(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 331 067**

(51) Int. Cl.:
C12N 9/42 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Número de solicitud europea: **03767478 .5**
(96) Fecha de presentación : **19.12.2003**
(97) Número de publicación de la solicitud: **1578964**
(97) Fecha de publicación de la solicitud: **28.09.2005**

(54) Título: **Polipéptidos con actividad celobiohidrolasa II y polinucleótidos que los codifican.**

(30) Prioridad: **20.12.2002 US 435100 P**

(45) Fecha de publicación de la mención BOPI:
21.12.2009

(45) Fecha de la publicación del folleto de la patente:
21.12.2009

(73) Titular/es: **Novozymes A/S**
Krogshojvej 36
2880 Bagsvaerd, DK

(72) Inventor/es: **Wu, Wenping;**
Lange, Lene;
Skovlund, Dominique, Aubert y
Liu, Ye

(74) Agente: **Tomás Gil, Tesifonte Enrique**

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).

Venta de fascículos: Oficina Española de Patentes y Marcas. Pº de la Castellana, 75 – 28071 Madrid

DESCRIPCIÓN

Polipéptidos con actividad celobiohidrolasa II y polinucleótidos que los codifican.

Campo de la invención

La presente invención se refiere a polipéptidos que tienen actividad celobiohidrolasa II (también referida como CBH II o CBH 2) y polinucleótidos que tienen una secuencia de nucleótidos que codifica para los polipéptidos. La invención también se refiere a constructos de ácidos nucleicos, vectores, y células huéspedes que comprenden los constructos de ácidos nucleicos al igual que métodos para producir y usar los polipéptidos.

Antecedentes de la invención

La celulosa es una materia prima industrial importante y una fuente de energía renovable. La estructura física y la morfología de la celulosa nativa son complejas y los detalles finos de su estructura han sido difíciles de determinar experimentalmente. No obstante, la composición química de la celulosa es simple, consistiendo en residuos de D-glucosa enlazados por enlaces beta-1,4-glicosídicos para formar polímeros lineales con longitud de cadenas de unos 10.000 residuos glicosídicos.

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 la celulosa en glucosa: endo (1,4)-beta-D-glucanasas (EC 3,2,1,4) que cortan las cadenas de celulosa al azar; celobiohidrolasas (EC 3,2,1,91) que seccionan unidades de celobiosilo 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, las celobiohidrolasas son las enzimas clave para la degradación de la celulosa cristalina nativa.

Exo-celobiohidrolasas (celobiohidrolasa II, o CBH II) se refieren a las celobiohidrolasas que degradan celulosa para hidrolizar la celobiosa del extremo no reducido de las cadenas poliméricas de celulosa. El grupo celobiohidrolasa II pertenece al mismo grupo EC, es decir EC 3.2.1.91, como el grupo celobiohidrolasa I, la diferencia es que la celobiohidrolasa I degrada la celulosa hidrolizando la celobiosa del extremo de reducción de las cadenas poliméricas de celulosa.

La celobiohidrolasa II y el uso de la misma son conocidas de p. ej. EERI *et al* (GENE. vol. 51,1, 1987, páginas 43-52), 6,114,158, WO 91/17244, EP 214 971, y EP 851 031.

Es un objeto de la presente invención proporcionar polipéptidos mejorados con actividad celobiohidrolasa II y polinucleótidos que codifican los polipéptidos. Los polipéptidos mejorados pueden tener actividad específica mejorada y/o estabilidad mejorada - en particular termostabilidad mejorada. Los polipéptidos pueden tener también una capacidad mejorada para resistir la inhibición por celobiosa.

Resumen de la invención

En un primer aspecto la presente invención se refiere a un polipéptido que tiene actividad celobiohidrolasa II, seleccionado del grupo que consiste en: (a) un polipéptido que comprende una secuencia de aminoácidos, dicha secuencia de aminoácidos tiene al menos un 90% de identidad con la secuencia de aminoácidos mostrada como aminoácidos 1 a 477 de SEC ID NO:2, (b) un polipéptido que comprende una secuencia de aminoácidos, dicha secuencia de aminoácidos tiene al menos un 90% de identidad con el polipéptido codificado por los nucleótidos 63 a 1493 de SEC ID NO:1, (c) un polipéptido que es codificado por una secuencia de nucleótidos que se híbrida bajo condiciones de astringencia muy alta, (i) la cadena complementaria de los nucleótidos 63 a 1493 de SEC ID NO:1, (ii) la cadena complementaria de los nucleótidos 63 a 563 de SEC ID NO:1, o (iii) la cadena complementaria de los nucleótidos seleccionados del grupo que consiste en los nucleótidos 63 a 263 de SEC ID NO:1, donde condiciones de astringencia muy alta comprenden prehibridación e hibridación a 42ºC en 5X SSPE, 1.0% SDS, 5X solución de Denhardt, 100 microg/ml de ADN de esperma de salmón cortado y desnaturalizado, y finalmente lavado tres veces del material portador cada vez durante 15 minutos usando 0,1 X SSC, 0.1% SDS a 68ºC, o (d) un fragmento de (a) o (b) con actividad celobiohidrolasa II.

En un segundo aspecto la presente invención se refiere a un polinucleótido que tiene una secuencia de nucleótidos que codifica para el polipéptido definido en el primer aspecto.

En un tercer aspecto la presente invención se refiere a un constructo de ácidos nucleicos que comprende la secuencia de nucleótidos definida en el segundo aspecto operativamente enlazada a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.

En un cuarto aspecto la presente invención se refiere a un vector de expresión recombinante que comprende el constructo de ácidos nucleicos definido en el tercer aspecto.

En un quinto aspecto la presente invención se refiere a una célula huésped recombinante que comprende el constructo de ácidos nucleicos definido en el tercer aspecto.

En un sexto aspecto la presente invención se refiere a un método para la producción de un polipéptido tal y como se define en el primer aspecto, el método comprendiendo: (a) cultivar una célula huésped recombinante tal y como se define en el quinto aspecto bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.

En un séptimo aspecto la presente invención se refiere a un método para la producción *in situ* de un polipéptido tal y como se define en el primer aspecto en cualquiera de las reivindicaciones, el método comprendiendo: (a) cultivo de una célula huésped recombinante tal y como se define en el quinto aspecto bajo condiciones propicias para la producción del polipéptido; y (b) puesta en contacto del polipéptido con un sustrato deseado sin recuperación previa del polipéptido.

En un octavo aspecto la presente invención se refiere a un método para redistribuir el ADN comprendiendo usar el polinucleótido tal y como se define en el segundo aspecto.

En un noveno aspecto la presente invención se refiere a un método para la producción de etanol a partir de la biomasa, comprendiendo la puesta en contacto de la biomasa con el polipéptido tal y como se define en el primer aspecto.

En un décimo aspecto la presente invención se refiere a un uso del polipéptido tal y como se define en el primer aspecto para producir etanol.

En un decimoprimer aspecto la presente invención se refiere a una planta transgénica, parte de la planta o célula vegetal, que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad celobiohidrolasa II tal y como se define en el primer aspecto.

En un decimosegundo aspecto la presente invención se refiere a una composición detergente que comprende un tensioactivo y el polipéptido según cualquiera del primer aspecto.

Otros aspectos de la presente invención serán aparentes a partir de la descripción más abajo y de las reivindicaciones anexas.

Definiciones

Antes de describir la presente invención con detalles adicionales, los términos siguientes y convenciones serán definidos en primer lugar:

Polipéptido substancialmente puro: en el contexto presente, el término "polipéptido sustancialmente puro" significa una preparación de polipéptido que contiene como máximo el 10% en peso de otro material polipeptídico con el cual está originalmente asociado (porcentajes inferiores de otro material polipeptídico son preferidos, p. ej. como mucho el 8% en peso, como mucho el 6% en peso, como mucho el 5% en peso, como mucho el 4% como mucho el 3% en peso, como mucho 2% en peso, como mucho el 1% en peso, y como mucho el 0.5% en peso). Por lo tanto, se prefiere que el polipéptido substancialmente puro tenga al menos el 92% de pureza, es decir que el polipéptido constituya al menos el 92% en peso del material polipeptídico total presente en la preparación, y porcentajes más altos son preferidos tales como al menos el 94% de pureza, al menos 95% de pureza, al menos el 96% de pureza, al menos el 96% de pureza, al menos el 97% de pureza, al menos el 98% de pureza, al menos 99%, y como mucho el 99.5% de pureza. 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 polipeptídica esté esencialmente libre de otro material polipeptídico con el cual está asociada originalmente. Esto puede ser realizado, por ejemplo, preparando el polipéptido mediante métodos recombinantes bien conocidos. Aquí, el término "polipéptido esencialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada".

Actividad celobiohidrolasa II: el término "actividad celobiohidrolasa II" se define aquí como una celulosa 1,4-beta-cellobiosidasa (también referida como actividad exo-glucanasa, Exocelobiohidrolasa o 1,4-beta-celobiohidrolasa), tal y como se define en la clase enzimática EC 3,2,1,91 o familia CAZy familia 6 de glucósido hidrolasa, que cataliza la hidrólisis de enlaces 1,4-beta-D-glucosídicos en celulosa y celotetraosa, liberando celobiosa de las extremidades reductoras de las cadenas.

Para objetivos de la presente invención, la actividad celobiohidrolasa II puede ser determinada según el procedimiento descrito en el ejemplo 3.

En una forma de realización, la actividad celobiohidrolasa II puede ser determinada según el procedimiento descrito en Deshpande MV *et al*., Methods in Enzymology, págs. 126-130 (1988): "Selective Assay for Exo-1,4-Beta-Glucanases". Según este procedimiento, una unidad de actividad celobiohidrolasa II (actividad de ruptura de enlaces aglucónicos) es definida como 1.0 micromol de p-nitrofenol producida por minuto a 50ºC, pH 5.0. Los polipéptidos de la presente invención deberían preferiblemente tener un mínimo del 20% de la actividad celobiohidrolasa II del polipéptido mostrado en la SEC ID NO:2. En una forma de realización preferida particular, los polipéptidos deberían tener al menos el 40%, tal como al menos el 50%, preferiblemente al menos el 60%, tal como al menos el 70%, más

preferiblemente al menos el 80%, tal como al menos el 90%, de la forma más preferible al menos el 95%, tal como aproximadamente o al menos el 100% de la actividad celobiohidrolasa II del polipéptido mostrado en los aminoácidos 1 a 477 de SEC ID NO:2.

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

Para los 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 de programa FASTA (véase W. R. Pearson y D. J. Lipman (1988), "Improved Tools for Biological Sequence Analysis", PNAS 85:2444-2448; y 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 de espacio fue -12, y penalización por extensión de espacio fue -2.

El grado de identidad entre dos secuencias de nucleótidos es determinado usando el mismo algoritmo y paquete de software que se ha descrito anteriormente. La matriz de marcado usada fue la matriz de identidad, penalización de espacio fue -16, y penalización por extensión de espacio fue -4.

Fragmento: cuando se usa aquí, un "fragmento" de SEC ID NO:2 es un polipéptido que tiene uno o más aminoácidos delecionados del término amino y/o carboxilo de esta secuencia de aminoácidos.

Variante alélica: en el contexto presente, el término "variante alélica" denota cualquiera de dos o más formas alternativas de un gen que ocupa el mismo locus cromosómico. La variación alélica surge naturalmente a través de la mutación, y puede suponer polimorfismo dentro de poblaciones. Las mutaciones de genes 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 sustancialmente puro" como se utiliza en este caso se refiere a una preparación de polinucleótido, donde el polinucleótido ha sido eliminado de su ambiente natural genético, y está por tanto libre de otras secuencias de codificación extrañas o secuencias de codificación indeseadas y está en una forma adecuada para el uso dentro de sistemas de producción de proteínas creadas genéticamente. Por lo tanto, un polinucleótido substancialmente puro contiene como mucho el 10% en peso de otro material polinucleótido con el cual está originalmente asociado (porcentajes inferiores de otro material polinucleótido son preferidos, p. ej. como mucho el 8% en peso, como mucho el 6% en peso, como mucho el 5% en peso, como mucho el 4% como mucho el 3% en peso, como mucho el 2% en peso, como mucho el 1% en peso, y como mucho el 0.5% en peso). Un polinucleótido substancialmente puro puede, no obstante, incluir regiones no codificantes de origen natural 5' y 3', tal como promotores y terminadores. Se prefiere que el polinucleótido substancialmente puro tenga al menos el 92% de pureza, es decir que el polinucleótido constituya al menos el 92% en peso del material de polinucleótido total presente en la preparación, y porcentajes más altos son preferidos tales como al menos el 94% de pureza, al menos 95% de pureza, al menos 96% de pureza, al menos 96% de pureza, al menos 97% de pureza, al menos 98% de pureza, al menos 99%, y como mucho el 99.5% de pureza. 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 de polinucleótidos esté esencialmente libre de otro material polinucleótido con el cual está originalmente asociada. Aquí, el término "polinucleótido sustancialmente 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)" se destina a significar cualquier modificación química de un polipéptido que consiste en la SEC ID NO:2, al igual que manipulación genética del ADN que codifica este polipéptido. La(s) modificación(es) puede(n) ser sustitución(es) de la(s) cadena(s) lateral(es) del aminoácido, sustitución(es), eliminación(s) y/o inserción(s) en el(los) aminoácido(s) de interés.

Variante artificial: cuando se usa aquí, el término "variante artificial" significa un polipéptido que tiene actividad celobiohidrolasa II, que ha sido producido por un organismo que está expresando un gen modificado en comparación con la SEC ID NO:1. El gen modificado, donde dicha variante es producida cuando se expresa en un huésped adecuado, es obtenida a través de la intervención humana por modificación de una secuencia de nucleótidos que consiste en la SEC ID NO:1.

ADNc: el término "ADNc" cuando se usa en el contexto presente, se destina a cubrir una molécula de ADN que puede ser preparada por transcripción inversa de una molécula de ARNm madura, dividida, derivada de una célula eucariótica. El ADNc carece de las secuencias de intrón que están normalmente presentes en el ADN genómico correspondiente. El transcrito de ARN inicial primario 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 el ADNc es derivado de ARNm en consecuencia carece de secuencias de intrones.

Constructo de ácidos nucleicos: cuando se usa aquí, el término "constructo de ácidos nucleicos" significa una molécula de ácido nucleico, bien mono o bicatenario, que es aislada de un gen de origen natural o que ha sido modificada para contener segmentos de ácidos nucleicos de tal modo que de lo contrario no existirían en la naturaleza. El término constructo de ácidos nucleicos es sinónimo del término "cassette de expresión" cuando el constructo de ácidos

nucleicos contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención.

Secuencia de control: el término "secuencias de control" se define 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 externa a la secuencia de nucleótidos que codifica el polipéptido. Tales secuencias de control incluyen, pero no se limitan a, una secuencia líder, de poliadenilación, secuencia de propéptido, de promotor, secuencia de péptido señal, y de terminador de la transcripción. Como mínimo, las secuencias de control incluyen un promotor, y señales de parada transcripcional y traduccional. Las secuencias de control pueden ser provistas de enlaces con el fin de introducir sitios de restricción específicos que facilitan 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" está definido aquí como una configuración donde una secuencia de control está apropiadamente colocada a una posición con respecto a la secuencia codificante de la secuencia de ADN de manera 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 directamente específica la secuencia de aminoácidos de su producto proteínico. Los bordes de la secuencia codificante son generalmente determinados por un marco de lectura abierto, que normalmente empieza con el codón de iniciación ATG. La secuencia codificante normalmente incluye ADN, ADNc, y secuencias de nucleótidos recombinantes.

Expresión: en el contexto presente, el término "expresión" incluye cualquier fase implicada en la producción del polipéptido incluyendo, pero no limitado a, transcripción, modificación postranscripcional, traducción, modificación postraduccional, y secreción.

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

Célula huésped: el término "célula huésped", como se utiliza en este caso, incluye cualquier tipo de célula que es susceptible de transformación con un constructo de ácidos nucleicos.

Los términos, "sonda de polinucleótidos", "hibridación" al igual que las distintas condiciones de astringencia están definidas en la sección titulada "Polipéptidos que tienen actividad Celobiohidrolasa II".

Descripción detallada de la invención

Polipéptidos que tienen actividad celobiohidrolasa II

En una primera forma de realización, la presente invención se refiere a polipéptidos que tienen actividad celobiohidrolasa II y donde los polipéptidos comprenden, preferiblemente consisten en, una secuencia de aminoácidos que tiene algo de identidad a la secuencia de aminoácidos mostrada en la SEC ID NO:2 de al menos el 90%, preferiblemente al menos el 95%, p. ej. al menos el 96%, tal como al menos el 97%, e incluso de la forma más preferible al menos el 98%, tal como al menos el 99% (de ahora en adelante "polipéptidos homólogos"). En una forma de realización interesante, la secuencia de aminoácidos difiere de como mucho diez aminoácidos (p. ej. de diez aminoácidos), en particular de como mucho cinco aminoácidos (p. ej. de cinco aminoácidos), tal como de como mucho cuatro aminoácidos (p. ej. de cuatro aminoácidos), p. ej. de como mucho tres aminoácidos (p. ej. de tres aminoácidos) de la secuencia de aminoácidos mostrada en la SEC ID NO:2. En una forma de realización particular interesante, la secuencia de aminoácidos difiere de como mucho dos aminoácidos (p. ej. de dos aminoácidos), tal como de un aminoácido de la secuencia de aminoácidos mostrada en la SEC ID NO:2.

Preferiblemente, los polipéptidos de la presente invención comprenden la secuencia de aminoácidos mostrada en la SEC ID NO:2, una variante alélica de la misma, o un fragmento de la misma que tiene actividad celobiohidrolasa II. En otra forma de realización preferida, el polipéptido de la presente invención es la secuencia de aminoácidos mostrada en la SEC ID NO:2.

El polipéptido de la invención puede ser una celobiohidrolasa II 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 se describe en el Ejemplo 1. Además, el polipéptido de la invención puede ser preparado por la técnica de redistribución de ADN, tal como se describe en J.E. Ness *et al*. Nature Biotechnology 17, 893-896 (1999). Además, el polipéptido de la invención puede ser una variante artificial que comprende, preferiblemente consiste en, una secuencia de aminoácidos que tiene al menos una sustitución, eliminación y/o inserción de un aminoácido en comparación con la secuencia de aminoácidos mostrada en la SEC ID NO:2. Tales variantes artificiales pueden ser construidas por técnicas estándar conocidas en la técnica, tal como por mutagénesis dirigida/aleatoria del polipéptido que comprende la secuencia de aminoácidos mostrada en la SEC ID NO:2. En una forma de realización de la invención, cambios de aminoácidos (en la variante artificial al igual que en polipéptidos tipo salvaje) son de una naturaleza menor, es decir sustituciones de aminoácidos conservadoras que significativamente

no afectan al pliegue y/o actividad de la proteína; deleciones pequeñas, normalmente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones amino- o carboxilo-terminales, tal como un residuo de metionina aminoterminal; un pequeño péptido de enlace de hasta aproximadamente 20-25 residuos; o una pequeña extensión que facilita la purificación cambiando la carga neta u otra función, tal como un tracto de polihistidina, un epítopo antigénico o un dominio de unión.

Ejemplos de sustituciones conservadoras están dentro del grupo de aminoácidos básicos (arginina, lisina y histidina), aminoácidos acídicos (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 son conocidas en la técnica y están descritas, por ejemplo, por H. Neurath y R.L. Hill, 1979, en, The Proteins, Academic Press, New York. Los cambios que ocurrían con más frecuencia son Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, y Asp/Gly al igual que los mismos a la inversa.

En una forma de realización de la invención interesante, 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, los cuales mejoran la termoestabilidad del polipéptido, los cuales alteran la especificidad del sustrato, los cuales cambian el pH óptimo, y similares.

Preferiblemente, el número de tales sustituciones, deleciones y/o inserciones en comparación con la secuencia de aminoácidos mostrada en la SEC ID NO:2. es como mucho 10, tal como como mucho 9, p. ej. como mucho 8, más preferiblemente como mucho 7, p. ej. como mucho 6, tal como como mucho 5, de la forma más preferible como mucho 4, p. ej. como mucho 3, tal como como mucho 2, en particular como mucho 1.

Los inventores presentes han aislado secuencias de nucleótidos que codifican polipéptidos que tienen actividad celobiohidrolasa II del microorganismo *Chaetomium thermophilum*. En una forma de realización de la invención interesante, el polipéptido comprende una secuencia de aminoácidos que tiene al menos el 90%, de la forma más preferible al menos el 95%, p. ej. al menos el 96%, tal como al menos el 97%, e incluso de la forma más preferible al menos el 98%, tal como al menos el 99% de identidad con el polipéptido codificado por la celobiohidrolasa II que codifica parte de la secuencia de nucleótidos presente en un organismo seleccionado del grupo que consiste en *Chaetomium thermophilum* CGMCC 0859 (de ahora en adelante "polipéptidos homólogos"). En una forma de realización interesante, la secuencia de aminoácidos difiere de como mucho diez aminoácidos (p. ej. por diez aminoácidos), en particular de como mucho cinco aminoácidos (p. ej. de cinco aminoácidos), tal como de como mucho cuatro aminoácidos (p. ej. de cuatro aminoácidos), p. ej. de como mucho tres aminoácidos (p. ej. de tres aminoácidos) del polipéptido codificado por la celobiohidrolasa II que codifica parte de la secuencia de nucleótidos presente en *Chaetomium thermophilum* CGMCC 0859. En una forma de realización interesante particular, la secuencia de aminoácidos difiere de como mucho dos aminoácidos (p. ej. de dos aminoácidos), tal como de un aminoácido del polipéptido codificado por la celobiohidrolasa II que codifica parte de la secuencia de nucleótidos presente en *Chaetomium thermophilum* CGMCC 0859.

En una segunda forma de realización, la presente invención se refiere a polipéptidos que tienen actividad celobiohidrolasa II que son codificados por secuencias de nucleótidos que se hibridan bajo condiciones de astringencia muy altas con una sonda de polinucleótidos que consiste en la cadena complementaria de nucleótidos 63 a 1493 de la SEC ID NO:1.

En otra forma de realización, la presente invención se refiere a polipéptidos que tienen actividad celobiohidrolasa II que son codificados por la celobiohidrolasa II que codifica parte de la secuencia de nucleótidos presente en un microorganismo de la familia *Chaetomiaceae*, preferiblemente al género *Chaetomium*, más preferiblemente a las especies *Chaetomium thermophilum[KeM3]*.

Una secuencia de nucleótidos seleccionada del grupo que consiste en la SEC ID NO:1 o una subsecuencia de la misma, al igual que la secuencia de aminoácidos mostrada en la SEC ID NO:2, o un fragmento de la misma, puede ser usada para diseñar una sonda de polinucleótidos para identificar y clonar ADN que codifica polipéptidos de ADN que tienen actividad celobiohidrolasa II de cepas de diferentes géneros o especies según métodos bien conocidos en la técnica. En particular, tales sondas pueden ser usadas para hibridación con el genómico o ADNc del género o especies de interés, según procedimientos de transferencia de Southern estándares, para identificar y aislar el gen correspondiente en su interior. Tales sondas pueden ser considerablemente más cortas que la secuencia entera, pero deberían tener al menos 15, preferiblemente al menos 25, más preferiblemente al menos 35 nucleótidos de longitud, tal como al menos 70 nucleótidos de longitud. Es, no obstante, preferido que la sonda de polinucleótidos tenga al menos 100 nucleótidos de longitud. Por ejemplo, la sonda de polinucleótidos puede tener al menos 200 nucleótidos de longitud, al menos 300 nucleótidos de longitud, al menos 400 nucleótidos de longitud o al menos 500 nucleótidos de longitud. Incluso sondas más largas pueden ser usadas, p. ej., sondas de polinucleótidos que tienen al menos 600 nucleótidos de longitud, al menos 700 nucleótidos de longitud, al menos 800 nucleótidos de longitud, o al menos 900 nucleótidos de longitud. Ambas sondas de ADN y ARN pueden ser usadas. Las sondas son normalmente marcadas para detectar el gen correspondiente (por ejemplo, con ^{32}P, ^{3}H, ^{35}S, biotina, o avidina).

Así, una biblioteca de ADN genómico o de ADNc obtenida a partir de estos otros organismos puede ser seleccionada para ADN que se hibrida con las sondas anteriormente descritas y que codifica un polipéptido que tiene actividad

celobiohidrolasa II. El ADN genómico u otro de estos otros organismos puede ser separado por agarosa o electroforesis en gel de poliacrilamida, u otras técnicas de separación. El ADN de las bibliotecas o el ADN separado puede ser transferido a, e inmovilizado, en nitrocelulosa u otros materiales portadores adecuados. Para identificar un clon o ADN que sea homólogo a la secuencia mostrada en la SEC ID NO:1, el material portador con el ADN inmovilizado se usa en una transferencia de Southern.

Para objetivos de la presente invención, la hibridación indica que la secuencia de nucleótidos se hibrida a una sonda de polinucleótidos marcada que se hibrida a la secuencia de nucleótidos mostrada en la SEC ID NO:1 bajo condiciones de astringencia muy alta. Las moléculas a las que la sonda de polinucleótidos se hibrida bajo estas condiciones pueden ser detectadas usando una película de rayos X o por cualquier otro método conocido en la técnica. Siempre que se usa el término "sonda de polinucleótidos" en el contexto presente, debe entenderse que tal sonda contiene al menos 15 nucleótidos.

En una forma de realización interesante, la sonda de polinucleótidos es la cadena complementaria de los nucleótidos 63 a 1493 de la SEC ID NO:1.

En otra forma de realización interesante, la sonda de polinucleótidos es la cadena complementaria de la secuencia de nucleótidos que codifica el polipéptido mostrado en la SEC ID NO:2. En una forma de realización adicional interesante, la sonda de polinucleótidos es la cadena complementaria de la secuencia de nucleótidos mostrada en la SEC ID NO:1.

Para sondas largas de al menos 100 nucleótidos de longitud, condiciones de astringencia de muy baja a muy alta son definidas como prehibridación e hibridación a 42ºC en 5X SSPE, 1.0% SDS, 5X solución de Denhardt, 100 microg/ml de ADN de esperma de salmón cortado y desnaturalizado, según procedimientos de transferencia de Southern estándares. Preferiblemente, las sondas largas de al menos 100 nucleótidos no contienen más de 1000 nucleótidos. Para sondas largas de al menos 100 nucleótidos de longitud, el material portador es finalmente lavado tres veces cada vez durante 15 minutos usando 2 X SSC, 0,1% SDS a 42ºC (astringencia muy baja), preferiblemente lavado tres veces cada vez durante 15 minutos usando 0,5 X SSC, 0,1% SDS a 42ºC (astringencia baja), más preferiblemente lavado tres veces cada vez durante 15 minutos usando 0,2 X SSC, 0,1% SDS a 42ºC (astringencia media), incluso más preferiblemente tres veces lavado cada vez durante 15 minutos usando 0,2 X SSC, 0,1% de SDS a 55ºC (astringencia alta a media), de la forma más preferible lavado tres veces cada vez durante 15 minutos usando 0,1 X SSC, 0,1% de SDS a 60ºC (astringencia alta), en particular lavado tres veces cada vez durante 15 minutos usando 0,1 X SSC, 0,1% de SDS a 68ºC (astringencia muy alta).

Aunque no particularmente preferidas, están contemplado que sondas más cortas, p. ej. sondas que tienen de aproximadamente 15 a 99 nucleótidos de longitud, tal como de aproximadamente 15 a aproximadamente 70 nucleótidos de longitud, pueden también ser usadas. Para estas sondas cortas, son definidas condiciones de astringencia como prehibridación, hibridación, y lavado post-hibridación a 5ºC a 10ºC debajo de la T_m calculada usando el cálculo según Bolton y McCarthy (1962, Proceedings of the National Academy of Sciences USA 48:1390) en 0,9 M de NaCl, 0,09 M de Tris-HCl pH 7.6, 6 mM de EDTA, 0,5% NP-40, IX solución de Denhardt, 1 mM de pirofosfato de sodio, 1 mM de fosfato de sodio monobásico, 0,1 mM de ATP, y 0,2 mg de ARN de levadura por ml siguiendo procedimientos de transferencia de Southern estándares.

Para sondas cortas que tienen aproximadamente de 15 nucleótidos a 99 nucleótidos de longitud, el material portador es lavado una vez en 6X SCC más 0,1% de SDS durante 15 minutos y dos veces cada vez durante 15 minutos usando 6X SSC a 5ºC a 10ºC debajo de la T_m calculada.

Fuentes para polipéptidos que tienen actividad celobiohidrolasa II

Un polipéptido de la presente invención puede ser obtenido de microorganismos de cualquier género. Para objetivos de la presente invención, el término "obtenido de" como se utiliza en este caso significará que el polipéptido codificado por la secuencia de nucleótidos es producido por una célula donde la secuencia de nucleótidos está naturalmente presente o en el que la secuencia de nucleótidos ha sido insertada. En una forma de realización preferida, el polipéptido es segregado extracelularmente.

Un polipéptido de la presente invención puede ser un polipéptido bacteriano. Por ejemplo, el polipéptido puede ser un polipéptido bacteriano gram positivo tal como un polipéptido de *Bacillus*, p. ej., un polipéptido de *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans, Bacillus lautus, Bacillus lentus, Bacillus licheniformis, Bacillus megaterium, Bacillus stearothermophilus, Bacillus subtilis*, o *Bacillus thuringiensis*; o un polipéptido de *Streptomyces*, p. ej., un polipéptido de *Streptomyces lividans* o *Streptomyces murinus*; o un polipéptido bacteriano gram negativo, p. ej., un polipéptido de *E. coli* o de *Pseudomonas* sp..

Un polipéptido de la presente invención puede ser un polipéptido fúngico, y preferiblemente un polipéptido de levadura tal como un polipéptido de *Candida, Kluyveromyces, Neocallimastix, Pichia, Piromyces, Saccharomyces, Schizosaccharomyces*, o de *Yarrowia*; o más preferiblemente un polipéptido filamentoso fúngico tal como un polipéptido de *Acremonium, Aspergillus, Chaetomium, Chaetomium, Gloeophyllum, Malbrancheae, Melanocarpus, Meripilus, Myceliophthora, Stilbella, Thielavia*, o *Trichophaea*.

En una forma de realización interesante, el polipéptido es un polipéptido de *Saccharomyces carlsbergensis, Saccharomyces cerevisiae, Saccharomyces diastaticus, Saccharomyces douglasii, Saccharomyces kluyveri, Saccharomyces norbensis* o *Saccharomyces oviformis*.

En una forma de realización preferida, el polipéptido es un *Chaetomium thermophilum[KeM4], Myceliophthora thermophila, Acremonium thermophilum, Thielavia australiensis, Aspergilli. Polipéptido de Thielavia microspore, Aspergillus tubingensis[KeM9], Gloephyllum trabeum, Meripilus giganteus, Trichophaea saccata, Stilbella annulata*, o *Malbrancheae cinnamomea*.

Será entendido que para las especies mencionadas, la invención comprende los estados perfecto e imperfecto, y otros equivalentes taxonómicos, p. ej., anamorfos, a pesar del nombre de especies por el que éstas son conocidas. Expertos en la técnica fácilmente reconocerán la identidad de equivalentes apropiados.

Además, estos polipéptidos pueden ser identificados y obtenidos de otras fuentes incluyendo microorganismos aislados de la naturaleza (p. ej., suciedad, agua, plantas, animales, etc.) usando las sondas mencionadas arriba. Técnicas para aislar microorganismos de hábitats naturales son conocidas en la técnica. La secuencia de nucleótidos puede luego ser derivada por selección de forma similar de una biblioteca genómica o de ADNc de otro microorganismo. Una vez que una secuencia de nucleótidos que codifica un polipéptido ha sido detectada con la(s) sonda(s), la secuencia puede ser aislada o clonada para utilizar técnicas que son conocidas por los expertos en la técnica (véase, p. ej., J. Sambrook, E. F. Fritsch, y T. Maniatus, 1989, Molecular Cloning, A Laboratory Manual, 2a edición, Cold Spring Harbor, New York).

Polipéptidos codificados por secuencias de nucleótidos de la presente invención también incluyen polipéptidos fusionados o polipéptidos de fusión divisibles donde otro polipéptido está fusionado en el N-término o el C-término del polipéptido o fragmento del mismo. Un polipéptido fusionado es producido por fusión de una secuencia de nucleótidos (o una parte de la misma) que codifica otro polipéptido a una secuencia de nucleótidos (o una parte de la misma) de la presente invención. Técnicas para producir polipéptidos de fusión son conocidos en la técnica, e incluyen ligar las secuencias codificantes que codifican los polipéptidos de modo que éstas estén en marco y que la expresión del polipéptido fusionado esté bajo el control del(los) mismo(s) promotor(es) y terminador.

Polinucleótidos y secuencias de nucleótidos

La presente invención también se refiere a polinucleótidos que tienen 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 nucleótidos es la SEC ID NO:1. La presente invención también comprende polinucleótidos, preferiblemente que consisten en, secuencias de nucleótidos que codifican el polipéptido mostrado en la SEC ID NO:2, que difiere de la secuencia de nucleótidos mostrada en la SEC ID NO:1 en virtud de la degeneración del código genético.

La presente invención también se refiere a polinucleótidos que comprenden, preferiblemente que consisten en, una subsecuencia de la secuencia de nucleótidos mostrada en la SEC ID NO:1 que codifican fragmentos de la secuencia de aminoácidos mostrada en la SEC ID NO:2 que tiene actividad celobiohidrolasa II. Una subsecuencia de la secuencia de nucleótidos mostrada en la SEC ID NO:1 es una secuencia de nucleótidos comprendida por la secuencia mostrada en la SEC ID NO:1 a excepción de que uno o más nucleótidos del extremo 5' y/o 3' ha sido delecionado.

La presente invención también se refiere a polinucleótidos que tienen, preferiblemente que consisten en, una secuencia de nucleótidos modificada que comprende al menos una modificación en la secuencia codificante del polipéptido maduro mostrada en la SEC ID NO:1, y donde la secuencia de nucleótidos modificada codifica un polipéptido que consiste en la secuencia de aminoácidos mostrada en la SEC ID NO:2.

Las técnicas usadas para aislar o clonar una secuencia de nucleótidos que codifica un polipéptido son conocidas 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 ADN genómico de este tipo puede ser efectuada, p. ej., usando la reacción en cadena de la polimerasa bien conocida (PCR) o selección de anticuerpos de bibliotecas de expresión para detectar fragmentos de ADN clonados con características estructurales compartidas. Véase, p. ej., Innis *et al*., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Otros procedimientos de amplificación tales como la reacción en cadena de la ligasa (LCR), la transcripción activada ligada (LAT) y la amplificación basada en secuencias de nucleótidos (NASBA) pueden ser usadas. La secuencia de nucleótidos puede ser clonada de una cepa seleccionada de una cepa de un género seleccionado del grupo que consiste en *Chaetomium, Myceliophthora, Melanocarpus, Acremonium, Thielavia, Aspergillus, Gloeophyllum, Meripilus, Trichophaea, Stilbella* y *Malbrancheae*, u otro organismo o relacionado y por lo tanto, por ejemplo, puede ser una variante alélica o de especies de la región codificante del polipéptido de la secuencia de nucleótidos.

La secuencia de nucleótidos puede ser obtenida por procedimientos de clonación estándares usados en ingeniería genética para recolocar la secuencia de nucleótidos desde su ubicación natural a un sitio diferente donde será reproducida. Los procedimientos de clonación pueden implicar escisión y aislamiento de un fragmento deseado que comprende la secuencia de nucleótidos que codifica el polipéptido, inserción del fragmento en una molécula de vector,

e incorporación del vector recombinante en una célula huésped donde copias múltiples o clones de la secuencia de nucleótidos serán replicados. La secuencia de nucleótidos puede ser de origen genómico, ADNc, ARN, semisintético, sintético, o cualquier combinación de los mismos.

La presente invención también se refiere a un polinucleótido comprendiendo, preferiblemente consistiendo en, una secuencia de nucleótidos que tiene un grado de identidad con la secuencia de nucleótidos mostrada en los nucleótidos 63 a 1493 de la SEC ID NO:1 de al menos un 90% de identidad, más preferiblemente al menos un 95% de identidad, tal como al menos un 96% de identidad, p. ej. al menos un 97% de identidad, incluso más preferiblemente al menos un 98% de identidad, tal como al menos un 99%. Preferiblemente, la secuencia de nucleótidos codifica un polipéptido que tiene actividad celobiohidrolasa II. El grado de identidad entre dos secuencias de nucleótidos es determinado como se ha descrito previamente (véase la sección titulada "Definiciones").

La modificación de una secuencia de nucleótidos que codifica un polipéptido de la presente invención puede ser necesaria para la síntesis de un polipéptido, que comprende una secuencia de aminoácidos que tiene al menos una sustitución, eliminación y/o inserción en comparación con la secuencia de aminoácidos mostrada en la SEC ID NO:2. Estas variantes artificiales pueden diferir en cierto modo creado genéticamente del polipéptido aislado de su fuente nativa, p. ej., variantes que difieren en actividad específica, termostabilidad o pH óptimo.

Será evidente para los expertos en la técnica que estas modificaciones pueden ser hechas fuera de las regiones críticas para la función de la molécula y además resultan en un polipéptido activo. Residuos de aminoácidos esenciales para la actividad del polipéptido codificado por la secuencia de nucleótidos de la invención, y en consecuencia preferiblemente no sometido a modificación, tal como sustitución, pueden ser identificados según procedimientos conocidos en la técnica, tales como mutagénesis dirigida o mutagénesis de barrido de alanina (véase, p. ej., Cunningham y Wells, 1989, Science 244: 1081-1085). En esta técnica, las mutaciones son introducidas en cada residuo cargado positivamente en la molécula, y las moléculas mutantes resultantes son evaluadas para actividad celobiohidrolasa II 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 según está determinado por tales técnicas como análisis de resonancia magnética nuclear, cristalografía o marcado por fotoafinidad (véase, p. ej., 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).

Además, una secuencia de nucleótidos que codifica un polipéptido de la presente invención puede ser modificada 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 corresponde al uso de codón del organismo huésped destinado a la producción de la enzima.

La introducción de una mutación en la secuencia de nucleótidos para intercambiar un nucleótido por otro nucleótido puede ser realizada por mutagénesis dirigida usando cualquiera de los métodos conocidos en la técnica. Particularmente útil es el procedimiento que utiliza un vector de ADN bicatenario superenrollado con un inserto de interés y dos cebadores sintéticos que contienen la mutación deseada. Los cebadores oligonucleótidos, cada uno complementario a cadenas opuestas del vector, se extienden durante la variación cíclica de la temperatura mediante *Pfu* ADN polimerasa. Al incorporar los cebadores, un plásmido mutado que contiene cortes en bisel es generado. Después de la variación cíclica de la temperatura, el producto es tratado con Dpnl que es específico para ADN metilado y hemimetilado para digerir el molde de ADN parental y para seleccionar para ADN sintetizado conteniendo la mutación. Otros procedimientos conocidos pueden también ser usados en la técnica. Para una descripción general de sustitución de nucleótidos, véase, p. ej., Ford *et al*., 1991, Protein Expression and Purification 2: 95-107.

La presente invención también se refiere a un polinucleótido comprendiendo, preferiblemente consistiendo en, una secuencia de nucleótidos que codifica un polipéptido que tiene actividad celobiohidrolasa II, y que se hibrida bajo condiciones de astringencia muy alta con la cadena complementaria de nucleótidos 63 a 1493 de la SEC ID NO:1.

Como será entendido, detalles y particularidades en lo que se refiere a la hibridación de las secuencias de nucleótidos serán los mismos o análogos para los aspectos de hibridación discutidos en la sección titulada "Polipéptidos con actividad Celobiohidrolasa II" en la presente.

Recombinación (redistribución) de ADN

Las secuencias de nucleótidos de la SEC ID NO:1 pueden ser usadas en un proceso de recombinación de ADN (o redistribución). Las nuevas secuencias de polinucleótidos obtenidas en tal proceso pueden codificar polipéptidos nuevos que tienen actividad de celobiasa con propiedades mejoradas, tal como estabilidad mejorada (estabilidad de almacenamiento, termostabilidad), actividad específica mejorada, pH óptimo mejorado, y/o tolerancia mejorada hacia compuestos específicos.

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 son cambiados en comparación con los polinucleótidos de entrada.

La recombinación de ADN o redistribución puede ser un proceso (parcialmente) aleatorio donde una biblioteca de genes quiméricos es generado a partir de dos o más genes iniciadores. Varios formatos conocidos pueden utilizarse para efectuar este proceso de redistribución o de recombinación.

El proceso puede implicar fragmentación aleatoria de ADN parental seguido de reensamblaje por PCR para genes nuevos de cuerpo entero, p. ej. como se presenta en US5605793, US5811238, US5830721, US6117679. La recombinación *in vitro* de genes pueden ser realizada, p. ej. como se describe en US6159687, WO98/41623, US6159688, US5965408 y US6153510. El proceso de recombinación puede ocurrir *in vivo* en una célula viva, p. ej. como se describe en WO 97/07205 y WO 98/28416.

El ADN parental puede ser fragmentado por tratamiento de DNA'sa I o por digeridos de endonucleasa de restricción como se describe por Kikuchi *et al* (2000a, Gene 236:159-167). La redistribución de dos padres puede ser hecha por redistribución de ADN parental monocatenario de los dos padres como se describe en Kikuchi *et al* (2000b, Gene 243:133-137).

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 sería los métodos descritos en US 6159687: Ejemplos 1 y 2.

Constructos de ácidos nucleicos

La presente invención también se refiere a constructos de ácidos nucleicos que comprenden una secuencia de nucleótidos de la presente invención operativamente enlazada 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.

Una secuencia de nucleótidos que codifica un polipéptido de la presente invención puede ser manipulada en una variedad de vías para proporcionar la expresión del polipéptido. La 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 recombinantes son bien conocidos en la técnica.

La secuencia de control puede ser una secuencia promotora apropiada, una secuencia de nucleótidos que es reconocida por una célula huésped para la expresión de la secuencia de nucleótidos. La secuencia promotora contiene secuencias de control transcripcionales, que median la expresión del polipéptido. El promotor puede ser cualquier secuencia de nucleótidos que muestra actividad transcripcional en la célula huésped de elección incluyendo promotores mutantes, truncados, e híbridos, y pueden ser obtenidos de genes que codifican polipéptidos extracelulares o intracelulares bien homólogos o heterólogos a la célula huésped.

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 (*dagA*) *Streptomyces coelicolor*, gen de levansucrasa (*sacB*) de *Bacillus subtilis*, gen de alfa-amilasa (*amyL*) de *Bacillus licheniformis*, gen de amilasa maltógénica (*amyM*) de *Bacillus stearothermophilus*, gen de alfa-amilasa (*amyQ*) de *Bacillus amyloliquefaciens*, gen de penicilinasa (*penP*) de *Bacillus licheniformis*, 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), al igual que el promotor tac (DeBoer *et al*., 1983, Proceedings of the National Academy of Sciences USA 80: 21-25). Otros promotores están descritos en "Useful proteins from recombinant bacteria" en Scientific American, 1980, 242: 74-94; y en Sambrook *et al*., 1989, *supra*.

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 ácido de *Aspergillus niger*, glucoamilasa (glaA) de *Aspergillus niger* o de *Aspergillus awamori*, 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.

En un huésped de levadura, promotores útiles son obtenidos de los genes para enolasa (ENO-1) de *Saccharomyces cerevisiae*, galactocinasa (GAL1) de *Saccharomyces cerevisiae*, alcohol deshidrogenasa/gliceraldehido-3-fosfato deshidrogenasa (ADH2/GAP) de *Saccharomyces cerevisiae*, y 3-fosfoglicerato quinasa de *Saccharomyces cerevisiae*. Otros promotores útiles para células huéspedes de levadura están descritos por Romanos *et al*., 1992, Yeast 8: 423-488.

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

Terminadores preferidos para células huéspedes filamentosas fúngicas son obtenidos de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, antranilato sintasa de *Aspergillus nidulans*, alfa-glucosidasa de *Aspergillus niger*, y proteasa de tipo tripsina de *Fusarium oxysporum*.

Terminadores preferidos para células huéspedes de levadura son obtenidos de los genes para enolasa de *Saccharomyces cerevisiae*, citocromo C de *Saccharomyces cerevisiae* (CYC1), y gliceraldehido-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae*. Otros terminadores útiles para células huéspedes de levadura están descritos por Romanos *et al*., 1992, *supra*.

La secuencia de control puede también ser una secuencia guía adecuada, una región no traducida de un ARNm que es importante para la traducción por la célula huésped. La secuencia líder está operativamente enlazada al término 5' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier secuencia guía que es funcional en la célula huésped de elección puede ser usada en la presente invención.

Líderes preferidos para células huéspedes filamentosas fúngicas son obtenidos de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.

Líderes adecuados para células huéspedes de levadura son obtenidos de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), 3-fosfoglicerato quinasa de *Saccharomyces cerevisiae*, alfa-factor de *Saccharomyces cerevisiae*, y alcohol deshidrogenasa/g liceraldehido-3-fosfato deshidrogenasa (ADH2/GAP) de *Saccharomyces cerevisiae*.

La secuencia de control puede también ser una secuencia de poliadenilación, una secuencia operativamente enlazada al término 3' de la secuencia de nucleótidos y que, cuando es transcrita, 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 puede ser usada en la presente invención.

Secuencias de poliadenilación preferidas para células huéspedes filamentosas fúngicas son obtenidas de los genes para TAKA amilasa de *Aspergillus oryzae*, glucoamilasa de *Aspergillus niger*, antranilato sintasa de *Aspergillus nidulans*, proteasa de tipo tripsina de *Fusarium oxysporum*, y alfa-glucosidasa de *Aspergillus niger*.

Secuencias de poliadenilación útiles para células huéspedes de levadura están descritas por Guo y Sherman, 1995, Molecular Cellular Biology 15: 5983-5990.

La secuencia de control puede también ser una región codificante del péptido señal que codifica para una secuencia de aminoácidos enlazada al término amino de un polipéptido y 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 codificante del péptido señal naturalmente enlazada en el marco de lectura de traducción con el segmento de la región codificante que codifica el polipéptido segregado. De forma alternativa, el extremo 5' de la secuencia codificante puede contener una región codificante del péptido señal que es externa a la secuencia codificante. La región codificante del péptido señal externo puede ser requerida donde la secuencia codificante naturalmente no contiene una región codificante del péptido señal. De forma alternativa, la región codificante del péptido señal externo puede simplemente reemplazar la región codificante del péptido señal natural para aumentar la secreción del polipéptido. No obstante, cualquier región codificante del péptido señal que dirige el polipéptido expresado en la vía secretora de una célula huésped de elección puede ser usada en la presente invención.

Regiones codificantes del péptido señal eficaces para células huéspedes bacterianas son las regiones codificantes del 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 *prsA* de *Bacillus subtilis*. Otros péptidos señal están descritos por Simonen y Palva, 1993, Microbiological Reviews 57: 109-137.

Regiones codificantes del péptido señal eficaces para células huéspedes filamentosas fúngicas son las regiones codificantes del 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*.

Péptidos señal útiles para células huéspedes de levadura son obtenidos de los genes para alfa factor de *Saccharomyces cerevisiae* e invertasa de *Saccharomyces cerevisiae*. Otras regiones codificantes del péptido señal útiles están descritas por Romanos *et al*., 1992, *supra*.

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

Cuando tanto el péptido señal como las regiones del propéptido están presentes en el término amino de un polipéptido, la región del propéptido está situada junto al término amino de un polipéptido y la región del péptido señal está situada junto al término amino de la región del propéptido.

Puede también ser deseable añadir secuencias reguladoras que permiten la regulación de la expresión del polipéptido con respecto 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 químico o físico, incluyendo la presencia de un compuesto regulador. Sistemas reguladores en sistemas procarióticos incluyen los sistemas del operador *lac, tac*, y *trp*. En levadura, el sistema ADH2 o sistema GAL1 puede ser usado. En hongos filamentosos, el promotor TAKA de alfa-amilasa, promotor de glucoamilasa de *Aspergillus niger*, y el promotor de glucoamilasa de *Aspergillus oryzae* pueden ser usados como secuencias reguladoras. Otros ejemplos de secuencias reguladoras son los que permiten la amplificación génica. En sistemas eucarióticos, estos incluyen el gen de dihidrofolato-reductasa que es amplificado en presencia de metotrexato, y los genes de metalotioneína que son amplificados con metales pesados. En estos casos, la secuencia de nucleótidos que codifica el polipéptido sería operativamente enlazada con la secuencia reguladora.

Vectores de expresión

La presente invención también se refiere a vectores de expresión recombinantes que comprenden el constructo de ácidos nucleicos de la invención. Las distintas secuencias de nucleótidos y de control anteriormente descritas pueden ser unidas 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 estos sitios. De forma alternativa, la secuencia de nucleótidos de la presente invención puede ser expresada insertando 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 está localizada en el vector de modo que la secuencia codificante está operativamente enlazada con las secuencias de control apropiadas para la expresión.

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 que puede provocar la expresión de la secuencia de nucleótidos. La elección del vector normalmente dependerá de la compatibilidad del vector con la célula huésped en la que será introducido el vector. Los vectores pueden ser plásmidos circulares lineales o cerrados.

El vector puede ser un vector de replicación autónoma, es decir, un vector que existe como una entidad extracromosómica, cuya replicación es independiente de la replicación cromosómica, p. ej., un plásmido, un elemento extracromosómico, un minicromosoma, o un cromosoma artificial.

El vector puede contener cualquier medio para asegurar la autorreplicación. De forma alternativa, el vector puede ser uno que, cuando se introduce en la célula huésped, es integrado en el genoma y se replica con el(los) cromosoma(s) en el que ha sido integrado. Además, un único vector o plásmido o dos o más vectores o plásmidos que juntos contienen el ADN total que debe ser introducido en el genoma de la célula huésped, o un transposón pueden ser usados.

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

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

Preferidos para el uso en una célula de *Aspergillus* son los genes *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen bar de *Streptomyces hygroscopicus*.

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

Para integración en el genoma de la célula huésped, el vector puede depender de la secuencia de nucleótidos que codifica el polipéptido o cualquier otro elemento del vector para la integración estable del vector en el genoma por recombinación homóloga o no homóloga. De forma alternativa, el vector puede contener secuencias de nucleótidos adicionales para dirigir la integración por recombinación homóloga en el genoma de la célula huésped. Las secuencias de nucleótidos adicionales permiten que el vector sea integrado en el genoma de la célula huésped a una(s) ubicación(es) precisa(s) en el(los) cromosoma(s). Para aumentar la probabilidad de integración a 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 a la secuencia diana correspondiente para aumentar la probabilidad de recombinación homóloga.

Los elementos integracionales pueden ser cualquier secuencia que sea homóloga a la secuencia diana en el genoma de la célula huésped. Además, los elementos integracionales pueden ser secuencias de nucleótidos no codificantes o codificantes. Por otro lado, el vector puede ser integrado en el genoma de la célula huésped por recombinación no homóloga.

Para replicación autónoma, el vector puede además comprender un origen de replicación que permite que el vector se replique 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 permitiendo la replicación en *E. coli*, y pUB110; pE194; pTA1060, y pAMβ1 permitiendo la replicación en *Bacillus*. Ejemplos de orígenes de replicación para el 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 su temperatura de funcionamiento sensible en la célula huésped (véase, p. ej., Ehrlich, 1978, Proceedings of the National Academy of Sciences USA 75: 1433).

Más de una copia de una secuencia de nucleótidos de la presente invención puede ser insertada en la célula huésped para aumentar la producción del producto genético. Un aumento en el número de copias de la secuencia de nucleótidos puede ser obtenido integrando al menos una copia adicional de la secuencia en el genoma de la célula huésped o mediante la inclusión de un gen marcador seleccionable amplificable con la secuencia de nucleótidos donde las células que contienen copias amplificadas del gen marcador seleccionable, y de ese modo copias adicionales de la secuencia de nucleótidos, pueden ser seleccionadas para cultivar las células en presencia del agente seleccionable apropiado.

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

Células huéspedes

La presente invención también se refiere a una célula huésped recombinante que comprende el constructo de ácidos nucleicos de la invención, que son ventajosamente usados en la producción recombinante de los polipéptidos. Un vector que comprende una secuencia de nucleótidos de la presente invención es introducido en una célula huésped de modo que el vector es mantenido como un integrante cromosómico o como un vector extracromosómico autorreplicante como se ha descrito anteriormente.

La célula huésped puede ser un microorganismo unicelular, p. ej., un procariota, o un microorganismo no unicelular, p. ej., un eucariota.

Células unicelulares útiles son células bacterianas tales como bacterias gram positivas incluyendo, pero no limitadas a, una célula de *Bacillus*, p. ej., *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*, p. ej., *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.

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

La célula huésped puede ser una eucariota, tal como una célula de mamífero, insecto, planta, o fúngica.

En una forma de realización preferida, la célula huésped es una célula fúngica. “Hongos” como se utiliza en este caso incluye los filos Ascomycota, Basidiomycota, Chytridiomycota, y Zygomycota (como se define por Hawksworth *et al*., en, Ainsworth and Bisby’s Dictionary of The Fungi, 8ª edición, 1995, CAB International, University Press, Cambridge, UK) al igual que Oomycota (como se cita en Hawksworth *et al*., 1995, *supra*, página 171) y todos los hongos mitospóricos (Hawksworth *et al*., 1995, *supra*).

En una forma de realización más preferida, la célula huésped fúngica es una célula de levadura. “Levadura” como se utiliza en este caso incluye levadura ascosporogénea (Endomycetales), levadura basidiosporogénea, y levadura de los *Fungi Imperfecti* (Blastomycetes). Puesto que la clasificación de levadura puede cambiar en el futuro, para los objetivos 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).

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*.

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 oviformes*. 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*.

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 se define por Hawksworth *et al*., 1995, *supra*). Los hongos filamentosos están caracterizados por una pared micelial compuesta por quitina, celulosa, glucano, quitosano, manano, y otros polisacáridos complejos. El crecimiento vegetativo es por alargamiento hifal y el catabolismo de carbono es estrictamente aeróbico. Por el contrario, el crecimiento vegetativo por levaduras tales como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y el catabolismo de carbono puede ser fermentativo.

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

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 parental fúngica filamentosa 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*.

Las células micóticas pueden ser transformadas por un proceso que implica formación de protoplastos, transformación de los protoplastos, y regeneración de la pared celular en cierto modo conocido *per se*. Procedimientos adecuados para la transformación de células huéspedes de *Aspergillus* están 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* están descritos por Malardier *et al*., 1989, Gene 78: 147-156 y WO 96/00787. La levadura puede ser transformada usando los procedimientos descritos por Becker y Guarente, en Abelson, J. N. and Simon, M. I., editors, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Volumen 194, págs. 182-187, Academic Press, Inc., New York; Ito *et al*., 1983, Journal of Bacteriology 153: 163; y Hinnen *et al*., 1978, Proceedings of the National Academy of Sciences USA 75: 1920.

Métodos de producción

La presente invención también se refiere a métodos para producir un polipéptido de la presente invención comprendiendo (a) cultivar una cepa, que en su forma tipo salvaje es capaz de producir el polipéptido; y (b) recuperar el polipéptido. Preferiblemente, la cepa es seleccionada de una especie dentro de un género comprendido en el grupo que consiste en *Acremonium, Aspergillus, Chaetomium, Chaetomium, Gloeophyllum, Malbrancheae, Melanocarpus, Meripilus, Myceliophthora, Stilbella, Thielavia*, o *Trichophaea*; más preferiblemente la cepa es seleccionada del grupo que consiste en *Chaetomium thermophilum, Myceliophthora thermophila, Thielavia australiensis, Thielavia microspore, Aspergillus sp[KeM30].*, *los Aspergilli negros, Aspergillus tubingensis syn. A. neotubingensis Frisvad sp.nov.[KeM31], Gloeophyllum trabeum, Meripilus giganteus, Trichophaea saccata, Stilbella annulata*, y *Malbrancheae cinnamomea*.

La presente invención también se refiere a métodos para producir un polipéptido de la presente invención comprendiendo (a) cultivar una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido.

La presente invención también se refiere a métodos para la producción *in situ* de un polipéptido de la presente invención comprendiendo (a) cultivo de una célula huésped bajo condiciones propicias para la producción del polipéptido; y (b) puesta en contacto del polipéptido con un sustrato deseado, tal como un sustrato celulósico, sin recuperación previa del polipéptido. El término "producción *in situ*" se destina a significar que el polipéptido es producido directamente en el locus donde está destinado a ser usado, tal como en un proceso de fermentación para la producción de etanol.

En los métodos de producción de la presente invención, las células son cultivadas en un medio nutritivo adecuado para la producción del polipéptido usando métodos conocidos en la técnica. Por ejemplo, la célula puede ser cultivada por cultivo en un matraz de agitación, fermentación a escala pequeña o a gran escala (incluyendo fermentaciones continuas, discontinuas, de flujo discontinuo, o de estado sólido) en fermentadores de laboratorio o industriales realizadas en un medio adecuado y bajo condiciones que permiten que el polipéptido sea expresado y/o aislado. El cultivo se desarrolla en un medio nutritivo adecuado comprendiendo fuentes de nitrógeno y carbono y sales inorgánicas, usando

procedimientos conocidos en la técnica. Los medios adecuados están disponibles de proveedores comerciales o pueden ser preparados según composiciones publicadas (p. ej., en catálogos de la American Type Culture Collection). Si el polipéptido es segregado en el medio nutritivo, el polipéptido puede ser recuperado directamente del medio. Si el polipéptido no es segregado, éste puede ser recuperado de lisatos celulares.

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

El polipéptido resultante puede ser recuperado por métodos conocidos en la técnica. Por ejemplo, el polipéptido puede ser recuperado del medio nutritivo por procedimientos convencionales incluyendo, pero no limitado a, centrifugado, filtración, extracción, secado por pulverización, evaporación, o precipitación.

Los polipéptidos de la presente invención pueden ser purificados por una variedad de procedimientos conocidos en la técnica incluyendo, pero no limitados a, cromatografía (p. ej., intercambio iónico, afinidad, hidrofóbico, cromatoenfoque, y exclusión por tamaños), procedimientos electroforéticos (p. ej., isoelectroenfoque preparatorio), solubilidad diferencial (p. ej., precipitación con sulfato amónico), SDS-PAGE, o extracción (véase, p. ej., Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Plantas

La presente invención también se refiere a una planta transgénica, parte de la planta, o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad celobiohidrolasa II de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido puede ser recuperado de la planta o parte de la planta. De forma alternativa, la planta o parte de la planta que contiene el polipéptido recombinante puede ser usado como tal para mejorar la calidad de un alimento o pienso, p. ej., mejorar el valor nutritivo, apetencia, y propiedades reológicas, o para destruir un factor antinutritivo.

La planta transgénica puede ser dicotiledónea (una dicot) o monocotiledónea (una monocot). Ejemplos de plantas monocot son hierbas, tal como poa de prados (césped azul, Poa), pasto de forraje tal como festuca, lolio, césped templado, tal como agrostis, y cereales, p. ej., trigo, avena, centeno, cebada, arroz, sorgo, mijo, y maíz (grano).

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

Ejemplos de partes de la planta son el tallo, callo, hojas, raíz, frutas, semillas, y tubérculos. También tejidos vegetales específicos, tales como cloroplasto, apoplasto, mitocondria, vacuola, peroxisomas, y citoplasma son considerados para ser una parte de la planta. Además, cualquier célula vegetal, cualquiera que sea el origen del tejido, está considerada una parte de la planta.

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

La planta transgénica o célula vegetal que expresa un polipéptido de la presente invención puede ser construida conforme a métodos conocidos en la técnica. Brevemente, la planta o célula vegetal es construida incorporando uno o más constructos de expresión que codifican un polipéptido de la presente invención en el genoma huésped de la planta y que propagan la resultante planta modificada o célula vegetal en una planta transgénica o célula vegetal.

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 requeridas para la expresión de la secuencia de nucleótidos en la planta o parte de la planta de elección. Además, el constructo de expresión puede comprender un marcador seleccionable útil para identificar células huéspedes donde se ha integrado el constructo de expresión y las secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (esto depende del método introducción de ADN que será usado).

La elección de secuencias reguladoras, tales como secuencias del promotor y del terminador y opcionalmente secuencias señal o de tránsito son determinadas, por ejemplo, basándose en cuándo, dónde, y cómo se desea que el polipéptido 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 específico, y el producto genético puede ser dirigido a un tejido específico o parte de la planta tal como semillas u hojas. Secuencias reguladoras son, por ejemplo, descritas por Tague *et al*., 1988, Plant Physiology 86: 506.

Para expresión constitutiva, puede ser usado el promotor 35S-CaMV (Franck *et al*., 1980, Cell 21: 285-294). Promotores específicos para un órgano pueden ser, por ejemplo, un promotor de tejidos sumidero de almacenamiento tales 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 las semillas tal como el promotor de glutelina, prolamina, globulina, o de albúmina de arroz (Wu *et al*., 1998, Plant and Cell Physiology 39: 885-889), un promotor de *Vicia faba* de la legúmina B4 y el gen de proteína de semilla desconocido de *Vicia faba* (Conrad *et al*., 1998, Journal of Plant Physiology 152: 708-711), un promotor de una proteína de cuerpo de aceite de semilla (Chen *et al*., 1998, Plant and Cell Physiology 39: 935-941), el promotor de napA de proteína de almacenamiento de *Brassica napus*, o cualquier otro promotor específico de semillas conocido en la técnica, p. ej., como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de las hojas tal como el promotor rbcs del arroz o tomate (Kyozuka *et al*., 1993, Plant Physiology 102: 991-1000, el promotor del gen de adenina metiltransferasa del virus de chlorella (Mitra y Higgins, 1994, Plant Molecular Biology 26: 85-93), o el promotor del gen *aldP* de arroz (Kagaya *et al*., 1995, Molecular and General Genetics 248: 668-674), o un promotor dañado inducible tal como el promotor pin2 de la patata (Xu *et al*., 1993, Plant Molecular Biology 22: 573-588).

Un elemento potenciador del promotor puede también ser usado para conseguir una expresión más alta de la enzima en la planta. Por ejemplo, el elemento potenciador del promotor puede ser un intrón que está colocado entre el promotor y la secuencia de nucleótidos que codifica un polipéptido de la presente invención. Por ejemplo, Xu *et al*., 1993, *supra* revelan el uso del primer intrón del gen de actina 1 de arroz para aumentar la expresión.

El gen marcador seleccionable y cualquier otra parte del constructo de expresión pueden ser elegidos entre aquellos disponibles en la técnica.

El constructo de ácidos nucleicos está incorporado en el genoma de planta según técnicas convencionales conocidas en la técnica, que incluyen la transformación mediada por *Agrobacterium*, transformación mediada por virus, microinyección, bombardeo de partículas, transformación biolística, y electroporación (Gasser *et al*., 1990, Science 244: 1293; Potrykus, 1990, Bio/Technology 8: 535; Shimamoto *et al*., 1989, Nature 338: 274).

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

Después de la transformación, los transformantes que han incorporado en ellos el constructo de expresión son seleccionados y regenerados en plantas enteras según métodos bien conocidos en la técnica.

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

La presente invención también se refiere a métodos para la producción *in situ* de un polipéptido de la presente invención comprendiendo (a) cultivo de una planta transgénica o una célula vegetal que comprende una secuencia de nucleótidos que codifica un polipéptido que tiene actividad celobiohidrolasa II de la presente invención 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.

Composiciones

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

La composición puede comprender un polipéptido de la invención como el componente enzimático más importante, p. ej., una composición monocomponente. De forma alternativa, la composición puede comprender actividades enzimáticas múltiples, tales como una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolitica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa, o xilanasa.

Las composiciones pueden ser preparadas conforme a métodos conocidos en la técnica y pueden estar en forma de una composición líquida o seca. Por ejemplo, la composición de polipéptido puede estar en forma de un granulado o un microgranulado. El polipéptido que será incluido en la composición puede ser estabilizado conforme a métodos conocidos en la técnica.

Más abajo están provistos ejemplos de usos preferidos de las composiciones de polipéptido de la invención. La dosificación de la composición de polipéptido de la invención y otras condiciones bajo las cuales se usa la composición pueden ser determinadas basándose en métodos conocidos en la técnica.

Composiciones detergentes

El polipéptido de la invención puede ser añadido y por lo tanto volverse un componente de una composición detergente.

La composición detergente de la invención puede por ejemplo ser formulada como una composición detergente para lavado a mano o a máquina incluyendo una composición de aditivo para el lavado de la ropa adecuada para el pretratamiento de telas manchadas y una composición de suavizante añadida al enjuague, o ser formulada como una composición detergente para el uso en operaciones de limpieza generales de superficies duras del hogar, o ser formulada para operaciones de lavado de la vajilla a mano o a máquina.

En un aspecto específico, la invención proporciona un aditivo detergente que comprende el polipéptido de la invención. El aditivo detergente al igual que la composición detergente pueden comprender una o varias otras enzimas tales como una proteasa, una lipasa, una cutinasa, una amilasa, una carbohidrasa, una celulasa, una pectinasa, una mananasa, una arabinasa, una galactanasa, una xilanasa, una oxidasa, p. ej., una lacasa, y/o una peroxidasa.

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

Proteasas: proteasas adecuadas incluyen aquellas de origen animal, vegetal o microbiano. El origen microbiano es preferido. Se incluyen mutantes químicamente modificados o creados genéticamente de proteínas. 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*, p. ej., 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.

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

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

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

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

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

Celulasas: celulasas adecuadas incluyen aquellas de origen bacteriano o fúngico. Se incluyen mutantes modificados químicamente o creados genéticamente de proteínas. Las celulasas adecuadas incluyen celulasas de los géneros *Bacillus, Pseudomonas, Humicola, Fusarium, Thielavia, Acremonium*, p. ej. las celulasas fúngicas producidas de *Humicola insolens, Myceliophthora thermophila* y *Fusarium oxysporum* descritas en US 4,435,307, US 5,648,263, US 5,691,178, US 5,776,757 y WO 89/09259.

Las celulasas especialmente adecuadas son las celulasas alcalinas o neutras que tienen beneficios de cuidado del color. Ejemplos de celulasas de este tipo son las celulasas descritas en EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasa tales como 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/oxidasas: peroxidasas/oxidasas adecuadas incluyen aquellas de origen vegetal, bacteriano o fúngico. Se incluyen mutantes modificados químicamente o creados genéticamente de proteínas. Ejemplos de peroxidasas útiles incluyen peroxidasas de *Coprinus*, p. ej. de *C. cinereus*, y variantes de las mismas como aquellas descritas en WO 93/24618, WO 95/10602, y WO 98/15257.

La(s) enzima(s) detergente(s) puede(n) ser incluida(s) en una composición detergente añadiendo aditivos separados que contienen una o más enzimas, o añadiendo un aditivo combinado comprendiendo todas estas enzimas. Un aditivo detergente de la invención, es decir un aditivo separado o un aditivo combinado, puede ser formulado p. ej. como un granulado, un líquido, un lodo, etc. Formulaciones de aditivo detergente preferidas son granulados, en particular granulados no pulverulentos, líquidos, en particular líquidos estabilizados, o lodos.

Granulados no pulverulentos pueden ser producidos, p. ej., como se describe en US 4,106,991 y 4,661,452 y pueden opcionalmente ser revestidos por métodos conocidos en la técnica. Ejemplos de materiales de revestimiento ceroso son productos de poli(etileno óxido) (polietilenglicol, PEG) con pesos molares medios de 1000 a 20000; nonilfenoles etoxilados que tienen de 16 a 50 unidades de óxido de etileno; alcoholes etoxilados grasos donde el alcohol contiene de 12 a 20 átomos de carbono y en el cual hay 15 a 80 unidades de óxido de etileno; alcoholes grasos; ácidos grasos; y mono- y di- y triglicéridos de ácidos grasos. Ejemplos de materiales que forman películas de revestimiento adecuados para aplicación por técnicas de lecho fluidificado están dados en GB 1483591. Preparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol tal como propilenoglicol, un azúcar o alcohol de azúcar, ácido láctico o ácido bórico según métodos establecidos. Enzimas protegidas pueden ser preparadas según el método descrito en EP 238,216.

La composición detergente de la invención puede estar en cualquier forma conveniente, p. ej., una barra, una pastilla, un polvo, un gránulo, una pasta o un líquido. Un detergente líquido puede ser acuoso, normalmente conteniendo hasta un 70% agua y un 0-30% de solvente orgánico, o no acuoso.

La composición detergente comprende uno o más tensioactivos, que pueden ser no iónicos incluyendo semipolares y/o aniónicos y/o catiónicos y/o zwitteriónicos. Los tensioactivos están normalmente presentes a un nivel del 0,1% al 60% en peso.

Cuando se incluye en la misma el detergente normalmente contiene de aproximadamente el 1% a aproximadamente el 40% de un surfactante aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, alquil-sulfato (sulfato de alcohol graso), etoxisulfato de alcohol, alcanosulfonato secundario, alfa-sulfo éster metílico de ácido graso, ácido o jabón alquil o alquenilsuccínico.

Cuando se incluye en la misma el detergente normalmente contiene de aproximadamente el 0,2% a aproximadamente el 40% de un surfactante no iónico tal como alcohol etoxilato, nonilfenol, etoxilato alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, polihidroxi alquil amida de ácido graso, o derivados N-acil N-alquil de glucosamina (“glucamidas”).

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 etilenodiaminatetraacético, ácido dietilenotriaminopentaacético, ácido alquil o alquenilsuccínico, silicatos solubles o silicatos estratificados (p. ej. SKS-6 de Hoechst).

El detergente puede comprender uno o más polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli (etilenglicol), poli(vinil alcohol), poli(vinilpiridina-N-óxido), poli(vinilimidazol), policarboxilatos tales como poliacrilatos, copolímeros de ácido maléico/acrílico y copolímeros de lauril metacrilato/ácido acrílico.

El detergente puede contener un sistema blanqueante que puede comprender una fuente de H_2O_2 tal como perborato o percarbonato que puede ser combinado con un activador blanqueante de formación de peracido tal como tetraacetiletilenodiamina o nonanoiloxibencenosulfonato. De forma alternativa, el sistema blanqueante puede comprender peroxiácidos de p. ej. tipo amida, imida, o sulfona.

La(s) enzima(s) de la composición detergente de la invención puede(n) ser estabilizada(s) usando agentes convencionales estabilizantes, p. ej., un poliol tal como propilenoglicol o glicerol, un azúcar o alcohol de azúcar, ácido láctico, ácido bórico, o un derivado de ácido bórico, p. ej., un éster de borato aromático, o un derivado de fenilo de ácido borónico tal como ácido 4-formilfenil borónico, y la composición puede ser formulada como se describe en p. ej. WO 92/19709 y WO 92/19708.

El detergente puede también contener otros ingredientes de detergentes convencionales tales como p. ej. acondicionadores de tejidos incluyendo arcillas, reforzadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de supensión de suciedad, agentes de antirreposición de suciedad, tintes, bactericidas, blanqueadores ópticos, hidrotropos, inhibidores de decoloración, o perfumes.

Está actualmente contemplado que en las composiciones detergentes cualquier enzima, en particular el polipéptido de la invención, puede ser añadida 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.

El polipéptido de la invención puede adicionalmente ser incorporado en las formulaciones de detergente descritas en WO 97/07202 que está incorporada por la presente como referencia.

Producción de Etanol a partir de Biomasa

La presente invención también se refiere a métodos para producir etanol a partir de biomasa, tales como materiales celulósicos, comprendiendo el contacto de la biomasa con los polipéptidos de la invención. El etanol puede posteriormente ser recuperado. Los polipéptidos de la invención pueden ser producidos "*in-situ*", es decir, como parte de, o directamente en un proceso de producción de etanol, para cultivar una célula huésped o una cepa, que en su forma tipo salvaje es capaz de producir los polipéptidos, bajo condiciones propicias para la producción de los polipéptidos.

El etanol puede ser producido por degradación enzimática de biomasa y conversión de los polisacáridos liberados a etanol. Esta especie de etanol es frecuentemente referida como bioetanol o biocombustible. Éste puede ser usado como un aditivo o extendedor de combustible en mezclas de menos del 1% y hasta el 100% (un sustituto de combustible). En algunos países, tal como Brasil, el etanol está substituyendo la gasolina a una extensión muy grande.

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

Tres clases más importantes de enzimas de celulasa se utilizan para descomponer la biomasa:

- Las "endo-1,4-beta-glucanasas" o 1,4-beta-D-glucan-4-glucanohidrolasas (EC 3,2,1,4), que actúan de forma aleatoria en sustratos de 1,4-beta-glucano solubles e insolubles.
- Las "exo-1,4-beta-D-glucanasas" incluyendo las 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 referida como celobiohidrolasa I y II, que libera D-celobiosa de 1,4-beta-glucanos.
- Las "beta-D-glucosidasas" o beta-D-glucósido glucohidrolasas (EC 3.2.1.21), que actúan para liberar unidades de D-glucosa de celobiosa y celodextrinas solubles, al igual que un conjunto de glicósidos.

Estas tres clases de enzimas trabajan juntas sinergísticamente en una interacción compleja que resulta en descristalización eficaz e hidrólisis de celulosa nativa de biomasa para producir los azúcares reductores que son convertidos a etanol por fermentación.

La presente invención está posteriormente descrita por los ejemplos siguientes que no deberían ser interpretados como limitando el ámbito de la invención.

Ejemplos

Productos químicos usados como tampones y sustratos fueron productos comerciales de al menos grado reactivo.

Ejemplo 1

Selección molecular de celobiohidrasa II de hongos termofílicos

Las cepas fúngicas fueron crecidas en 80 ml de medio líquido (2.5% Avicel, 0.5% Glucosa, 0.14% $(NH_4)_2SO_4$) en matraces de Erlenmeyer de 500 ml. Los matraces fueron incubados durante 72 horas a 45ºC en un agitador giratorio a 165 r.p.m. El micelio fue cosechado por centrifugado a 7000 r.p.m. durante 30 minutos y almacenado a -80ºC antes del uso para extracción del ARN.

El ARN total fue extraído de 100 mg de micelio de cada cepa usando el RNeasy Mini Kit (QIAGEN, Cat.No.74904).

Cebadores degenerados fueron diseñados basándose en el alineamiento de secuencias de proteínas de CBHII ya conocidas. Los cebadores siguientes fueron diseñados.

SEC ID NO:3	CBHII 1S:	5' TGG GGN CA(A/G) TG(T/C) GGN GG 3'
SEC ID NO:4	CBHII 2S:	5' TGG (T/C)TN GGN TGG CCN GC 3'
SEC ID NO:5	CBHII 2AS:	5' GCN GGC CAN CCN A(A/G)C CA 3'(inverso)
SEC ID NO:6	CBHII 3AS:	5' TT(A/G) CAC CA(A/G) TCN CCC CA 3' (inverso)
SEC ID NO:7	CBHII 4AS:	5' GG(T/C) TTN ACC CAN AC(A/G) AA 3' (inverso)
SEC ID NO:8	CBHII 5AS:	5' AA(A/G) TAN GC(T/C) TG(A/G) AAC CA 3' (inverso)

El sistema 3' RACE (Gibco., Cat.No.18373-019) fue usado para sintetizar ADNc del ARN total. Aproximadamente 5 microgramos del ARN total fueron usados como molde y Cebador adaptador (proporcionado por sistema 3'RACE) fue usado para sintetizar la primera cadena de ADNc. Después el ADNc fue amplificado usando combinaciones diferentes de cebadores degenerados. La mezcla reactiva comprendía 2,5 microL10 de tampón PCR, 1,5 microL 25 mM de $MgCl_2$,1,5 microL 25 mM de $MgCl_2$, 0,5 microL 10 mM de mezcla dNTP, 0,5 microL, 10 microM de Cebador3', 0,5 microL AUAP (10 microM, proporcionado por sistema 3'RACE), 0,5 microL taqDNA Polimerasa (5 u/microL, Promega), 1 microL ADNc reacción de síntesis y sometida a autoclave, agua destilada a 25 microL. La PCR fue realizada según las condiciones siguientes: la reacción fue sometida a 94ºC durante 3 minutos seguido de 30 ciclos de 94ºC durante 30 seg, 50ºC durante 30 seg y extensión a 72ºC durante 1 minuto. Una fase de extensión final a 72ºC durante 10 minutos seguida de una fase de retención a 4ºC completó el programa.

Productos de PCR del tamaño adecuado para cada par de cebador fueron recuperados de 1% gel de agarosa (1 X TBE), después purificados por incubación en un baño maría a 60ºC seguido de purificación usando GFXTMPCR ADN y Gel Band Purification Kit. (Amersham Pharmacia Biotech Inc., Cat. Nº. 27-9602-01). Las concentraciones de productos purificados fueron determinadas midiendo la absorbencia de A260 y A280 en un espectrofotómetro. Después estos fragmentos purificados fueron ligados al vector pGEM-T (Promega, Cat.No.A3600) según el kit de Promega (Cat.No.A3600).

Usando el método "choque por calor" 1 microL de productos de ligadura fueron transformados en 50 microL de células competentes JM109 de alta eficiencia. Los cultivos de transformación fueron colocados en placas sobre placas LB con ampicilina/IPTG/X-gal, y las placas fueron incubadas durante toda la noche a 37ºC. Los clones recombinantes fueron identificados por selección de color en placas indicadoras y selección por PCR de las colonias. Los clones positivos fueron inoculados en 3 ml de medio líquido LB e incubados durante toda la noche a 37ºC en un agitador giratorio a 250 rpm. Las células fueron granuladas por centrifugado durante 5 min a 10,000xg y muestra de plásmido fueron preparadas a partir del granulado celular usando el sistema de purificación de ADN Minipreps (Promega, Cat.No.A7100). Finalmente los plásmidos fueron secuenciados con el kit de reacción BigDye Terminator Cycle Sequencing Ready (PE) usando un secuenciador ABI377. La reacción de secuenciación fue la siguiente: 4 microL de mezcla de reacción Terminator Ready, 1.0-1.5 microgramos de ADN plásmido, 3.2 pmol de cebador y dH_2O a un volumen final de 10 microL.

Análisis de secuencias de los clones de ADNc de distintos pares de cebador mostraron que las secuencias contienen regiones codificantes del gen CBHII. Los cebadores fueron usados exitosamente para selección molecular del gen CBHII de *Chaetomium thermophilum*. Las secuencias de ADN codificantes de CBH II identificadas están mostradas como SEC ID NO:1.

De forma alternativa al método aplicado arriba, la biblioteca de ADNc podría ser seleccionada para el ADNc de longitud entera usando técnicas de hibridación estándares y la secuencia de ADNc parcial como una sonda. Los clones que dan una señal de hibridación positiva con la sonda son luego purificados y secuenciados para determinar la secuencia de ADNc más larga. La búsqueda y comparación de homología confirma que el ADNc de longitud entera corresponde a la secuencia de ADNc parcial de CBH II que fue originalmente usada como una sonda.

Los dos enfoques anteriormente descritos dependen de la presencia del ADNc de longitud entera de CBH II en la biblioteca de ADNc o en los ADNcs usados para su construcción. De forma alternativa, las técnicas RACE 5' y 3' (amplificación rápida de extremidades ADNc, por sus siglas en inglés) o técnicas derivadas identificadas podrían ser usadas para las regiones 5' y 3' perdidas. Para este propósito, mARNs de son aislados y utilizados para sintetizar ADNcs de primera cadena usando un cebador adaptador conteniendo oligo(dT) o un cebador específico del gen 5' (GSP).

El ADNc de longitud entera del gen CBH II puede también ser obtenido usando ADN genómico. El gen CBH II puede ser identificado por técnicas de PCR tales como aquella descrita más arriba o por selección de bibliotecas genómicas estándares usando técnicas de hibridación y el ADNc de CBH II parcial como una sonda. La búsqueda y comparación de homología con el ADNc de CBH II parcial se utiliza para que la secuencia genómica corresponda al gen CBH II. La identificación de secuencias de consenso tales como el sitio de iniciación de la transcripción, codones de inicio y de parada o sitios polyA podrían ser usados para definir la región que comprende el ADNc de longitud entera. Cebadores construidos a partir de las extremidades 5' y 3' de esta región podrían luego ser usados para amplificar el ADNc de longitud entera de la biblioteca de ARNm o de ADNc (ver arriba).

Por expresión del gen de longitud entera en un constructo de huésped de expresión adecuada la enzima CBH II es cosechada como una enzima intra celular o extra celular del caldo de cultivo.

Ejemplo 2

Usando Blast la secuencia de proteínas fue comparada con SWall, ERDBP, y GenSeqP. El núcleo catalítico fue predicho usando PFAMM HMM y sólo esta región de núcleo fue usada para buscar las bases de datos.

Chaetomium thermophilum NP000980 tiene un 83% de identidad a la glicosil hidrolasa de avicelasa II de *Humicola insolens* en SWALL:Q9C1S9, dominio de familia 6.

Ejemplo 3

Transformación de Aspergillus oryzae

Protoplastos son obtenidos a partir de cepa JAL 250 de *A. oryzae* donde el gen *pyrG* de la cepa huésped es deleccionado. La preparación y transformación de protoplastos se hace tal y como se ha descrito anteriormente (Christensen *et al*., *supra*). Los transformantes de *A. oryzae* que expresan orotidina monofosfato descarboxilasa son seleccionados basándose en su capacidad para crecer en la ausencia de uracilo. Los transformantes, son purificados de esporas dos veces en placas selectivas y los transformantes purificados de esporas son usados para análisis adicional.

Ejemplo 3

Actividad celobiohidrolasa

Una celobiohidrolasa está caracterizada por la capacidad para hidrolizar celulosa altamente cristalina muy eficazmente en comparación con otras celulasas. La celobiohidrolasa puede tener una actividad catalítica más alta usando PASC (celulosa hinchada por ácido fosfórico) como sustrato que usando CMC como sustrato. Para los objetivos de la presente invención, cualquiera de los ensayos siguientes puede utilizarse para identificar la actividad celobiohidrolasa:

Actividad en Azo-Avicel

Azo-Avicel (Megazyme, Bray Business Park, Bray, Wicklow, Irlanda) fue usada según las instrucciones del fabricante.

Actividad en PNP-beta-celobiosa

50 microL de solución de sustrato de CBH (5 mM de PNP beta-d-celobiosa (p-nitrofenil β-D-Cellobiosida Sigma N- 5759) en 0.1 M de tampón Na-acetato, pH 5.0) fue mezclado con 1 ml de solución de sustrato e incubado 20 minutos a 40ºC. La reacción fue detenida por adición de 5 ml de reactivo de parada (0.1 M de Na-carbonato, pH 11.5). La absorbencia fue medida a 404 nm.

Actividad en PASC y CMC

El sustrato es degradado con celobiohidrolasa para formar azúcares reductores. Una carbohidrato oxidasa de *Microdochium nivale* (rMnO) u otra oxidasa equivalente actúa en los azúcares reductores para formar H_2O_2 en presencia de O_2. El H_2O_2 formado 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 puede ser cuantificado por su absorbencia a 550 nm.

Cuando todos los componentes excepto la celobiohidrolasa están en exceso, el nivel de aumento en absorbencia es proporcional a la actividad de celobiohidrolasa. La reacción es una reacción de una fase cinética y puede ser realizada automáticamente en un analizador de centrifugación de Cobas Fara (Hoffmann La Roche) u otro espectrofotómetro equivalente que puede medir la cinética de estado de equilibrio.

Tampón: 50 mM tampón Na-acetato (pH 5.0);

Reactivos: Oxidasa rMnO, carbohidrato oxidasa purificada de *Microdochium nivale*, 2 mg/l

Peroxidasa, Sigma P-8125 (96 U/mg), 25 mg/l

4-aminoantipirina, SIGMA A-4382, 200 mg/l

TOPS, SIGMA E-8506, 600 mg/L

PASC o CMC (ver abajo), 5 g/l.

Todos los reactivos fueron añadidos al tampón en las concentraciones indicadas arriba y esta solución de reactivo fue mezclada íntegramente.

50 microL de muestra de celobiohidrolasa II (en una dilución adecuada) fue mezclada con 300 microL de solución de reactivo e incubada 20 minutos a 40ºC. Se detectó una formación de color morado y se midió como absorbencia a 550 nm.

El coeficiente de absorción de AA/TOPS-condensado es 0.01935 A_{550}/(microM cm). El nivel es calculado como micromoles de azúcar reductor producidos por minuto de OD_{550}/minuto y el coeficiente de absorción.

PASC

Materiales: 5 g Avicel® (Art. 2331 Merck);

150 ml de ácido Orto-fosfórico al 85% (Art. 573 Merck);

800 ml de acetona (Art. 14 Merck);

Aprox. 2 litros de agua desionizada (Milli-Q);

Vaso de precipitados de cristal de 1 L;

Embudo de filtro de cristal de 1 L;

Matraz de succión de 2 L;

Homogenizador Ultra Turrax.

Acetona y ácido orto-fosfórico son enfriados en hielo. Avicel® es humedecido con agua, y después se añaden 150 ml de ácido orto-fosfórico al 85% enfriado en hielo. La mezcla es colocada en un baño de hielo con agitación débil durante una hora.

Añadir 500 ml de acetona enfriada en hielo con agitación, y transferir la mezcla a un embudo de filtro de cristal y lavar con 3 x 100 ml de acetona enfriada en hielo, succionar hasta secarse todo lo posible en cada lavado. Lavar con 2 x 500 ml de agua (o hasta que no haya ningún olor de acetona), succionar hasta secarse todo lo posible en cada lavado.

Resuspender los sólidos en agua a un volumen total de 500 ml, y mezclar hasta homogeneidad usando un homogenizador Ultra Turrax. Almacenar mojado en frigorífico y equilibrar con tampón por centrifugado y resuspensión antes del uso.

CMC

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 corriente. Este agua fue sustituido por agua dulce dos veces al día durante 4 días.

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

La celulosa fue microfibrilada y se obtuvo una suspensión de microfibrillas de celulosa bacteriana individuales por homogenización de las microfibrillas de celulosa purificada en una mezcladora Waring durante 30 min. Las microfi-

brillas de celulosa fueron adicionalmente purificadas dializando 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 4ºC.

Depósito de material biológico

Centro de la Colección de Cultivos Microbiológicos Generales de China (CGMCC)

El material biológico siguiente ha sido depositado el 19 de diciembre de 2002 según las condiciones del tratado de Budapest con el Centro de la Colección de Cultivos Microbiológicos Generales de China (CGMCC), Instituto de Microbiología, Academia China de Ciencias, Haidian, Beijing 100080, China:

Número de acceso:	CGMCC 0859
Referencia del solicitante:	NP000980
Descripción:	*Chaetomium thermophilum*
Clasificación:	*Chaetomiaceae, Sordariales, Ascomycota*
Secuencia(s) relacionada(s):	SEC ID NO:1, SEC ID NO:2

Referencias citadas en la descripción

Esta lista de referencias citada por el solicitante ha sido recopilada exclusivamente para la información del lector. No forma parte del documento de patente europea. La misma ha sido confeccionada con la mayor diligencia; la OEP sin embargo no asume responsabilidad alguna por eventuales errores u omisiones.

Documentos de patente citados en la descripción

- US 6114158 A [0005]
- WO 9117244 A [0005]
- EP 214971 A [0005]
- EP 851031 A [0005]
- US 5605793 A [0081]
- US 5811238 A [0081]
- US 5830721 A [0081]
- US 6117679 A [0081]
- US 6159687 A [0081] [0083]
- WO 9841623 A [0081]
- US 6159688 A [0081]
- US 5965408 A [0081]
- US 6153510 A [0081]
- WO 9707205 A [0081]
- WO 9828416 A [0081]
- WO 9600787 A [0088] [0130]
- WO 9533836 A [0103]
- EP 238023 A [0130]
- WO 9114772 A [0146]

- WO 8906279 A [0162]
- WO 8906270 A [0162]
- WO 9425583 A [0162]
- WO 9219729 A [0163]
- WO 9820115 A [0163]
- WO 9820116 A [0163]
- WO 9834946 A [0163]
- EP 258068 A [0164]
- EP 305216 A [0164]
- WO 9613580 A [0164]
- EP 218272 A [0164]
- EP 331376 A [0164]
- GB 1372034 A [0164]
- WO 9506720 A [0164]
- WO 9627002 A [0164]
- WO 9612012 A [0164]
- JP 64744992 B [0164]
- WO 9116422 A [0164]
- WO 9205249 A [0165]
- WO 9401541 A [0165]
- EP 407225 A [0165]
- EP 260105 A [0165]
- WO 9535381 A [0165]
- WO 9600292 A [0165]
- WO 9530744 A [0165]
- WO 9425578 A [0165]
- WO 9514783 A [0165]
- WO 9522615 A [0165]
- WO 9704079 A [0165]
- WO 9707202 A [0165] [0183]
- GB 1296839 A [0166]
- WO 9402597 A [0167]
- WO 9418314 A [0167]
- WO 9623873 A [0167]

- WO 9743424 A [0167]
- US 4435307 A [0168]
- US 5648263 A [0168]
- US 5691178 A [0168]
- US 5776757 A [0168]
- WO 8909259 A [0168]
- EP 0495257 A [0169]
- EP 0531372 A [0169]
- WO 9611262 A [0169]
- WO 9629397 A [0169]
- WO 9808940 A [0169]
- WO 9407998 A [0169]
- EP 0531315 A [0169]
- US 5457046 A [0169]
- US 5686593 A [0169]
- US 5763254 A [0169]
- WO 9524471 A [0169]
- WO 9812307 A [0169]
- DK 9800299 W [0169]
- WO 9324618 A [0170]
- WO 9510602 A [0170]
- WO 9815257 A [0170]
- US 4106991 A [0172]
- US 4661452 A [0172]
- GB 1483591 A [0172]
- EP 238216 A [0172]
- WO 9219709 A [0180]
- WO 9219708 A [0180]

Bibliografía distinta de patentes citada en la descripción

- EERI *et al*. *GENE*, 1987, vol. 51,1, 43-52 [0005]
- **Deshpande** M V *et al*. *Methods in Enzymology*, 1988, 126-130 [0023]
- W. R. **Pearson**; D. J. **Lipman**. Improved Tools for Biological Sequence Analysis. *PNAS*, 1988, vol. 85, 2444-2448 [0025]
- W. R. **Pearson**. Rapid and Sensitive Sequence Comparison with FASTP and FASTA. *Methods in Enzymology*, 1990, vol. 183, 63-98 [0025]

- J. E. **Ness** *et al*. *Nature Biotechnology*, 1999, vol. 17, 893-896 [0043]
- H. **Neurath**; R. L. **Hill**. *The Proteins Academic Press* 1979. [0044]
- **Bolton**; **McCarthy**. *Proceedings of the National Academy of Sciences USA*, 1962, vol. 48, 1390- [0056]
- J. **Sambrook**; E. F. **Fritsch**; T. **Maniatus**. *Molecular Cloning, A Laboratory Manual Cold Spring Harbor* 1989. [0064]
- **Innis** *et al*. PCR: A Guide to Methods and Application *Academic Press*. 1990. [0069]
- **Cunningham**; **Wells**. *Science*, 1989, vol. 244, 1081-1085 [0073]
- de **Vos** *et al*. *Science*, 1992, vol. 255, 306-312 [0073]
- **Smith** *et al*. *Journal of Molecular Biology*, vol. 224, 899-904 [0073]
- **Wlodaver** *et al*. *FEBS Letters*, 1992, vol. 309, 59-64 [0073]
- **Ford** *et al*. *Protein Expression and Purification*, 1991, vol. 2, 95-107 [0075]
- **Kikuchi** *et al*. *Gene*, 2000, vol. 236, 159-167 [0082]
- **Kikuchi** *et al*. *Gene*, 2000, vol. 243, 133-137 [0082]
- **Crameri** *et al*. *Nature*, 1998, vol. 391, 288-291 [0083]
- **Ness** *et al*. *Nature Biotechnology*, vol. 17, 893-896 [0083]
- **Villa-Kamaroff** *et al*. *Proceedings of the National Academy of Sciences USA*, 1978, vol. 75, 3727-3731 [0087]
- **DeBoer** *et al*. *Proceedings of the National Academy of Sciences USA*, 1983, vol. 80, 21-25 [0087]
- Useful proteins from recombinant bacteria. *Scientific American*, 1980, vol. 242, 74-94 [0087]
- **Romanos** *et al*. *Yeast*, 1992, vol. 8, 423-488 [0089]
- **Guo**; **Sherman**. *Molecular Cellular Biology*, 1995, vol. 15, 5983-5990 [0098]
- **Simonen**; **Palva**. *Microbiological Reviews*, 1993, vol. 57, 109-137 [0100]
- **Ehrlich**. *Proceedings of the National Academy of Sciences USA*, 1978, vol. 75, 1433- [0115]
- **Chang**; **Cohen**. *Molecular General Genetics*, 1979, vol. 168, 111-115 [0121]
- **Young**; **Spizizin**. *Journal of Bacteriology*, 1961, vol. 81, 823-829 [0121]
- **Dubnau**; **Davidoff-Abelson**. *Journal of Molecular Biology*, 1971, vol. 56, 209-221 [0121]
- **Shigekawa**; **Dower**. *Biotechniques*, 1988, vol. 6, 742-751 [0121]
- **Koehler**; **Thorne**. *Journal of Bacteriology*, 1987, vol. 169, 5771-5278 [0121]
- **Hawksworth** *et al*. Ainsworth and Bisby's Dictionary of The Fungi CAB International, *University Press* [0123]
- Soc. App. Bacteriol. *Symposium Series* 1980. [0124]
- **Yelton** *et al*. *Proceedings of the National Academy of Sciences USA*, 1984, vol. 81, 1470-1474 [0130]
- **Malardier** *et al*. *Gene*, 1989, vol. 78, 147-156 [0130]
- Guide to Yeast Genetics and Molecular Biology. **Becker**; **Guarente**. *Methods in Enzymology Academic Press, Inc*. vol. 194, 182-187 [0130]
- **Ito** *et al*. *Journal of Bacteriology*, 1983, vol. 153, 163- [0130]
- **Hinnen** *et al*. *Proceedings of the National Academy of Sciences USA*, 1978, vol. 75, 1920- [0130]

- Protein Purification. *VCH Publishers* 1989. [0137]
- **Tague** *et al*. *Plant Physiology*, 1988, vol. 86, 506- [0145]
- **Franck** *et al*. *Cell*, 1980, vol. 21, 285-294 [0146]
- **Edwards**; **Coruzzi**. *Ann. Rev. Genet*., 1990, vol. 24, 275-303 [0146]
- **Ito** *et al*. *Plant Mol. Biol*., 1994, vol. 24, 863-878 [0146]
- **Wu** *et al*. *Plant and Cell Physiology*, 1998, vol. 39, 885-889 [0146]
- **Conrad** *et al*. *Journal of Plant Physiology*, 1998, vol. 152, 708-711 [0146]
- **Chen** *et al*. *Plant and Cell Physiology*, 1998, vol. 39, 935-941 [0146]
- **Kyozuka** *et al*. *Plant Physiology*, 1993, vol. 102, 991-1000 [0146]
- **Mitra**; **Higgins**. *Plant Molecular Biology*, 1994, vol. 26, 85-93 [0146]
- **Kagaya** *et al*. *Molecular and General Genetics*, 1995, vol. 248, 668-674 [0146]
- **Xu** *et al*. *Plant Molecular Biology*, 1993, vol. 22, 573-588 [0146]
- **Gasser** *et al*. *Science*, 1989, vol. 244, 1293- [0149]
- **Potrykus**. *Bio/Technology*, 1990, vol. 8, 535- [0149]
- **Shimamoto** *et al*. *Nature*, 1989, vol. 338, 274- [0149]
- **Hooykas**; **Schilperoort**. *Plant Molecular Biology*, 1992, vol. 19, 15-38 [0150]
- **Christou**. *Plant Journal*, 1992, vol. 2, 275-281 [0150]
- **Shimamoto**. *Current Opinion Biotechnology*, 1994, vol. 5, 158-162 [0150]
- **Vasil** *et al*. *Bio/Technology*, 1992, vol. 10, 667-674 [0150]
- **Omirulleh** *et al*. *Plant Molecular Biology*, 1993, vol. 21, 415-428 [0150]
- **Dartois** *et al*. *Biochemica et Biophysica Acta*, vol. 1131, 253-360 [0164]

REIVINDICACIONES

1. Polipéptido que tiene actividad celobiohidrolasa II, seleccionado del grupo que consiste en:

(a) un polipéptido que comprende una secuencia de aminoácidos, dicha secuencia de aminoácidos tiene al menos un 90% de identidad con la secuencia de aminoácidos mostrada como aminoácidos 1 a 477 de la SEC ID NO:2,

(b) un polipéptido que comprende una secuencia de aminoácidos, dicha secuencia de aminoácidos tiene al menos un 90% de identidad con el polipéptido codificado por los nucleótidos 63 a 1493 de la SEC ID NO:1,

(c) un polipéptido que es codificado por una secuencia de nucleótidos que se hibrida bajo condiciones de astringencia muy alta con,

(i) la cadena complementaria de los nucleótidos 63 a 1493 de la SEC ID NO:1,

(ii) la cadena complementaria de los nucleótidos 63 a 563 de la SEC ID NO:1, o

(iii) la cadena complementaria de los nucleótidos consistiendo en los nucleótidos 63 a 263 de SEC ID NO:1,

donde condiciones de astringencia muy alta comprenden prehibridación e hibridación a 42ºC en 5X SSPE, 1.0% SDS, 5X solución de Denhardt, 100 microg/ml de ADN de esperma de salmón cortado y desnaturalizado, y finalmente lavado del material portador tres veces cada vez durante 15 minutos usando 0,1 X SSC, 0,1% SDS a 68ºC, o,

(d) un fragmento de (a) o (b) que tiene actividad celobiohidrolasa II.

2. Polipéptido según la reivindicación 1, que comprende una secuencia de aminoácidos que tiene al menos un 95% identidad con la secuencia de aminoácidos mostrada coma aminoácidos 1 a 477 de SEC ID NO:2,

3. Polinucleátido que tiene una secuencia de nucleótidos que codifica para el polipéptido definido en cualquiera de las reivindicaciones 1 o 2.

4. Constructo de ácidos nucleicos que comprende la secuencia de nucleótidos definida en la reivindicación 3 operativamente enlazada a una o más secuencias de control que dirigen la producción del polipéptido en un huésped adecuado.

5. Vector de expresión recombinante que comprende el constructo de ácidos nucleicos definido en la reivindicación 4.

6. Célula huésped recombinante que comprende el constructo de ácidos nucleicos definido en la reivindicación 4.

7. Método para la producción de un polipéptido tal y como se define en cualquiera de las reivindicaciones 1 o 2, el método comprendiendo:

(a) cultivo de una célula huésped recombinante tal y como se define en la reivindicación 6 bajo condiciones propicias para la producción del polipéptido; y

(b) recuperación del polipéptido.

8. Método para producción *in situ* de un polipéptido tal y como se define en cualquiera de las reivindicaciones 1 o 2, el método comprendiendo:

(a) cultivo de una célula huésped recombinante tal y como se define en la reivindicación 6 bajo condiciones propicias para la producción del polipéptido; y

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

9. Método para redistribuir ADN comprendiendo el uso del polinucleótido tal y como se define en la reivindicación 3.

10. Método para la producción de etanol a partir de biomasa, comprendiendo la puesta en contacto de la biomasa con el polipéptido tal y como se define en cualquiera de las reivindicaciones 1 o 2.

11. Uso del polipéptido tal y como se define en cualquiera de las reivindicaciones 1 o 2 para producir etanol.

12. Planta transgénica, parte de planta o célula de planta, que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad celobiohidrolasa II tal y como se define en cualquiera de las reivindicaciones 1 o 2.

13. Composición detergente que comprende un tensioactivo y el polipéptido según cualquiera de las reivindicaciones 1 o 2.

LISTA DE SECUENCIAS

<110> Novozymes AS

<120> CBHII

<130> 10377.204-WO

<160> 36

<170> Versión de patentIn 3.2

<210> 1
<211> 1731
<212> ADN
<213> *Chaetomium thermophilum* NP000980

<220>
<221> CDS
<222> (63)..(1493)

<400> 1

```
cgggggggggg ggacagcaca acagagtcaa gacaagcttg gtcgctttgt cagaagttca      60

tc atg gct aag cag ctg ctg ctc act gcc gct ctt gcg gcc act tcg          107
   Met Ala Lys Gln Leu Leu Leu Thr Ala Ala Leu Ala Ala Thr Ser
   1               5                  10                  15

ctg gct gcc cct ctc ctt gag gag cgc cag agc tgc tcc tcc gtc tgg         155
Leu Ala Ala Pro Leu Leu Glu Glu Arg Gln Ser Cys Ser Ser Val Trp
                20                  25                  30

ggt caa tgc ggt ggc atc aat tac aac ggc ccg acc tgc tgc cag tcc         203
Gly Gln Cys Gly Gly Ile Asn Tyr Asn Gly Pro Thr Cys Cys Gln Ser
            35                  40                  45

ggc agt gtt tgc act tac ctg aat gac tgg tac agc cag tgc att ccc         251
Gly Ser Val Cys Thr Tyr Leu Asn Asp Trp Tyr Ser Gln Cys Ile Pro
        50                  55                  60

ggt cag gct cag ccc ggc acg act agc acc acg gct cgg acc acc agc         299
Gly Gln Ala Gln Pro Gly Thr Thr Ser Thr Thr Ala Arg Thr Thr Ser
    65                  70                  75

acc agc acc acc agc act tcg tcg gtc cgc ccg acc acc tcg aat acc         347
Thr Ser Thr Thr Ser Thr Ser Ser Val Arg Pro Thr Thr Ser Asn Thr
80                  85                  90                  95

cct gtg acg act gct ccc ccg acg acc acc atc ccg ggc ggc gcc tcg         395
Pro Val Thr Thr Ala Pro Pro Thr Thr Thr Ile Pro Gly Gly Ala Ser
                100                 105                 110

agc acg gcc agc tac aac ggc aac ccg ttt tcg ggt gtt caa ctt tgg         443
Ser Thr Ala Ser Tyr Asn Gly Asn Pro Phe Ser Gly Val Gln Leu Trp
            115                 120                 125

gcc aac acc tac tac tcg tcc gag gtg cac act ttg gcc atc ccc agc         491
Ala Asn Thr Tyr Tyr Ser Ser Glu Val His Thr Leu Ala Ile Pro Ser
        130                 135                 140

ttg tct cct gag ctg gct gcc aag gcc gcc aag gtc gct gag gtt ccc         539
Leu Ser Pro Glu Leu Ala Ala Lys Ala Ala Lys Val Ala Glu Val Pro
    145                 150                 155

agc ttc cag tgg ctc gac cgc aat gtg act gtt gac act ctc ttc tcc         587
Ser Phe Gln Trp Leu Asp Arg Asn Val Thr Val Asp Thr Leu Phe Ser
160                 165                 170                 175
```

```
ggc act ctt gcc gaa atc cgc gcc gcc aac cag cgc ggt gcc aac ccg      635
Gly Thr Leu Ala Glu Ile Arg Ala Ala Asn Gln Arg Gly Ala Asn Pro
                180                 185                 190

cct tat gcc ggc att ttc gtg gtt tat gac tta cca gac cgt gat tgc      683
Pro Tyr Ala Gly Ile Phe Val Val Tyr Asp Leu Pro Asp Arg Asp Cys
            195                 200                 205

gcg gct gct gct tcg aac ggc gag tgg tct atc gcc aac aat ggt gcc      731
Ala Ala Ala Ala Ser Asn Gly Glu Trp Ser Ile Ala Asn Asn Gly Ala
        210                 215                 220

aac aac tac aag cgc tac atc gac cgg atc cgt gag ctc ctt atc cag      779
Asn Asn Tyr Lys Arg Tyr Ile Asp Arg Ile Arg Glu Leu Leu Ile Gln
    225                 230                 235

tac tcc gat atc cgc act att ctg gtc att gaa cct gat tcc ctg gcc      827
Tyr Ser Asp Ile Arg Thr Ile Leu Val Ile Glu Pro Asp Ser Leu Ala
240                 245                 250                 255

aac atg gtc acc aac atg aac gtc cag aag tgc tcg aac gct gcc tcc      875
Asn Met Val Thr Asn Met Asn Val Gln Lys Cys Ser Asn Ala Ala Ser
                260                 265                 270

act tac aag gag ctt act gtc tat gcc ctc aaa cag ctc aat ctt cct      923
Thr Tyr Lys Glu Leu Thr Val Tyr Ala Leu Lys Gln Leu Asn Leu Pro
            275                 280                 285

cac gtt gcc atg tac atg gat gct ggc cac gct ggc tgg ctt ggc tgg      971
His Val Ala Met Tyr Met Asp Ala Gly His Ala Gly Trp Leu Gly Trp
        290                 295                 300

ccc gcc aac atc cag cct gct gct gag ctc ttt gct caa atc tac cgc     1019
Pro Ala Asn Ile Gln Pro Ala Ala Glu Leu Phe Ala Gln Ile Tyr Arg
    305                 310                 315

gac gct ggc agg ccc gct gct gtc cgc ggt ctt gcg acc aac gtt gcc     1067
Asp Ala Gly Arg Pro Ala Ala Val Arg Gly Leu Ala Thr Asn Val Ala
320                 325                 330                 335

aac tac aat gct tgg tcg atc gcc agc cct ccg tcc tac acc tct cct     1115
Asn Tyr Asn Ala Trp Ser Ile Ala Ser Pro Pro Ser Tyr Thr Ser Pro
                340                 345                 350

aac ccg aac tac gac gag aag cac tat att gag gcc ttt gct cct ctt     1163
Asn Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu Ala Phe Ala Pro Leu
            355                 360                 365

ctc cgc aac cag ggc ttc gac gca aag ttc atc gtc gac acc ggc cgt     1211
Leu Arg Asn Gln Gly Phe Asp Ala Lys Phe Ile Val Asp Thr Gly Arg
        370                 375                 380

aac ggc aag cag ccc act ggc cag ctt gaa tgg ggt cac tgg tgc aat     1259
Asn Gly Lys Gln Pro Thr Gly Gln Leu Glu Trp Gly His Trp Cys Asn
    385                 390                 395

gtc aag gga act ggc ttc ggt gtg cgc cct act gct aac act ggg cat     1307
Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ala Asn Thr Gly His
400                 405                 410                 415

gaa ctt gtt gat gct ttc gtg tgg gtc aag ccc ggt ggc gag tcc gac     1355
Glu Leu Val Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp
                420                 425                 430

ggc acc agt gcg gac acc agc gct gct cgt tat gac tat cac tgc ggc     1403
Gly Thr Ser Ala Asp Thr Ser Ala Ala Arg Tyr Asp Tyr His Cys Gly
            435                 440                 445
```

```
ctt tcc gac gca ctg act ccg gcg cct gag gct ggc caa tgg ttc cag     1451
Leu Ser Asp Ala Leu Thr Pro Ala Pro Glu Ala Gly Gln Trp Phe Gln
        450                 455                 460

gct tat ttc gaa cag ctg ctc atc aat gcc aac cct ccg ctc             1493
Ala Tyr Phe Glu Gln Leu Leu Ile Asn Ala Asn Pro Pro Leu
    465                 470                 475

tgaacggaag cggagatacc ggaaggcggt gagaagagcg gaattcaagt ctgcttatca   1553

aaatccactc accaagtgga ttaaagcgga tttatacatc tgagaaacaa cctgctttaa   1613

actcttcttg tacatatttc acttcgagac gtgcctcttt ctcaggagca ctgtagatac   1673

caatatatct gtcacatttc atataaaaaa aaaaaaaaag aaaaaaagta ctagtcga     1731
```

<210> 2

<211> 477

<212> PRT

<213> *Chaetomium thermophilum* NP000980

<400> 2

```
Met Ala Lys Gln Leu Leu Leu Thr Ala Ala Leu Ala Ala Thr Ser Leu
1               5                   10                  15

Ala Ala Pro Leu Leu Glu Glu Arg Gln Ser Cys Ser Ser Val Trp Gly
            20                  25                  30

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

Ser Val Cys Thr Tyr Leu Asn Asp Trp Tyr Ser Gln Cys Ile Pro Gly
    50                  55                  60

Gln Ala Gln Pro Gly Thr Thr Ser Thr Thr Ala Arg Thr Thr Ser Thr
65                  70                  75                  80

Ser Thr Thr Ser Thr Ser Ser Val Arg Pro Thr Thr Ser Asn Thr Pro
                85                  90                  95

Val Thr Thr Ala Pro Pro Thr Thr Thr Ile Pro Gly Gly Ala Ser Ser
            100                 105                 110

Thr Ala Ser Tyr Asn Gly Asn Pro Phe Ser Gly Val Gln Leu Trp Ala
        115                 120                 125

Asn Thr Tyr Tyr Ser Ser Glu Val His Thr Leu Ala Ile Pro Ser Leu
    130                 135                 140

Ser Pro Glu Leu Ala Ala Lys Ala Ala Lys Val Ala Glu Val Pro Ser
145                 150                 155                 160

Phe Gln Trp Leu Asp Arg Asn Val Thr Val Asp Thr Leu Phe Ser Gly
                165                 170                 175
```

Thr Leu Ala Glu Ile Arg Ala Ala Asn Gln Arg Gly Ala Asn Pro Pro
180 185 190

Tyr Ala Gly Ile Phe Val Val Tyr Asp Leu Pro Asp Arg Asp Cys Ala
195 200 205

Ala Ala Ala Ser Asn Gly Glu Trp Ser Ile Ala Asn Asn Gly Ala Asn
210 215 220

Asn Tyr Lys Arg Tyr Ile Asp Arg Ile Arg Glu Leu Leu Ile Gln Tyr
225 230 235 240

Ser Asp Ile Arg Thr Ile Leu Val Ile Glu Pro Asp Ser Leu Ala Asn
245 250 255

Met Val Thr Asn Met Asn Val Gln Lys Cys Ser Asn Ala Ala Ser Thr
260 265 270

Tyr Lys Glu Leu Thr Val Tyr Ala Leu Lys Gln Leu Asn Leu Pro His
275 280 285

Val Ala Met Tyr Met Asp Ala Gly His Ala Gly Trp Leu Gly Trp Pro
290 295 300

Ala Asn Ile Gln Pro Ala Ala Glu Leu Phe Ala Gln Ile Tyr Arg Asp
305 310 315 320

Ala Gly Arg Pro Ala Ala Val Arg Gly Leu Ala Thr Asn Val Ala Asn
325 330 335

Tyr Asn Ala Trp Ser Ile Ala Ser Pro Pro Ser Tyr Thr Ser Pro Asn
340 345 350

Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu Ala Phe Ala Pro Leu Leu
355 360 365

Arg Asn Gln Gly Phe Asp Ala Lys Phe Ile Val Asp Thr Gly Arg Asn
370 375 380

Gly Lys Gln Pro Thr Gly Gln Leu Glu Trp Gly His Trp Cys Asn Val
385 390 395 400

Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ala Asn Thr Gly His Glu
405 410 415

Leu Val Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly
420 425 430

Thr Ser Ala Asp Thr Ser Ala Ala Arg Tyr Asp Tyr His Cys Gly Leu
435 440 445

Ser Asp Ala Leu Thr Pro Ala Pro Glu Ala Gly Gln Trp Phe Gln Ala
450 455 460

Tyr Phe Glu Gln Leu Leu Ile Asn Ala Asn Pro Pro Leu
465 470 475

<210> 3
<211> 1272
<212> ADN
<213> *Myceliophthora thermophila*

<220>
<221> característica_misc
<222> (61)..(61)
<223> n es a, c, g, o t

<220>
<221> característica_misc
<222> (87)..(87)
<223> n es a, c, g, o t

<220>
<221> característica_misc
<222> (420)..(420)
<223> n es a, c, g, o t

<400> 3

```
acgtggatcc gaattcaagc ttcatggggt tctcctgcga tccactagta acggctcgcc     60
ngagagctgg aaagccagca gcacctnttg gagaggcagc tctctctctc caccacgccc    120
ccgcccgtct ccagccctcg tgaccagcat tcccggcggt gcgacctcca cggcgagcta    180
ctctggcaac cccttctcgg gcgtccggct cttcgccaac gactactaca ggtccgaggt    240
ccacaatctc gccattccta gcatgactgg tactctggcg gctcaaggct tccgccgtcg    300
cgcgaagtcc ctagcttcca gtggctcgac acggaacgtg cactcatcag acaccctgat    360
ggtccagact ctgtacccag gtccgggctc tcaataaggc acggtgaaca atcctacccn    420
tatgctgccc aactcgtcgt ctacgacctc cccgaccgtg actgtgccgc cgctgcgtcc    480
aacggygagt tttcgattgc aaacggcggc gccgccaact acaggagcta catcgacgct    540
atccgcaagc acatcattga gtactcggac atccggatca tcctggttat cgagcccgac    600
tcgatggcca acatggtgac caacatgaac gtggccaagt gcagcaacgc cgcgtcgacg    660
taccacgagt tgaccgtgta cgcgctcaag cagctgaacc tgcccaacgt cgccatgtat    720
ctcgacgccg gccacgccgg ctggctcggc tggcccgcca acatccagcc cgccgccgag    780
ctgtttgccg gcatctacaa tgatgccggc aagccggctg ccgtccgcgg cctggccact    840
aacgtcgcca actacaacgc ctggagcatc gcttcggccc cgtcgtacac gtcggctaac    900
cctaactacg acgagaagca ctacatcgag gccttcagcc cgctcttgaa ctcggccggc    960
ttccccgcac gcttcattgt cgacactggc cgcaacggca aacaacctac cggccaacaa   1020
cagtggggcg actggtgcaa tgtcaagggc accggctttg gcgtgcgccc gacggccaac   1080
acgggccacg agctggtcga tgcctttgtc tgggtcaagc ccggcggcga gtccgacggc   1140
acaagcgaca ccagcgccgc ccgctacgac taccactgcg gcctgtccga tgccctgcag   1200
cctgcccccg aggctggaca gtggttccag gcctacttcg agcagctgct caccaacgcc   1260
aacccgccct tc                                                       1272
```

<210> 4
<211> 420
<212> PRT
<213> *Myceliophthora thermophila*

<400> 4

```
Thr Trp Ile Arg Ile Gln Ala Ser Trp Gly Ser Pro Ala Ile His Arg
1               5                  10                  15

Leu Ala Glu Leu Glu Ser Gln Gln His Leu Leu Glu Arg Gln Leu Ser
            20                  25                  30

Leu Ser Thr Thr Pro Pro Pro Val Ser Ser Pro Arg Asp Gln His Ser
        35                  40                  45

Arg Arg Cys Asp Leu His Gly Glu Leu Leu Trp Gln Pro Leu Leu Gly
    50                  55                  60

Arg Pro Ala Leu Arg Gln Arg Leu Leu Gln Val Arg Gly Pro Gln Ser
65                  70                  75                  80

Arg His Ser His Asp Trp Tyr Ser Gly Gly Ser Arg Leu Pro Pro Ser
                85                  90                  95

Arg Glu Val Pro Ser Phe Gln Trp Leu Asp Thr Glu Arg Ala Leu Ile
            100                 105                 110

Arg His Pro Asp Gly Pro Asp Ser Val Pro Arg Ser Gly Leu Ser Ile
        115                 120                 125

Arg His Gly Glu Gln Ser Tyr Pro Tyr Ala Ala Gln Leu Val Val Tyr
    130                 135                 140

Asp Leu Pro Asp Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe
145                 150                 155                 160

Ser Ile Ala Asn Gly Gly Ala Ala Asn Tyr Arg Ser Tyr Ile Asp Ala
                165                 170                 175

Ile Arg Lys His Ile Ile Glu Tyr Ser Asp Ile Arg Ile Ile Leu Val
            180                 185                 190

Ile Glu Pro Asp Ser Met Ala Asn Met Val Thr Asn Met Asn Val Ala
        195                 200                 205

Lys Cys Ser Asn Ala Ala Ser Thr Tyr His Glu Leu Thr Val Tyr Ala
```

```
      210                 215                 220

Leu Lys Gln Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala Gly
225                 230                 235                 240

His Ala Gly Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Glu
                245                 250                 255

Leu Phe Ala Gly Ile Tyr Asn Asp Ala Gly Lys Pro Ala Ala Val Arg
            260                 265                 270

Gly Leu Ala Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Ile Ala Ser
        275                 280                 285

Ala Pro Ser Tyr Thr Ser Ala Asn Pro Asn Tyr Asp Glu Lys His Tyr
    290                 295                 300

Ile Glu Ala Phe Ser Pro Leu Leu Asn Ser Ala Gly Phe Pro Ala Arg
305                 310                 315                 320

Phe Ile Val Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln
                325                 330                 335

Gln Trp Gly Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg
            340                 345                 350

Pro Thr Ala Asn Thr Gly His Glu Leu Val Asp Ala Phe Val Trp Val
        355                 360                 365

Lys Pro Gly Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala Arg
    370                 375                 380

Tyr Asp Tyr His Cys Gly Leu Ser Asp Ala Leu Gln Pro Ala Pro Glu
385                 390                 395                 400

Ala Gly Gln Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr Asn Ala
                405                 410                 415

Asn Pro Pro Phe
            420
```

<210> 5

<211> 419

<212> ADN

<213> *Acremonium* sp. T178-4 NP001132

<220>

<221> CDS

<222> (1)..(417)

<400> 5

```
tac gca agt gtc tac tcg gac gcc gga tca ccg gct gca ctc cgc ggt       48
Tyr Ala Ser Val Tyr Ser Asp Ala Gly Ser Pro Ala Ala Leu Arg Gly
1               5                  10                  15

ctc gct acc aac gtc gcc aat tac aac gcc tgg aca atc gat acc tgc       96
Leu Ala Thr Asn Val Ala Asn Tyr Asn Ala Trp Thr Ile Asp Thr Cys
            20                  25                  30

cct tca tac aca cag ggt aac tcc att tgc gac gag aag gac tac atc      144
Pro Ser Tyr Thr Gln Gly Asn Ser Ile Cys Asp Glu Lys Asp Tyr Ile
        35                  40                  45

aat gcg ctt gct ccc ctg ctt cgc agc tca ggg ctt acg gac gct cat      192
Asn Ala Leu Ala Pro Leu Leu Arg Ser Ser Gly Leu Thr Asp Ala His
    50                  55                  60

ttc atc act gat acc ggc cgc aac ggc aag caa cca aca ggc caa caa      240
Phe Ile Thr Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln
65                  70                  75                  80

gcc tgg ggc gac tgg tgc aat gtc atc ggc acg ggc ttt ggc gtg cgc      288
Ala Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly Val Arg
                85                  90                  95

ccg tcc acg aac aca ggt gat tct tta ctt gac gcc ttc gtc tgg gtt      336
Pro Ser Thr Asn Thr Gly Asp Ser Leu Leu Asp Ala Phe Val Trp Val
            100                 105                 110

aaa ccc ggt ggc gag agt gac ggg act tct gat act tgt gcg gcg cgg      384
Lys Pro Gly Gly Glu Ser Asp Gly Thr Ser Asp Thr Cys Ala Ala Arg
        115                 120                 125

tat gat gcg cat tgc ggg tat agc gat gcg ctg ca                      419
Tyr Asp Ala His Cys Gly Tyr Ser Asp Ala Leu
    130                 135
```

<210> 6

<211> 139

<212> PRT

<213> *Acremonium* sp. T178-4 NP001132

<400> 6

```
Tyr Ala Ser Val Tyr Ser Asp Ala Gly Ser Pro Ala Ala Leu Arg Gly
1               5                  10                  15

Leu Ala Thr Asn Val Ala Asn Tyr Asn Ala Trp Thr Ile Asp Thr Cys
            20                  25                  30

Pro Ser Tyr Thr Gln Gly Asn Ser Ile Cys Asp Glu Lys Asp Tyr Ile
        35                  40                  45

Asn Ala Leu Ala Pro Leu Leu Arg Ser Ser Gly Leu Thr Asp Ala His
    50                  55                  60

Phe Ile Thr Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln
65                  70                  75                  80

Ala Trp Gly Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly Val Arg
                85                  90                  95

Pro Ser Thr Asn Thr Gly Asp Ser Leu Leu Asp Ala Phe Val Trp Val
            100                 105                 110

Lys Pro Gly Gly Glu Ser Asp Gly Thr Ser Asp Thr Cys Ala Ala Arg
        115                 120                 125

Tyr Asp Ala His Cys Gly Tyr Ser Asp Ala Leu
    130                 135
```

<210> 7
<211> 306
<212> ADN
<213> *Melanocarpus* sp. AT181-3 NP001133

<220>
<221> CDS
<222> (1)..(306)

<400> 7

```
gat gcc ggc aag ccg cac tcg gtc cgc ggt ctc gcc acc aac gtc gcc        48
Asp Ala Gly Lys Pro His Ser Val Arg Gly Leu Ala Thr Asn Val Ala
1               5                   10                  15

aac tac aat gcc tgg agc gtc gcc tcg gcc ccg cct tac acc agc ccc        96
Asn Tyr Asn Ala Trp Ser Val Ala Ser Ala Pro Pro Tyr Thr Ser Pro
            20                  25                  30

aac ccc aac tac gat gag aag cac tac att gag gcc ttc agc cct ctc       144
Asn Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu Ala Phe Ser Pro Leu
        35                  40                  45

ctt gag gcc cgc ggc ttc cct gcc cgc ttc atc gtc gac cag ggc cgc       192
Leu Glu Ala Arg Gly Phe Pro Ala Arg Phe Ile Val Asp Gln Gly Arg
    50                  55                  60

agc ggc aag cag ccc acc ggc cag aag gag tgg ggc cac tgg tgc aac       240
Ser Gly Lys Gln Pro Thr Gly Gln Lys Glu Trp Gly His Trp Cys Asn
65                  70                  75                  80

gct atc ggc acc ggc ttc ggc att cgc ccg acc gcc aac acc ggc cac       288
Ala Ile Gly Thr Gly Phe Gly Ile Arg Pro Thr Ala Asn Thr Gly His
                85                  90                  95

aac ctg gtt gat gcc ttc                                               306
Asn Leu Val Asp Ala Phe
            100
```

<210> 8

<211> 102

<212> PRT

<213> *Melanocarpus* sp. AT181-3 NP001133

<400> 8

```
Asp Ala Gly Lys Pro His Ser Val Arg Gly Leu Ala Thr Asn Val Ala
1               5                   10                  15

Asn Tyr Asn Ala Trp Ser Val Ala Ser Ala Pro Pro Tyr Thr Ser Pro
            20                  25                  30

Asn Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu Ala Phe Ser Pro Leu
        35                  40                  45

Leu Glu Ala Arg Gly Phe Pro Ala Arg Phe Ile Val Asp Gln Gly Arg
    50                  55                  60

Ser Gly Lys Gln Pro Thr Gly Gln Lys Glu Trp Gly His Trp Cys Asn
65                  70                  75                  80

Ala Ile Gly Thr Gly Phe Gly Ile Arg Pro Thr Ala Asn Thr Gly His
                85                  90                  95

Asn Leu Val Asp Ala Phe
            100
```

<210> 9

<211> 432

<212> ADN

<213> *Thielavia* cf. microspora T046-1 NP001134

<220>

<221> CDS

<222> (1)..(432)

<400> 9

```
gcc aac atc cag ccc gct gcc acc ctg ttc gcc ggc atc tac agc gac      48
Ala Asn Ile Gln Pro Ala Ala Thr Leu Phe Ala Gly Ile Tyr Ser Asp
1               5                  10                  15

gct ggc aag ccc gcc tcg gtc cgc ggt ttg gcc acc aac gtg gcc aac      96
Ala Gly Lys Pro Ala Ser Val Arg Gly Leu Ala Thr Asn Val Ala Asn
            20                  25                  30

tac aac gcc tgg agc ctg tcg tcg gcg ccg tcg tac acg agc ccc aac     144
Tyr Asn Ala Trp Ser Leu Ser Ser Ala Pro Ser Tyr Thr Ser Pro Asn
        35                  40                  45

gcc aac tac gac gag aag cac tac gtc gag gcc ttt gcc ccg ctc ctc     192
Ala Asn Tyr Asp Glu Lys His Tyr Val Glu Ala Phe Ala Pro Leu Leu
    50                  55                  60

cag gcg gcc ggc ttc ccc gcc aag ttc atc acc gac acg ggc cgc aac     240
Gln Ala Ala Gly Phe Pro Ala Lys Phe Ile Thr Asp Thr Gly Arg Asn
65                  70                  75                  80

ggc aag cag ccc acg ggc cag agc gcg tgg ggc gac tgg tgc aac gtc     288
Gly Lys Gln Pro Thr Gly Gln Ser Ala Trp Gly Asp Trp Cys Asn Val
                85                  90                  95

aag ggc acc ggc ttc ggt gtc cgc ccg acc tcg gag acg ggc cac gac     336
Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ser Glu Thr Gly His Asp
            100                 105                 110

ctc ctc gac gcc ttc gtc tgg gtc aag ccc ggt ggc gag tcg gac ggc     384
Leu Leu Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly
        115                 120                 125

acc agc gac acc agc gcc gcc cgc tac gac tac cac tgc ggt ctg tcg     432
Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp Tyr His Cys Gly Leu Ser
    130                 135                 140
```

<210> 10

<211> 144

<212> PRT

<213> *Thielavia* cf. microspora T046-1 NP001134

<400> 10

```
Ala Asn Ile Gln Pro Ala Ala Thr Leu Phe Ala Gly Ile Tyr Ser Asp
1               5                  10                  15

Ala Gly Lys Pro Ala Ser Val Arg Gly Leu Ala Thr Asn Val Ala Asn
            20                  25                  30

Tyr Asn Ala Trp Ser Leu Ser Ser Ala Pro Ser Tyr Thr Ser Pro Asn
        35                  40                  45

Ala Asn Tyr Asp Glu Lys His Tyr Val Glu Ala Phe Ala Pro Leu Leu
    50                  55                  60

Gln Ala Ala Gly Phe Pro Ala Lys Phe Ile Thr Asp Thr Gly Arg Asn
65                  70                  75                  80

Gly Lys Gln Pro Thr Gly Gln Ser Ala Trp Gly Asp Trp Cys Asn Val
                85                  90                  95

Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ser Glu Thr Gly His Asp
            100                 105                 110

Leu Leu Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly
        115                 120                 125

Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp Tyr His Cys Gly Leu Ser
    130                 135                 140
```

<210>11
<211> 297
<212> ADN
<213> *Aspergillus* sp. T186-2 NP001136

<220>
<221> CDS
<222> (1)..(297)

<400> 11

```
gga ttg ttt ggg tgg cct gcc aac ctt act cct tcc gct cgt ctc ttc        48
Gly Leu Phe Gly Trp Pro Ala Asn Leu Thr Pro Ser Ala Arg Leu Phe
1               5                   10                  15

gcc caa atc tac aag gat gcc ggc agg tct gcc ttc atc cgt ggt ctt        96
Ala Gln Ile Tyr Lys Asp Ala Gly Arg Ser Ala Phe Ile Arg Gly Leu
            20                  25                  30

gcc acc aac gtc tcc aac tac aac gcc ctc agt gca acc acc cgt gat       144
Ala Thr Asn Val Ser Asn Tyr Asn Ala Leu Ser Ala Thr Thr Arg Asp
        35                  40                  45

ccc gtc acc cag ggc aat gac aac tac gat gag ctc cgc ttc atc aac       192
Pro Val Thr Gln Gly Asn Asp Asn Tyr Asp Glu Leu Arg Phe Ile Asn
    50                  55                  60

gct ctt gct cct ctc ctc cga aat gaa ggc tgg gac gcc aag ttc atc       240
Ala Leu Ala Pro Leu Leu Arg Asn Glu Gly Trp Asp Ala Lys Phe Ile
65                  70                  75                  80

gtc gac cag ggt cgt tct ggt gtc cag aac atc cga cag gag tgg ggc       288
Val Asp Gln Gly Arg Ser Gly Val Gln Asn Ile Arg Gln Glu Trp Gly
                85                  90                  95

gac tgg tgc                                                          297
Asp Trp Cys
```

<210> 12

<211> 99

<212> PRT

<213> *Aspergillus* sp. T186-2 NP001136

<400> 12

```
Gly Leu Phe Gly Trp Pro Ala Asn Leu Thr Pro Ser Ala Arg Leu Phe
1               5                   10                  15

Ala Gln Ile Tyr Lys Asp Ala Gly Arg Ser Ala Phe Ile Arg Gly Leu
            20                  25                  30

Ala Thr Asn Val Ser Asn Tyr Asn Ala Leu Ser Ala Thr Thr Arg Asp
        35                  40                  45

Pro Val Thr Gln Gly Asn Asp Asn Tyr Asp Glu Leu Arg Phe Ile Asn
    50                  55                  60

Ala Leu Ala Pro Leu Leu Arg Asn Glu Gly Trp Asp Ala Lys Phe Ile
65                  70                  75                  80

Val Asp Gln Gly Arg Ser Gly Val Gln Asn Ile Arg Gln Glu Trp Gly
                85                  90                  95

Asp Trp Cys
```

<210> 13

<211> 420

<212> ADN

<213> *Thielavia cf. australiensis* T55-15 NP001000

<220>

<221> CDS

<222> (1)..(420)

<400> 13

```
tgg ctg ggg tgg ccc gcc aac atc cag ccc gct gct acc ctg ttc gcc      48
Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Thr Leu Phe Ala
1               5                  10                  15

ggc atc tac aac gac gct ggc aag ccc gcc tcg gtc cgt ggt ctg gcc      96
Gly Ile Tyr Asn Asp Ala Gly Lys Pro Ala Ser Val Arg Gly Leu Ala
            20                  25                  30

acc aac gtt gcc aac tac aac gcc tgg agc ctg tcc tcg gcc ccg tcg     144
Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Leu Ser Ser Ala Pro Ser

        35                  40                  45

tac acg acc ccc aac gcc aac tac gac gag aag cac tac gtc gag gcc     192
Tyr Thr Thr Pro Asn Ala Asn Tyr Asp Glu Lys His Tyr Val Glu Ala
    50                  55                  60

ttt gcc ccg ctt ctc tcg gcc gct ggc ttc ccc gcc aag ttc atc acc     240
Phe Ala Pro Leu Leu Ser Ala Ala Gly Phe Pro Ala Lys Phe Ile Thr
65                  70                  75                  80

gac act ggc cgc aac ggc aag cag ccc acc ggc cag agc cag tgg ggc     288
Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Ser Gln Trp Gly
                85                  90                  95

gat tgg tgc aac gtc aag ggc acc ggc ttc ggt gtc cgc ccg acc tcc     336
Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ser
            100                 105                 110

gag acg ggc cac gag ctc ctg gat gcc ttt gtc tgg gcc aag ccc ggt     384
Glu Thr Gly His Glu Leu Leu Asp Ala Phe Val Trp Ala Lys Pro Gly
        115                 120                 125

ggc gag tcc gac ggt acc agc gac acc agc gct gcc                     420
Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala
    130                 135                 140
```

<210> 14

<211> 140

<212> PRT

<213> *Thielavia cf. australiensis* T55-15 NP001000

<400> 14

```
Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Thr Leu Phe Ala
1               5                  10                  15

Gly Ile Tyr Asn Asp Ala Gly Lys Pro Ala Ser Val Arg Gly Leu Ala
            20                  25                  30

Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Leu Ser Ser Ala Pro Ser
        35                  40                  45

Tyr Thr Thr Pro Asn Ala Asn Tyr Asp Glu Lys His Tyr Val Glu Ala
    50                  55                  60

Phe Ala Pro Leu Leu Ser Ala Ala Gly Phe Pro Ala Lys Phe Ile Thr
65                  70                  75                  80

Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Ser Gln Trp Gly
                85                  90                  95

Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ser
            100                 105                 110

Glu Thr Gly His Glu Leu Leu Asp Ala Phe Val Trp Ala Lys Pro Gly
        115                 120                 125

Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala
    130                 135                 140
```

<210> 15
<211> 1221
<212> ADN
<213> *Aspergillus tubingensis* NP001143

<220>
<221> CDS
<222> (1)..(1221)

<220>
<221> característica_misc
<222> (903)..(903)
<223> n es a, c, g, o t

<220>
<221> característica_misc
<222> (1011)..(1011)

<223> n es a, c, g, o t

<220>
<221> característica_misc
<222> (1017)..(1017)
<223> n es a, c, g, o t

<400> 15

```
atg aat atg cac tcc atc aac atg cga gcc atc tgg ccc ctc gtc tct        48
Met Asn Met His Ser Ile Asn Met Arg Ala Ile Trp Pro Leu Val Ser
1               5                  10                  15

ctc ttc tct gcc gtt aag gcc ctc ccc gcc gca agc gcg act gct tca        96
Leu Phe Ser Ala Val Lys Ala Leu Pro Ala Ala Ser Ala Thr Ala Ser
            20                  25                  30

gcg tct gtt gcg gcc tcg agc tct ccg gcg ccg act gcc tct gct acc       144
Ala Ser Val Ala Ala Ser Ser Ser Pro Ala Pro Thr Ala Ser Ala Thr
        35                  40                  45

ggc aat ccc ttt gag gga tac cag ctc tat gtg aac ccc tac tat aag       192
Gly Asn Pro Phe Glu Gly Tyr Gln Leu Tyr Val Asn Pro Tyr Tyr Lys
    50                  55                  60

tcg caa gtg gag agt tcg gcc att cca tca ttg tct gct agt tcg ctg       240
Ser Gln Val Glu Ser Ser Ala Ile Pro Ser Leu Ser Ala Ser Ser Leu
65                  70                  75                  80

gtc gcg cag gcg agt gct gca gcc gat gtg cct tca ttt tac tgg cta       288
Val Ala Gln Ala Ser Ala Ala Ala Asp Val Pro Ser Phe Tyr Trp Leu
                85                  90                  95

gac acg gcc gac aag gtg cct acc atg ggt gaa tat ctg gat gac atc       336
Asp Thr Ala Asp Lys Val Pro Thr Met Gly Glu Tyr Leu Asp Asp Ile
            100                 105                 110

cag acg caa aac gcc gct gga gcg aat cct ccc att gct ggt atc ttc       384
Gln Thr Gln Asn Ala Ala Gly Ala Asn Pro Pro Ile Ala Gly Ile Phe
        115                 120                 125

gtc gtc tat gac ctg ccg gat cgg gat tgc gct gcc ttg gct agt aat       432
Val Val Tyr Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn
    130                 135                 140

ggg gaa tac gcg atc agt gat gga ggc gtg gag aag tat aag gcg tac       480
Gly Glu Tyr Ala Ile Ser Asp Gly Gly Val Glu Lys Tyr Lys Ala Tyr
145                 150                 155                 160
```

```
att gat tct att cgc gag cag gtc gag acg tac tcg gat gtt cag act      528
Ile Asp Ser Ile Arg Glu Gln Val Glu Thr Tyr Ser Asp Val Gln Thr
                165                 170                 175

att ttg att atc gaa ccg gat agc tta gct aac ctg gtg acg aat ctc      576
Ile Leu Ile Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu
            180                 185                 190

gat gtg gct aaa tgc gcc aat gct caa tct gct tac ctg gaa tgc acc      624
Asp Val Ala Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Thr
        195                 200                 205

aat tat gca ctt gag cag ttg aat ctr ccg aac gtg gct atg tat ctt      672
Asn Tyr Ala Leu Glu Gln Leu Asn Xaa Pro Asn Val Ala Met Tyr Leu
    210                 215                 220

gat gct ggc cat gct gga tgg ctg gga tgg cct gcc aac atc ggt ccc      720
Asp Ala Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn Ile Gly Pro
225                 230                 235                 240

gcg gcg gaa ctc tac gca tcg gtg tat aag aat gcg tcg tct cca gca      768
Ala Ala Glu Leu Tyr Ala Ser Val Tyr Lys Asn Ala Ser Ser Pro Ala
                245                 250                 255

gct gtt cgt gga ctc gct aca rac gta gct aac ttc aat gcc tgg agc      816
Ala Val Arg Gly Leu Ala Thr Xaa Val Ala Asn Phe Asn Ala Trp Ser
            260                 265                 270

atc gac act tgc ccc tcc tat acw tcg ggt aac gat gtc tgt gat gaa      864
Ile Asp Thr Cys Pro Ser Tyr Xaa Ser Gly Asn Asp Val Cys Asp Glu
        275                 280                 285

aaa agc tac atc aat gcc ttt gca ccg gag ctc tct agn gct gga ttt      912
Lys Ser Tyr Ile Asn Ala Phe Ala Pro Glu Leu Ser Xaa Ala Gly Phe
    290                 295                 300

gat gcc cac ttt att acc gat acg ggt cgc aat gga aag cag cct act      960
Asp Ala His Phe Ile Thr Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr
305                 310                 315                 320

gga caa agc gcg tgg ggt gac tgg ggc aat gtc aag gat act ggc ttc     1008
Gly Gln Ser Ala Trp Gly Asp Trp Gly Asn Val Lys Asp Thr Gly Phe
                325                 330                 335

ggn gct can ccg aca acc gat act gga aac gag ctg gct gat gcc ttt     1056
Gly Ala Xaa Pro Thr Thr Asp Thr Gly Asn Glu Leu Ala Asp Ala Phe
            340                 345                 350

gtc tgg gyc aac cct ggc gga aag agt gat ggg acg tcg gac act agc     1104
Val Trp Xaa Asn Pro Gly Gly Lys Ser Asp Gly Thr Ser Asp Thr Ser
        355                 360                 365

tct tct cgc tac gat gcg cat tgc gga tat agt gat gct ttg cag cct     1152
Ser Ser Arg Tyr Asp Ala His Cys Gly Tyr Ser Asp Ala Leu Gln Pro
    370                 375                 380

gcc ccg gag gct ggt act tgg ttc cag gca tac ttt gag cag ctt ttg     1200
Ala Pro Glu Ala Gly Thr Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu
385                 390                 395                 400

acc aat gcc aac cct tcc ctg                                        1221
Thr Asn Ala Asn Pro Ser Leu
                405
```

<210> 16

<211> 407

<212> PRT

<213> *Aspergillus tubingensis* NP001143

<220>

<221> característica_misc

<222> (217)..(217)

<223> El “Xaa” en la ubicación 217 simboliza Leu.

<220>
<221> característica_misc
<222> (264)..(264)
<223> El “Xaa” en la ubicación 264 simboliza Asp, o Asn.

<220>
<221> característica_misc
<222> (280)..(280)
<223> El “Xaa” en la ubicación 280 simboliza Thr.

<220>
<221> característica_misc
<222> (301) ..(301)
<223> El “Xaa” en la ubicación 301 simboliza Arg, o Ser.

<220>
<221> característica_misc
<222> (339)..(339)
<223> El “Xaa” en la ubicación 339 simboliza Gln, o His.

<220>
<221> característica_misc
<222> (355)..(355)
<223> El “Xaa” en la ubicación 355 simboliza Ala, o Val.

<400> 16

```
Met Asn Met His Ser Ile Asn Met Arg Ala Ile Trp Pro Leu Val Ser
1               5                   10                  15

Leu Phe Ser Ala Val Lys Ala Leu Pro Ala Ala Ser Ala Thr Ala Ser
            20                  25                  30

Ala Ser Val Ala Ala Ser Ser Ser Pro Ala Pro Thr Ala Ser Ala Thr
        35                  40                  45

Gly Asn Pro Phe Glu Gly Tyr Gln Leu Tyr Val Asn Pro Tyr Tyr Lys
    50                  55                  60

Ser Gln Val Glu Ser Ser Ala Ile Pro Ser Leu Ser Ala Ser Ser Leu
65                  70                  75                  80

Val Ala Gln Ala Ser Ala Ala Ala Asp Val Pro Ser Phe Tyr Trp Leu
                85                  90                  95

Asp Thr Ala Asp Lys Val Pro Thr Met Gly Glu Tyr Leu Asp Asp Ile
            100                 105                 110

Gln Thr Gln Asn Ala Ala Gly Ala Asn Pro Pro Ile Ala Gly Ile Phe
        115                 120                 125

Val Val Tyr Asp Leu Pro Asp Arg Asp Cys Ala Ala Leu Ala Ser Asn
    130                 135                 140
```

Gly Glu Tyr Ala Ile Ser Asp Gly Gly Val Glu Lys Tyr Lys Ala Tyr
145 150 155 160

Ile Asp Ser Ile Arg Glu Gln Val Glu Thr Tyr Ser Asp Val Gln Thr
165 170 175

Ile Leu Ile Ile Glu Pro Asp Ser Leu Ala Asn Leu Val Thr Asn Leu
180 185 190

Asp Val Ala Lys Cys Ala Asn Ala Gln Ser Ala Tyr Leu Glu Cys Thr
195 200 205

Asn Tyr Ala Leu Glu Gln Leu Asn Xaa Pro Asn Val Ala Met Tyr Leu
210 215 220

Asp Ala Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn Ile Gly Pro
225 230 235 240

Ala Ala Glu Leu Tyr Ala Ser Val Tyr Lys Asn Ala Ser Ser Pro Ala
245 250 255

Ala Val Arg Gly Leu Ala Thr Xaa Val Ala Asn Phe Asn Ala Trp Ser
260 265 270

Ile Asp Thr Cys Pro Ser Tyr Xaa Ser Gly Asn Asp Val Cys Asp Glu
275 280 285

Lys Ser Tyr Ile Asn Ala Phe Ala Pro Glu Leu Ser Xaa Ala Gly Phe
290 295 300

Asp Ala His Phe Ile Thr Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr
305 310 315 320

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

Gly Ala Xaa Pro Thr Thr Asp Thr Gly Asn Glu Leu Ala Asp Ala Phe
340 345 350

Val Trp Xaa Asn Pro Gly Gly Lys Ser Asp Gly Thr Ser Asp Thr Ser
355 360 365

Ser Ser Arg Tyr Asp Ala His Cys Gly Tyr Ser Asp Ala Leu Gln Pro
370 375 380

Ala Pro Glu Ala Gly Thr Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu
385 390 395 400

Thr Asn Ala Asn Pro Ser Leu
405

<210> 17
<211> 429
<212> ADN
<213> *Gloeophyllum trabeum* NP001144

<220>
<221> CDS
<222> (1)..(429)

<400> 17

```
gca tcg tct cca gca gct gtt cgt gga ctc gct aca aac gta gct aac        48
Ala Ser Ser Pro Ala Ala Val Arg Gly Leu Ala Thr Asn Val Ala Asn
1               5                  10                  15

ttc aat gcc tgg agc atc gac act tgc ccc tcc tat aca tcg ggt aac        96
Phe Asn Ala Trp Ser Ile Asp Thr Cys Pro Ser Tyr Thr Ser Gly Asn
            20                  25                  30

gat gtc tgt gat gag aag agc tac atc aat gcc ttt gca ccg gag ctc       144
Asp Val Cys Asp Glu Lys Ser Tyr Ile Asn Ala Phe Ala Pro Glu Leu
        35                  40                  45

tct agt gct gga ttt gat gcc cac ttt att acc gat acg ggt cgc aat       192
Ser Ser Ala Gly Phe Asp Ala His Phe Ile Thr Asp Thr Gly Arg Asn
    50                  55                  60

gga aag cag cct act gga cag agc gcg tgg ggt gac tgg tgc aat gtc       240
Gly Lys Gln Pro Thr Gly Gln Ser Ala Trp Gly Asp Trp Cys Asn Val
65                  70                  75                  80

aag gat act ggc ttc ggt gct cag ccg acg acc gat act gga gac gag       288
Lys Asp Thr Gly Phe Gly Ala Gln Pro Thr Thr Asp Thr Gly Asp Glu
                85                  90                  95

ctg gct gat gcc ttt gtc tgg gtc aag cct ggc gga gag agt gat ggg       336
Leu Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly
            100                 105                 110

acg tcg gac act agc tct tct cgc tac gat gcg cat tgc gga tat agt       384
Thr Ser Asp Thr Ser Ser Ser Arg Tyr Asp Ala His Cys Gly Tyr Ser
        115                 120                 125

gat gct ttg cag cct gcc ccg gag gct ggt act tgg ttc caa ggc           429
Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly Thr Trp Phe Gln Gly
    130                 135                 140
```

<210> 18
<211> 143
<212> PRT
<213> *Gloeophyllum trabeum* NP001144

<400> 18

```
Ala Ser Ser Pro Ala Ala Val Arg Gly Leu Ala Thr Asn Val Ala Asn
1               5                  10                  15

Phe Asn Ala Trp Ser Ile Asp Thr Cys Pro Ser Tyr Thr Ser Gly Asn
            20                  25                  30

Asp Val Cys Asp Glu Lys Ser Tyr Ile Asn Ala Phe Ala Pro Glu Leu
        35                  40                  45

Ser Ser Ala Gly Phe Asp Ala His Phe Ile Thr Asp Thr Gly Arg Asn
    50                  55                  60

Gly Lys Gln Pro Thr Gly Gln Ser Ala Trp Gly Asp Trp Cys Asn Val
65                  70                  75                  80

Lys Asp Thr Gly Phe Gly Ala Gln Pro Thr Thr Asp Thr Gly Asp Glu
                85                  90                  95

Leu Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly
            100                 105                 110

Thr Ser Asp Thr Ser Ser Ser Arg Tyr Asp Ala His Cys Gly Tyr Ser
        115                 120                 125

Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly Thr Trp Phe Gln Gly
    130                 135                 140
```

<210> 19
<211> 213
<212> ADN
<213> *Meripilus giganteus* ND001631

<220>
<221> característica misc
<222> (1)..(213)
<223> y es cor t

<400> 19

```
aaactcggag tacgtctgga acaagtgccg gcctcnggcr gctgggtacc gtcttcgtcc   60
gacagacagc asgtgctgtc gtaccggtrg cgacgagctg ttcgaggtay cgtcggactc  120
ryctccgggm ttcacccaga tkatcacgtc tatgasyggg ttgcccgtgt tcgtcctcat  180
ggcgcgtgcc gaagccgttg cccttgatgt tgc                               213
```

<210> 20
<211> 71
<212> PRT
<213> Enzima *Meripilus giganteus* ND001631

<220>
<221> característica_misc
<222> (1)..(71)
<223> x es cualquier aminoácido

<400> 20

```
Ala Thr Ser Arg Ala Thr Ala Ser Ala Arg Ala Met Arg Thr Asn Thr
1               5                  10                  15

Gly Asn Pro Xaa Ile Asp Val Ile Ile Trp Val Xaa Pro Gly Xaa Glu
            20                  25                  30

Ser Asp Xaa Thr Ser Asn Ser Ser Ser Xaa Pro Val Arg Gln His Xaa
        35                  40                  45

Leu Ser Val Gly Arg Arg Arg Tyr Pro Ala Ala Xaa Gly Arg His Leu
    50                  55                  60

Phe Gln Thr Tyr Ser Glu Phe
65                  70
```

<210> 21
<211> 782
<212> ADN
<213> *Trichophaea saccata* NP000960

<220>
<221> CDS
<222> (43)..(702)

<400> 21

```
ggcacgaggg cagatcgatc gactcgagga ccacatcgca tc atg aag aac ttc          54
                                                 Met Lys Asn Phe
                                                 1

ctt ctg gcg tcc gcg ctg atc gcg gtt gcc gca gct cag cag agt gct          102
Leu Leu Ala Ser Ala Leu Ile Ala Val Ala Ala Ala Gln Gln Ser Ala
5                   10                  15                  20

tgg gga cag tgc ggt gga att ggc tgg act ggc gcg acg act tgt atc          150
Trp Gly Gln Cys Gly Gly Ile Gly Trp Thr Gly Ala Thr Thr Cys Ile
                25                  30                  35

tct ggc tac acg tgc tca aag atc aac gac tac tat tcc cag tgc att          198
Ser Gly Tyr Thr Cys Ser Lys Ile Asn Asp Tyr Tyr Ser Gln Cys Ile
            40                  45                  50

ccg ggt acg gct tca acc acc act caa ggc ggc ggc aat ggc gga gga          246
Pro Gly Thr Ala Ser Thr Thr Thr Gln Gly Gly Gly Asn Gly Gly Gly
        55                  60                  65

aac ggc ggt aca acg act act ccc act acc act cca gcg gcc agt aac          294
Asn Gly Gly Thr Thr Thr Thr Pro Thr Thr Thr Pro Ala Ala Ser Asn
    70                  75                  80

acc aac aac ccg ttc tcc ggc aag acc caa tgg gcg aac cct tac tac          342
Thr Asn Asn Pro Phe Ser Gly Lys Thr Gln Trp Ala Asn Pro Tyr Tyr
85                  90                  95                  100

gct tcc gag gtc tcg agc atc gcc atc ccg tcc ctc gtt gcc gcc gga          390
Ala Ser Glu Val Ser Ser Ile Ala Ile Pro Ser Leu Val Ala Ala Gly
                105                 110                 115

aac acc cac tac atc gtc gac caa ggc cgc agc ggc aag cag ccg acc          438
Asn Thr His Tyr Ile Val Asp Gln Gly Arg Ser Gly Lys Gln Pro Thr
            120                 125                 130

ggc cag ctc cag cag ggc gat tgg tgc aac gcc ctg gga acc ggc ttt          486
Gly Gln Leu Gln Gln Gly Asp Trp Cys Asn Ala Leu Gly Thr Gly Phe
        135                 140                 145

gga att cgt cct gat aca acc ccg gat gat ccc aac ctt gat gct ttc          534
Gly Ile Arg Pro Asp Thr Thr Pro Asp Asp Pro Asn Leu Asp Ala Phe
    150                 155                 160

gtg tgg gtt aag ccg ggt ggt gaa tcg gat ggt acc agc aat act tcc          582
Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Thr Ser Asn Thr Ser


 165                 170                 175                 180

 tcg acc cgc tat gat tat cat tgt gga cag agc gat gcg cta caa ccg          630
 Ser Thr Arg Tyr Asp Tyr His Cys Gly Gln Ser Asp Ala Leu Gln Pro
                 185                 190                 195

 gcc ccg gag gcg gga acg tgg ttc cag gcg tat ttt gtg cag ttg ctg          678
 Ala Pro Glu Ala Gly Thr Trp Phe Gln Ala Tyr Phe Val Gln Leu Leu
             200                 205                 210

 cag aat gct aat cct agc ttc acg taagcttggg agcgtggggg ttggaagatg          732
 Gln Asn Ala Asn Pro Ser Phe Thr
         215                 220

 tgtattgtat gtgtagatag agaaaaactg ttggcctatt caggactaag                    782
```

<210> 22

<211> 220

<212> PRT

<213> *Trichophaea saccata* NP000960

<400> 22

```
Met Lys Asn Phe Leu Leu Ala Ser Ala Leu Ile Ala Val Ala Ala Ala
1               5                  10                  15

Gln Gln Ser Ala Trp Gly Gln Cys Gly Gly Ile Gly Trp Thr Gly Ala
            20                  25                  30

Thr Thr Cys Ile Ser Gly Tyr Thr Cys Ser Lys Ile Asn Asp Tyr Tyr
        35                  40                  45

Ser Gln Cys Ile Pro Gly Thr Ala Ser Thr Thr Thr Gln Gly Gly Gly
    50                  55                  60

Asn Gly Gly Gly Asn Gly Gly Thr Thr Thr Thr Pro Thr Thr Thr Pro
65                  70                  75                  80

Ala Ala Ser Asn Thr Asn Asn Pro Phe Ser Gly Lys Thr Gln Trp Ala
                85                  90                  95

Asn Pro Tyr Tyr Ala Ser Glu Val Ser Ser Ile Ala Ile Pro Ser Leu
            100                 105                 110

Val Ala Ala Gly Asn Thr His Tyr Ile Val Asp Gln Gly Arg Ser Gly
        115                 120                 125

Lys Gln Pro Thr Gly Gln Leu Gln Gln Gly Asp Trp Cys Asn Ala Leu
    130                 135                 140

Gly Thr Gly Phe Gly Ile Arg Pro Asp Thr Thr Pro Asp Asp Pro Asn
145                 150                 155                 160

Leu Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Thr
                165                 170                 175

Ser Asn Thr Ser Ser Thr Arg Tyr Asp Tyr His Cys Gly Gln Ser Asp
            180                 185                 190

Ala Leu Gln Pro Ala Pro Glu Ala Gly Thr Trp Phe Gln Ala Tyr Phe
        195                 200                 205

Val Gln Leu Leu Gln Asn Ala Asn Pro Ser Phe Thr
    210                 215                 220
```

<210> 23

<211> 1587

<212> ADN

<213> *Stibella annulata* NP001040

<220>

<221> CDS

<222> (21)..(1394)

<400> 23

```
ggcacgaggc gcgcatcaca atg gcc ggt cga ttc ttc ctc tct gct gcc ttc      53
                       Met Ala Gly Arg Phe Phe Leu Ser Ala Ala Phe
                       1               5                  10

ctg gct agc gcg gct ttg gcc gtc cct ctc gag gag agg cag aac tgc      101
Leu Ala Ser Ala Ala Leu Ala Val Pro Leu Glu Glu Arg Gln Asn Cys
               15                  20                  25

tcc ccg cag tgg gcc cag tgc ggt gga aat gga tgg agc ggt ccg acg      149
Ser Pro Gln Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Pro Thr
           30                  35                  40

tgc tgc gcc tcc ggc agc aac tgc cag gtc acc aac gag tgg tac tct      197
Cys Cys Ala Ser Gly Ser Asn Cys Gln Val Thr Asn Glu Trp Tyr Ser
       45                  50                  55

cag tgt gtt ccg ggc gcg gcg cct ccc cct ccc ccc gtc acc acg acg      245
Gln Cys Val Pro Gly Ala Ala Pro Pro Pro Pro Pro Val Thr Thr Thr
60                     65                  70                  75

cgg tcg acc acc acg ccc ccg acg acg acg acc agg acc acc gct gat      293
Arg Ser Thr Thr Thr Pro Pro Thr Thr Thr Thr Arg Thr Thr Ala Asp
                   80                  85                  90

gcc cct cct ccc acc ggc ggc gct act tac acc ggc aac ccc ttc ctc      341
Ala Pro Pro Pro Thr Gly Gly Ala Thr Tyr Thr Gly Asn Pro Phe Leu
               95                  100                 105

ggt gtc aac cag tgg gcc aac aac ttc tac cgg tct gag atc atg aac      389
Gly Val Asn Gln Trp Ala Asn Asn Phe Tyr Arg Ser Glu Ile Met Asn
           110                 115                 120

atc gcc gtc ccg tcc ctg tcc ggt gcc atg gct acc gcc gcc gcc aag      437
Ile Ala Val Pro Ser Leu Ser Gly Ala Met Ala Thr Ala Ala Ala Lys
       125                 130                 135

gtc gcc gat gtg ccc acc ttc cag tgg att gac aag atg gac aag ctc      485
Val Ala Asp Val Pro Thr Phe Gln Trp Ile Asp Lys Met Asp Lys Leu
140                    145                 150                 155

ccc ttg atc gat gag gct ctc gcc gac gtc cgc gct gcc aac gcc cgt      533
Pro Leu Ile Asp Glu Ala Leu Ala Asp Val Arg Ala Ala Asn Ala Arg
                   160                 165                 170

ggc ggc aac tac gct tcc atc ctg gtc gtc tac aac ctg ccc gac cgt      581
```

```
Gly Gly Asn Tyr Ala Ser Ile Leu Val Val Tyr Asn Leu Pro Asp Arg
                175                 180                 185

gac tgc gcc gcc gcc gcc tcg aac ggc gag ttc gcc atc gcc gac ggc      629
Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe Ala Ile Ala Asp Gly
        190                 195                 200

ggt gtt gct aag tac aag aac tac att gac gag att cgc aag ctc gtc      677
Gly Val Ala Lys Tyr Lys Asn Tyr Ile Asp Glu Ile Arg Lys Leu Val
    205                 210                 215

atc aag tac aac gac ctc cgt atc atc ctg gtc atc gag ccc gac tcc      725
Ile Lys Tyr Asn Asp Leu Arg Ile Ile Leu Val Ile Glu Pro Asp Ser
220                 225                 230                 235

ctc gcc aac atg gtg acc aac atg aac gtc gcc aag tgc cag aac gcc      773
Leu Ala Asn Met Val Thr Asn Met Asn Val Ala Lys Cys Gln Asn Ala
                240                 245                 250

gcc tcg gcc tac cgg gag tgc acc aac tat gcc ctg acg aac ctc gac      821
Ala Ser Ala Tyr Arg Glu Cys Thr Asn Tyr Ala Leu Thr Asn Leu Asp
            255                 260                 265

ctg ccc aac gtc gcc cag tac atg gat gcc gga cat gct ggc tgg ctc      869
Leu Pro Asn Val Ala Gln Tyr Met Asp Ala Gly His Ala Gly Trp Leu
        270                 275                 280

ggc tgg ccc gcc aac atc acc ccc gcc gcc cag ctc ttc gcc gag gtc      917
Gly Trp Pro Ala Asn Ile Thr Pro Ala Ala Gln Leu Phe Ala Glu Val
    285                 290                 295

tac aag cag gcc ggc agc ccc aag tcg gtc cgt ggt ctg gcc atc aac      965
Tyr Lys Gln Ala Gly Ser Pro Lys Ser Val Arg Gly Leu Ala Ile Asn
300                 305                 310                 315

gtc tcc aac tac aac gcg tgg agc gtt tcg tcc cct cct ccc tac acc     1013
Val Ser Asn Tyr Asn Ala Trp Ser Val Ser Ser Pro Pro Pro Tyr Thr
                320                 325                 330

tct ccc aac ccc aac tac gac gag cgc cac ttc gtt gag gcc ttt gcg     1061
Ser Pro Asn Pro Asn Tyr Asp Glu Arg His Phe Val Glu Ala Phe Ala
            335                 340                 345

ccc ctc ctg cgc cag aac ggc tgg gat gcc aag ttc atc gtc gac cag     1109
Pro Leu Leu Arg Gln Asn Gly Trp Asp Ala Lys Phe Ile Val Asp Gln
        350                 355                 360

ggc cgc tcc ggc agg cag ccc acc ggc cag cag gag tgg gga cac tgg     1157
Gly Arg Ser Gly Arg Gln Pro Thr Gly Gln Gln Glu Trp Gly His Trp
    365                 370                 375

tgc aac gcc atc ggc act ggc ttc ggc cag cgc ccg acg tcc aac acc     1205
Cys Asn Ala Ile Gly Thr Gly Phe Gly Gln Arg Pro Thr Ser Asn Thr
380                 385                 390                 395

ggc cac gcc gat gtt gac gct ttc gtc tgg atc aag ccg ggc ggt gag     1253
Gly His Ala Asp Val Asp Ala Phe Val Trp Ile Lys Pro Gly Gly Glu
                400                 405                 410

tgc gac ggc acc agc gac acc tcg gcc gcc cgc tac gac cac ttc tgt     1301
Cys Asp Gly Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp His Phe Cys
            415                 420                 425

ggc aac cct gat gcc ctc aag ccg gcc ccc gaa gcc gga gag tgg ttc     1349
Gly Asn Pro Asp Ala Leu Lys Pro Ala Pro Glu Ala Gly Glu Trp Phe
        430                 435                 440

cag gcc tac ttc gag cag ctt ctg cgc aac gcc aac ccc gcc ttc         1394
```

Gln Ala Tyr Phe Glu Gln Leu Leu Arg Asn Ala Asn Pro Ala Phe
445 450 455

taagtgctgg atgagctttt ctgagagggt acttccgcgg tcttgggttt cactcttctc 1454

agcctttcag ggcagcagtt ttggtttctt ggggtaggac ctccgggttt atgtagacgg 1514

agttaggaag ccaaacctac tatgaatgta gtattcaaga agataatgac ttgaaaaaaa 1574

aaaaaaaaaa aaa 1587

<210> 24

<211> 458

<212> PRT

<213> *Stibella annulata* NP001040

<400> 24

Met Ala Gly Arg Phe Phe Leu Ser Ala Ala Phe Leu Ala Ser Ala Ala
1 5 10 15

Leu Ala Val Pro Leu Glu Glu Arg Gln Asn Cys Ser Pro Gln Trp Ala
20 25 30

Gln Cys Gly Gly Asn Gly Trp Ser Gly Pro Thr Cys Cys Ala Ser Gly
35 40 45

Ser Asn Cys Gln Val Thr Asn Glu Trp Tyr Ser Gln Cys Val Pro Gly
50 55 60

Ala Ala Pro Pro Pro Pro Pro Val Thr Thr Thr Arg Ser Thr Thr Thr
65 70 75 80

Pro Pro Thr Thr Thr Thr Arg Thr Thr Ala Asp Ala Pro Pro Pro Thr
85 90 95

Gly Gly Ala Thr Tyr Thr Gly Asn Pro Phe Leu Gly Val Asn Gln Trp
100 105 110

Ala Asn Asn Phe Tyr Arg Ser Glu Ile Met Asn Ile Ala Val Pro Ser
115 120 125

Leu Ser Gly Ala Met Ala Thr Ala Ala Ala Lys Val Ala Asp Val Pro
130 135 140

Thr Phe Gln Trp Ile Asp Lys Met Asp Lys Leu Pro Leu Ile Asp Glu
145 150 155 160

Ala Leu Ala Asp Val Arg Ala Ala Asn Ala Arg Gly Gly Asn Tyr Ala
165 170 175

Ser Ile Leu Val Val Tyr Asn Leu Pro Asp Arg Asp Cys Ala Ala Ala
180 185 190

Ala Ser Asn Gly Glu Phe Ala Ile Ala Asp Gly Gly Val Ala Lys Tyr

195 200 205

Lys Asn Tyr Ile Asp Glu Ile Arg Lys Leu Val Ile Lys Tyr Asn Asp
210 215 220

Leu Arg Ile Ile Leu Val Ile Glu Pro Asp Ser Leu Ala Asn Met Val
225 230 235 240

Thr Asn Met Asn Val Ala Lys Cys Gln Asn Ala Ala Ser Ala Tyr Arg
245 250 255

Glu Cys Thr Asn Tyr Ala Leu Thr Asn Leu Asp Leu Pro Asn Val Ala
260 265 270

Gln Tyr Met Asp Ala Gly His Ala Gly Trp Leu Gly Trp Pro Ala Asn
275 280 285

Ile Thr Pro Ala Ala Gln Leu Phe Ala Glu Val Tyr Lys Gln Ala Gly
290 295 300

Ser Pro Lys Ser Val Arg Gly Leu Ala Ile Asn Val Ser Asn Tyr Asn
305 310 315 320

Ala Trp Ser Val Ser Ser Pro Pro Pro Tyr Thr Ser Pro Asn Pro Asn
325 330 335

Tyr Asp Glu Arg His Phe Val Glu Ala Phe Ala Pro Leu Leu Arg Gln
340 345 350

Asn Gly Trp Asp Ala Lys Phe Ile Val Asp Gln Gly Arg Ser Gly Arg
355 360 365

Gln Pro Thr Gly Gln Gln Glu Trp Gly His Trp Cys Asn Ala Ile Gly
370 375 380

Thr Gly Phe Gly Gln Arg Pro Thr Ser Asn Thr Gly His Ala Asp Val
385 390 395 400

Asp Ala Phe Val Trp Ile Lys Pro Gly Gly Glu Cys Asp Gly Thr Ser
405 410 415

Asp Thr Ser Ala Ala Arg Tyr Asp His Phe Cys Gly Asn Pro Asp Ala
420 425 430

Leu Lys Pro Ala Pro Glu Ala Gly Glu Trp Phe Gln Ala Tyr Phe Glu
435 440 445

Gln Leu Leu Arg Asn Ala Asn Pro Ala Phe
450 455

<210> 25

<211> 1559

<212> ADN

<213> *Malbrancheae cinnamonea* NP001045

<220>

<221> CDS

<222> (41)..(1210)

<400> 25

```
cgaggcgtcc atcccacgcc gcacgcgtca gccagtcatt atg aga gac tct ttg       55
                                                Met Arg Asp Ser Leu
                                                1               5

ttc act ttg cta tct ctt gcc ttg ggc tcg gcc tct gcc agc cct ttc      103
Phe Thr Leu Leu Ser Leu Ala Leu Gly Ser Ala Ser Ala Ser Pro Phe
                10                  15                  20

ctt ctt cca agg caa gcg aac tcc tcc aac ccg ttt gct gga cac acg      151
Leu Leu Pro Arg Gln Ala Asn Ser Ser Asn Pro Phe Ala Gly His Thr
            25                  30                  35

atc tat cca aac ccg tac tac tcc aac gag att gac gag ttt gcc att      199
Ile Tyr Pro Asn Pro Tyr Tyr Ser Asn Glu Ile Asp Glu Phe Ala Ile
        40                  45                  50

ccc gcg ctg caa gag acc gat cct gca ctt gtg gag aag gcc gct tta      247
Pro Ala Leu Gln Glu Thr Asp Pro Ala Leu Val Glu Lys Ala Ala Leu
    55                  60                  65

gta aaa gaa gtt gga act ttc ttc tgg att gat gtc gtc gcc aag gtc      295
Val Lys Glu Val Gly Thr Phe Phe Trp Ile Asp Val Val Ala Lys Val
70                  75                  80                  85

cca gat atc ggc cct tac ctg cag ggg atc caa gaa gca aac gcc gca      343
Pro Asp Ile Gly Pro Tyr Leu Gln Gly Ile Gln Glu Ala Asn Ala Ala
                90                  95                  100

ggc cag aat ccg ccg tac atc ggc gcg att gtt gtc tat gac ctc ccc      391
Gly Gln Asn Pro Pro Tyr Ile Gly Ala Ile Val Val Tyr Asp Leu Pro
            105                 110                 115

aac cgt gac tgc gct gcc gca gct tcc aac gga gag ttc agc ctc gag      439
Asn Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe Ser Leu Glu
        120                 125                 130

gat ggc ggc gag gag aag tac cgc ggt tat atc gac ggt atc cgg gag      487
Asp Gly Gly Glu Glu Lys Tyr Arg Gly Tyr Ile Asp Gly Ile Arg Glu
    135                 140                 145

caa att gag aaa tac cca gac gtc cgt gtc gcg ctg gtt atc gag ccc      535
Gln Ile Glu Lys Tyr Pro Asp Val Arg Val Ala Leu Val Ile Glu Pro
150                 155                 160                 165

gat tcg ctc gcg aac atg gtc acc aat ctc aat gtc ccc aag tgc gct      583
Asp Ser Leu Ala Asn Met Val Thr Asn Leu Asn Val Pro Lys Cys Ala
                170                 175                 180

gag tcg gag cag gct tat cga gat ggc gtc gcg tat gca ctg aaa cag      631
Glu Ser Glu Gln Ala Tyr Arg Asp Gly Val Ala Tyr Ala Leu Lys Gln
            185                 190                 195

cta gac ctc ccc aac gtc tgg aca tat atc gat gct ggt cat tca ggt      679
Leu Asp Leu Pro Asn Val Trp Thr Tyr Ile Asp Ala Gly His Ser Gly
        200                 205                 210

tgg ctt ggc tgg ccc gcc aac atc gag cct gcc gca gaa att ttt gtt      727
Trp Leu Gly Trp Pro Ala Asn Ile Glu Pro Ala Ala Glu Ile Phe Val
```

```
    215                 220                 225

gag gtc tgg aat gca gct ggc agg cca aag tcc act cga ggg ttt gct      775
Glu Val Trp Asn Ala Ala Gly Arg Pro Lys Ser Thr Arg Gly Phe Ala
230                 235                 240                 245

acg aac gtt tcc aac tac aac ggt tat tcc ctc agc acc gct cct ccc      823
Thr Asn Val Ser Asn Tyr Asn Gly Tyr Ser Leu Ser Thr Ala Pro Pro
                250                 255                 260

tac act gag ccc aac ccc aat ttc gac gaa gtg cgt tat atc aat gca      871
Tyr Thr Glu Pro Asn Pro Asn Phe Asp Glu Val Arg Tyr Ile Asn Ala
            265                 270                 275

ttc cgc cca ctc ctc gag gca cgg ggt ttc cca gca tac ttc atc gtc      919
Phe Arg Pro Leu Leu Glu Ala Arg Gly Phe Pro Ala Tyr Phe Ile Val
        280                 285                 290

gac caa ggc cgc agc ggt gtc cag ccc act gcg cag att gag caa gga      967
Asp Gln Gly Arg Ser Gly Val Gln Pro Thr Ala Gln Ile Glu Gln Gly
    295                 300                 305

cac tgg tgc aat gtg atc gac acc ggt ttt gga act cgc ccc act act     1015
His Trp Cys Asn Val Ile Asp Thr Gly Phe Gly Thr Arg Pro Thr Thr
310                 315                 320                 325

gac act ggt aat gag tac gtt gac tcg atc gtg tgg gtg aag cct ggc     1063
Asp Thr Gly Asn Glu Tyr Val Asp Ser Ile Val Trp Val Lys Pro Gly
                330                 335                 340

ggc gaa tcg gac gga acc agc gat acc tct gct gag aga tat gac tac     1111
Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Glu Arg Tyr Asp Tyr
            345                 350                 355

cac tgc gga ctt gag gat gca ttg aag cca gct cct gaa gcg gga cag     1159
His Cys Gly Leu Glu Asp Ala Leu Lys Pro Ala Pro Glu Ala Gly Gln
        360                 365                 370

tgg ttc cag gcc tac ttc gag caa ctg ctc aga aat gcc aac ccc cca     1207
Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Arg Asn Ala Asn Pro Pro
    375                 380                 385

ttc taaatcagat gaaggacgga cccaattgat gacggcctgt cttcgtgatc          1260
Phe
390

cgacgaaagc aatgtcaggg tgaaaatgac cgagagattg gagagtcatg aggataggta   1320

gtcaatgatt tcacccgagt ttccacgttt tacccttctt gtacatagtt tggagtcgcc   1380

tgttggtttc agtagtacat cttatccgac agagtctatc gtttgattac cccagtcaaa   1440

agcgttattg caatcttttc ctagggattt attgtttgct gcggatgtcg tggctatggg   1500

cagctgactg aattaaactg gaactcttgg tatccaaaaa aaaaaaaaaa aaaaaaaaa    1559
```

<210> 26

<211> 390

<212> PRT

<213> *Malbrancheae cinnamonea* NP001045

<400> 26

```
Met Arg Asp Ser Leu Phe Thr Leu Leu Ser Leu Ala Leu Gly Ser Ala
1               5                   10                  15
```

Ser Ala Ser Pro Phe Leu Leu Pro Arg Gln Ala Asn Ser Ser Asn Pro
20 25 30

Phe Ala Gly His Thr Ile Tyr Pro Asn Pro Tyr Tyr Ser Asn Glu Ile
35 40 45

Asp Glu Phe Ala Ile Pro Ala Leu Gln Glu Thr Asp Pro Ala Leu Val
50 55 60

Glu Lys Ala Ala Leu Val Lys Glu Val Gly Thr Phe Phe Trp Ile Asp
65 70 75 80

Val Val Ala Lys Val Pro Asp Ile Gly Pro Tyr Leu Gln Gly Ile Gln
85 90 95

Glu Ala Asn Ala Ala Gly Gln Asn Pro Pro Tyr Ile Gly Ala Ile Val
100 105 110

Val Tyr Asp Leu Pro Asn Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly
115 120 125

Glu Phe Ser Leu Glu Asp Gly Gly Glu Glu Lys Tyr Arg Gly Tyr Ile
130 135 140

Asp Gly Ile Arg Glu Gln Ile Glu Lys Tyr Pro Asp Val Arg Val Ala
145 150 155 160

Leu Val Ile Glu Pro Asp Ser Leu Ala Asn Met Val Thr Asn Leu Asn
165 170 175

Val Pro Lys Cys Ala Glu Ser Glu Gln Ala Tyr Arg Asp Gly Val Ala
180 185 190

Tyr Ala Leu Lys Gln Leu Asp Leu Pro Asn Val Trp Thr Tyr Ile Asp
195 200 205

Ala Gly His Ser Gly Trp Leu Gly Trp Pro Ala Asn Ile Glu Pro Ala
210 215 220

Ala Glu Ile Phe Val Glu Val Trp Asn Ala Ala Gly Arg Pro Lys Ser
225 230 235 240

Thr Arg Gly Phe Ala Thr Asn Val Ser Asn Tyr Asn Gly Tyr Ser Leu
245 250 255

Ser Thr Ala Pro Pro Tyr Thr Glu Pro Asn Pro Asn Phe Asp Glu Val
260 265 270

Arg Tyr Ile Asn Ala Phe Arg Pro Leu Leu Glu Ala Arg Gly Phe Pro
275 280 285

Ala Tyr Phe Ile Val Asp Gln Gly Arg Ser Gly Val Gln Pro Thr Ala
290 295 300

Gln Ile Glu Gln Gly His Trp Cys Asn Val Ile Asp Thr Gly Phe Gly
305 310 315 320

Thr Arg Pro Thr Thr Asp Thr Gly Asn Glu Tyr Val Asp Ser Ile Val
325 330 335

Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala
340 345 350

Glu Arg Tyr Asp Tyr His Cys Gly Leu Glu Asp Ala Leu Lys Pro Ala
355 360 365

Pro Glu Ala Gly Gln Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Arg
370 375 380

Asn Ala Asn Pro Pro Phe
385 390

<210> 27
<211> 17
<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1)..(17)
<223> r es a o g
y es t o c
n es a, g, t, o c

<400> 27

tggggncart gyggngg 17

<210> 28
<211> 17
<212> ADN
<213> artificial

<220>
<223> cebador

<220>

<221> característica_misc
<222> (1)..(17)
<223> y es t o c
n es a, g, t, o c

<400> 28

tggytnggnt ggccngc 17

<210> 29
<211> 17
<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1)..(17)
<223> n es a, g, t; o g
r es a o g

<400> 29

gcnggccanc cnarcca 17

<210> 30
<211> 17
<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1) .. (17)
<223> r es a o g
n es a, g, t, o c

<400> 30

ttrcaccart cncccca 17

<210> 31
<211> 17
<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1)..(17)
<223> r es a o g
y es t o c
n es a, g, t o c

<400> 31

ggyttnaccc anacraa 17

<210> 32
<211> 17
<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1)..(17)
<223> r es a o g
y es t o c
n es a, g, t o c

<400> 32

aartangcyt graacca 17

<210> 33
<211> 25
<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1)..(25)

<400> 33

gggtcatgag agactctttg ttcac 25

<210> 34
<211> 34

<212> ADN
<213> artificial

<220>
<223> cebador

<220>
<221> característica_misc
<222> (1)..(34)

<400> 34

gggttaatta attagaatgg ggggttggca tttc 34

<210> 35
<211> 31
<212> ADN
<213> artificial

<220>
<223> cebador

<400> 35

actggattta ccatggccgg tcgattcttc c 31

<210> 36
<211> 31
<212> ADN
<213> artificial

<220>
<223> cebador

<400> 36

agtcacctct agttattaga aggcggggtt g 31