

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 397 267**

51 Int. Cl.:

**C21C 1/00** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **26.06.2002 E 02745180 (6)**

97 Fecha y número de publicación de la concesión europea: **17.10.2012 EP 1421224**

54 Título: **Polipéptidos con actividad de celobiohidrolasa 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:

**05.03.2013**

73 Titular/es:

**NOVOZYMES A/S (100.0%)  
KROGSHOEJVEJ 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 2 397 267 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 con 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 nucleótida que codifica para los polipéptidos. La invención también se refiere a vectores y células huésped que incluyen constructos de ácidos nucleicos al igual que métodos para producir y usar los polipéptidos.

10 Antecedentes de la invención

[0002] Celulosa es una materia prima industrial importante y una fuente de energía renovable. La estructura física y la morfología de 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.

[0003] Para ser eficaz, la digestión de 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 las extremidades de 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, 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 hidrolización de la celobiosa del extremo no-reducido de las cadenas de polímero de celulosa.

[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 pueden también tener una capacidad mejorada para resistir 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:

(a) un polipéptido que comprende una secuencia de aminoácidos seleccionados del grupo que consiste en:

40 una secuencia de aminoácido que tiene como mínimo 80% de identidad con aminoácidos 1 a 526 de SEC ID nº 2, y

(b) un polipéptido que incluye una secuencia de aminoácidos seleccionados del grupo que consiste en:

una secuencia de aminoácidos que tiene al menos 80% de identidad con el polipéptido codificado por nucleótidos 1 a 1578 de SEC ID nº 1,

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

[0008] En un tercer aspecto, la presente invención se refiere a un vector de expresión recombinante que incluye un constructo de ácidos nucleicos que incluye la secuencia de nucleótidos, que codifica para 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 una célula huésped recombinante que incluye un vector de expresión recombinante que incluye un constructo de ácidos nucleicos, dicho constructo comprende una secuencia de nucleótidos según la invención operativamente enlazada a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado, donde la célula huésped es seleccionada del grupo que consiste en Aspergillum, Fusarium, Humicola, Mucor, Myceliophthora, Neurospora, Penicillium, Thielavia, Tolipocladium y Trichoderma.

[0010] En un quinto aspecto, la presente invención se refiere a un método para la producción de un polipéptido de la invención, el método incluye:

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

(b) recuperación el polipéptido.

[0011] En un sexto aspecto, la presente invención se refiere a un método para producción in situ de un polipéptido de la

invención, el método incluye:

- (a) cultivo de una célula huésped recombinante de la invención bajo condiciones propicias para la producción del polipéptido; y
- (b) contacto del polipéptido con un sustrato deseado sin recuperación previa del polipéptido.

[0012] Otros aspectos de la presente invención será aparentes de la descripción que se realizará más adelante y de las reivindicaciones anexas.

#### Definiciones

[0013] Antes de discutir la presente invención en detalles adicionales, los siguientes términos y convenios serán definidos primeramente:

**Polipéptido substancialmente puro:** en el presente contexto, el término "polipéptido substancialmente puro" significa una preparación de polipéptido que contiene como mucho 10% en peso de otro material polipéptido con el cual es originalmente asociado (porcentajes inferiores de otro material polipéptido son preferidos, por ejemplo como mucho 8% en peso, como mucho 6% en peso, como mucho 5% en peso, como mucho 4%, como mucho 3% en peso, como mucho 2% en peso, como mucho 1% en peso y como mucho ½% en peso). Así, se prefiere que el polipéptido substancialmente puro sea al menos 92% puro, es decir que el polipéptido constituya al menos 92% en peso del material polipéptido total presente en la preparación, y porcentajes más altos son preferidos tal como al menos 94% puro, al menos 95% puro, al menos 96% puro, al menos 96% puro, al menos 97% puro, al menos 98% puro, al menos 99% y como mucho 99.5% puro. Los polipéptidos descritos aquí están preferiblemente en una forma substancialmente 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 polipéptido con el cual es originalmente asociado. Esto puede ser realizado, por ejemplo, por preparación del polipéptido mediante bien conocidos métodos recombinantes. Aquí, el término "polipéptido substancialmente 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 las extremidades no-reducidas de las cadenas.

Para fines de la presente invención, actividad de celobiohidrolasa I puede ser determinada según al procedimiento descrito en el ejemplo 2. En una forma de realización, 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 μmol de p-nitrofenol producido por minuto a 50°C, pH 5,0. Los polipéptidos de la presente invención deberían preferiblemente tener al menos 20% de la actividad de celobiohidrolasa I de un polipéptido que consiste en una secuencia de aminoácidos seleccionados del grupo que consiste en SEC ID n° 2.

En una forma de realización particular preferida, los polipéptidos deberían tener al menos 40%, tal como al menos 50%, preferiblemente al menos 60%, tal como al menos 70%, más preferiblemente al menos 80%, tal como al menos 90%, muchos preferiblemente al menos 95%, tal como al rededor de o al menos 100% de la actividad de celobiohidrolasa I del polipéptido que consiste en la secuencia de aminoácidos seleccionados de los aminoácidos 1 a 526 de SEC ID n° 2.

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

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 del programa FASTA (ver 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 marcado usada fue BLOSUM50, penalización del espacio fue -12, y penalización de extensión del espacio fue -2.

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 marcado usada fue la matriz de identidad, la penalización de gap fue -16 y penalización de extensión de gap fue -4.

**Fragmento:** cuando se usa aquí, un "fragmento" de una secuencia seleccionada de SEC ID n° 2 es un polipéptido con uno o más aminoácidos eliminados del terminal amino y/o carboxilo 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 accesión de SWISS-PROT P00725. Más preferiblemente, un fragmento incluye la secuencia de aminoácidos que corresponde al "dominio catalítico" de Celobiohidrolasa I de *Trichoderma reesei* como está descrito en el número de accesión de SWISS-PROT P00725. De la forma más preferible, un fragmento contiene al menos 434 residuos de aminoácidos, por ejemplo, los residuos de aminoácido seleccionados de aminoácidos 1 a 434 de SEC ID n° 2. En particular, un fragmento contiene al menos 215 residuos de aminoácidos, por ejemplo, los residuos de aminoácidos seleccionados de los aminoácidos 200 a 434 de SEC ID n° 2.

**Variante alélica:** en el presente contexto, el término "variante alélica" denota dos o más formas cualquiera alternativas de un gen que ocupa la misma localización cromosómica. Variante alélica surge naturalmente a través de mutación y puede suponer polimorfismo dentro de poblaciones. Mutaciones de gen pueden ser silenciosas (ningún cambio en el polipéptido codificado) o pueden codificar polipéptidos con secuencias de aminoácidos alteradas. Una variante alélica de un polipéptido es un polipéptido codificado por una variante alélica de un gen.

**Polinucleótido substancialmente puro:** el término "polinucleótido substancialmente puro" como se utiliza en este caso se refiere a una preparación polinucleótida, donde el polinucleótido ha sido quitado de su ambiente natural genético, y está así libre de otras secuencias de codificación indeseadas o extrañas y está en una forma adecuada para su uso dentro de sistemas de producción de proteína creada genéticamente. Así, un polinucleótido substancialmente puro contiene como mucho 10% en peso de otro material polinucleótido con el cual es originalmente asociado (porcentajes inferiores de otro material polinucleótido son preferidos, por ejemplo como mucho 8% en peso, como mucho 6% en peso, como mucho 5% en peso, como mucho 4% como mucho 3% en peso, como mucho 2% en peso, como mucho 1% en peso, y como mucho ½% por peso). Un polinucleótido substancialmente puro puede, no obstante, incluir regiones no traducidas 5' y 3' de origen natural, tales como promotores y terminadores. Se prefiere que el polinucleótido substancialmente puro sea al menos 92% puro, es decir que el polinucleótido constituya al menos 92% en peso del material de polinucleótido total presente en la preparación, y porcentajes más altos son preferidos tal como al menos 94% puro, al menos 95% puro, al menos 96% puro, al menos 96% puro, al menos 97% puro, al menos 98% puro, a mínimo 99% y, como mucho, 99,5% puro. Los polinucleótidos descritos aquí están preferiblemente en una forma substancialmente 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 sea originalmente asociada. Aquí, el término "polinucleótido substancialmente puro" es sinónimo de los términos "polinucleótido aislado" y "polinucleótido en forma aislada".

**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 SEC ID n°2, al igual que manipulación genética del ADN que codifica ese polipéptido. La modificación(es) puede ser sustitución(es) de la cadena(s) lateral de aminoácido, sustitución(es), deleción(es) y/o inserciones(s) en o al ácido(s) de amino de interés.

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 SEC ID n° 1. El gen modificado, donde dicha variante se produce cuando expresada en un huésped adecuado, se obtiene a través de intervención humana por modificación de una secuencia de nucleótido seleccionada de la SEC ID n° 1.

**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 de una molécula de ARNm dividida madura derivada de una célula eucariota. A ADNc le faltan las secuencias de intrón que están normalmente presentes en el ADN genómico correspondiente. La transcripción de ARN primario inicial es un precursor para ARNm y va a través de una serie de eventos de tratamiento antes de aparecer como ARNm maduro dividido. Estos eventos incluyen la eliminación de secuencias de intrón por un proceso llamado empalme. Cuando ADNc se deriva de ARNm a éste, por lo tanto, le faltan secuencias de intrón.

**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 sinónimos del término "cassette de expresión" cuando el constructo de ácidos nucleicos contiene las secuencias de control requerida para la expresión de una secuencia codificante de la presente invención.

**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 foránea a la secuencia de nucleótidos que codifica el polipéptido. Tales secuencias de control incluyen, pero de forma no limitativa, una secuencia de poliadenilación guía, 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.

**Operativamente enlazado:** el término "operativamente enlazado" es definido aquí como una configuración en la que una secuencia de control es 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.

**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 bordes 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.

**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.

**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 es operativamente enlazado a segmentos adicionales que proveen para su transcripción.

**Célula huésped:** el término "célula huésped", como se utiliza en este caso, incluye cualquier tipo de célula que sea susceptible de transformación con un constructo de ácidos nucleicos. Los términos "sonda polinucleótida", "hibridación", al igual que las varias condiciones de astringencia se definen en la sección titulada "polipéptidos con actividad de celobiohidrolasa I".

**Termoestabilidad:** el término "termoestabilidad", como se utiliza en este caso, es medido como está descrito en el ejemplo 2.

## Descripción detallada de la invención

## Polipéptidos con actividad de celobiohidrolasa

5 [0014] En una primera forma de realización, la presente invención se refiere a polipéptidos con actividad de celobiohidrolasa I y donde los polipéptidos incluyen, preferiblemente, una secuencia de aminoácidos que tiene un grado de identidad con una secuencia de aminoácido de SEC ID nº2, (es decir, el polipéptido maduro) de al menos 80%, tal como al menos 85%, incluso más preferiblemente al menos 90%, más preferiblemente al menos 95%, por ejemplo al menos 96%, tal como al menos 97% e incluso más preferiblemente al menos 98%, tal como al menos 99% (de ahora en adelante "polipéptidos homólogos"). La secuencia de aminoácidos puede diferir en, como mucho, diez aminoácidos (p. ej. en diez aminoácidos), en particular en 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 seleccionada de SEC ID nº 2. La secuencia de aminoácidos puede diferir en como mucho dos aminoácidos (p. ej. en dos aminoácidos), tal como en un aminoácido de una secuencia de aminoácidos seleccionados de SEC ID nº 2.

[0015] Preferiblemente, los polipéptidos de la presente invención comprenden una secuencia de aminoácidos seleccionados de la SEC ID nº 2. En otra forma de realización preferida, el polipéptido de la presente invención consiste de una secuencia de aminoácidos seleccionados de la SEC ID nº 2.

20 [0016] 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 descrito en J.E. Ness et al. Nature Biotechnology 17, 893-896 (1999). Los cambios de aminoácidos (en la variante artificial al igual que en los polipéptidos de tipo salvaje) son de naturaleza menor, esto es, sustituciones de aminoácido conservador que no afectan significativamente al pliegue y/o actividad de la proteína; deleciones pequeñas, típicamente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones de terminal amino o carboxilo, tales como un residuo de metionina terminal amino; un pequeño péptido enlazador de hasta aproximadamente 20-25 residuos; o una pequeña extensión que facilita purificación por carga de red cambiante u otra función, tal como un tracto de poli-histidina, un epítipo antigénico o un dominio de unión.

[0017] Ejemplos de sustituciones conservadoras están dentro del 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 pequeños aminoácidos (glicina, alanina, serina y treonina). Sustituciones de aminoácidos que generalmente no alteran la actividad específica se conocen en la técnica y son 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/Glyal igual que éstos a la inversa.

40 [0018] 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, lo que mejora la termoestabilidad del polipéptido, altera la especificidad de sustrato, cambia el pH óptimo y similares.

45 [0019] Preferiblemente, el número de tales sustituciones, deleciones y/o inserciones comparadas con una secuencia de aminoácido de SEC ID nº 2 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.

50 [0020] Los presentes inventores han aislado secuencias de nucleótidos que codifican polipéptidos con actividad de celobiohidrolasa I del microorganismo *Acremonium thermophilum*. Así, la presente divulgación se refiere a polipéptidos que incluyen una secuencia de aminoácido que tiene al menos 80%, tal como al menos 85%, más preferiblemente al menos 90%, muchos preferiblemente al menos 95%, por ejemplo al menos 96%, tal como al menos 97%, e incluso más preferiblemente al menos 98%, tal como al menos 99% de identidad con el polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo de *Acremonium thermophilum*. 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. 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) del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado del grupo que consiste en *Acremonium thermophilum*. En una forma de realización particular interesante, la secuencia de aminoácidos difiere en como mucho dos aminoácidos (p. ej. en dos aminoácidos), tal como en un aminoácido del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos presentes en un organismo seleccionado del grupo que consiste en *Acremonium thermophilum*.

[0021] Preferiblemente, los polipéptidos de la presente invención comprenden la secuencia de aminoácidos del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótido insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n° 0584. En otra forma de realización preferida, el polipéptido de la presente invención consiste en la secuencia de aminoácidos del polipéptido codificado por la celobiohidrolasa I que codifica parte de la secuencia de nucleótidos insertada en un plásmido presente en un microorganismo depositado seleccionado de CGMCC n° 0584.

[0022] 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 los nucleótidos seleccionados de:

nucleótidos 1 a 1578 de SEC ID n° 1;

(ii) la cadena complementaria de los nucleótidos seleccionados de nucleótidos 1 a 500 de SEC ID n° 1; y

(iii) la cadena complementaria de los nucleótidos seleccionados de nucleótidos 1 a 200 de SEC ID n° 1

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

[0023] Una secuencia de nucleótidos seleccionada de SEC ID n° 1 o una subsecuencia de la misma, al igual que una secuencia de aminoácidos seleccionada de SEC ID n° 2 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 genómico o ADNc del género o especie de interés, siguiendo procedimientos de transferencia de Southern estándar, para identificar y aislar el gen correspondiente en este. Tales sondas puede ser considerablemente más cortas que la secuencia entera, pero deberían ser 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 sea de al menos 100 nucleótidos de longitud. Por ejemplo, la sonda polinucleótida puede ser de 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. Incluso sondas más largas pueden ser utilizadas, por ejemplo, sondas polinucleótidas que son de 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. Ambas sondas de ADN y ARN pueden ser usadas. Las sondas son típicamente etiquetadas para detectar del gen correspondiente (por ejemplo, con <sup>32</sup>P, <sup>3</sup>H, <sup>35</sup>S, biotina o avidina).

[0024] 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. ADN genómico u otro de tales otros organismos se puede separar por agarosa o electroforesis en gel de poliacrilamida u otras técnicas de separación. ADN de las genotecas o ADN separado puede ser transferido a e inmovilizado en nitrocelulosa u otros materiales portadores adecuados. Para identificar un clon o ADN que es homólogo con SEC ID n° 1 el material portador con el ADN inmovilizado es usado en un *Southern blot*.

[0025] Para fines de la presente invención, hibridación indica que la secuencia de nucleótidos hibridiza a una sonda de polinucleótido etiquetada que hibridiza a la secuencia de nucleótidos mostrada en la SEC ID n° 1 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.

[0026] En una forma de realización interesante, la sonda polinucleótida es la cadena complementaria de los nucleótidos seleccionados del grupo que consiste en:

nucleótidos 1 a 1578 de SEC ID n° 1,

nucleótidos 1 a 1302 de SEC ID n° 1;

o la cadena complementaria de los nucleótidos seleccionados de:

nucleótidos 1 a 500 de SEC ID n° 1;

o la cadena complementaria de los nucleótidos seleccionados de:

nucleótidos 1 a 200 de SEC ID n° 1.

[0027] En otra forma de realización interesante, la sonda polinucleótida es la cadena complementaria de la secuencia de nucleótido que codifica un polipéptido seleccionado de SEC ID n° 2. En una forma de realización aún más interesante, la sonda polinucleótida es la cadena complementaria de una secuencia de nucleótido seleccionado de SEC ID n° 1. En otra forma de realización interesante, la sonda polinucleótida es la cadena complementaria de la secuencia de nucleótido contenida en un plásmido que se contiene en un microorganismo depositado seleccionado de CGMCC n°.

0584.

[0028] 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 5X SSPE, 1,0% SDS, 5X solución de Denhardt, 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 1.000 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 2 x SSC, 0,1% SDS a 42°C (astringencia muy baja), preferiblemente lavado tres veces cada uno durante 15 minutos usando 0,5 x SSC, 0,1% SDS a 42°C (astringencia baja), más preferiblemente lavado tres veces cada uno durante 15 minutos usando 0,2 x SSC, 0,1% SDS a 42°C (astringencia media), incluso más preferiblemente lavado tres veces cada uno durante 15 minutos usando 0,2 x SSC, 0,1% SDS a 55°C (astringencia media alta), de la forma más preferible lavado tres veces cada uno durante 15 minutos usando 0,1 x SSC, 0,1% SDS a 60°C (astringencia alta), en particular lavado tres veces cada uno durante 15 minutos usando 0,1 x SSC, 0,1% SDS a 68°C (astringencia altísima).

[0029] Aunque no particularmente preferido, se contempla que sondas más cortas, por ejemplo, sondas que son 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 de 5°C a 10°C por debajo de la T<sub>m</sub> calculada usando el cálculo según Bolton and McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) en 0,9 M NaCl, 0,09 M Tris-HCl pH 7,6, 6 mM EDTA, 0,5% NP-40, 1X solución de Denhardt, 1 mM pirofosfato de sodio, 1 mM fosfato monobásico de sodio, 0,1 mM ATP y 0,2 mg de ARN de levadura por ml siguiendo los procedimientos de *Southern blot* estándar.

[0030] Para sondas cortas que son de aproximadamente 15 nucleótidos a 99 nucleótidos de longitud, el material portador se lava una vez en 6X SSC más 0,1% SDS durante 15 minutos y dos veces cada uno durante 15 minutos uso 6X SSC a 5°C a 10°C por debajo del calculado T<sub>m</sub>.

Fuentes para polipéptidos con actividad de celobiohidrolasa I

[0031] Un polipéptido de la presente invención se puede obtener de *Acremonium thermophilum*, por ejemplo, el polipéptido que consiste en una secuencia de aminoácidos seleccionados de SEC ID n° 2.

[0032] Cepas de estas especies son fácilmente accesibles para el público en una cantidad de colecciones de cultivo, tal como la 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).

[0033] 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 microorganismos de aislamiento de hábitats naturales son bien conocidas en la técnica. La secuencia de nucleótidos puede luego ser derivada mediante la selección de forma similar de un genómico o genoteca de ADNc de otro microorganismo. Una vez una secuencia de nucleótidos que codifica un polipéptido ha sido detectada con la sonda(s), la secuencia puede ser aislada o clonada utilizando de técnicas que se conocen por aquellos de habilidad común en la técnica (ver, por ejemplo, Sambrook et al., 1989, supra).

[0034] Polipéptidos codificados por secuencias de nucleótidos de la presente divulgación también incluyen polipéptidos fundidos o polipéptidos divisibles de fusión en los que otro polipéptido se funde al N-término o al C-término del polipéptido o fragmento del mismo. Un polipéptido fusionado se produce por fundición 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 mismo promotor(es) y terminador.

Polinucleótidos y secuencias de nucleótidos

[0035] La presente invención también se refiere a polinucleótidos con una secuencia de nucleótidos que codifica para un polipéptido de la invención. En particular, la presente invención se refiere a polinucleótidos que consisten en una secuencia de nucleótidos que codifica para un polipéptido de la invención. En una forma de realización preferida, la secuencia de nucleótido es SEC ID n° 1. En una forma de realización más preferida, la secuencia de nucleótido es el polipéptido maduro que codifica la región contenida en un plásmido que está contenido en un microorganismo depositado seleccionado de CGMCC n°. 0584. La presente invención también encierra 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° 2, que difiere de una secuencia de nucleótidos seleccionada de SEC ID n° 1 en virtud de la degeneración del código genético.

[0036] La presente divulgación también se refiere a polinucleótidos que comprenden, preferiblemente que consisten en, una subsecuencia de una secuencia de nucleótidos seleccionada de SEC ID n° 1 que codifica fragmentos de una secuencia de aminoácidos de SEC ID n° 2 que tienen actividad de celobiohidrolasa I. Una subsecuencia de una secuencia de nucleótidos de SEC ID n° 1 es una secuencia de nucleótidos encerrada por una secuencia de SEC ID n° 1 excepto que uno o más nucleótidos del final 5' y/o 3' han sido eliminados.

[0037] La presente divulgación también se refiere a polinucleótidos que tienen, preferiblemente que consisten en, una secuencia de nucleótidos modificada que comprende como mínimo una modificación en el polipéptido maduro que codifica la secuencia SEC ID n° 1, y dónde la secuencia de nucleótido modificada codifica un polipéptido que consiste en una secuencia de aminoácido de SEC ID n° 2.

[0038] Las técnicas usadas para aislar o clonar una secuencia de nucleótidos que codifican un polipéptido se conocen en la técnica e incluyen 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. Ver, por ejemplo, Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York . Otros procedimientos de amplificación tales como reacción en cadena de ligasa (LCR), transcripción activada ligada (LAT) y amplificación basada en secuencias de nucleótidos (NASBA) puede 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 un alélico o variante de especies del polipéptido que codifica la región de la secuencia de nucleótidos.

[0039] La secuencia de nucleótidos se pueden obtener por procedimientos de clonación estándar usados en ingeniería genética para recolocar la secuencia de nucleótidos de su ubicación natural a un sitio diferente donde ésta será reproducida. Los procedimientos de clonación pueden implicar escisión y aislamiento de un fragmento deseado que incluye 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, semisintético sintético o cualquier combinación de las mismas.

[0040] La presente invención también se refiere a un polinucleótido que comprende, preferiblemente que consiste en, una secuencia de nucleótidos que tiene un grado de identidad con una secuencia nucleotídica de nucleótidos 1 a 1578 de SEC ID n° 1, de al menos 80% de identidad, por ejemplo al menos 85% de identidad, tal como al menos 90% de identidad, más preferiblemente al menos 95% de identidad, tal como al menos 96% de identidad, por ejemplo al menos 97% de identidad, incluso más preferiblemente al menos 98% de identidad, tal como al menos 99%. Preferiblemente, la secuencia de nucleótidos codifica un polipéptido con actividad de celobiohidrolasa I. El grado de identidad entre dos secuencias de nucleótidos se determina como se ha descrito previamente (ver la sección titulada "Definiciones").

[0041] 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°. 0584 es al menos del 80%, tal como al menos del 90%, más preferiblemente al menos del 95%, tal como al menos del 96%, por ejemplo al menos del 97%, incluso más preferiblemente al menos del 98%, tal como al menos del 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°. 0584. 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°. 0584.

[0042] 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° 2. Estas variantes artificiales puede 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.

[0043] 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 mutagénesis dirigida o mutagénesis de escaneado de alanina (ver, 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 (ver, 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).

5 [0044] 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.

10 [0045] 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  
15 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, ver, por ejemplo, Ford et al., 1991, Protein Expression and Purification 2: 95-107 .

20 [0046] 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 baja, más preferiblemente bajo condiciones de astringencia media, más preferiblemente bajo condiciones de astringencia medio-alta, incluso más  
25 preferiblemente bajo condiciones de astringencia alta, y de la forma más preferible bajo condiciones de astringencia altísima con una sonda polinucleótida seleccionada del grupo que consiste en

(i) la cadena complementaria de los nucleótidos seleccionados de:  
nucleótidos 1 a 1578 de SEC ID nº 1,  
nucleótidos 1 a 1302 de SEC ID nº 1,;

30 (ii) la cadena complementaria de los nucleótidos seleccionados de:  
nucleótidos 1 a 500 de SEC ID nº 1 y

(iii) la cadena complementaria de los nucleótidos seleccionados de:  
nucleótidos 1 a 200 de SEC ID nº 1.

35 [0047] 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 ácidos nucleicos

40 [0048] 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.

45 [0049] Una secuencia de nucleótidos que codifica un polipéptido de la presente invención se puede manipular en una variedad de formas para proveer expresión del polipéptido. Manipulación de la secuencia de nucleótidos antes de su inserción 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.

50 [0050] 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 del promotor 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  
55 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.

[0051] 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.

[0052] 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.

[0053] 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 son descritos por Romanos et al., 1992, Yeast 8: 423-488.

[0054] La secuencia de control puede también ser una secuencia del terminador de transcripción adecuada, una secuencia reconocida por una célula huésped para terminar transcripción. La secuencia del terminador es operativamente enlazada al terminal 3' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier terminador que es funcional en la célula huésped de elección se puede utilizar en la presente invención.

[0055] Terminadores preferidos para células huésped fúngicas filamentosas se obtienen de los genes para TAKA amilasa 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*.

[0056] 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.

[0057] 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 terminal 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.

[0058] Los líderes preferidos para células huésped fúngicas filamentosas se obtienen de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.

[0059] Líderes adecuados para células huésped de levadura se obtienen de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), quinasa de 3-fosfoglicerato de *Saccharomyces cerevisiae*, alfa-factor de *Saccharomyces cerevisiae* y de alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH2/GAP).

[0060] La secuencia de control puede también ser una secuencia de poliadenilación, una secuencia operativamente enlazada al terminal 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 es funcional en la célula huésped de elección se puede utilizar en la presente invención.

[0061] Secuencias de poliadenilación preferida 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-glucosidasa de *Aspergillus niger*.

[0062] Secuencias de poliadenilación útil para células huésped de levadura son descritas por Guo and Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.

[0063] 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 foránea a la secuencia codificante. La región de codificación de péptido señal foránea puede ser requerida donde la secuencia codificante no contenga naturalmente una región de codificación de péptido señal. Alternativamente, la región de codificación de péptido señal foráneo puede simplemente reemplazar la región de codificación de péptido señal natural para mejorar 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.

[0064] 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 son descritos por Simonen and Palva, 1993, Microbiological Reviews 57: 109-137 .

[0065] 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*.

[0066] 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.

[0067] La secuencia de control puede también ser una región de codificación de propéptido que codifica para una secuencia de aminoácidos situados en el terminal de amino 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 de codificación de propéptido se puede obtener de los genes para proteasa alcalina de *Bacillus subtilis* (*aprE*), proteasa neutra de *Bacillus subtilis* (*nprT*), alfa-factor de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei*, y lacasa de *Myceliophthora thermophila* (WO 95/33836).

[0068] Donde ambos péptidos señal y regiones de propéptido están presentes en el terminal amino de un polipéptido, la región de propéptido está situada junto al terminal amino de un polipéptido y la región de péptido señal está situada junto al terminal amino de la región de propéptido.

[0069] Puede también ser deseable añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido relativa 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 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 presencia de nucleótidos que codifica el polipéptido sería operativamente enlazada con la secuencia reguladora.

#### Vectores de expresión

[0070] La presente invención también se refiere a vectores de expresión recombinantes que incluyen el constructo de ácidos nucleicos de la descripción. Las varias secuencias de nucleótidos y de control anteriormente descritas pueden ser juntas para producir un vector de expresión recombinante que puede incluir uno o más sitios de restricción convenientes para 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. Al crear el vector de expresión, la secuencia codificante se localiza en el vector de modo que la secuencia codificante es operativamente enlazada con las secuencias de control apropiadas para la expresión.

[0071] El vector de expresión recombinante puede ser cualquier vector (p. ej., un plásmido o virus) que puede ser convenientemente sometido a procedimientos de ADN recombinante y puede 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 puede ser plásmidos lineales o circulares cerrados.

[0072] 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 replicación cromosómica, por ejemplo, un plásmido, un elemento extracromosómico, un minicromosoma o un cromosoma artificial.

[0073] El vector puede contener cualquier medio para asegurar 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 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 a ser introducido en el genoma de la célula huésped o un transposón pueden ser utilizados.

[0074] Los vectores de la presente invención preferiblemente contienen uno o más marcadores seleccionables que permitir selección fácil de células transformadas. Un marcador seleccionable es un gen cuyo producto proporciona biocida o resistencia vírica, resistencia a metales pesados, prototrofia para auxótrofos y similares.

[0075] Ejemplos de marcadores seleccionables bacterianos son los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia antibiótica tal como resistencia a ampicilina, canamicina, cloranfenicol o resistencia. Marcadores adecuados para células huésped de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1 y URA3. Marcadores seleccionables para uso en una célula huésped filamentosa fúngica incluyen, pero de forma no limitativa, *amdS* (acetamidasa), *argB* (ornitina-carbamoiltransferasa), *bar* (fosfonitricina acetiltransferasa), *hygB* (higromicina fosfotransferasa), *niaD* (nitrato-reductasa), *pyrG* (orotidina-5'-fosfato-descarboxilasa), *sC* (adeniltransferasa de sulfato), *trpC* (sintasa de antranilato), al igual que equivalentes de los mismos.

[0076] 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 hygroscopicus*.

[0077] Los vectores de la presente invención preferiblemente contienen un elemento(s) que permite integración estable del vector en el genoma de la célula huésped o replicación autónoma del vector en la célula independiente del genoma.

[0078] Para 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 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 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 1.500 pares de bases, preferiblemente 400 a 1.500 pares de bases y, de la forma más preferible 800 a 1.500 pares de bases, que son altamente homólogos de la secuencia objetivo correspondiente para mejorar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga de 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.

[0079] Para replicación autónoma, el vector puede comprender además un origen de replicación permitiendo al vector replicar 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 pAM $\beta$ 1 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 (ver, por ejemplo, Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75: 1433).

[0080] Más de una copia de una secuencia de nucleótidos de la presente invención se puede insertar en la célula huésped para aumentar producción del producto genético. 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.

[0081] 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 (ver, por ejemplo, Sambrook et al., 1989, supra).

#### 50 Células huésped

[0082] 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 duplica como se describe anteriormente.

[0083] La célula huésped puede ser un microorganismo unicelular, por ejemplo, un microorganismo procariota o uno no unicelular, por ejemplo, un eucariota.

[0084] 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.

[0085] 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 Thorne, 1987, Journal of Bacteriology 169: 5771-5278).

[0086] La célula huésped puede ser un eucariota, tal como un mamífero, insecto, planta o célula fúngica.

[0087] 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).

[0088] En una forma de realización más preferida, la célula huésped fúngica es una célula de levadura. "Yeast" como se utiliza en este caso incluye levadura ascosporógena (Endomicetales), levadura basidioesporogénea, y levadura de los hongos imperfectos (Blastomicetos). 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).

[0089] 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*.

[0090] 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*.

[0091] 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, crecimiento vegetativo por levaduras tal como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y catabolismo de carbono puede ser fermentativo.

[0092] 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*.

[0093] 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*.

[0094] Células fúngicas se pueden transformar por un proceso implicando 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 de *Aspergillus* son descritos en EP 238 023 y Yelton et al., 1984, 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 . Levadura puede ser transformada usando 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

[0095] La presente divulgación también se refiere a métodos para producir un polipéptido de la presente invención que incluye (a) cultivo de una cepa, que en su forma de tipo salvaje es capaz de producir el polipéptido; y (b) recuperación del polipéptido. Preferiblemente, la cepa es seleccionada del grupo que consiste en infestantes de *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 la cepa es seleccionada 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 pruinosum*, *Thielavia* cf. *microspora*, *Pseudoplectania nigrella* y *Phytophthora*.

[0096] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que incluye (a) cultivo de una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

[0097] La presente invención también se refiere a métodos para producción in situ de un polipéptido de la presente invención que incluye (a) cultivo de una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) contacto del polipéptido con un sustrato deseado, tal como un sustrato celulósico, 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.

[0098] 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 está segregado, se puede recuperar de lisatos celulares.

[0099] Los polipéptidos pueden ser detectados usando métodos conocidos en la técnica que son específicos para los polipéptidos. Estos métodos de detección pueden incluir 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.

[0100] 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.

[0101] Los polipéptidos de la presente invención se pueden purificar por una variedad de procedimientos conocidos en la técnica, incluidos, pero no limitados a, cromatografía (p. ej., de intercambio iónico, de afinidad, hidrofóbica, de cromatografía de exclusión de tamaño), procedimientos electroforéticos (p. ej., isoelectroenfoque preparativo), solubilidad diferencial (p. ej., precipitación de sulfato amónico), SDS-PAGE o extracción (ver, por ejemplo, Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

## Plantas

[0102] 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.

[0103] 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.

[0104] 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*.

[0105] 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.

[0106] También están incluidas dentro del campo de la presente invención la progenie (clonal o semilla) de tales plantas, partes de planta y células vegetales.

[0107] 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 propagando la resultante planta modificada o célula vegetal en una planta transgénica o célula vegetal.

[0108] 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 con 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 secuencias de ADN necesarias para introducción del constructo en la planta en cuestión (esto depende del método de introducción de ADN que va a ser usado).

[0109] 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 .

[0110] Para expresión constitutiva, se puede usar el promotor 35S-CaMV (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 (Kyoizuka 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).

[0111] 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 expresión.

[0112] El gen marcador seleccionable y cualquier otra parte del constructo de expresión se puede elegir de aquéllas disponibles en la técnica.

[0113] 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).

[0114] Actualmente, 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 monocotiledóneas de transformación, 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 transformación de monocotiledóneas se basa en transformación de protoplasto como se describe por Omirulleh et al., 1993, Plant Molecular Biology 21: 415-428.

5 [0115] Tras la transformación, los transformantes que tienen el constructo de expresión incorporado se seleccionan y son regenerados en plantas enteras según métodos bien conocidos en la técnica.

10 [0116] La presente invención también se refiere a métodos para producir un polipéptido de la presente invención que consisten en (a) cultivar una planta transgénica o una célula vegetal que comprende una secuencia de nucleótidos que codifican un polipéptido con actividad de celobiohidrolasa I de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.

15 [0117] La presente invención también se refiere a métodos para producción in situ de un polipéptido de la presente invención que consisten en (a) cultivar una planta transgénica o una célula vegetal que incluye 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) contactar el polipéptido con un sustrato deseado, tal como un sustrato celulósico, sin recuperación previa del polipéptido.

#### Composiciones

20 [0118] En otro aspecto ulterior, la presente invención se refiere a composiciones que comprenden un polipéptido de la presente invención.

25 [0119] La composición puede comprender un polipéptido de la invención como el componente enzimático mayor, por ejemplo, una composición monocomponente. Alternativamente, la composición puede comprender múltiples actividades enzimáticas, tales como aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, glicosiltransferasa de ciclodextrina, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, glucosidasa alfa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa o xilanasas.

30 [0120] Las composiciones se pueden preparar conforme a métodos conocidos en la técnica y pueden ser en forma de un líquido o una composición seca. Por ejemplo, la composición del polipéptido puede ser en forma de un granulado o un microgranulado. El polipéptido a ser incluido en la composición se puede estabilizar conforme a métodos conocidos en la técnica.

35 [0121] 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

40 [0122] El polipéptido de la invención se puede adicionar a y así convertirse en un componente de una composición detergente.

45 [0123] 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 pretratamiento de tejidos manchados y una composición de suavizante adicionado 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.

50 [0124] En un aspecto específico, la invención proporciona un aditivo detergente que incluye el polipéptido de la invención. El aditivo detergente al igual que la composición detergente puede comprender uno o más enzimas, tales 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.

55 [0125] En general las propiedades de la enzima(s) elegida deberían ser compatibles con el detergente seleccionado, (es decir, pH óptimo, compatibilidad con otros ingredientes no enzimáticos y enzimáticos, etc.), y la enzima(s) debería estar presente en cantidades eficaces.

60 [0126] Proteasas: proteasas adecuadas incluyen aquellas de origen animal, vegetal o microbiano. El origen microbiano es preferido. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. La proteasa puede ser una serina proteasa o una metalo proteasa, preferiblemente una proteasa alcalina microbiana o una proteasa de tipo tripsina. Ejemplos de proteasas alcalinas son subtilisinas, especialmente aquellas derivadas de *Bacillus*, por ejemplo, subtilisina Novo, subtilisina Carlsberg, subtilisina 309, subtilisina 147 y subtilisina 168 (descritas en WO 89/06279). Ejemplos de proteasas de tipo tripsina son tripsina (p. ej. de origen porcino o bovino) y la proteasa de

*Fusarium* descrita en WO 89/06270 y WO 94/25583.

[0127] 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 posiciones siguientes: 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 proteína son 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).

[0128] Otros ejemplos son variantes de lipasa tales como aquellos 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 (beta y/o alfa) incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son 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.

[0129] 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 posiciones siguientes: 15, 23, 105, 106, 124, 128, 133, 154, 156, 181, 188, 190, 197, 202, 208, 209, 243, 264, 304, 305, 391, 408, y 444.

Celulasas: celulasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Mutantes químicamente modificados o creados genéticamente de proteína son incluidos. Celulasas adecuadas incluyen celulasas del género *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium*, por ejemplo, las celulasas fúngicas producido 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.

[0130] Celulasas especialmente adecuadas son las celulasas neutrales o alcalinas con beneficios de cuidado de color. Ejemplos de tales celulasas son 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 aquellas 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.

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.

[0131] La enzima(s) detergente se puede incluir en una composición detergente añadiendo aditivos separados con una o más enzimas, o por adición de un aditivo combinado que incluya todas estas enzimas. Un aditivo detergente de la invención, es decir, un aditivo separado o un aditivo combinado, se puede formular por ejemplo como un granulado, un líquido, un compuesto acuoso, etc. Las formulaciones de aditivo detergente preferidas son granulados, en particular granulados no polvorientos, líquidos, en particular líquidos estabilizados, o suspensiones acuosas.

[0132] 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 conocido en la técnica. Ejemplos de materiales de recubrimiento ceroso son productos de poli(etileno óxido) (polietilenglicol; PEG) con pesos molares medios de 1.000 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 película adecuados para aplicación por técnicas de lecho fluido se dan en GB 1483591. 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 al método descrito en EP 238,216.

[0133] La composición detergente de la invención puede ser 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 70 % agua y 0-30 % solvente orgánico o no acuoso.

[0134] La composición detergente comprende uno o más tensioactivos, que pueden ser no iónico, 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.

[0135] Cuando se incluye en esto, el detergente normalmente contiene aproximadamente de 1% a aproximadamente 40% de un tensioactivo aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, sulfato de alquilo (sulfato de

alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, éster metílico de ácido graso alfa-sulfo, alquil- o ácido alquenilsuccínico o jabón.

5 [0136] Cuando se incluye en esto, el detergente normalmente contiene aproximadamente de 0,2% a aproximadamente 40% de un tensioactivo no iónico tal como alcohol etoxilato, nonilfenol, etoxilato alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, amida de ácido graso de alquilo de polihidroxi, o derivados de n-alquilo de n-acilo de glucosamina ("glucamidas").

10 [0137] El detergente puede contener 0-65 % de un constructor de detergente o agente complejante tal como zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilendiaminatetraacético, ácido dietileno-triaminopentaacético, ácido alquenilsuccínico o de alquilo, silicatos solubles o silicatos estratificados (p. ej. SKS-6 de Hoechst).

15 [0138] El detergente puede comprender uno o más polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli (etilenglicol), alcohol polivinílico), poli(vinilpiridina-N-óxido), poli(vinilimidazol), policarboxilatos tal como poli-acrilatos, copolímeros de ácido maléico/acrílico y copolímeros de ácido acrílico/lauril metacrilato.

20 [0139] 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 tetraacetil-etilendiamina o nonanoiloxibencenosulfonato. Alternativamente, el sistema blanqueante puede comprender peroxiácidos de, por ejemplo, amida, imida o tipo sulfona.

25 [0140] La enzima(s) de la composición detergente de la invención puede ser estabilizada usando agentes estabilizantes convencionales, por ejemplo, un poliol tal como propilenoglicol 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 borónico de fenilo, tal como ácido borónico de 4-formilfenil, y la composición se puede formular como se describe en, por ejemplo, WO 92/19709 y WO 92/19708.

30 [0141] 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.

35 [0142] Es 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.

40 [0143] El polipéptido de la invención puede adicionalmente ser incorporado en las formulaciones detergentes descritas en WO 97/07202 que es por la presente incorporada como referencia.

#### Recombinación de ADN (redistribución)

45 [0144] Las secuencias de nucleótidos de SEC ID n° 1 se pueden usar en un proceso de recombinación (o de redistribución) de ADN. Las nuevas secuencias polinucleótidas obtenidas en tal proceso pueden codificar polipéptidos nuevos 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 a compuestos específicos.

50 [0145] 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.

55 [0146] Recombinación o redistribución de ADN puede ser un proceso (parcial) 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 este proceso de redistribución o de recombinación.

60 [0147] El proceso puede implicar fragmentación aleatoria de ADN progenitor seguida de reensamblaje por PCR para nuevos genes en toda su longitud, por ejemplo como presentado en US5605793, US5811238, US5830721, US6117679. Recombinación in vitro de genes puede ser realizada, por ejemplo como descrito en US6159687, WO98/41623, US6159688, US5965408, US6153510. El proceso de recombinación puede ocurrir in vivo en una célula viva, por ejemplo como descrito en WO 97/07205 y WO 98/28416.

65 [0148] El ADN progenitor se puede fragmentar por tratamiento de 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).

5 [0149] 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.

Producción de etanol a partir de biomasa

10 [0150] La presente invención también se refiere a métodos para producir etanol a partir de biomasa, tal como materiales celulósicos, incluyendo contacto de la biomasa con los polipéptidos de la invención. 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.

15 [0151] Etanol se puede producir por degradación enzimática de biomasa y conversión de los polisacáridos liberados en etanol. Esta especie de etanol es frecuentemente llamada bioetanol o biocombustible. Esto se puede usar como un aditivo de combustible o suplemento en mezclas de menos de 1% y hasta 100% (un sustituto de combustible). En algunos países, tal como Brasil, etanol está sustituyendo a la gasolina muy ampliamente.

20 [0152] El polisacárido predominante en la pared celular primaria de biomasa es celulosa, el segundo más abundante es hemicelulosa y el tercero es 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. Celulosa es un homopolímero de anhidrocelobiosa y así un beta-(1-4)-D-glucano lineal, mientras 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 de cristalino insoluble de cadenas de glucano paralelo. Hemicelulosas normalmente enlazadas por hidrógeno a celulosa, al igual que a otras hemicelulosas, lo que ayuda a estabilizar la matriz de pared celular.

30 [0153] 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 insoluble y soluble.  
 35 - las "exo-1,4-beta-D-glucanasas" que incluye 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.

40 [0154] Estas tres clases de enzimas trabajan juntas sinérgicamente en una interacción compleja que produce decristalización eficaz e hidrólisis de celulosa nativa de biomasa para producir los azúcares reductores que se convierten en etanol por fermentación.

45 [0155] La presente invención es posteriormente descrita por los siguientes ejemplos.

EJEMPLOS

50 [0156] 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)**

55 [0157] Una genoteca de ADNc de *Diplodia gossypina* fue seleccionada mediante PCR para presencia del gen CBH1. Para este propósito conjuntos de cebadores fueron contruidos, basados en alineación de secuencia e identificación de regiones conservadas entre proteínas de CBH1. La banda de PCR de una electrofóresis 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).

60 [0158] El gen CBH1 en toda su longitud de *Diplodia gossypina* se obtiene por acceso del depósito de patente CBS 247,96, hacer una preparación de ADN o ADNc, usar la secuencia parcial como base para construcción de cebadores específicos y usar técnicas de clonación de PCR estándar para paso a paso obtener el gen entero.

65 [0159] Otros métodos diferentes puede 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 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*. Búsqueda de homología y 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 las extremidades 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. Búsqueda de homología y comparación confirma que el ADNc en toda su longitud corresponde a la secuencia parcial de ADNc de CBH1 que fue originalmente usada como una sonda.

[0160] 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, se aíslan preferiblemente ADNm de *Diplodia gossypina* y se utilizan para sintetizar primero cepas de ADNc usando cebador de adaptador conteniendo oligo(dT)- o un 5'- cebador específico de gen (GSP).

[0161] 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. Búsqueda de homología y comparación con el ADNc de CBH1 parcial confirma 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).

[0162] 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.

[0163] 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)

[0164] 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 puede utilizarse para identificar una celobiohidrolasa I:

Actividad en Azo-Avicel

[0165] Azo-Avicel (Megazyme, Bray Business Park, Bray, Wicklow, Ireland) fue usado según las instrucciones de los fabricantes.

Actividad en PNP-beta-celobiosa

[0166]

Solución de sustrato: 5 mM PNP beta-D-celobiosa (p-nitrofenil  $\beta$ -D-Celobiosida Sigma N-5759) en 0,1 M tampón de Na-acetato, pH 5,0;  
Reactivo de parada: 0.1 M Na-carbonato, pH 11.5.

[0167] Solución de 50  $\mu$ L CBH I fue mezclada con 1 mL de solución de sustrato e incubada 20 minutos a 40°C. La reacción fue detenida por adición de 5 mL de reactivo de parada. Absorbancia fue medida a 404 nm.

Actividad en PASC y CMC

[0168] El sustrato se degrada con celobiohidrolasa I (CBH I) para formar azúcares reductores. Un Microdochiumnival (rMnO) de carbohidrato oxidasa 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 formado H<sub>2</sub>O<sub>2</sub> activa en presencia de peroxidasa en exceso la condensación oxidante de 4-

aminoantipirina (AA) y N-etil-N-sulfopropil-m-toluidina (TOPS) para formar un producto morado que se puede cuantificar por su absorbancia a 550 nm.

5 [0169] 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.

Tampón: 50 mM tampón de Na-acetato (pH 5,0);

10 Reactivos: rMnO oxidasa, carbohidrato oxidasa purificada de *Microdochium nivale*, 2 mg/L (concentración final);  
 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).

15 [0170] Todos los reactivos fueron adicionados al tampón en las concentraciones indicadas por encima y esta solución reactiva fue mezclada íntegramente.

[0171] 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. Formación de color morado fue detectada y medida como absorbancia a 550 nm.

20 [0172] El coeficiente de absorción de AA-TOPS-condensado es 0,01935 A550 / (µM cm). El índice se calcula como µmols que reducen azúcar producido por minuto de OD550 / minuto y el coeficiente de absorción.

PASC:

25 [0173]

30 Materiales: 5 g de Avicel® (Art. 2331 Merck);  
 150 mL 85% de Orto-fosfórico-ácido (Art. 573 Merck);  
 800 mL de acetona (Art. 14 Merck);  
 Aprox. 2 litros de agua desionizada (Milli-Q);  
 Vaso de precipitado de vidrio de 1 litro;  
 Embudo de filtro de vidrio de 1 litro;  
 Matraz de succión de 2 litros;  
 35 Homogeneizador UltraTurrax.

[0174] Acetona y orto-fosfórico-ácido se enfría en el hielo. Avicel® se humedece con agua y luego 150 mL 85% orto-fosfórico-ácido helado se añade. La mezcla se coloca en un baño de hielo con agitación débil durante una hora.

40 [0175] Añadir 500 mL de acetona helada con agitación y transferir la mezcla a un embudo de filtro de vidrio y lavar con 3 x 100 mL de acetona helada, aspirar lo más seco posible en cada lavado. Lavar con 2 x 500 mL de agua (o hasta no haya olor de acetona), aspirar lo más seco posible en cada lavado.

45 [0176] Resuspender los sólidos en el agua a un volumen total de 500 mL y mezclar para homogeneidad usando un homogeneizador Ultra Turrax. Guardar húmedo en el frigorífico y equilibrar con tampón por centrifugado y resuspensión antes de uso.

CMC:

50 [0177] 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.

55 [0178] Luego 1% (p/v) de NaOH fue usado en vez de agua y el producto fue resuspendido en la solución de alcali dos veces al día durante 4 días. Neutralización fue hecha por aclarado de la celulosa purificada con agua destilada hasta que el pH de la superficie del producto fue neutral (pH 7).

60 [0179] La celulosa fue microfibrilada y una suspensión de microfibrillas de celulosa bacteriana individual 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.

Depósito de material biológico

65 China General Microbiological Culture Collection Center (CGMCC)

[0180] 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:

5  
 Número de accesión: CGMCC nº. 0584  
 Referencia de solicitantes: ND000575  
 Fecha de depósito: 29-05-2001  
 Descripción: gen de CBH I de *Acremonium thermophilum* en plásmido  
 10 Clasificación: *Ascomycota; Sordariomycetes; Hypocrales; Hypocreaceae*  
 Origen: China, 1999  
 Secuencia(s) relacionada: SEC ID nº 1 y SEC ID nº 2 (secuencia de ADN que codifica una celobiohidrolasa I de *Acremonium thermophilum* y la secuencia de proteína correspondiente)

15  
 Listado de secuencias

[0181]

20  
 <110> Novozymes A/S  
 <120> Polipéptidos con actividad de celobiohidrolasa I y polinucleótidos que codifican la misma

<130> 10129-WO

25  
 <160> 2

<170> Versión de patentIn 3.1

<210> 1

30  
 <211> 1581

<212> ADN

<213> *Acremonium thermophilum*

<220>

35  
 <221> CDS

<222> (1)..(1581)

<223>

<400> 1

40

ES 2 397 267 T3

atg cac gcc aag ttc gcg acc ctc gcc gcc ctt gtg gog tcc gcc gcg	48
Met His Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ser Ala Ala	
1 5 10 15	
gcc cag cag gcc tgc aca ctc acg gct gag aac cac ccc acc ctg tgg	96
Ala Gln Gln Ala Cys Thr Leu Thr Ala Glu Asn His Pro Thr Leu Ser	
20 25 30	
tgg tcc aag tgc acg tcc gcc gcc agc tgc acc agc gtc tgg ggc tcc	144
Trp Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Ser Gly Ser	
35 40 45	
gtc acc atc gat gcc aac tgg cgg tgg act cac cag gtc tgg agc tgg	192
Val Thr Ile Asp Ala Asn Trp Arg Trp Thr His Gln Val Ser Ser Ser	
50 55 60	
acc aac tgc tac acg gcc aat gag tgg gac acg tcc atc tgc acc gac	240
Thr Asn Cys Tyr Thr Gly Asn Glu Trp Asp Thr Ser Ile Cys Thr Asp	
65 70 75 80	
ggt gct tgg tgc gcc gcc gcc tgc tgc ctc gat gcc gcc gac tac tgg	288
Gly Ala Ser Cys Ala Ala Ala Cys Cys Leu Asp Gly Ala Asp Tyr Ser	
85 90 95	
ggc acc tat ggc atc acc acc agc gcc aac gcc ctc agc ctc cag ttc	336
Gly Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Ser Leu Gln Phe	
100 105 110	
gtc act cag gcc ccc tac tgg acc aac att gcc tgg cgt acc tac ctg	384
Val Thr Gln Gly Pro Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu	
115 120 125	
atg gcc tgg gac acc aag tac cag atg ttc act ctg ctc gcc aac gag	432
Met Ala Ser Asp Thr Lys Tyr Gln Met Phe Thr Leu Leu Gly Asn Glu	
130 135 140	
ttc acc ttc gac gtg gac gtc aca gcc ctc gcc tgc ggt ctg aac gcc	480

ES 2 397 267 T3

Phe Thr Phe Asp Val Asp Val Thr Gly Leu Gly Cys Gly Leu Asn Gly	
145	150 155 160
gcc ctc tac ttc gtc tcc atg gac gag gac ggt ggt ctt tcc aag tac	528
Ala Leu Tyr Phe Val Ser Met Asp Glu Asp Gly Gly Leu Ser Lys Tyr	
	165 170 175
tcg ggc aac aag gct ggc gcc aag tac ggc acc ggc tac tgc gac tcg	576
Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser	
	180 185 190
cag tgc ccc cgc gac ctc aag ttc atc aac ggc gag gct aac aac gtt	624
Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Asn Val	
	195 200 205
ggc tgg acc ccg tcg tcc aac gac aag aac gcc ggc ttg ggc aac tac	672
Gly Trp Thr Pro Ser Ser Asn Asp Lys Asn Ala Gly Leu Gly Asn Tyr	
	210 215 220
ggc agc tgc tgc tcc gag atg gat gtc tgg gag gcc aac agc atc tcg	720
Gly Ser Cys Cys Ser Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser	
	225 230 235 240
gcg gcc tac acg ccc cat cct tgc act acc atc ggc cag acg cgc tgc	768
Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Thr Arg Cys	
	245 250 255
gag ggc gac gac tgc ggt ggt acc tac agc act gac cgc tac gcc ggc	816
Glu Gly Asp Asp Cys Gly Gly Thr Tyr Ser Thr Asp Arg Tyr Ala Gly	
	260 265 270
gag tgc gac cct gac gga tgc gac ttc aac tcg tac cgc atg gcc aac	864
Glu Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asn	
	275 280 285
acg acc ttc tac ggc aag ggc atg acc gtc gac acc agc aag aag ttc	912
Thr Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Ser Lys Lys Phe	
	290 295 300
acg gtg gtg acc cag ttc ctg acg gac tcg tct ggc aac ctg tcc gag	960
Thr Val Val Thr Gln Phe Leu Thr Asp Ser Ser Gly Asn Leu Ser Glu	
	305 310 315 320
atc aag cgc ttc tac gtc cag aac ggc gtc gtc att ccc aac tcg aac	1008
Ile Lys Arg Phe Tyr Val Gln Asn Gly Val Val Ile Pro Asn Ser Asn	
	325 330 335
tcc aac atc gcg ggc gtc tcg ggc aac tcc atc acc cag gcc ttc tgc	1056
Ser Asn Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Gln Ala Phe Cys	
	340 345 350
gat gct cag aag acc gct ttc ggc gac acc aac gtc ttc gac caa aag	1104
Asp Ala Gln Lys Thr Ala Phe Gly Asp Thr Asn Val Phe Asp Gln Lys	
	355 360 365
ggc ggc ctg gcc cag atg ggc aag gct ctt gcc cag ccc atg gtc ctc	1152
Gly Gly Leu Ala Gln Met Gly Lys Ala Leu Ala Gln Pro Met Val Leu	
	370 375 380
gtc atg tcc ctc tgg gac gac cac gcc gtc aac atg ctc tgg ctc gac	1200
Val Met Ser Leu Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp	
	385 390 395 400

ES 2 397 267 T3

tcc acc tac ccg acc aac ggc gcc ggc aag ccg ggc gcc gcc cgc ggt 1248  
 Ser Thr Tyr Pro Thr Asn Ala Ala Gly Lys Pro Gly Ala Ala Arg Gly  
 405 410 415  
 acc tgc ccc acc acc tcc ggc gtc ccc gcc gac gtc gag tcc cag gcg 1296  
 Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu Ser Gln Ala  
 420 425 430  
 ccc aac tcc aag gtc atc tac tcc aac atc cgc ttc ggc ccc atc ggc 1344  
 Pro Asn Ser Lys Val Ile Tyr Ser Asn Ile Arg Phe Gly Pro Ile Gly  
 435 440 445  
 tcc acc gtc tcc ggc ctg ccc ggc ggc ggc agc aac ccc ggc ggc ggc 1392  
 Ser Thr Val Ser Gly Leu Pro Gly Gly Gly Ser Asn Pro Gly Gly Gly  
 450 455 460  
 tcc agc tcc acc acc acc acc acc acc aga ccc gcc acc tcc acc acc tcc 1440  
 Ser Ser Ser Thr Thr Thr Thr Thr Arg Pro Ala Thr Ser Thr Thr Ser  
 465 470 475 480  
 tcc gcc agc tcc ggc ccg acc ggc ggt ggc acg gct gcc cac tgg ggc 1488  
 Ser Ala Ser Ser Gly Pro Thr Gly Gly Gly Thr Ala Ala His Trp Gly  
 485 490 495  
 cag tgc ggc ggc atc ggc tgg acc ggc ccg acc gtc tgc gcc tcc ccc 1536  
 Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Val Cys Ala Ser Pro  
 500 505 510  
 tac acc tgc cag aag ctg aac gac tgg tac tac cag tgc ctc taa 1581  
 Tyr Thr Cys Gln Lys Leu Asn Asp Trp Tyr Tyr Gln Cys Leu  
 515 520 525

<210> 2  
 <211> 526  
 <212> PRT  
 <213> Acremonium thermophilum

<400> 2

Met His Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Ser Ala Ala  
 1 5 10 15  
 Ala Gln Gln Ala Cys Thr Leu Thr Ala Glu Asn His Pro Thr Leu Ser  
 20 25 30  
 Trp Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Ser Gly Ser  
 35 40 45  
 Val Thr Ile Asp Ala Asn Trp Arg Trp Thr His Gln Val Ser Ser Ser  
 50 55 60  
 Thr Asn Cys Tyr Thr Gly Asn Glu Trp Asp Thr Ser Ile Cys Thr Asp  
 65 70 75 80  
 Gly Ala Ser Cys Ala Ala Ala Cys Cys Leu Asp Gly Ala Asp Tyr Ser  
 85 90 95  
 Gly Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ala Leu Ser Leu Gln Phe  
 100 105 110

Val Thr Gln Gly Pro Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu  
 115 120 125  
 Met Ala Ser Asp Thr Lys Tyr Gln Met Phe Thr Leu Leu Gly Asn Glu  
 130 135 140  
 Phe Thr Phe Asp Val Asp Val Thr Gly Leu Gly Cys Gly Leu Asn Gly  
 145 150 155 160  
 Ala Leu Tyr Phe Val Ser Met Asp Glu Asp Gly Gly Leu Ser Lys Tyr  
 165 170 175  
 Ser Gly Asn Lys Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser  
 180 185 190  
 Gln Cys Pro Arg Asp Leu Lys Phe Ile Asn Gly Glu Ala Asn Asn Val  
 195 200 205  
 Gly Trp Thr Pro Ser Ser Asn Asp Lys Asn Ala Gly Leu Gly Asn Tyr  
 210 215 220  
 Gly Ser Cys Cys Ser Glu Met Asp Val Trp Glu Ala Asn Ser Ile Ser  
 225 230 235 240  
 Ala Ala Tyr Thr Pro His Pro Cys Thr Thr Ile Gly Gln Thr Arg Cys  
 245 250 255  
 Glu Gly Asp Asp Cys Gly Gly Thr Tyr Ser Thr Asp Arg Tyr Ala Gly  
 260 265 270  
 Glu Cys Asp Pro Asp Gly Cys Asp Phe Asn Ser Tyr Arg Met Gly Asn  
 275 280 285  
 Thr Thr Phe Tyr Gly Lys Gly Met Thr Val Asp Thr Ser Lys Lys Phe  
 290 295 300  
 Thr Val Val Thr Gln Phe Leu Thr Asp Ser Ser Gly Asn Leu Ser Glu  
 305 310 315 320  
 Ile Lys Arg Phe Tyr Val Gln Asn Gly Val Val Ile Pro Asn Ser Asn  
 325 330 335  
 Ser Asn Ile Ala Gly Val Ser Gly Asn Ser Ile Thr Gln Ala Phe Cys  
 340 345 350  
 Asp Ala Gln Lys Thr Ala Phe Gly Asp Thr Asn Val Phe Asp Gln Lys  
 355 360 365  
 Gly Gly Leu Ala Gln Met Gly Lys Ala Leu Ala Gln Pro Met Val Leu  
 370 375 380  
 Val Met Ser Leu Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp  
 385 390 395 400  
 Ser Thr Tyr Pro Thr Asn Ala Ala Gly Lys Pro Gly Ala Ala Arg Gly  
 405 410 415  
 Thr Cys Pro Thr Thr Ser Gly Val Pro Ala Asp Val Glu Ser Gln Ala  
 420 425 430  
 Pro Asn Ser Lys Val Ile Tyr Ser Asn Ile Arg Phe Gly Pro Ile Gly  
 435 440 445

ES 2 397 267 T3

Ser Thr Val Ser Gly Leu Pro Gly Gly Gly Ser Asn Pro Gly Gly Gly  
 450 455 460

Ser Ser Ser Thr Thr Thr Thr Thr Arg Pro Ala Thr Ser Thr Thr Ser  
 465 470 475 480

Ser Ala Ser Ser Gly Pro Thr Gly Gly Gly Thr Ala Ala His Trp Gly  
 485 490 495

Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Val Cys Ala Ser Pro  
 500 505 510

Tyr Thr Cys Gln Lys Leu Asn Asp Trp Tyr Tyr Gln Cys Leu  
 515 520 525

**REIVINDICACIONES**

1. Polipéptido con actividad de celobiohidrolasa I, seleccionado del grupo que consiste en:  
5 (a) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 80% de identidad con los aminoácidos 1 a 526 de SEC ID nº 2, y  
(b) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos 80% de identidad con el polipéptido codificado por los nucleótidos 1 a 1.578 de SEC ID nº 1.
2. Polipéptido según la reivindicación 1, que incluye una secuencia de aminoácidos que tiene al menos 85% de  
10 identidad, preferiblemente al menos 90% de identidad, más preferiblemente al menos 95% de identidad, con los aminoácidos 1 a 526 de SEC ID nº 2.
3. Polipéptido según cualquiera de las reivindicaciones 1 ó 2, que consiste en los aminoácidos 1 a 526 de SEC ID nº 2.
- 15 4. Polinucleótido aislado con una secuencia de nucleótidos que codifica para el polipéptido definido en cualquiera de las reivindicaciones 1-3.
5. Vector de expresión recombinante que incluye un constructo de ácidos nucleicos, dicho constructo incluye la  
20 secuencia de nucleótidos definida en la reivindicación 4 operativamente enlazada a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.
6. Célula huésped recombinante que incluye un constructo de ácidos nucleicos, dicho constructo incluye la secuencia de  
25 nucleótidos definidos en la reivindicación 4 operativamente enlazada a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado, donde la célula huésped es seleccionada del grupo que consiste en *Aspergillus*, *Fusarium*, *Humicola*, *Mucor*, *Myceliophthora*, *Neurospora*, *Penicillium*, *Thielavia*, *Tolyposcladium* y *Trichoderma*.
7. Método para la producción de un polipéptido como está definido en cualquiera de las reivindicaciones 1-3, el método  
30 incluye:  
(a) cultivar una célula huésped recombinante tal y como se define en la reivindicación 8 bajo condiciones propicias para la producción del polipéptido; y  
(b) recuperar el polipéptido.
8. Método para la producción in situ de un polipéptido como está definido en cualquiera de las reivindicaciones 1-3, el  
35 método incluye:  
(a) cultivar una célula huésped recombinante tal y como se define en la reivindicación 8 bajo condiciones propicias para la producción del polipéptido; y  
(b) contactar el polipéptido con un sustrato deseado sin recuperación previa del polipéptido.
- 40 9. Método para la producción de etanol de biomasa, que incluye contacto de la biomasa con el polipéptido como se define en cualquiera de las reivindicaciones 1-3.
10. Uso del polipéptido tal y como se define en cualquiera de las reivindicaciones 1-3 para producir etanol.
- 45 11. Composición detergente que comprende un tensioactivo y el polipéptido según cualquiera de las reivindicaciones 1-3.