nutritivo adecuado

que incluye fuentes de nitrógeno y carbono y sales inorgánicas, utilizando procedimientos conocidos en la técnica. Medios adecuados están disponibles de proveedores comerciales o se pueden preparar según las composiciones publicadas (p. ej., en catálogos de la American Type Culture Collection). Si el polipéptido se segrega en el medio nutritivo, el polipéptido se puede recuperar directamente del medio. Si el polipéptido no es segregado, este se puede recuperar de lisatos celulares.

[0126] Los polipéptidos pueden ser detectados usando métodos conocidos en la técnica que son específicos para los polipéptidos. Estos métodos de detección pueden incluir el uso de anticuerpos específicos, formación de un producto enzimático o desaparición de un sustrato enzimático. Por ejemplo, un ensayo enzimático se puede utilizar para determinar la actividad del polipéptido como se describe en este caso.

[0127] El polipéptido resultante se puede recuperar por métodos conocidos en la técnica. Por ejemplo, el polipéptido puede ser recuperado del medio nutritivo por procedimientos convencionales, que incluyen, pero no están limitados a, centrifugado, filtración, extracción, secado por pulverización, evaporación o precipitación.

[0128] Los polipéptidos de la presente invención se pueden purificar por una variedad de procedimientos conocidos en la técnica, incluyendo, pero no limitados a, cromatografía (p. ej., de intercambio iónico, de afinidad, hidrofóbica, de cromatofoco y de exclusión de tamaño), procedimientos electroforéticos (p. ej., isoelectroenfoque preparativo), solubilidad diferencial (p. ej., precipitación de sulfato amónico), SDS-PAGE o extracción (véase, por ejemplo, Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

## Plantas

[0129] La presente divulgación también se refiere a una planta transgénica, parte de planta o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido se puede recuperar de la planta o parte de planta. Alternativamente, la planta o parte de planta con el polipéptido recombinante se puede utilizar como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorar el valor nutritivo, palatabilidad y propiedades reológicas, o para destruir un factor antinutritivo.

[0130] La planta transgénica puede ser dicotiledónea (un dicto) o monocotiledónea (un monocot). Ejemplos de plantas monocotiledóneas son hierbas, tales como césped de pradera (*Poa pratense*, *Poa*), hierba forrajera tal como *Festuca*, *Lolium*, césped templado, tal como *Agrostis* y cereales, por ejemplo, trigo, avena, centeno, cebada, arroz, sorgo, mijo y maíz.

[0131] Ejemplos de plantas dicot son tabaco, altramuces, patata, remolacha azucarera, leguminosas, tales como guisante, judía y semilla de soja y plantas de crucífero (de familia *Brassicaceae*) tales como coliflor, colza, canola y el organismo de modelo cercanamente relacionado *Arabidopsis thaliana*.

[0132] Ejemplos de partes de planta son tallos, callos, hojas, raíces, frutas, semillas, y tubérculos. También tejidos de planta específicos, tales como cloroplasto, apoplasto, mitocondria, vacuola, peroxisomas y citoplasma se consideran una parte de planta. Además, cualquier célula vegetal, sea cual sea el origen del tejido, se considera una parte de planta.

[0133] También están incluidas dentro del campo de la presente invención la progenie (clonal o semilla) de tales plantas, partes de planta y células vegetales.

[0134] La planta transgénica o célula vegetal que expresa un polipéptido de la presente invención se puede construir conforme a métodos conocido en la técnica. Brevemente, la planta o célula vegetal se construye por incorporación de uno o más constructos de expresión que codifican un polipéptido de la presente invención en el genoma de huésped de planta y propagación de la resultante planta modificada o célula vegetal en una planta transgénica o célula vegetal.

[0135] Convenientemente, el constructo de expresión es un constructo de ácidos nucleicos que comprende una secuencia de nucleótidos que codifica un polipéptido de la presente invención operativamente enlazado a secuencias reguladoras apropiadas requerido para la expresión de la secuencia de nucleótidos en la planta o parte de planta elegida. Además, el constructo de expresión puede incluir una etiqueta seleccionable útil para identificar células huésped en las que el constructo de expresión ha sido integrado y las secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (esto depende del método de introducción de ADN que va a ser usado).

[0136] La elección de secuencias reguladoras, tales como secuencias promotoras y terminadoras y, opcionalmente, secuencias de señal o de tránsito es determinada, por ejemplo, basándose en cuándo, dónde y cómo el polipéptido se desea que sea expresado. Por ejemplo, la expresión del gen que codifica un polipéptido de la presente invención puede ser constitutiva o inducible, o puede ser desarrollable, fase o tejido específica, y el producto genético puede ser previsto para un

tejido específico o parte de planta, tal como semillas u hojas. Secuencias reguladoras están, por ejemplo, descritas por Tague et al., 1988, *Plant Physiology* 86: 506.

5 [0137] Para la expresión constitutiva, el promotor 35S-CaMV se puede utilizar (Franck et al., 1980, *Cell* 21: 285-294). Promotores específicos de un órgano pueden ser, por ejemplo, un promotor de tejidos de depósito de almacenamiento, tal como semillas, tubérculos de patata y frutas (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), o de tejidos sumidero metabólicos tales como meristemas (Ito et al., 1994, *Plant Mol. Biol.* 24: 863-878), un promotor específico de semilla tal como la glutelina, prolamina, globulina o promotor de albúmina de arroz (Wu et al., 1998, *Plant and Cell Physiology* 39: 885-889), un promotor de *Vicia faba* de la legumina B4 y el gen de proteína de semilla desconocida de *Vicia faba* (Conrad et al., 1998, *Journal of Plant Physiology* 152: 708-711), un promotor de una proteína del cuerpo oleoso de la semilla (Chen et al., 1998, *Plant and Cell Physiology* 39: 935-941), el promotor de la proteína de almacenamiento napA de *Brassica napus* o cualquier otro promotor específico de semilla conocido en la técnica, por ejemplo, como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de hoja tal como el promotor *rbcS* de arroz o tomate (Kyojuzuka et al., 1993, *Plant Physiology* 102: 991-1000), el promotor genético de metiltransferasa de adenina del virus de chlorella (Mitra and Higgins, 1994, *Plant Molecular Biology* 26: 85-93) o el promotor genético *aldP* de arroz (Kagaya et al., 1995, *Molecular and General Genetics* 248: 668-674) o un promotor inducible por lesión tal como el promotor de patata pin2 (Xu et al., 1993, *Plant Molecular Biology* 22: 573-588).

20 [0138] Un elemento promotor intensificador puede también ser usado para conseguir expresión más alta de la enzima en la planta. Por ejemplo, el elemento promotor intensificador puede ser un intrón que es colocado entre el promotor y la secuencia de nucleótidos que codifica un polipéptido de la presente invención. Por ejemplo, Xu et al., 1993, *supra* revela el uso del primer intrón del gen de actina 1 de arroz para mejorar la expresión.

25 [0139] El gen marcador seleccionable y cualquiera de las otras partes del constructo de expresión se pueden elegir de aquellos disponibles en la técnica.

30 [0140] El constructo de ácidos nucleicos se incorpora en el genoma de la planta según técnicas convencionales conocidas en la técnica, incluyendo transformación mediada por *Agrobacterium*, transformación mediada por virus, microinyección, bombardeo de partícula, transformación biolística y electroporación (Gasser et al., 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto et al., 1989, *Nature* 338: 274).

35 [0141] Actualmente, la transferencia genética por medio de *Agrobacterium tumefaciens* es el método de elección para generar dicotiledóneas transgénicas (para una revisión, véase Hooykas and Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38). No obstante esto puede también ser usado para transformar monocotiledóneas, aunque otros métodos de transformación son generalmente preferidos para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas es bombardeo de partícula (oro microscópico o partículas de tungsteno revestidas con el ADN transformante) de callos embrionarios o embriones en desarrollo (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil et al., 1992, *Bio/Technology* 10: 667-674). Un método alternativo para la transformación de monocotiledóneas se basa en la transformación de protoplasto como se describe por Omirulleh et al., 1993, *Plant Molecular Biology* 21: 415-428.

40 [0142] Después de la transformación, los transformantes que tienen el constructo de expresión incorporado se seleccionan y regeneran en plantas enteras según métodos bien conocidos en la técnica.

45 [0143] La presente divulgación también se refiere a métodos para producir un polipéptido de la presente invención que comprende (a) cultivo de una planta transgénica o una célula vegetal que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

50 [0144] La presente divulgación también se refiere a métodos para la producción in situ de un polipéptido de la presente invención que comprende (a) cultivo de una planta transgénica o una célula vegetal que comprende una secuencia de nucleótidos que codifica un polipéptido con actividad de celobiohidrolasa I de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) poner en contacto el polipéptido con un sustrato deseado, tal como un sustrato celulósico, sin recuperación previa del polipéptido.

## 55 Composiciones

[0145] En otro aspecto ulterior, la presente invención se refiere a composiciones que comprenden un polipéptido de la presente invención.

60 [0146] La composición puede comprender un polipéptido de la invención como el componente enzimático mayor, por

ejemplo, una composición monocomponente. Alternativamente, la composición puede comprender actividades enzimáticas múltiples, tal como una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa, o xilanasas.

[0147] Las composiciones se pueden preparar conforme a métodos conocidos en la técnica y pueden estar en forma de líquido o composición seca. Por ejemplo, la composición de polipéptido puede estar en forma de granulado o un microgranulado. El polipéptido que debe ser incluido en la composición se puede estabilizar conforme a métodos conocidos en la técnica.

[0148] Más abajo se dan ejemplos de usos preferidos de las composiciones de polipéptidos de la invención. La dosificación de la composición de polipéptidos de la invención y otras condiciones bajo las cuales la composición es usada pueden ser determinadas basándose en métodos conocidos en la técnica.

### Composiciones Detergentes

[0149] El polipéptido de la invención se puede adicionar y así convertirse en un componente de una composición de detergente.

[0150] La composición detergente de la invención puede por ejemplo ser formulada como una composición de detergente de lavado de ropa a máquina o a mano que incluye una composición de aditivo de lavandería adecuada para el pretratamiento de tejidos manchados y una composición de suavizante adicionada para el enjuague, o ser formulado como una composición detergente para usar en operaciones de limpieza de superficies duras del hogar generales, o ser formulado para operaciones de lavado de la vajilla a máquina o a mano.

[0151] En un aspecto específico, la invención proporciona un aditivo de detergente que comprende el polipéptido de la invención. El aditivo de detergente al igual que la composición de detergente puede comprender unas o varias otras enzimas tal como una proteasa, una lipasa, una cutinasa, una amilasa, una carbohidrasa, una celulasa, una pectinasa, una mananasa, una arabinasa, una galactanasa, una xilanasas, una oxidasa, por ejemplo, una lacasa, y/o una peroxidasa.

[0152] En general las propiedades de la enzima(s) elegida(s) deberían ser compatibles con el detergente seleccionado, (es decir, pH óptimo, compatibilidad con otros ingredientes no enzimáticos y enzimáticos, etc.), y la(s) enzima(s) debería(n) estar presente(s) en cantidades eficaces.

[0153] Proteasas: proteasas adecuadas incluyen aquellas de origen animal, vegetal o microbiano. Origen microbiano es preferido. Mutantes químicamente modificados o creados genéticamente de proteínas están incluidos. La proteasa puede ser una serina proteasa o una metalo proteasa, preferiblemente una proteasa microbiana alcalina o una proteasa de tipo tripsina. Ejemplos de proteasas alcalinas son subtilisin, especialmente aquellas derivadas de *Bacillus*, por ejemplo, subtilisina Novo, subtilisina Carlsberg, subtilisina 309, subtilisina 147 y subtilisina 168 (descrita en WO 89/06279). Ejemplos de proteasas de tipo tripsina son tripsina (p. ej. de origen bovino o porcino) y la Proteasa de *Fusarium* descrita en WO 89/06270 y WO 94/25583.

[0154] Ejemplos de proteasas útiles son las variantes descritas en WO 92/19729, WO 98/20115, WO 98/20116, y WO 98/34946, especialmente las variantes con sustituciones en una o más de las siguientes posiciones: 27, 36, 57, 76, 87, 97, 101, 104, 120, 123, 167, 170, 194, 206, 218, 222, 224, 235 y 274.

Lipasas: lipasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteínas están incluidos. Ejemplos de lipasas útiles incluyen lipasas de *Humicola* (sinónimo *Thermomyces*), por ejemplo de *H. lanuginosa* (*T. Lanuginosus*) como descrito en EP 258 068 y EP 305 216 o de *H. insolens* como descrito en WO 96/13580, una lipasa de *Pseudomonas*, por ejemplo de *P. alcaligenes* o *P. pseudoalcaligenes* (EP 218 272), *P. cepacia* (EP 331 376), *P. stutzeri* (GB 1,372,034), *P. fluorescens*, cepas de *Pseudomonas* sp. SD 705 (WO 95/06720 y WO 96/27002), *P. wisconsinensis* (WO 96/12012), una lipasa de *Bacillus*, por ejemplo de *B. subtilis* (Dartois et al. (1993), Biochemica et Biophysica Acta, 1131, 253-360), *B. stearothermophilus* (JP 64/744992) o *B. pumilus* (WO 91/16422).

[0155] Otros ejemplos son variantes de lipasa tales como los descritos en WO 92/05249, WO 94/01541, EP 407 225, EP 260 105, WO 95/35381, WO 96/00292, WO 95/30744, WO 94/25578, WO 95/14783, WO 95/22615, WO 97/04079 y WO 97/07202.

Amilasas: amilasas adecuadas (alfa y/o beta) incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteínas están incluidos. Amilasas incluyen, por ejemplo, alfa-amilasas obtenidas de *Bacillus*, por ejemplo una cepa especial de *B. licheniformis*, descrita con más detalle en GB 1,296,839.

[0156] Ejemplos de amilasas útiles son las variantes descritas en WO 94/02597, WO 94/18314, WO 96/23873, y WO 97/43424, especialmente las variantes con sustituciones en una o más de las siguientes posiciones: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408, y 444.

5 Celulasas: celulasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteínas están incluidos. Celulasas adecuadas incluyen celulasas del género *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, por ejemplo, las celulasas fúngicas producidas de *Humicola insolens*, *Myceliophthora thermophila* y *Fusarium oxysporum* descritas en US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 y WO 89/09259.

10 [0157] Celulasas especialmente adecuadas son las celulasas neutras o alcalinas con beneficios de cuidado de color. Ejemplos de tales celulasas son las celulasas descritas en EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasa tales como las descritas en WO 94/07998, EP 0 531 315, US 5,457,046, US 5,686,593, US 5,763,254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

15 Peroxidasas/oxidadas: peroxidasas/oxidadas adecuadas incluyen aquellas de origen vegetal, fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. Ejemplos de peroxidasas útiles incluyen peroxidasas de *Coprinus*, por ejemplo de *C. cinereus*, y variantes de las mismas como aquellas descritas en WO 93/24618, WO 95/10602, y WO 98/15257.

20 [0158] La(s) enzima(s) detergente(s) se puede(n) incluir en una composición de detergente añadiendo aditivos separados que contienen una o más enzimas, o por adición de un aditivo combinado comprendiendo todas estas enzimas. Un aditivo de detergente de la invención, es decir un aditivo separado o un aditivo combinado, se puede formular por ejemplo como un granulado, un líquido, una suspensión etc. Formulaciones de aditivo de detergente preferidas son granulados, en particular granulados no polvorientos, líquidos, en particular líquidos estabilizados, o suspensiones.

25 [0159] Los granulados no polvorientos pueden ser producidos, por ejemplo, como descritos en US 4,106,991 y 4,661,452 y pueden opcionalmente ser revestidos por métodos conocidos en la técnica. Ejemplos de materiales de recubrimiento ceroso son productos de poli(etileno óxido) (polietilenglicol; PEG) con pesos molares medios de 1000 a 20 000, nonilfenoles etoxilados con de 16 a 50 unidades de óxido de etileno; alcoholes grasos etoxilados en los que el alcohol contiene de 12 a 20 átomos de carbono y en el que hay de 15 a 80 unidades de óxido de etileno; alcoholes grasos; ácidos grasos; y mono- y di- y triglicéridos de ácidos grasos. Ejemplos de materiales de recubrimiento que forman películas adecuadas para aplicaciones por técnicas de lecho fluidificado se dan en GB 1483591. Las preparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol tal como propilenglicol, un azúcar o alcohol de azúcar, ácido láctico o ácido bórico según métodos establecidos. Enzimas protegidas se pueden preparar según el método descrito en EP 238,216.

30 [0160] La composición detergente de la invención puede estar en cualquier forma conveniente, por ejemplo, una barra, una pastilla, un polvo, un gránulo, una pasta o un líquido. Un detergente líquido puede ser acuoso, típicamente con hasta un 70 % agua y un 0-30 % solvente orgánico o no acuoso.

35 [0161] La composición detergente comprende uno o más tensioactivos, que pueden ser no iónicos, incluyendo semipolar y/o aniónico y/o zwitteriónico y/o catiónico. Los tensioactivos están típicamente presentes a un nivel de 0,1 % a 60% en peso.

40 [0162] Cuando se incluye aquí el detergente normalmente contendrá de aproximadamente un 1% a aproximadamente un 40% de un surfactante aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, sulfato de alquilo (sulfato de alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, alfa-sulfo éster metílico de ácido graso, ácido alquil- o alquenilsuccínico o jabón.

45 [0163] Cuando se incluye aquí el detergente normalmente contendrá de aproximadamente un 0,2% a aproximadamente un 40% de un surfactante no iónico tal como etoxilato de alcohol, etoxilato de nonilfenol, alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, amida de ácido graso de alquilo polihidroxilado, o derivados N-alquilo N-acilo de glucosamina ("glucamidas").

50 [0164] El detergente puede contener un 0-65 % de un constructor de detergente o agente complejante tal como zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilendiaminatetraacético, ácido dietilentriaminopentaacético, ácido alquil- o alquenilsuccínico, silicatos solubles o silicatos estratificados (p. ej. SKS-6 de Hoechst).

55 [0165] El detergente puede comprender uno o más polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli(etilenglicol), alcohol polivinílico, poli(vinilpiridina-N-óxido), poli(vinilimidazol), policarboxilatos tales como poliácridatos, copolímeros de ácido maleico/acrílico y copolímeros de ácido acrílico/lauril metacrilato.

60

[0166] El detergente puede contener un sistema blanqueante que puede comprender una fuente de H<sub>2</sub>O<sub>2</sub> tal como perborato o percarbonato, que se puede combinar con un activador blanqueante de formación de perácido tal como tetraacetililenodiamina o nonanoiloxibencenosulfonato. Alternativamente, el sistema blanqueante puede comprender peroxiácidos de por ejemplo amida, imida, o tipo sulfona.

[0167] La(s) enzima(s) de la composición de detergente de la invención se puede(n) estabilizar utilizando agentes estabilizantes convencionales, por ejemplo, un poliol tal como propilenglicol o glicerol, un azúcar o alcohol de azúcar, ácido láctico, ácido bórico, o un derivado de ácido bórico, por ejemplo, un éster de borato aromático, o un derivado de ácido fenilborónico tal como ácido 4-formilfenil borónico, y la composición se puede formular como se describe en por ejemplo WO 92/19709 y WO 92/19708.

[0168] El detergente puede también contener otros ingredientes de detergentes convencionales tales como por ejemplo acondicionadores de tejidos, incluidas arcillas, potenciadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de suspensión de suciedad, agentes de reposición antisuciedad, colorantes, bactericidas, blanqueadores ópticos, hidrotropos, inhibidores de decoloración o perfumes.

[0169] Está actualmente contemplado que en las composiciones detergentes cualquier enzima, en particular el polipéptido de la invención, se puede adicionar en una cantidad correspondiente a 0,01-100 mg de proteína enzimática por litro de solución de lavado, preferiblemente 0,05-5 mg de proteína enzimática por litro de solución de lavado, en particular 0,1-1 mg de proteína enzimática por litro de solución de lavado.

[0170] El polipéptido de la invención puede adicionalmente ser incorporado en las formulaciones detergentes descritas en WO 97/07202.

#### 25 **Recombinación de ADN (redistribución)**

[0171] Las secuencias de nucleótidos de SEC ID n.º:7 se pueden utilizar en un proceso de recombinación de ADN (o redistribución). Las nuevas secuencias polinucleótidas obtenidas en tal proceso pueden codificar nuevos polipéptidos con actividad de celobiasa con propiedades mejoradas, tales como estabilidad mejorada (estabilidad de almacenamiento, termoestabilidad), actividad específica mejorada, pH óptimo mejorado, y/o tolerancia mejorada hacia compuestos específicos.

[0172] Redistribución entre dos o más polinucleótidos de entrada homólogos (polinucleótidos de punto de inicio) implica fragmentación de los polinucleótidos y recombinación de los fragmentos para obtener polinucleótidos de salida (es decir, polinucleótidos que han sido sometidos a un ciclo de redistribución) donde un número de fragmentos de nucleótidos se cambian en comparación con los polinucleótidos de entrada.

[0173] Recombinación o redistribución de ADN puede ser un proceso (parcialmente) aleatorio en el que una genoteca de genes quiméricos se genera de dos o más genes de inicio. Varios formatos conocidos pueden utilizarse para efectuar esta redistribución o proceso de recombinación.

[0174] El proceso puede implicar la fragmentación aleatoria del ADN progenitor seguida del reensamblaje por PCR para nuevos genes en toda su longitud, por ejemplo como presentado en US5605793, US5811238, US5830721, US6117679. La recombinación in vitro de genes puede ser realizada, por ejemplo como se describe en US6159687, WO98/41623, US6159688, US5965408, US6153510. El proceso de recombinación puede ocurrir in vivo en una célula viva, por ejemplo como se describe en WO 97/07205 y WO 98/28416.

[0175] El ADN progenitor se puede fragmentar por tratamiento con ADNasa I o por digeridos de endonucleasa de restricción como descrito por Kikuchi et al (2000a, Gene 236:159-167). La redistribución de dos progenitores puede realizarse por redistribución del ADN progenitor monocatenario de los dos progenitores como descrito en Kikuchi et al (2000b, Gene 243:133-137).

[0176] Un método particular de redistribución es seguir los métodos descritos en Crameri et al, 1998, Nature, 391: 288- 291 y Ness et al. Nature Biotechnology 17: 893-896. Otro formato serían los métodos descritos en US 6159687: Ejemplos 1 y 2.

#### 55 **Producción de Etanol a partir de Biomasa**

[0177] La presente invención también se refiere a métodos para producir etanol a partir de biomasa, tal como materiales celulósicos, incluyendo poner en contacto la biomasa con los polipéptidos de la invención. El etanol puede posteriormente ser recuperado. Los polipéptidos de la invención pueden ser producidos "in-situ", es decir, como parte de o directamente en un proceso de producción de etanol, por cultivo de una célula huésped o una cepa, que en su forma de tipo salvaje es capaz

de producir los polipéptidos, bajo condiciones propicias para la producción de los polipéptidos.

[0178] El etanol se puede producir por degradación enzimática de biomasa y conversión de los polisacáridos liberados en etanol. Esta especie de etanol es frecuentemente denominada bioetanol o biocombustible. Este se puede usar como un aditivo de combustible o suplemento en mezclas de menos de un 1% y hasta un 100% (un sustituto de combustible). En algunos países, tales como Brasil, el etanol está sustituyendo a la gasolina en una extensión muy grande.

[0179] El polisacárido predominante en la pared celular primaria de la biomasa es la celulosa, el segundo más abundante es la hemicelulosa, y el tercero es la pectina. La pared celular secundaria, producida después de que la célula haya detenido su crecimiento, también contiene polisacáridos y se refuerza a través de lignina polimérica de manera covalente reticulada para hemicelulosa. La celulosa es un homopolímero de anhidrocelobiosa y así un beta-(1-4)-D-glucano lineal, mientras que las hemicelulosas incluyen una variedad de compuestos tales como xilanos, xiloglucanos, arabinoxilanas y mananos en estructuras ramificadas complejas con un espectro de sustituyentes. Aunque generalmente poliforme, se encuentra celulosa en tejido vegetal principalmente como una matriz cristalina insoluble de cadenas de glucano paralelas. Las hemicelulosas normalmente establecen enlaces de hidrógeno con la celulosa, al igual que con otras hemicelulosas, que ayudan a estabilizar la matriz de la pared celular.

[0180] Tres clases principales de enzimas de celulasa se utilizan para descomponer biomasa:

- las "endo-1,4-beta-glucanasas" o 1,4-beta-D-glucano-4-glucanohidrolasas (EC 3.2.1.4), que actúan de forma aleatoria en sustratos de 1,4-beta-glucano insolubles y solubles.
- las "exo-1,4-beta-D-glucanasas" que incluyen 1,4-beta-D-glucano glucohidrolasas (EC 3.2.1.74), que liberan D-glucosa de 1,4-beta-D-glucanos e hidrolizan D-celobiosa lentamente, y 1,4-beta-D-glucano celobiohidrolasa (EC 3.2.1.91), también llamada celobiohidrolasa I, que libera D-celobiosa de 1,4- beta-glucanos.
- las "beta-D-glucosidasas" o glucohidrolasas de beta-d-glucósido (EC 3.2.1.21), que actúan para liberar unidades de D-glucosa de celobiosa y celodextrinas solubles, al igual que un conjunto de glucósidos.

[0181] Estas tres clases de enzimas trabajan juntas sinérgicamente en una interacción compleja que produce una decristalización eficaz e hidrólisis de celulosa nativa de biomasa para producir los azúcares reductores que se convierten en etanol por fermentación.

[0182] La presente invención está posteriormente descrita por los siguientes ejemplos que no deberían ser interpretados como limitativos del ámbito de la invención.

## EJEMPLOS

[0183] Productos químicos usados como tampones y sustratos fueron productos comerciales de al menos calidad reactiva.

### EJEMPLO 1

#### Clonación de una secuencia de ADN de celobiohidrolasa I parcial y en toda su longitud (CBH1)

[0184] Una genoteca de ADNc de *Diplodia gossypina* fue seleccionada mediante PCR para presencia del gen CBH1. Para este propósito conjuntos de cebadores fueron construidos, basados en alineación de secuencia e identificación de regiones conservadas entre proteínas de CBH1. La banda de PCR de una electroforesis en gel fue usada para obtener una secuencia parcial del gen CBH1 de *Diplodia gossypina*. La búsqueda de homología confirmó que la secuencia parcial era una secuencia parcial del gen CBH1 (EC 3.2.1.91).

[0185] El gen CBH1 de *Diplodia gossypina* en toda su longitud se obtiene accediendo al depósito de la patente CBS 247.96, haciendo una preparación de ADN o de ADNc, usando la secuencia parcial como base para la construcción de cebadores específicos, y usando técnicas de clonación de PCR estándar para paso a paso coger el gen entero.

[0186] Diferentes otros métodos pueden ser tomados:

- Podría realizarse una selección PCR de la genoteca de ADNc o de los ADNc que fueron usados para la construcción de la genoteca. Para hacer así, cebadores específicos de gen (GSP) y cebadores de vector/adaptador se construyen de la secuencia de ADNc parcial del gen CBH1 y de secuencia de vector/adaptador respectivamente; ambos conjuntos de cebadores diseñados para ir externos en las regiones faltantes 5' y 3' del ADNc de CBH1. Los

productos de PCR más largos obtenidos usando combinaciones de GSP y cebador de vector/adaptador representan las regiones 5' y 3' en toda su longitud del ADNc de CBH1 de *Diplodia gossypina*. La búsqueda de homología y la comparación con la secuencia de ADNc parcial confirman que los productos de PCR 5' y 3' pertenecen al mismo ADNc de CBH1 de *Diplodia gossypina*. El ADNc de longitud completa puede entonces ser obtenido por PCR usando un conjunto de cebadores construidos de los extremos 5' y 3'.

- Alternativamente, la genoteca de ADNc podría ser seleccionada para el ADNc en toda su longitud usando técnicas de hibridación estándar y la secuencia de ADNc parcial como una sonda. Los clones que dan una señal de hibridación positiva con la sonda son entonces purificados y ordenados para determinar la secuencia de ADNc más larga. La búsqueda de homología y la comparación confirman que el ADNc en toda su longitud corresponde a la secuencia parcial de ADNc de CBH1 que fue originalmente usada como una sonda.

[0187] Los dos métodos anteriormente descritos cuentan con la presencia del ADNc de CBH1 en toda su longitud en la genoteca de ADNc o en los ADNc usados para su construcción. Alternativamente, las técnicas de RACE (amplificación rápida de extremidades de ADNc) 5' y 3' o técnicas derivadas podrían ser usadas para identificar las regiones 5' y 3' faltantes. Para este propósito, preferiblemente se aísla ARNm de *Diplodia gossypina* y se utiliza para sintetizar la primera cadena de ADNc utilizando un cebador adaptador que contiene oligo(dT)- o un cebador específico de gen 5' (GSP).

[0188] El ADNc en toda su longitud del gen CBH1 de *Diplodia gossypina* puede también ser obtenido usando ADN genómico de *Diplodia gossypina*. El gen CBH1 se puede identificar por técnicas de PCR tales como la descrita anteriormente o por selección de genoteca genómica estándar usando técnicas de hibridación y el ADNc de CBH1 parcial como una sonda. La búsqueda de homología y la comparación con el ADNc de CBH1 parcial confirman que la secuencia genómica corresponde al gen CBH1 de *Diplodia gossypina*. Identificación de secuencias de consenso tales como sitio de iniciación de transcripción, inicio y parada de codones o sitios poliA podrían ser usados para definir la región que comprende el ADNc en toda su longitud. Cebadores construidos de las extremidades 5' y 3' de esta región podrían entonces ser usados para amplificar el ADNc en toda su longitud de ADNm o la genoteca de ADNc de *Diplodia gossypina* (ver por encima).

[0189] Por expresión del gen en toda su longitud en un constructo huésped de expresión adecuada la enzima CBH1 se cosecha como una enzima intra celular o extra celular del caldo de cultivo.

[0190] Los métodos anteriormente descritos recurren a la clonación de secuencias de ADN de celobiohidrolasa I de todos organismos y no sólo de *Diplodia gossypina*.

## EJEMPLO 2

### Actividad de Celobiohidrolasa I (CBH I)

[0191] Una celobiohidrolasa I se caracteriza por la capacidad para hidrolizar celulosa altamente cristalina muy eficazmente en comparación con otras celulasas. Celobiohidrolasa I puede tener una actividad catalítica más alta usando PASC (celulosa hinchada de ácido fosfórico) como sustrato que usando CMC como sustrato. Para los fines de la presente invención, cualquiera de los siguientes ensayos pueden utilizarse para identificar una celobiohidrolasa I:

#### Actividad en el Azo-Avicel

[0192] Azo-Avicel (Megazyme, Bray Business Park, Bray, Wicklow, Irlanda) fue usado según las instrucciones de los fabricantes.

#### Actividad en PNP-beta-celobiosa

[0193]  
Solución de sustrato: beta-D-Celobiosa PNP (p-nitrofenilo β-d-Cellobiosida Sigma N-5759) 5 mM en tampón de Na-acetato 0,1 M, pH 5,0  
Reactivo de parada: carbonato de Na 0,1 M, pH 11,5.

[0194] 50 µL de solución de CBH I fueron mezclados con 1 mL de solución de sustrato e incubados 20 minutos a 40 °C. La reacción fue detenida por adición de 5 mL reactivo de parada. La absorbancia fue medida a 404 nm.

#### Actividad en PASC y CMC

[0195] El sustrato se degrada con celobiohidrolasa I (CBH I) para formar azúcares reductores. Una carbohidrato oxidasa de *Microdochium nivale* (rMnO) u otra oxidasa equivalente actúa en los azúcares reductores para formar H<sub>2</sub>O<sub>2</sub> en presencia de

O<sub>2</sub>. El H<sub>2</sub>O<sub>2</sub> formado activa, en presencia de exceso de peroxidasa, la condensación oxidante de 4- aminoantipirina (AA) y N-etilo-N-sulfopropilo-m-toluidina (TOPS) para formar un producto morado que se puede cuantificar por su absorbancia a 550 nm.

5 [0196] Cuando todos los componentes excepto CBH I están en exceso, el índice de aumento en la absorbancia es proporcional a la actividad de CBH I. La reacción es una reacción de un paso cinético y se puede realizar automáticamente en un analizador por centrifugación Cobas Fara (Hoffmann La Roche) u otro espectrofotómetro equivalente que pueda medir cinética estable.

10 Tampón: 50 mM tampón de Na-acetato (pH 5,0);  
 Reactivos: rMnO oxidasa, carbohidrato oxidasa purificada de *Microdochium nivale*, 2 mg/L (concentración final);

15 Peroxidasa, SIGMA P-8125 (96 U/mg), 25 mg/L (concentración final);  
 4-aminoantipirina, SIGMA A-4382, 200 mg/L (concentración final);  
 TOPS, SIGMA E-8506, 600 mg/L (concentración final);  
 PASC o CMC (ver más abajo), 5 g/L (concentración final).

[0197] Todos los reactivos fueron adicionados al tampón en las concentraciones indicadas por encima y esta solución reactiva fue mezclada íntegramente.

20 [0198] Una muestra de 50 µL de celobiohidrolasa I (en una dilución adecuada) fue mezclada con 300 µL de solución reactiva e incubada 20 minutos a 40 °C. La formación de color morado fue detectada y medida como absorbancia a 550 nm.

25 [0199] El coeficiente de absorción AA-TOPS-condensado es 0,01935 A<sub>550</sub>/(µM cm). El índice se calcula como µmoles de azúcar reductor producido por minuto a partir de OD<sub>550</sub>/minuto y el coeficiente de absorción.

**PASC:**

30 [0200]

35 Materiales: 5 g Avicel® (técnica. 2331 Merck);  
 150 mL 85% ácido ortofosfórico (Art. 573 Merck);  
 800 mL acetona (Art. 14 Merck);  
 Aprox. 2 litros de agua desionizada (milli-Q);  
 Vaso de precipitados de cristal de 1 L;  
 Embudo de filtro de vidrio de 1 L;  
 Matraz de succión de 2 L;  
 Homogenizadores de Ultra Turrax.

40 [0201] La acetona y el ácido ortofosfórico se enfrían en hielo. Avicel® se humedece con agua, y luego se añaden los 150 mL de ácido ortofosfórico 85% enfriado en hielo. La mezcla se coloca en un baño de hielo con agitación débil durante una hora.

45 [0202] Añadir 500 mL de acetona enfriada con hielo con agitación, y transferir la mezcla a un embudo de filtro de vidrio y lavar con 3 x 100 mL acetona enfriada con hielo, aspirar lo más seco posible en cada lavado. Lavado con 2 x 500 mL de agua (o hasta que no haya olor de acetona), aspirar lo más seco posible en cada lavado.

[0203] Resuspender los sólidos en el agua a un volumen total de 500 mL, y mezclar hasta homogeneidad utilizando un homogenizador Ultra Turrax. Guardar húmedo en el frigorífico y equilibrar con tampón por centrifugado y resuspensión antes del uso.

50 **CMC:**

55 [0204] Microfibrillas de celulosa bacteriana en una forma impura fueron obtenidas del producto alimenticio japonés "nata de coco" (Fujico Company, Japón). La celulosa en 350 g de este producto fue purificada por suspensión del producto en aproximadamente 4 L de agua del grifo. Esta agua fue sustituida por agua dulce dos veces al día durante 4 días.

[0205] Luego NaOH al 1% (p/v) fue usado en vez de agua y el producto fue resuspendido en la solución de álcali dos veces al día durante 4 días. La neutralización fue hecha por aclarado de la celulosa purificada con agua destilada hasta que el pH de la superficie del producto fue neutro (pH 7).

60 [0206] La celulosa fue microfibrilada y una suspensión de microfibrillas de celulosa bacteriana individuales se obtuvo por

homogenización de las microfibrillas de celulosa purificada en una mezcladora Waring durante 30 min. Las microfibrillas de celulosa fueron de nuevo purificadas por dialización de esta suspensión a través de una membrana de poro contra agua destilada y las microfibrillas de celulosa aisladas y purificadas fueron almacenadas en una suspensión de agua a 4 °C.

5 **Depósito de Material biológico**

China General Microbiological Culture Collection Center (CGMCC)

10 [0207] El siguiente material biológico ha sido depositado según las condiciones del tratado de Budapest con el China General Microbiological Culture Collection Center (CGMCC), Institute of Microbiology, Chinese Academy of Sciences, Haidian, Beijing 100080, China:

Número de acceso: **CGMCC n.º 0582**

15 Referencia de solicitantes: ND000549

Fecha de depósito: 2001-05-29

Descripción: gen de CBH I de *Acremonium thermophilum* en plásmido

20

Clasificación: *Eurotiomycetes*; *Eurotiales*; *Trichocomaceae*

Origen: China

25 Secuencia(s) relacionada(s): SEC ID n.º: 7 y SEC ID n.º: 8 (secuencia de ADN que codifica una celobiohidrolasa I de *Thermoascus aurantiacus* y la secuencia de proteína correspondiente)

LISTADO DE SECUENCIAS

30 [0208]

<110> Novozymes A/S

<120> Polipéptidos con actividad de celobiohidrolasa I y polinucleótidos que codifican los mismos

35

<130> 10129-WO

<160> 67

40 <170> BiSSAP 1.2

<210> 1

<400> 1

45

000

<210> 2

<400> 2

50

000

<210> 3

<400> 3

55

000

<210> 4

<400> 4

60

000

```

<210> 5
<400> 5
000
5 <210> 6
<400> 6
000
10 <210> 7
<211> 1374
<212> ADN
<213> Thermoascus aurantiacus
15 <220>
<221> Fuente
<222> 1..1374
<223> /organism="Thermoascus aurantiacus" /mol_type="unassigned DNA"
20 <220>
<221> CDS
<222> 1..1374
25 <400> 7

      atg tat cag cgc gct ctt ctc ttc tct ttc ttc ctc tcc gcc gcc cgc      48
      Met Tyr Gln Arg Ala Leu Leu Phe Ser Phe Phe Leu Ser Ala Ala Arg
      1           5           10           15

      gcg cag cag gcc ggt acc cta acc gca gag aat cac cct tcc ctg acc      96
  
```

Ala	Gln	Gln	Ala	Gly	Thr	Leu	Thr	Ala	Glu	Asn	His	Pro	Ser	Leu	Thr		
			20					25					30				
tgg	cag	caa	tgc	tcc	agc	ggc	ggg	agt	tgt	acc	acg	cag	aat	gga	aaa	144	
Trp	Gln	Gln	Cys	Ser	Ser	Gly	Gly	Ser	Cys	Thr	Thr	Gln	Asn	Gly	Lys		
			35				40					45					
gtc	ggt	atc	gat	gcg	aac	tgg	cgt	tgg	gtc	cat	acc	acc	tct	gga	tac	192	
Val	Val	Ile	Asp	Ala	Asn	Trp	Arg	Trp	Val	His	Thr	Thr	Ser	Gly	Tyr		
			50				55					60					
acc	aac	tgc	tac	acg	ggc	aat	acg	tgg	gac	acc	agt	atc	tgt	ccc	gac	240	
Thr	Asn	Cys	Tyr	Thr	Gly	Asn	Thr	Trp	Asp	Thr	Ser	Ile	Cys	Pro	Asp		
					70					75					80		
gac	gtg	acc	tgc	gct	cag	aat	tgt	gcc	ttg	gat	gga	gcg	gat	tac	agt	288	
Asp	Val	Thr	Cys	Ala	Gln	Asn	Cys	Ala	Leu	Asp	Gly	Ala	Asp	Tyr	Ser		
				85					90					95			
ggc	acc	tat	ggg	ggt	acg	acc	agt	ggc	aac	gcc	ctg	aga	ctg	aac	ttt	336	
Gly	Thr	Tyr	Gly	Val	Thr	Thr	Ser	Gly	Asn	Ala	Leu	Arg	Leu	Asn	Phe		
				100				105						110			
gtc	acc	caa	agc	tca	ggg	aag	aac	att	ggc	tcg	cgc	ctg	tac	ctg	ctg	384	
Val	Thr	Gln	Ser	Ser	Gly	Lys	Asn	Ile	Gly	Ser	Arg	Leu	Tyr	Leu	Leu		
				115				120						125			
cag	gac	gac	acc	act	tat	cag	atc	ttc	aag	ctg	ctg	ggg	cag	gag	ttt	432	
Gln	Asp	Asp	Thr	Thr	Tyr	Gln	Ile	Phe	Lys	Leu	Leu	Gly	Gln	Glu	Phe		
						135						140					
acc	ttc	gat	gtc	gac	gtc	tcc	aat	ctc	cct	tgc	ggg	ctg	aac	ggc	gcc	480	
Thr	Phe	Asp	Val	Asp	Val	Ser	Asn	Leu	Pro	Cys	Gly	Leu	Asn	Gly	Ala		
					150						155				160		
ctc	tac	ttt	gtg	gcc	atg	gac	gcc	gac	ggc	gga	ttg	tcc	aaa	tac	cct	528	
Leu	Tyr	Phe	Val	Ala	Met	Asp	Ala	Asp	Gly	Gly	Leu	Ser	Lys	Tyr	Pro		
				165					170						175		
ggc	aac	aag	gca	ggc	gct	aag	tat	ggc	act	ggg	tac	tgc	gac	tct	cag	576	
Gly	Asn	Lys	Ala	Gly	Ala	Lys	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp	Ser	Gln		
			180					185						190			
tgc	cct	egg	gat	ctc	aag	ttc	atc	aac	ggg	cag	gcc	aac	ggt	gaa	ggc	624	
Cys	Pro	Arg	Asp	Leu	Lys	Phe	Ile	Asn	Gly	Gln	Ala	Asn	Val	Glu	Gly		
				195				200					205				
tgg	cag	ccg	tct	gcc	aac	gac	cca	aat	gcc	ggc	ggt	ggg	aac	cac	ggg	672	
Trp	Gln	Pro	Ser	Ala	Asn	Asp	Pro	Asn	Ala	Gly	Val	Gly	Asn	His	Gly		
				210				215				220					
tcc	tgc	tgc	gct	gag	atg	gat	gtc	tgg	gaa	gcc	aac	agc	atc	tct	act	720	
Ser	Cys	Cys	Ala	Glu	Met	Asp	Val	Trp	Glu	Ala	Asn	Ser	Ile	Ser	Thr		
					230						235				240		
gcg	gtg	acg	cct	cac	cca	tgc	gac	acc	ccc	ggc	cag	acc	atg	tgc	cag	768	
Ala	Val	Thr	Pro	His	Pro	Cys	Asp	Thr	Pro	Gly	Gln	Thr	Met	Cys	Gln		
				245							250				255		

ES 2 438 146 T3

gga gac gac tgt ggt gga acc tac tcc tcc act cga tat gct ggt acc	816
Gly Asp Asp Cys Gly Gly Thr Tyr Ser Ser Thr Arg Tyr Ala Gly Thr	
260 265 270	
tgc gac cct gat ggc tgc gac ttc aat cct tac cgc cag ggc aac cac	864
Cys Asp Pro Asp Gly Cys Asp Phe Asn Pro Tyr Arg Gln Gly Asn His	
275 280 285	
tcg ttc tac ggc ccc ggg aag atc gtc gac act agc tcc aaa ttc acc	912
Ser Phe Tyr Gly Pro Gly Lys Ile Val Asp Thr Ser Ser Lys Phe Thr	
290 295 300	
gtc gtc acc cag ttc atc acc gac gac ggg acc ccc tcc ggc acc ctg	960
Val Val Thr Gln Phe Ile Thr Asp Asp Gly Thr Pro Ser Gly Thr Leu	
305 310 315 320	
acg gag atc aaa cgc ttc tac gtc cag aac ggc aag gtg atc ccc cag	1008
Thr Glu Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Gln	
325 330 335	
tcg gag tcg acg atc agc ggc gtc acc ggc aac tca atc acc acc gag	1056
Ser Glu Ser Thr Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu	
340 345 350	
tat tgc acg gcc cag aag gcc gcc ttc ggc gac aac acc ggc ttc ttc	1104
Tyr Cys Thr Ala Gln Lys Ala Ala Phe Gly Asp Asn Thr Gly Phe Phe	
355 360 365	
acg cac ggc ggg ctt cag aag atc agt cag gct ctg gct cag ggc atg	1152
Thr His Gly Gly Leu Gln Lys Ile Ser Gln Ala Leu Ala Gln Gly Met	
370 375 380	
gtc ctc gtc atg agc ctg tgg gac gat cac gcc gcc aac atg ctc tgg	1200
Val Leu Val Met Ser Leu Trp Asp Asp His Ala Ala Asn Met Leu Trp	
385 390 395 400	
ctg gac agc acc tac ccg act gat gcg gac ccg gac acc cct ggc gtc	1248
Leu Asp Ser Thr Tyr Pro Thr Asp Ala Asp Pro Asp Thr Pro Gly Val	
405 410 415	
gcg cgc ggt acc tgc ccc acg acc tcc ggc gtc ccg gcc gac gtt gag	1296
Ala Arg Gly Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu	
420 425 430	
tcg cag aac ccc aat tca tat gtt atc tac tcc aac atc aag gtc gga	1344
Ser Gln Asn Pro Asn Ser Tyr Val Ile Tyr Ser Asn Ile Lys Val Gly	
435 440 445	
ccc atc aac tcg acc ttc acc gcc aac taa	1374
Pro Ile Asn Ser Thr Phe Thr Ala Asn	
450 455	

<210> 8  
<211> 457

<212> PRT

<213> *Thermoascus aurantiacus*

<220>

5 <223> [CDS]:1..1374 de SEC ID n.º 7

<400> 8

Met Tyr Gln Arg Ala Leu Leu Phe Ser Phe Phe Leu Ser Ala Ala Arg  
 1 5 10 15  
 Ala Gln Gln Ala Gly Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr  
 20 25 30  
 Trp Gln Gln Cys Ser Ser Gly Gly Ser Cys Thr Thr Gln Asn Gly Lys  
 35 40 45  
 Val Val Ile Asp Ala Asn Trp Arg Trp Val His Thr Thr Ser Gly Tyr  
 50 55 60  
 Thr Asn Cys Tyr Thr Gly Asn Thr Trp Asp Thr Ser Ile Cys Pro Asp  
 65 70 75 80  
 Asp Val Thr Cys Ala Gln Asn Cys Ala Leu Asp Gly Ala Asp Tyr Ser  
 85 90 95  
 Gly Thr Tyr Gly Val Thr Thr Ser Gly Asn Ala Leu Arg Leu Asn Phe  
 100 105 110  
 Val Thr Gln Ser Ser Gly Lys Asn Ile Gly Ser Arg Leu Tyr Leu Leu  
 115 120 125  
 Gln Asp Asp Thr Thr Tyr Gln Ile Phe Lys Leu Leu Gly Gln Glu Phe  
 130 135 140  
 Thr Phe Asp Val Asp Val Ser Asn Leu Pro Cys Gly Leu Asn Gly Ala  
 145 150 155 160  
 Leu Tyr Phe Val Ala Met Asp Ala Asp Gly Gly Leu Ser Lys Tyr Pro  
 165 170 175  
 Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln  
 180 185 190  
 Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly  
 195 200 205  
 Trp Gln Pro Ser Ala Asn Asp Pro Asn Ala Gly Val Gly Asn His Gly  
 210 215 220  
 Ser Cys Cys Ala Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser Thr  
 225 230 235 240  
 Ala Val Thr Pro His Pro Cys Asp Thr Pro Gly Gln Thr Met Cys Gln  
 245 250 255  
 Gly Asp Asp Cys Gly Gly Thr Tyr Ser Ser Thr Arg Tyr Ala Gly Thr  
 260 265 270  
 Cys Asp Pro Asp Gly Cys Asp Phe Asn Pro Tyr Arg Gln Gly Asn His  
 275 280 285  
 Ser Phe Tyr Gly Pro Gly Lys Ile Val Asp Thr Ser Ser Lys Phe Thr  
 290 295 300  
 Val Val Thr Gln Phe Ile Thr Asp Asp Gly Thr Pro Ser Gly Thr Leu  
 305 310 315 320  
 Thr Glu Ile Lys Arg Phe Tyr Val Gln Asn Gly Lys Val Ile Pro Gln  
 325 330 335  
 Ser Glu Ser Thr Ile Ser Gly Val Thr Gly Asn Ser Ile Thr Thr Glu  
 340 345 350  
 Tyr Cys Thr Ala Gln Lys Ala Ala Phe Gly Asp Asn Thr Gly Phe Phe  
 355 360 365  
 Thr His Gly Gly Leu Gln Lys Ile Ser Gln Ala Leu Ala Gln Gly Met  
 370 375 380  
 Val Leu Val Met Ser Leu Trp Asp Asp His Ala Ala Asn Met Leu Trp  
 385 390 395 400  
 Leu Asp Ser Thr Tyr Pro Thr Asp Ala Asp Pro Asp Thr Pro Gly Val  
 405 410 415  
 Ala Arg Gly Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu  
 420 425 430

ES 2 438 146 T3

Ser Gln Asn Pro Asn Ser Tyr Val Ile Tyr Ser Asn Ile Lys Val Gly  
          435                          440                          445  
Pro Ile Asn Ser Thr Phe Thr Ala Asn  
          450                          455

- <210> 9
- 5 <400> 9  
000
- <210> 10
- 10 <400> 10  
000
- <210> 11
- 15 <400> 11  
000
- <210> 12
- 20 <400> 12  
000
- <210> 13
- 25 <400> 13  
000
- <210> 14
- 30 <400> 14  
000
- <210> 15
- 35 <400> 15  
000
- <210> 16
- 40 <400> 16  
000
- <210> 17
- 45 <400> 17  
000
- <210> 18
- 50 <400> 18  
000
- <210> 19
- 55 <400> 19  
000

# ES 2 438 146 T3

<210> 20  
5 <400> 20  
000  
<210> 21  
10 <400> 21  
000  
<210> 22  
15 <400> 22  
000  
<210> 23  
20 <400> 23  
000  
<210> 24  
25 <400> 24  
000  
<210> 25  
30 <400> 25  
000  
<210> 26  
35 <400> 26  
000  
<210> 27  
40 <400> 27  
000  
<210> 28  
45 <400> 28  
000  
<210> 29  
50 <400> 29  
000  
<210> 30  
55 <400> 30  
000  
<210> 31  
60 <400> 31  
000

<210> 32  
<400> 32  
000  
5 <210> 33  
<400> 33  
000  
10 <210> 34  
<400> 34  
000  
15 <210> 35  
<400> 35  
000  
20 <210> 36  
<400> 36  
000  
25 <210> 37  
<400> 37  
000  
30 <210> 38  
<400> 38  
000  
35 <210> 39  
<400> 39  
000  
40 <210> 40  
<400> 40  
000  
45 <210> 41  
<400> 41  
000  
50 <210> 42  
<400> 42  
000  
55 <210> 43  
<400> 43  
000  
60 <210> 44

# ES 2 438 146 T3

<400> 44  
000

5 <210> 45

<400> 45  
000

10 <210> 46

<400> 46  
000

15 <210> 47

<400> 47  
000

20 <210> 48

<400> 48  
000

25 <210> 49

<400> 49  
000

30 <210> 50

<400> 50  
000

35 <210> 51

<400> 51  
000

40 <210> 52

<400> 52  
000

45 <210> 53

<400> 53  
000

50 <210> 54

<400> 54  
000

55 <210> 55

<400> 55  
000

60 <210> 56

# ES 2 438 146 T3

<400> 56  
000

5 <210> 57

<400> 57  
000

10 <210> 58

<400> 58  
000

15 <210> 59

<400> 59  
000

20 <210> 60

<400> 60  
000

25 <210> 61

<400> 61  
000

30 <210> 62

<400> 62  
000

35 <210> 63

<400> 63  
000

40 <210> 64

<400> 64  
000

45 <210> 65

<400> 65  
000

50 <210> 66

<400> 66  
000

55 <210> 67

<400> 67  
000

60

**REIVINDICACIONES**

1. Polipéptido con actividad de celobiohidrolasa I, seleccionado del grupo consistiendo en:
- 5 (a) un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 90% de identidad con los aminoácidos 1 a 457 de SEC ID n.º: 8,
- (b) un polipéptido que incluye una secuencia de aminoácidos que tiene al menos un 90% de identidad con el polipéptido codificado por los nucleótidos 1 a 1371 de SEC ID n.º: 7.
- 10 2. Polipéptido según la reivindicación 1, que incluye una secuencia de aminoácidos que tiene al menos un 90% de identidad, más preferiblemente al menos un 95% de identidad, con los aminoácidos 1 a 457 de SEC ID n.º: 8.
- 15 3. Polipéptido según cualquiera de las reivindicaciones 1-2, que consiste en los aminoácidos 1 a 457 de SEC ID n.º: 8.
4. Polinucleótido con una secuencia de nucleótidos que codifica el polipéptido definido en cualquiera de las reivindicaciones 1-3.
5. Constructo de ácidos nucleicos que comprende la secuencia de nucleótidos definida en la reivindicación 4 operativamente enlazado a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.
- 20 6. Vector de expresión recombinante que comprende el constructo de ácidos nucleicos definido en la reivindicación 5.
7. Célula huésped recombinante que comprende el constructo de ácidos nucleicos definido en la reivindicación 5.
- 25 8. Método para producir un polipéptido tal y como se define en cualquiera de las reivindicaciones 1-3, el método comprende:
- (a) cultivar una célula huésped recombinante tal y como se define en la reivindicación 7 bajo condiciones propicias para la producción del polipéptido; y
- 30 (b) recuperar el polipéptido.
9. Método para la producción in situ de un polipéptido tal y como se define en cualquiera de las reivindicaciones 1-3, el método comprende:
- 35 (a) cultivar una célula huésped recombinante tal y como se define en la reivindicación 7 bajo condiciones propicias para la producción del polipéptido; y
- (b) poner en contacto el polipéptido con un sustrato deseado sin recuperación previa del polipéptido.
- 40 10. Método para producir etanol a partir de biomasa, que comprende poner en contacto la biomasa con el polipéptido tal y como se define en cualquiera de las reivindicaciones 1-3.
- 45 11. Uso del polipéptido tal y como se define en cualquiera de las reivindicaciones 1-3 para producir etanol.
12. Planta transgénica, parte de planta o célula vegetal, que tiene incorporado en el genoma de la planta huésped uno o más constructos de expresión que codifican un polipéptido con actividad de celobiohidrolasa I tal y como se define en cualquiera de las reivindicaciones 1-3.
- 50 13. Composición de detergente que comprende un surfactante y el polipéptido según cualquiera de las reivindicaciones 1-3.