



19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA

11 Número de publicación: **2 358 243**

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

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Número de solicitud europea: **06733650 .3**

96 Fecha de presentación : **06.01.2006**

97 Número de publicación de la solicitud: **1836299**

97 Fecha de publicación de la solicitud: **26.09.2007**

54

Título: **Polipéptidos con actividad celobiohidrolasa y polinucleótidos que los codifican.**

30

Prioridad: **06.01.2005 US 642274 P**

45

Fecha de publicación de la mención BOPI:
06.05.2011

45

Fecha de la publicación del folleto de la patente:
06.05.2011

73

Titular/es: **NOVOZYMES, Inc.**
1445 Drew Avenue
Davis, California 95616, US

72

Inventor/es: **Brown, Kimberly;**
Harris, Paul;
López de León, Alfredo y
Merino, Sandra

74

Agente: **Tomás Gil, Tesifonte Enrique**

ES 2 358 243 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Polipéptidos con actividad celobiohidrolasa y polinucleótidos que los codifican.

5 Declaración de derechos de invenciones realizadas con la subvención del Gobierno Federal en materia de Investigación y Desarrollo.

Esta invención fue realizada con el soporte del Gobierno bajo el subcontrato NREL n°. ZCO-30017-02, contrato principal DE-AC36-98GO10337, otorgado por el Departamento de Energía. El Gobierno tiene ciertos derechos sobre esta invención.

Antecedentes de la invención

Campo de la invención

15 La presente invención se refiere a polipéptidos aislados que tienen actividad de celobiohidrolasa y polinucleótidos aislados que codifican los polipéptidos. La invención también se refiere a constructos de ácidos nucleicos, vectores, y células huéspedes comprendiendo los polinucleótidos al igual que a métodos para producir y usar los polipéptidos.

Descripción de las técnicas relacionadas

20 La celulosa es un polímero de la simple glucosa del azúcar unida de manera covalente por enlaces beta-1,4. Muchos microorganismos producen enzimas que hidrolizan glucanos enlazados en beta. Estas enzimas incluyen endoglucanasas, celobiohidrolasas, y beta-glucosidasas. Las endoglucanasas digieren el polímero de celulosa en lugares al azar, abriéndose para el ataque por celobiohidrolasas. Las celobiohidrolasas liberan consecutivamente moléculas de celobiosa de las extremidades del polímero de celulosa. La celobiohidrolasa I es una actividad de 1,4-D-glucano celobiohidrolasa (E.C. 3,2,1,91) que cataliza la hidrólisis de enlaces 1,4-beta-D-glucosídicos en celulosa, celotetriosas, o cualquier glucosa enlazada en beta-1,4 conteniendo polímero, liberando celobiosa de las extremidades reductoras de la cadena. La celobiohidrolasa II es una actividad 1,4-D-glucano celobiohidrolasa (E.C. 3,2,1,91) que cataliza la hidrólisis de enlaces 1,4-beta-D-glucosídicos en celulosa, celotetriosas, o cualquier polímero conteniendo glucosa enlazada en beta 1,4, liberando celobiosa de las extremidades no reductoras de la cadena. La celobiosa es un dímero enlazado en beta-1,4 hidrosoluble de glucosa. Las beta-glucosidasas hidrolizan celobiosa a glucosa. La celobiosa es un dímero enlazado en beta-1,4 hidrosoluble de glucosa. Las beta-glucosidasas hidrolizan celobiosa a glucosa.

35 La conversión de materias primas celulósicas en etanol tiene las ventajas de la disponibilidad preparada de grandes cantidades de materia prima, la conveniencia de evitar la combustión o vertido de residuo de los materiales, y la limpieza del combustible de etanol. Madera, residuos agrícolas, brotes herbáceos, y residuos sólidos municipales han sido considerados como materias primas para la producción de etanol. Estos materiales principalmente consisten en celulosa, hemicelulosa, y lignina. Una vez la celulosa se convierte a glucosa, la glucosa es fácilmente fermentada por levadura en etanol.

WO 04/56981 divulga una celobiohidrolasa II de *Chaetomium thermophilum*.

45 Es un objeto de la presente invención proporcionar polipéptidos con actividad de celobiohidrolasa y polinucleótidos que codifican los polipéptidos.

Resumen de la invención

50 La presente invención se refiere a polipéptidos aislados que tienen actividad de celobiohidrolasa seleccionada del grupo que consiste en:

(a) un polipéptido comprendiendo una secuencia de aminoácidos que tiene al menos 80% de identidad con el polipéptido maduro de SEC ID n°: 2;

55 (b) un polipéptido que se codifica por una secuencia de nucleótidos que se hibrida bajo al menos condiciones de astringencia media-alta con (i) la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (ii) la secuencia de ADN genómico comprendiendo la secuencia codificante del polipéptido maduro de SEC ID n°: 1, o (iii) una cadena complementaria de (i) o (ii); y

60 (c) una variante comprendiendo una sustitución conservadora, deleción, y/o inserción de uno o más aminoácidos del polipéptido maduro de SEC ID n°: 2.

65 La presente invención también se refiere a polipéptidos aislados que codifican polinucleótidos que tienen actividad de celobiohidrolasa, seleccionados del grupo que consiste en:

(a) un polinucleótido que codifica un polipéptido comprendiendo una secuencia de aminoácidos que tiene al menos 80% de identidad con el polipéptido maduro de SEC ID n°: 2;

ES 2 358 243 T3

(b) un polinucleótido que tiene al menos 60% de identidad con la secuencia codificante del polipéptido maduro de SEC ID n°: 1 y

(c) un polinucleótido que se hibrida bajo al menos condiciones de astringencia media-alta con (i) la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (ii) la secuencia de ADN genómico comprendiendo la secuencia codificante del polipéptido maduro de SEC ID n°: 1, o (iii) una cadena complementaria de (i) o (ii).

En un aspecto preferido, el polipéptido maduro es los aminoácidos 18 a 481 de SEC ID n°: 2. En otro aspecto preferido, la secuencia codificante del polipéptido maduro es los nucleótidos 52 a 1443 de SEC ID n°: 1.

La presente invención también se refiere a constructos de ácidos nucleicos, vectores de expresión recombinantes, y células huéspedes recombinantes comprendiendo los polinucleótidos.

La presente invención también se refiere a métodos para producir tal polipéptido que tiene actividad de celobiohidrolasa comprendiendo: (a) cultivo de una célula huésped recombinante comprendiendo un constructo de ácidos nucleicos comprendiendo un polinucleótido que codifica el polipéptido bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

La presente invención también se refiere a métodos de uso de los polipéptidos que tienen actividad de celobiohidrolasa en detergentes y en la conversión de celulosa a glucosa.

La presente invención además se refiere a constructos de ácidos nucleicos comprendiendo un gen que codifica una proteína, donde el gen está enlazado operativamente a una secuencia de nucleótidos que codifica un péptido señal que comprende o que consiste en los aminoácidos 1 a 17 de SEC ID n°: 2, donde el gen es extranjero a la secuencia de nucleótidos.

Breve descripción de las figuras

Las Figuras 1A y 1B muestran la secuencia de ADNc y la secuencia de aminoácidos deducida de una celobiohidrolasa de *Thielavia terrestris* NRRL 8126 (Cel6A) (SEC ID N°s: 1 y 2, respectivamente).

La Figura 2 muestra un mapa de restricción de pAILo1.

La Figura 3 muestra un mapa de restricción de pBANE10.

La Figura 4 muestra un mapa de restricción de pAILo2.

La Figura 5 muestra un mapa de restricción de pAILo21.

La Figura 6 muestra la hidrólisis de PASC a glucosa por celobiohidrolasa Cel6A de *Thielavia terrestris* o *Humicola insolens*. β -glucosidasa de *Aspergillus oryzae* fue incluida en el ensayo para convertir celobiosa a glucosa.

La Figura 7 muestra un mapa de restricción de pCW076.

La Figura 8 muestra un mapa de restricción de pCW085.

Definiciones

Actividad de celobiohidrolasa: el término “actividad de celobiohidrolasa” se define aquí como una actividad de 1,4-D-glucano celobiohidrolasa (E.C. 3,2,1,91) que cataliza la hidrólisis de los enlaces 1,4-beta-D-glucosídicos en celulosa, celotetrosa, o cualquier glucosa enlazada en beta-1,4 conteniendo polímero, liberando celobiosa de las extremidades de la cadena. Para objetivos de la presente invención, la actividad de celobiohidrolasa se determina por la liberación de azúcar de reducción hidrosoluble de la celulosa como se mide por el método PHBAH de Lever *et al.*, 1972, Anal. Biochem. 47: 273-279. Una distinción entre el modo de ataque de la exoglucanasa a una celobiohidrolasa y el modo de ataque de la endoglucanasa se hace por una medición similar de la liberación de azúcar de reducción de celulosa sustituida tal como carboximetilcelulosa o hidroxietil celulosa (Ghose, 1987 Pure & Appl. Chem. 59: 257-268). Una celobiohidrolasa real tendrá una proporción muy alta de actividad en celulosa no sustituida contra celulosa sustituida (Bailey *et al.*, 1993, Biotechnol. Appl. Biochem. 17: 65-76).

Los polipéptidos de la presente invención tienen al menos 20%, preferiblemente al menos 40%, más preferiblemente al menos 50%, más preferiblemente al menos 60%, más preferiblemente al menos 70%, más preferiblemente al menos 80%, incluso más preferiblemente al menos 90%, de la forma más preferible al menos 95%, e incluso de la forma más preferible al menos 100% de la actividad de celobiohidrolasa del polipéptido maduro de SEC ID n°: 2.

Familia 6 de glucósido hidrolasa o Familia GH6: el término “Familia 6 de glucósido hidrolasa” o “Familia GH6” o “CeI6” se define aquí como un polipéptido que cae en la Familia 6 de glucósido hidrolasa según Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, Biochem. J. 280: 309-316, y Henrissat y Bairoch, 1996, Updating the sequence-based classification of glycosyl hydrolases, Biochem. J. 316: de 695-696.

ES 2 358 243 T3

Polipéptido aislado: el término “polipéptido aislado” como se usa aquí se refiere a un polipéptido que es al menos 20% puro, preferiblemente al menos 40% puro, más preferiblemente al menos 60% puro, incluso más preferiblemente al menos 80% puro, de la forma más preferible al menos 90% puro, e incluso de la forma más preferible al menos 95% puro, según está determinado por SDS-PAGE.

Polipéptido substancialmente puro: el término “polipéptido substancialmente puro” denota aquí una preparación de polipéptido que contiene como mucho 10%, preferiblemente como mucho 8%, más preferiblemente como mucho 6%, más preferiblemente como mucho 5%, más preferiblemente como mucho 4%, más preferiblemente como mucho 3%, incluso más preferiblemente como mucho 2%, de la forma más preferible como mucho 1%, e incluso de la forma más preferible como mucho 0,5% en peso de otro material polipeptídico con el que está asociado originalmente. Es, por lo tanto, preferido que el polipéptido substancialmente puro sea al menos 92% puro, preferiblemente al menos 94% puro, más preferiblemente al menos 95% puro, más preferiblemente al menos 96% puro, más preferiblemente al menos 96% puro, más preferiblemente al menos 97% puro, más preferiblemente al menos 98%, puro incluso más preferiblemente al menos 99%, de la forma más preferible al menos 99,5% puro, e incluso de la forma más preferible 100% puro por peso del material polipeptídico total presente en la preparación.

Los polipéptidos de la presente invención están preferiblemente en una forma substancialmente pura. En particular, se prefiere que los polipéptidos estén en “forma esencialmente pura”, es decir, que la preparación de polipéptido esté esencialmente libre de otro material polipeptídico con el cual está asociado originalmente. Esto puede ser realizado, por ejemplo, preparando el polipéptido mediante métodos recombinantes bien conocidos o por métodos de purificación tradicionales.

Aquí, el término “polipéptido esencialmente puro” es sinónimo de los términos “polipéptido aislado” y “polipéptido en forma aislada”.

Polipéptido maduro: el término “polipéptido maduro” es definido aquí como un polipéptido que tiene actividad de celobiohidrolasa que está en su forma final después de su traducción y cualquier modificación postraduccional, tal como tratamiento N-terminal, truncamiento C-terminal, glicosilación, etc.

Identidad: la relación entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos está descrita por el parámetro “identidad”.

Para objetivos de la presente invención, el grado de identidad entre dos secuencias de aminoácidos se determina por el método Clustal (Higgins, 1989, CABIOS 5: 151-153) usando el Software de LASERGENE™ MEGALIGN™ (DNASTAR, Inc., Madison, WI) con una tabla de peso de residuo PAM250 y los siguientes parámetros de alineación múltiples: penalización de espacio de 10 y penalización de longitud de espacio de 10. Parámetros de alineación de parejas son Ktuple=1, penalización de espacio=3, ventanas=5, y diagonales=5.

Para objetivos de la presente invención, el grado de identidad entre dos secuencias de nucleótidos se determina por el método de Wilbur-Lipman (Wilbur y Lipman, 1983, Proceedings of the National Academy of Science USA 80: 726-730) usando el Software de LASERGENE™ MEGALIGN™ (DNASTAR, Inc., Madison, WI) con una tabla de identidad y los siguientes parámetros de alineación múltiple: Penalización de espacio de 10 y penalización de longitud de espacio de 10. Parámetros de alineación de parejas son Ktuple=3, penalización de espacio=3, y ventanas=20.

Secuencia homóloga: el término “secuencia homóloga” es definido aquí como una proteína predicha que da un valor E (o puntuación de expectativa) inferior a 0,001 en una búsqueda tfasty (Pearson, W.R., 1999, en Bioinformatics Methods and Protocols, S. Misener and S. A. Krawetz, ed., págs. 185-219) con la celobiohidrolasa de *Thielavia terrestris* de la presente invención.

Fragmento de polipéptido: el término “fragmento polipeptídico” es definido aquí como un polipéptido que tiene uno o más aminoácidos delecionados del termino amino y/o carboxilo del polipéptido maduro de SEC ID n°: 2 o una secuencia homóloga de la misma; donde el fragmento tiene actividad de celobiohidrolasa. En un aspecto preferido, un fragmento contiene al menos 390 residuos de aminoácidos, más preferiblemente al menos 415 residuos de aminoácidos, y de la forma más preferible al menos 440 residuos de aminoácidos del polipéptido maduro de SEC ID n°: 2 o una secuencia homóloga de la misma.

Subsecuencia: el término “subsecuencia” es definido aquí como una secuencia de nucleótidos que tiene uno o más nucleótidos delecionados de la extremidad 5' y/o 3' de la secuencia codificante del polipéptido maduro de SEC ID n°: 1 o una secuencia homóloga de la misma; donde la subsecuencia codifica un fragmento de polipéptido que tiene actividad de celobiohidrolasa. En un aspecto preferido, una subsecuencia contiene al menos 1170 nucleótidos, más preferiblemente al menos 1245 nucleótidos, y de la forma más preferible al menos 1320 nucleótidos de la secuencia codificante del polipéptido maduro de SEC ID n°: 1 o una secuencia homóloga de la misma.

Variante alélica: el término “variante alélica” denota aquí 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 resultar en polimorfismo dentro de las poblaciones. Las mutaciones de gen pueden ser silenciosas (ningún cambio en el polipéptido codificado) o pueden codificar polipéptidos que tienen secuencias de aminoácidos alteradas. Una variante alélica de un polipéptido es un polipéptido codificado por una variante alélica de un gen.

ES 2 358 243 T3

Polinucleótido aislado: el término “polinucleótido aislado” como se usa aquí se refiere a un polinucleótido que es al menos 20% puro, preferiblemente al menos 40% puro, más preferiblemente al menos 60%, puro incluso más preferiblemente al menos 80% puro, de la forma más preferible al menos 90% puro, e incluso de la forma más preferible al menos 95% puro, como es determinado por electroforesis con agarosa.

5

Polinucleótido substancialmente puro: el término “polinucleótido sustancialmente puro” como se usa aquí se refiere a una preparación polinucleótida libre de otros nucleótidos indeseados o extraños y en una forma adecuada para el uso dentro de sistemas de producción de proteínas creadas genéticamente. Así, un polinucleótido substancialmente puro contiene como mucho 10%, preferiblemente como mucho 8%, más preferiblemente como mucho 6%, más preferiblemente como mucho 5%, más preferiblemente como mucho 4%, más preferiblemente como mucho 3%, incluso más preferiblemente como mucho 2%, de la forma más preferible como mucho 1%, e incluso de la forma más preferible como mucho 0.5% en peso de otro material polinucleótido con el cual está originalmente asociado. Un polinucleótido substancialmente puro puede, no obstante, incluir regiones no traducidas de origen natural 5' y 3', tales como promotores y terminadores. Se prefiere que el polinucleótido substancialmente puro sea al menos 90% puro, preferiblemente al menos 92% puro, más preferiblemente al menos 94% puro, más preferiblemente al menos 95% puro, más preferiblemente al menos 96% puro, más preferiblemente al menos 97% puro, incluso más preferiblemente al menos 98% puro, de la forma más preferible al menos 99%, e incluso de la forma más preferible al menos 99,5% puro en peso. Los polinucleótidos de la presente invención son preferiblemente en una forma substancialmente pura. En particular, se prefiere que los polinucleótidos descritos aquí estén en “forma esencialmente pura”, es decir, que la preparación polinucleótida esté esencialmente libre de otro material polinucleótido con el cual está originalmente asociado. Aquí, el término “polinucleótido sustancialmente puro” es sinónimo de los términos “polinucleótido aislado” y “polinucleótido en forma aislada”. Los polinucleótidos pueden ser de origen genómico, ADNc, ARN, semisintético, sintético, o cualquier combinación de los mismos.

Secuencia codificante de polipéptido maduro: el término “secuencia codificante de polipéptido maduro” es definido aquí como una secuencia de nucleótidos que codifica un polipéptido maduro que tiene actividad de celobiohidrolasa.

ADNc: el término “ADNc” es definido aquí como una molécula de ADN que se puede preparar por transcripción inversa de una molécula de ARNm madura, dividida, obtenida de una célula eucariótica. ADNc carece de secuencias de intrones que están normalmente presentes en el ADN genómico correspondiente. La transcripción de ARN inicial primaria es un precursor de ARNm que se procesa a través de una serie de pasos antes de aparecer como ARNm maduro dividido. Estos pasos incluyen la eliminación de secuencias de intrones por un proceso llamado empalme. ADNc derivado de ARNm carece, por lo tanto, de cualquier secuencia de intrones.

35

Construido de ácidos nucleicos: el término “construido de ácidos nucleicos” como se utiliza en este caso se refiere a una molécula de ácido nucleico, sea mono o bicatenario, que se aísla de un gen de origen natural o que se modifica para contener segmentos de ácidos nucleicos en cierto modo que de otra manera no existiría en naturaleza. El término construido de ácidos nucleicos es sinónimo del término “cassette de expresión” cuando el construido 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” es definido aquí para incluir todos los componentes, que son necesarios o ventajosos para la expresión de un polinucleótido que codifica un polipéptido de la presente invención. Cada secuencia de control puede ser nativa o extranjera a la secuencia de nucleótidos que codifica el polipéptido o nativa o extranjera una de otra. Tales secuencias de control incluyen, pero no de forma limitativa, una secuencia líder, secuencia de poliadenilación, secuencia de propéptido, promotor, secuencia de péptido señal, y terminador de la transcripción. Como mínimo, las secuencias de control incluyen un promotor, y señales de parada traduccionales y transcripcionales. Las secuencias de control pueden ser provistas de enlazadores con el fin de introducir sitios de restricción específicos que facilitan la ligadura de las secuencias de control con la región codificante de la secuencia de nucleótidos que codifica un polipéptido.

Operativamente enlazado: el término “operativamente enlazado” denota aquí una configuración en la que una secuencia de control se coloca en una posición apropiada relativa a la secuencia de codificación de la secuencia polinucleótida de manera que la secuencia de control dirige la expresión de la secuencia codificante de un polipéptido.

Secuencia codificante: cuando se usa aquí el término “secuencia codificante” significa una secuencia de nucleótidos, que especifica directamente la secuencia de aminoácidos de su producto proteínico. Los límites de la secuencia codificante son generalmente determinados por un marco de lectura abierto, que comienza normalmente con el codón de inicio ATG o codones de inicio alternativos tales como GTG y TTG y extremidades con un codón de terminación tal como TAA, TAG, y TGA. La secuencia codificante puede ser una secuencia de ADN, ADNc, o de nucleótidos recombinante.

Expresión: el término “expresión” incluye cualquier paso implicado en la producción del polipéptido incluyendo, pero no limitado a, transcripción, modificación postranscripcional, traducción, modificación postraduccional, y secreción.

Vector de expresión: el término “vector de expresión” es definido aquí como una molécula de ADN circular o lineal que comprende un polinucleótido que codifica un polipéptido de la invención, y que está enlazado operativamente a nucleótidos adicionales que proveen su expresión.

- 5 Célula huésped: el término “célula huésped”, como se utiliza en este caso, incluye cualquier tipo célula que es susceptible de transformación, transfección, transducción, y similares con un constructo de ácidos nucleicos o vector de expresión que comprende un polinucleótido de la presente invención.

10 Modificación: el término “modificación” significa aquí cualquier modificación química del polipéptido que consiste en el polipéptido maduro de SEC ID n°: 2 o una secuencia homóloga de la misma; así como manipulación genética del ADN que codifica tal polipéptido. La modificación puede ser sustituciones, deleciones y/o inserciones de uno o más aminoácidos al igual que sustituciones de una o más cadenas laterales de aminoácido.

15 Variante artificial: cuando se usa aquí, el término “variante artificial” significa un polipéptido que tiene actividad de celobiohidrolasa producida por un organismo que expresa una secuencia de nucleótidos modificada de la secuencia codificante del polipéptido maduro de SEC ID n°: 1 o una secuencia homóloga de la misma. La secuencia de nucleótidos modificada se obtiene a través de la intervención humana por modificación de la secuencia de nucleótidos descrita en SEC ID n°: 1 o una secuencia homóloga de la misma.

20 Descripción detallada de la invención

Polipéptidos que tienen actividad de celobiohidrolasa

25 En un primer aspecto, la presente invención se refiere a polipéptidos aislados comprendiendo una secuencia de aminoácidos que tiene un grado de identidad al polipéptido maduro de SEC ID n°: 2, de al menos 60%, preferiblemente al menos 65%, más preferiblemente al menos 70%, más preferiblemente al menos 75%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, incluso más preferiblemente al menos 90%, de la forma más preferible al menos 95%, e incluso de la forma más preferible al menos 97%, 98%, o 99%, que tiene actividad de celobiohidrolasa (de ahora en adelante “polipéptidos homólogos”). En un aspecto preferido, los polipéptidos homólogos tienen una
30 secuencia de aminoácidos que difiere por diez aminoácidos, preferiblemente por cinco aminoácidos, más preferiblemente por cuatro aminoácidos, incluso más preferiblemente por tres aminoácidos, de la forma más preferible por dos aminoácidos, e incluso de la forma más preferible por un aminoácido del polipéptido maduro de SEC ID n°: 2.

35 Un polipéptido de la presente invención comprende preferiblemente la secuencia de aminoácidos de la SEC ID n°: 2 o una variante alélica de la misma; o un fragmento de la misma que tiene actividad de celobiohidrolasa. En un aspecto preferido, un polipéptido comprende la secuencia de aminoácidos de SEC ID n°: 2. En otro aspecto preferido, un polipéptido comprende el polipéptido maduro de SEC ID n°: 2. En otro aspecto preferido, un polipéptido comprende los aminoácidos 18 a 481 de SEC ID n°: 2, o una variante alélica de la misma; o un fragmento de la misma que tiene actividad de celobiohidrolasa. En otro aspecto preferido, un polipéptido comprende los aminoácidos 18 a 481 de SEC
40 ID n°: 2. En otro aspecto preferido, un polipéptido consiste en la secuencia de aminoácidos de SEC ID n°: 2 o una variante alélica de la misma; o un fragmento de la misma que tiene actividad de celobiohidrolasa. En otro aspecto preferido, un polipéptido consiste en la secuencia de aminoácidos de SEC ID n°: 2. En otro aspecto preferido, un polipéptido consiste en el polipéptido maduro de SEC ID n°: 2. En otro aspecto preferido, un polipéptido consiste en los aminoácidos 18 a 481 de SEC ID n°: 2 o una variante alélica de la misma; o un fragmento de la misma que tiene
45 actividad de celobiohidrolasa. En otro aspecto preferido, un polipéptido consiste en los aminoácidos 18 a 481 de SEC ID n°: 2.

50 En un segundo aspecto, la presente invención se refiere a polipéptidos aislados que tienen actividad de celobiohidrolasa que se codifican por polinucleótidos que se hibridan bajo condiciones de astringencia muy baja, preferiblemente condiciones de astringencia baja, más preferiblemente condiciones de astringencia media, más preferiblemente condiciones de astringencia media-alta, incluso más preferiblemente condiciones de astringencia alta, y de la forma más preferible condiciones de astringencias muy alta con (i) la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (ii) la secuencia de ADN genómico que comprende la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (iii) una subsecuencia de (i) o (ii), o (iv) una cadena complementaria de (i); (ii), o (iii) (J. Sambrook, E.F. Fritsch, y T. Maniatis, 1989, Molecular Cloning, A Laboratory Manual, 2ª edición, Cold Spring Harbor, New York). Una subsecuencia de la secuencia codificante del polipéptido maduro de SEC ID n°: 1 contiene al menos 100 nucleótidos contiguos o preferiblemente al menos 200 nucleótidos contiguos. Por otra parte, la subsecuencia puede codificar un fragmento de polipéptido que tiene actividad de celobiohidrolasa. En un aspecto preferido, la secuencia codificante del polipéptido maduro es los nucleótidos 52 a 1443 de SEC ID n°: 1.
60

65 La secuencia de nucleótidos de SEC ID n°: 1 o una subsecuencia de la misma; así como la secuencia de aminoácidos de SEC ID n°: 2 o un fragmento de la misma; se puede usar para diseñar una sonda de ácidos nucleicos para identificar y clonar ADN que codifica los polipéptidos que tienen actividad de celobiohidrolasa de cepas de géneros diferentes o especies según métodos bien conocidos en la técnica. En particular, tales sondas se pueden usar para hibridación con el ADN genómico o ADNc del género o especie de interés, después de procedimientos de Southern blot estándares, para identificar y aislar el gen correspondiente en éstas. Tales sondas pueden ser considerablemente más cortas que la secuencia entera, pero deberían ser al menos 14, preferiblemente al menos 25, más preferiblemente al menos 35, y de la forma más preferible al menos 70 nucleótidos de longitud. Es, no obstante, preferido que la sonda

ES 2 358 243 T3

de ácidos nucleicos sea al menos 100 nucleótidos de longitud. Por ejemplo, la sonda de ácidos nucleicos puede ser al menos 200 nucleótidos, preferiblemente al menos 300 nucleótidos, más preferiblemente al menos 400 nucleótidos, o de la forma más preferible al menos 500 nucleótidos de longitud. Sondas incluso más largas pueden ser utilizadas, por ejemplo, sondas de ácidos nucleicos que son al menos 600 nucleótidos, al menos preferiblemente al menos 700 nucleótidos, más preferiblemente al menos 800 nucleótidos, o de la forma más preferible al menos 900 nucleótidos de longitud. Ambas sondas de ADN y ARN pueden ser usadas. Las sondas son típicamente marcadas para la detección del gen correspondiente (por ejemplo, con ³²P, ³H, ³⁵S, biotina, o avidina). Tales sondas están incluidas por la presente invención.

Un ADN genómico o biblioteca de ADNc obtenidos a partir de tales otros organismos pueden, por lo tanto, ser seleccionados para ADN que se hibrida con las sondas anteriormente descritas y que codifica un polipéptido que tiene actividad de celobiohidrolasa. ADN genómico u otro ADN de tales otros organismos se pueden separar por agarosa o electroforesis en gel de poliacrilamida, u otras técnicas de separación. ADN de las bibliotecas o el ADN separado se puede transferir a e inmovilizar en nitrocelulosa u otro material portador adecuado. Para identificar un clon o ADN que es homólogo a la SEC ID n°: 1 o una subsecuencia de la misma; el material portador se usa en un Southern blot.

Para objetivos de la presente invención, la hibridación indica que la secuencia de nucleótidos se hibrida a una sonda de ácidos nucleicos marcada correspondiente a la secuencia codificante del polipéptido maduro de SEC ID n°: 1, la secuencia de ADN genómico comprendiendo la secuencia codificante del polipéptido maduro de SEC ID n°: 1; su cadena complementaria; o una subsecuencia de la misma; bajo condiciones de astringencia muy baja a muy alta. Moléculas a las que la sonda de ácidos nucleicos se hibrida bajo estas condiciones pueden ser detectadas usando, por ejemplo, película radiográfica.

En un aspecto preferido, la sonda de ácidos nucleicos es la secuencia codificante del polipéptido maduro de SEC ID n°: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es los nucleótidos 52 a 1443 de SEC ID n°: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es una secuencia polinucleótida que codifica el polipéptido de SEC ID n°: 2, o una subsecuencia de la misma. En otro aspecto preferido, la sonda de ácidos nucleicos es SEC ID n°: 1. En otro aspecto preferido, la sonda de ácidos nucleicos es la secuencia de polinucleótidos contenida en el plásmido pTter6A que está contenido en *E. coli* NRRL B-30802, donde la secuencia de polinucleótidos de la misma codifica un polipéptido que tiene actividad de celobiohidrolasa. En otro aspecto preferido, la sonda de ácidos nucleicos es la región codificante del polipéptido maduro contenida en el plásmido pTter6A que está contenido en *E. coli* NRRL B-30802.

Para sondas largas de al menos 100 nucleótidos de longitud, condiciones de astringencia muy baja a muy alta son definidas como prehibridación e hibridación a 42°C en 5X SSPE, 0,3% SDS, 200 µg/ml de ADN de esperma de salmón desnaturalizado y cortado, y bien 25% formamida para astringencias bajas y muy bajas, 35% de formamida para astringencias media y media-alta, o 50% de formamida para astringencias alta y muy alta, después de procedimientos de Southern blot estándares durante 12 a 24 horas óptimamente.

Para sondas largas de al menos 100 nucleótidos de longitud, el material portador es finalmente lavado tres veces cada una durante 15 minutos usando 2X SSC, 0,2% SDS preferiblemente al menos a 45°C (astringencia muy baja), más preferiblemente al menos a 50°C (astringencia baja), más preferiblemente al menos a 55°C (astringencia media), más preferiblemente al menos a 60°C (astringencia media-alta), incluso más preferiblemente al menos a 65°C (astringencia alta), y de la forma más preferible al menos a 70°C (astringencia muy alta).

Para sondas cortas que son de aproximadamente 15 nucleótidos a aproximadamente 70 nucleótidos de longitud, condiciones de astringencia son definidas como prehibridación, hibridación, y lavado post-hibridación a aproximadamente 5°C a aproximadamente 10°C por 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 NaCl, 0,09 M tris-HCl pH 7,6, 6 mM EDTA, 0,5% NP-40, 1X solución del Denhardt, 1 mM pirofosfato de sodio, 1 mM fosfato monobásico de sodio, 0,1 mM ATP, y 0,2 mg de ARN de levadura por ml después de procedimientos de Southern blot estándares durante 12 a 24 horas óptimamente.

Para sondas cortas que son de aproximadamente 15 nucleótidos a aproximadamente 70 nucleótidos de longitud, el material portador se lava una vez en 6X SSC más 0,1% SDS durante 15 minutos y dos veces cada una durante 15 minutos usando 6X SSC a 5°C a 10°C por debajo de la T_m calculada.

En un tercer aspecto, la presente invención se refiere a variantes artificiales comprendiendo una sustitución conservadora, delección, y/o inserción de uno o más aminoácidos del polipéptido maduro de SEC ID n°: 2 o una secuencia homóloga de la misma. Preferiblemente, cambios de aminoácidos son de una naturaleza menor, es decir sustituciones de aminoácidos conservadoras o inserciones que no afectan significativamente al pliegue y/o actividad de la proteína; delecciones pequeñas, típicamente de uno a aproximadamente 30 aminoácidos; pequeñas extensiones amino o carboxilo terminales, tales 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 epítipo antigénico o un dominio de unión.

Ejemplos de sustituciones conservadoras están en el grupo de aminoácidos básicos (arginina, lisina e histidina), aminoácidos ácidos (ácido glutámico y ácido aspártico), aminoácidos polares (glutamina y asparagina), aminoácidos hidrofóbicos (leucina, isoleucina y valina), aminoácidos aromáticos (fenilalanina, triptófano y tirosina), y aminoácidos pequeños (glicina, alanina, serina, treonina y metionina). Sustituciones de aminoácidos que generalmente no alteran la actividad específica son conocidas en la técnica y son descritas, por ejemplo, por H. Neurath y R.L. Hill, 1979, en, *The Proteins*, Academic Press, New York. Los intercambios que ocurren 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.

Además de los 20 aminoácidos estándares, aminoácidos no estándares (tales como 4-hidroxi prolina, 6-N-metil lisina, ácido 2-aminoisobutírico, isovalina, y alfa-metil serina) se pueden sustituir por residuos de aminoácidos de un polipéptido de tipo salvaje. Un número limitado de aminoácidos no conservadores, aminoácidos que no son codificados por el código genético, y aminoácidos innaturales se pueden sustituir por residuos de aminoácidos. "Aminoácidos innaturales" han sido modificados después de la síntesis de proteína, y/o tienen una estructura química en sus cadena(s) lateral(es) diferente de los aminoácidos estándares. Aminoácidos innaturales pueden ser químicamente sintetizados, y preferiblemente, están disponibles comercialmente, e incluyen ácido pipercolico, ácido carboxílico de tiazolidina, deshidroprolina, 3- y 4-metilprolina, y 3,3-dimetilprolina.

Alternativamente, los cambios de aminoácidos son de tal naturaleza que las propiedades fisicoquímicas de los polipéptidos son alteradas. Por ejemplo, cambios de aminoácidos pueden mejorar la termoestabilidad del polipéptido, alterar la especificidad del sustrato, cambiar el pH óptimo, y similares.

Aminoácidos esenciales en el polipéptido progenitor se pueden identificar según procedimientos conocidos en la técnica, tales como mutagénesis dirigida o mutagénesis de barrido de alanina (Cunningham and Wells, 1989, *Science* 244: 1081-1085). En esta técnica, se introducen mutaciones de alanina únicas en cada residuo en la molécula, y las moléculas mutantes resultantes se evalúan por su actividad biológica (es decir, actividad de celobiohidrolasa) para identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Véase también, Hilton *et al.*, 1996, *J. Biol. Chem.*, 271: 4699-4708. El sitio activo de la enzima u otra interacción biológica puede también ser determinada por análisis físico de estructura, como es determinado por técnicas tales como resonancia magnética nuclear, cristalografía, difracción electrónica, o marcado por fotoafinidad, conjuntamente con mutación de aminoácidos de sitio de contacto putativo. Véase, por ejemplo, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *J. Mol. Biol.* 224: 899-904; Wlodaven *et al.*, 1992, *FEBS Lett.* 309: 59-64. Las identidades de aminoácidos esenciales pueden también ser inferidas del análisis de identidades con polipéptidos que se relacionan con un polipéptido según la invención.

Sustituciones de aminoácido múltiples o únicas pueden ser hechas y evaluadas usando métodos conocidos de mutagénesis, recombinación, y/o redistribución, seguido de un procedimiento de selección pertinente, tales como aquellos descritos por Reidhaar-Olson and Sauer, 1988, *Science* 241: 53-57; Bowie and Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO 95/17413; o WO 95/22625. Otros métodos que pueden ser usados incluyen PCR con tendencia al error, exposición en fago (p. ej., Lowman *et al.*, 1991, *Biochem.* 30: 10832-10837; patente estadounidense n°. 5,223,409; WO 92/06204), y mutagénesis dirigida a la región (Derbyshire *et al.*, 1986, *Gene* 46: 145; Ner *et al.*, 1988, *DNA* 7: 127).

Métodos de mutagénesis/redistribución se pueden combinar con métodos de selección automatizados de alto rendimiento, para detectar actividad de polipéptidos clonados, mutagenizados expresados por células huéspedes (Ness *et al.*, 1999, *Nature Biotechnology* 17: 893-896). Moléculas mutagenizadas de ADN que codifican polipéptidos activos se pueden recuperar de las células huéspedes y secuenciar rápidamente usando métodos estándares en la técnica. Estos métodos permiten la determinación rápida de la importancia de residuos de aminoácidos individuales en un polipéptido de interés, y se pueden aplicar a polipéptidos de estructura desconocida.

El número total de sustituciones de aminoácidos, deleciones y/o inserciones del polipéptido maduro de SEC ID n°: 2, tal como aminoácidos 18 a 481 de SEC ID n°: 2, es 10, preferiblemente 9, más preferiblemente 8, más preferiblemente 7, más preferiblemente como mucho 6, más preferiblemente 5, más preferiblemente 4, incluso más preferiblemente 3, de la forma más preferible 2, e incluso de la forma más preferible 1.

Fuentes de polipéptidos que tienen actividad de celobiohidrolasa

Un polipéptido de la presente invención se puede obtener de microorganismos de cualquier género. Para objetivos de la presente invención, el término "obtenido de" como se utiliza en este caso en relación con una fuente dada debe significar que el polipéptido codificado por una secuencia de nucleótidos se produce por la fuente o por una cepa en donde la secuencia de nucleótidos de la fuente ha sido insertada. En un aspecto preferido, el polipéptido obtenido de una fuente dada 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* que tiene actividad de celobiohidrolasa, por ejemplo, un polipéptido de *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus*

ES 2 358 243 T3

stearothermophilus, *bacillus subtilis*, o de *Bacillus thuringiensis* que tiene actividad de celobiohidrolasa; o un polipéptido de *Streptomyces* que tiene actividad de celobiohidrolasa, por ejemplo, un polipéptido de *Streptomyces lividans* o *Streptomyces murinus* que tiene actividad de celobiohidrolasa; o un polipéptido bacteriano gram negativo que tiene actividad de celobiohidrolasa, por ejemplo, un polipéptido de *E. coli* o de *Pseudomonas sp.* que tiene actividad de celobiohidrolasa.

Un polipéptido de la presente invención puede también ser un polipéptido fúngico, y más preferiblemente un polipéptido de levadura tal como un polipéptido de *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, o *Yarrowia* que tiene actividad de celobiohidrolasa; o más preferiblemente un polipéptido filamentoso fúngico tal como un polipéptido de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Mycelophthora*, *Neocallimastix*, *Neumspora*, *Paecilomyces*, *Penicillium*, *Piromyces*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolyposcladum*, o *Trichoderma* que tiene actividad de celobiohidrolasa.

En un aspecto preferido, el polipéptido es un polipéptido de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, o *Saccharomyces oviformis* que tiene actividad de celobiohidrolasa.

En otro aspecto preferido, el polipéptido es un polipéptido de *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium gramineum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophylla*, *Neurospora crassa*, *Penicillium purpurogenum*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, o *Trichoderma viride* que tiene actividad de celobiohidrolasa.

En otro aspecto preferido, el polipéptido es un polipéptido de *Thielavia achromatica*, *Thielavia albomyces*, *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovlspora*, *Thielavia peruviana*, *Thielavia spededonium*, *Thielavia setosa*, *Thielavia subthermophila*, *Thielavia terrestris*, *Thielavia terricola*, *Thielavia thermophylla*, *Thielavia variospora*, o *Thielavia warengii* que tiene actividad de celobiohidrolasa.

En un aspecto más preferido; el polipéptido es un polipéptido de *Thielavia terrestris*, y más preferiblemente *Thielavia terrestris* NRRL 8126, por ejemplo, el polipéptido de SEC ID n°: 2, o el polipéptido maduro del mismo.

Será entendido que para las especies mencionadas la invención incluye los estados imperfectos y perfectos, y otros equivalentes taxonómicos, por ejemplo, anamorfos, independientemente del nombre de las especies por el que se conocen. Expertos en la técnica fácilmente reconocen la identidad de los equivalentes apropiados.

Cepas de estas especies son fácilmente accesibles al público en un número de colecciones de cultivo, tal como la Colección Americana de Cultivos Tipo (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), y Colección de Patentes de Cultivos del Servicio de Investigación para la Agricultura, Centro de investigación Regional del Norte (NRRL).

Además, tales polipéptidos se pueden identificar y obtener de otras fuentes incluyendo microorganismos aislados de la naturaleza (p. ej., tierra, abonos, agua, etc.) usando las sondas mencionadas arriba. Técnicas para aislar microorganismos de hábitats naturales se conocen en la técnica. El polinucleótido puede luego ser obtenido por selección de forma similar de una biblioteca genómica o de ADNc de tal microorganismo. Una vez que una secuencia polinucleotídica que codifica un polipéptido ha sido detectada con la(s) sonda(s), el polinucleótido puede ser aislado o clonado utilizando técnicas que son bien conocidas por los expertos en la técnica (véase, por ejemplo, Sambrook *et al.*, 1989, *supra*).

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

Polinucleótidos

La presente invención también se refiere a polinucleótidos aislados que comprenden o que consisten en una secuencia de nucleótidos que codifica un polipéptido de la presente invención que tiene actividad de celobiohidrolasa.

En un aspecto preferido, la secuencia de nucleótidos comprende o consiste en SEC ID n°: 1. En otro aspecto más preferido, la secuencia de nucleótidos comprende o consiste en la secuencia contenida en el plásmido pTter6A que está

5 contenida en *E. coli* NRRL B-30802. En otro aspecto preferido, la secuencia de nucleótidos comprende o consiste en la región codificante del polipéptido maduro de SEC ID n°: 1. En otro aspecto preferido, la secuencia de nucleótidos comprende o consiste en nucleótidos 52 a 1443 de SEC ID n°: 1. En otro aspecto más preferido, la secuencia de nucleótidos comprende o consiste en la región codificante del polipéptido maduro contenida en el plásmido pTter6A
10 que está contenido en *E. coli* NRRL B-30802. La presente invención también incluye secuencias de nucleótidos que codifican un polipéptido que comprende o que consiste en la secuencia de aminoácidos de SEC ID n°: 2 o el polipéptido maduro de la misma, que difiere de SEC ID n°: 1 o la secuencia codificante del polipéptido maduro de la misma en virtud de la degeneración del código genético. La presente invención también se refiere a subsecuencias de SEC ID n°: 1 que codifican fragmentos de SEC ID n°: 2 que tienen actividad de celobiohidrolasa.

15 La presente invención también se refiere a polinucleótidos mutantes que comprenden o que consisten al menos en una mutación en la secuencia codificante del polipéptido maduro de SEC ID n°: 1, en la que la secuencia de nucleótidos mutante codifica el polipéptido maduro de SEC ID n°: 2. En un aspecto preferido, el polipéptido maduro es los aminoácidos 18 a 481 de SEC ID n°: 2.

20 Las técnicas usadas para aislar o clonar un polinucleótido que codifica un polipéptido se conocen en la técnica e incluyen aislamiento de ADN genómico, preparación de ADNc, o una combinación de ambos. La clonación de los polinucleótidos de la presente invención de tal ADN genómico puede ser efectuada, por ejemplo, usando la reacción en cadena de polimerasa (PCR) bien conocida o selección de anticuerpos de bibliotecas de expresión para detectar fragmentos de ADN clonados con características estructurales compartidas. Véase, por ejemplo, Innis *et al.*, 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Otros procedimientos de amplificación de ácidos nucleicos tal como la reacción en cadena de la ligasa (LCR), la transcripción activada ligada (LAT) y la ampli-
25 ficación basada en secuencias de nucleótidos (NASBA) puede ser utilizada. Los polinucleótidos se pueden clonar de una cepa de *Thielavia*, u otro organismo relacionado y así, por ejemplo, puede ser una variante alélica o de especies de la región codificante del polipéptido de la secuencia de nucleótidos.

30 La presente invención también se refiere a polinucleótidos que comprenden o que consisten en secuencias de nucleótidos que tienen un grado de identidad a la secuencia codificante del polipéptido maduro de SEC ID n°: 1 de al menos 60%, preferiblemente al menos 65%, más preferiblemente al menos 70%, más preferiblemente al menos 75%, más preferiblemente al menos 80%, más preferiblemente al menos 85%, más preferiblemente al menos 90%, incluso más preferiblemente al menos 95%, y de la forma más preferible al menos 97% de identidad, que codifican un polipéptido activo. En un aspecto preferido, la secuencia codificante del polipéptido maduro es los nucleótidos 52 a 1443 de SEC ID n°: 1.

35 La modificación de una secuencia de nucleótidos que codifica un polipéptido de la presente invención puede ser necesaria para la síntesis de polipéptidos sustancialmente similares al polipéptido. El término "sustancialmente similar" al polipéptido se refiere a las formas de origen no natural del polipéptido. Estos polipéptidos pueden diferir en alguna manera creada genéticamente del polipéptido aislado de su fuente nativa, por ejemplo, variantes artificiales que difieren en actividad específica, termostabilidad, pH óptimo, o similar. La secuencia variante puede ser construida basándose en la secuencia de nucleótidos presentada como la región codificante del polipéptido de SEC ID n°: 1, por ejemplo, una subsecuencia de la misma, y/o por introducción de sustituciones de nucleótidos que no dan lugar a otra secuencia de aminoácidos del polipéptido codificado por la secuencia de nucleótidos, pero que corresponden al uso del codón del organismo huésped destinado para la producción de la enzima, o por introducción de sustituciones de nucleótidos que pueden dar lugar a una secuencia de aminoácidos diferente. Para una descripción general de sustitución
45 de nucleótidos, véase, por ejemplo, Ford *et al.*, 1991, Protein Expression and Purification 2: 95-107.

50 Será aparente para los expertos en la técnica que tales sustituciones pueden ser hechas fuera de las regiones críticas para la función de la molécula y todavía resultan en un polipéptido activo. Residuos de aminoácidos esenciales para la actividad del polipéptido codificados por un polinucleótido aislado de la invención, y por lo tanto preferiblemente no sujeto a sustitución, se pueden identificar según procedimientos conocidos en la técnica, tal como mutagénesis dirigida o mutagénesis de barrido de alanina (véase, por ejemplo, Cunningham y Wells, 1989, *supra*). En esta técnica, se introducen mutaciones en cada residuo cargado positivamente en la molécula, y las moléculas mutantes resultantes se evalúan para actividad de celobiohidrolasa para identificar residuos de aminoácidos que son críticos para la actividad de la molécula. Sitios de interacción de enzima-sustrato pueden también ser determinados por análisis de la estructura tridimensional como está determinado por técnicas tales como el análisis de resonancia magnética nuclear, cristalografía o marcado por fotoafinidad (véase, por ejemplo, de Vos *et al.*, 1992, *supra*; Smith *et al.*, 1992, *supra*; Wlodaver
55 *et al.*, 1992, *supra*).

60 La presente invención también se refiere a polinucleótidos aislados que codifican un polipéptido de la presente invención, que se hibridan bajo condiciones de astringencia muy baja, preferiblemente condiciones de astringencia baja, más preferiblemente condiciones de astringencia media, más preferiblemente condiciones de astringencia media-alta, incluso más preferiblemente condiciones de astringencia alta, y de la forma más preferible condiciones de astringencia muy alta con (i) la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (ii) la secuencia de ADN genómico comprendiendo la secuencia codificante del polipéptido maduro de SEC ID n°: 1, o (iii) una cadena complementaria de (i) o (ii); o variantes alélicas y subsecuencias de las mismas (Sambrook *et al.*, 1989, *supra*), tal y como se define
65 aquí. En un aspecto preferido, la secuencia codificante del polipéptido maduro de SEC ID n°: 1 es nucleótidos 52 a 1443.

La presente invención también se refiere a polinucleótidos aislados obtenidos por (a) hibridación de una población de ADN bajo condiciones de astringencia muy baja, baja, media, media-alta, alta, o muy alta con (i) la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (ii) la secuencia de ADN genómico comprendiendo la secuencia codificante del polipéptido maduro de SEC ID n°: 1, o (iii) una cadena complementaria de (I) o (ii); y (b) aislamiento del polinucleótido hibridante, que codifica un polipéptido que tiene actividad de celobiohidrolasa. En un aspecto preferido, la secuencia codificante del polipéptido maduro de SEC ID n°: 1 es los nucleótidos 52 a 1443.

Constructos de ácidos nucleicos

La presente invención también se refiere a constructos de ácidos nucleicos que comprenden un polinucleótido aislado de la presente invención operativamente enlazado 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.

Un polinucleótido aislado que codifica un polipéptido de la presente invención se puede manipular en una variedad de maneras para proveer la expresión del polipéptido. La manipulación de la secuencia del polinucleótido antes de su inserción en un vector puede ser deseable o necesaria dependiendo del vector de expresión. Las técnicas para modificar secuencias polinucleótidas utilizando métodos de ADN recombinante se conocen bien en la técnica.

La secuencia de control puede ser una secuencia del promotor apropiada, una secuencia de nucleótidos que se reconoce por una célula huésped para la expresión de un polinucleótido que codifica un polipéptido de la presente invención. La secuencia del promotor 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, mutante truncado, y promotores híbridos, y se pueden obtener de genes que codifican polipéptidos intracelulares o extracelulares 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 de *Streptomyces coelicolor* (*dagA*), gen de levansucrasa de *Bacillus subtilis* (*sacB*), gen de alfa-amilasa de *Bacillus licheniformis* (*amyL*), gen de amilasa maltogénica de *Bacillus stearothermophilus* (*amyM*), gen de alfa-amilasa de *Bacillus amyloliquefaciens* (*amyQ*), gen de penicilinas de *Bacillus licheniformis* (*penP*), genes *xylA* y *xylB* de *Bacillus subtilis*, y gen de beta-lactamasa procariótica (Villa-Kamaroff *et al.*, 1978, Proceedings of the National Academy of Sciences USA 75: 3727-3731), 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 de *Aspergillus niger* o de *Aspergillus awamori* (*glaA*), lipasa de *Rhizomucor miehei*, proteasa alcalina de *Aspergillus oryzae*, triosa fosfato isomerasa de *Aspergillus oryzae*, acetamidasa de *Aspergillus nidulans*, amiloglucosidasa de *Fusarium venenatum* (WO 00/56900), *Fusarium venenatum* Daria (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), proteasa de tipo tripsina de *Fusarium oxysporum* (WO 96/00787), beta-glucosidasa de *Trichoderma reesei*, celobiohidrolasa I de *Trichoderma reesei*, celobiohidrolasa II de *Trichoderma reesei*, endoglucanasa I de *Trichoderma reesei*, endoglucanasa II de *Trichoderma reesei*, endoglucanasa III de *Trichoderma reesei*, endoglucanasa IV de *Trichoderma reesei*, endoglucanasa V de *Trichoderma reesei*, xilanasasa I de *Trichoderma reesei*, xilanasasa II de *Trichoderma reesei*, beta-xilosidasa de *Trichoderma reesei*, 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, mutante truncado, y promotores híbridos de los mismos.

En un huésped de levadura, promotores útiles se obtienen de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactocinasa de *Saccharomyces cerevisiae* (GAL1), alcohol deshidrogenasa/gliceraldehído 3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH1, ADH2/GAP), triosa fosfato isomerasa de *Saccharomyces cerevisiae* (TPI), metalotionina de *Saccharomyces cerevisiae* (CUP1), 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 vinculada al término 3' de la secuencia de nucleótidos que codifica el polipéptido. Cualquier terminador que es funcional en la célula huésped de elección se puede utilizar en la presente invención.

Terminadores preferidos para células huéspedes filamentosas fúngicas se obtienen 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*.

ES 2 358 243 T3

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

La secuencia de control puede también ser una secuencia líder 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 líder que es funcional en la célula huésped de elección se puede utilizar en la presente invención.

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

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

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 se transcribe, es reconocida por la célula huésped como una señal para añadir residuos de poliadenosina a ARNm transcrito. Cualquier secuencia de poliadenilación que es funcional en la célula huésped de elección se puede utilizar en la presente invención.

Secuencias de poliadenilación preferidas para células huéspedes filamentosas fúngicas se obtienen 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 son descritas por Guo y Sherman, 1995, *Molecular Cellular Biology* 15: 5983-5990.

La secuencia de control puede también ser una región de codificación de péptido señal que codifica para una secuencia de aminoácidos vinculada al término amino de un polipéptido y dirige el polipéptido codificado en la vía secretora de la célula. La extremidad 5' de la secuencia codificante de la secuencia de nucleótidos puede intrínsecamente contener una región codificante del péptido señal naturalmente vinculado en el marco de lectura de traducción con el segmento de la región codificante que codifica el polipéptido segregado. Alternativamente, la extremidad 5' de la secuencia codificante puede contener una región codificante del péptido señal que es extranjera a la secuencia codificante. La región codificante del péptido señal extranjera puede ser requerida cuando la secuencia codificante no contiene naturalmente una región codificante del péptido señal. Alternativamente, la región codificante del péptido señal extranjera puede simplemente reemplazar la región codificante del péptido señal natural para mejorar 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, es decir, segregado en un medio de cultivo, se puede utilizar 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 son 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*, endoglucanasa de *Humicola insolens* V, y lipasa de *Humicola lanuginosa*.

En un aspecto preferido, el péptido señal es los aminoácidos 1 a 17 de SEC ID n°: 2. En otro aspecto preferido, la región codificante del péptido señal es los nucleótidos 1 a 51 de SEC ID n°: 1 que codifican los aminoácidos 1 a 17 de SEC ID n°: 2.

Péptidos señal útiles para células huéspedes de levadura se obtienen de los genes para el factor alfa 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 se puede convertir a un polipéptido maduro activo por escisión autocatalítica o catalítica del propéptido del propolipéptido. La región codificante del propéptido se puede obtener de los genes para proteasa alcalina de *Bacillus subtilis* (aprE), proteasa neutra de *Bacillus subtilis* (nprT), factor alfa de *Saccharomyces cerevisiae*, proteinasa aspártica de *Rhizomucor miehei*, y lacasa *Myceliophthora thermophila* (WO 95/33836).

Donde ambas regiones del péptido señal y 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.

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

Vectores de expresión

20 La presente invención también se refiere a vectores de expresión recombinantes que comprenden un polinucleótido de la presente invención, un promotor, y señales de parada traduccionales y transcripcionales. Los varios ácidos nucleicos y secuencias de control que se describen aquí se pueden unir 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 codifican el polipéptido en tales sitios. Alternativamente, una secuencia de nucleótidos de la presente invención se puede expresar por inserción de la secuencia de nucleótidos o un constructo de ácidos nucleicos comprendiendo la secuencia en un vector apropiado para la expresión. En la creación del vector de expresión, la secuencia codificante se localiza en el vector de modo que la secuencia codificante está enlazada operativamente con las secuencias de control apropiadas para la expresión.

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

35 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, por ejemplo, un plásmido, un elemento extracromosómico, un minicromosoma, o un cromosoma artificial. El vector puede contener cualquier medio para asegurar la autorreplicación. Alternativamente, el vector puede ser uno que, introducido en la célula huésped, se integra en el genoma y replicado con el(los) cromosoma(s) en que ha sido integrado. Además, un único vector o plásmido o dos o más vectores o plásmidos que juntos contienen el ADN total para ser introducido en el genoma de la célula huésped, o un transposón puede ser utilizado.

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

50 Ejemplos de marcadores seleccionables bacterianos son los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia antibiótica tal como resistencia a 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 de forma limitativa, *amdS* (acetamidasa), *argB* (ornitina-carbamoiltransferasa), *bar* (fosfonitricina acetiltransferasa), *hph* (higromicina fosfotransferasa), *niaD* (nitrato reductasa), *pyrG* (orotidina-5'-fosfato descarboxilasa), *sC* (sulfato adeniltransferasa), y *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*.

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

65 Para la integración en el genoma de la célula huésped, el vector puede depender de la secuencia del polinucleótido que codifica el polipéptido o cualquier otro elemento del vector para integración en el genoma por recombinación homóloga o no homóloga. Alternativamente, el vector puede contener secuencias de nucleótidos adicionales para dirigir la integración por recombinación homóloga en el genoma de la célula huésped en una(s) ubicación(es) precisa(s) en el(los) cromosoma(s). Para aumentar la probabilidad de integración a una ubicación precisa, los elementos integrales deberían contener preferiblemente un número suficiente de ácidos nucleicos, tal como 100 a 10.000 pares de bases, preferiblemente 400 a 10.000 pares de bases, y de la forma más preferible 800 a 10.000 pares de bases, que tienen un alto grado de identidad con la secuencia blanco correspondiente para mejorar la probabilidad de recombinación

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

Para la replicación autónoma, el vector puede comprender además un origen de replicación permitiendo que el vector se replique de manera autónoma en la célula huésped en cuestión. El origen de replicación puede ser cualquier replicador plásmido que medie la replicación autónoma que funciona en una célula. El término “origen de replicación” o “replicador plásmido” se define aquí como una secuencia de nucleótidos que permite que un plásmido o vector se replique *in vivo*.

Ejemplos de orígenes bacterianos de replicación son los orígenes de replicación de plásmidos pBR322; pUC19; pACiC177, y pACYC184 permitiendo la replicación en *E. coli*, y pUB110; pE194; pTA1060, y pAMB1 permitiendo la replicación en *Bacillus*.

Ejemplos de orígenes de replicación para el uso en una célula huésped de levadura son el origen de replicación de 2 micra, ARS1, ARS4, la combinación de ARS1 y CEN3, y la combinación de ARS4 y CEN6.

Ejemplos de orígenes de replicación útiles en una célula micótica filamentosa son AMA1 y ANS1 (Gems *et al.*, 1991, Gene 98: 61-67; Cullen *et al.*, 1987, Nucleic Acids Research 15: 9163-9175; WO 00/24883). Aislamiento del gen AMA1 y construcción de plásmidos o vectores comprendiendo el gen se puede realizar según los métodos descritos en WO 00/24883.

Más de una copia de un polinucleótido de la presente invención se puede insertar en la célula huésped para aumentar la producción del producto genético. Un aumento en el número de copias del polinucleótido se puede obtener integrando al menos una copia adicional de la secuencia en el genoma de la célula huésped o incluyendo un gen marcador seleccionable amplificable con el polinucleótido donde las células que contienen copias amplificadas del gen del marcador seleccionable, y por tanto copias adicionales del polinucleótido, se pueden seleccionar cultivando 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, por ejemplo, Sambrook *et al.*, 1989, *supra*).

Células huéspedes

La presente invención también se refiere a células huéspedes recombinantes, que comprenden un polinucleótido de la presente invención, que es ventajosamente usado en la producción recombinante de los polipéptidos. Un vector que comprende un polinucleótido de la presente invención se introduce en una célula huésped de modo que el vector es mantenido como un integrante cromosómico o como un vector extracromosómico que se duplica como se ha descrito anteriormente. El término “célula huésped” incluye cualquier descendiente de una célula madre que no es idéntico a la célula madre debido a mutaciones que ocurren durante la replicación. La elección de una célula huésped en gran parte depende del gen que codifica el polipéptido y su fuente.

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

Microorganismos unicelulares útiles son células bacterianas tales como bacterias gram positivas incluyendo, pero no limitado a, una célula de *Bacillus*, por ejemplo, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, y *Bacillus thuringiensis*, o una célula de *Streptomyces*, por ejemplo, *Streptomyces lividans* y *Streptomyces murinus*, o bacterias gram negativas tal como *E. coli* y *Pseudomonas sp.* En un aspecto preferido, la célula huésped bacteriana es una célula de *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus*, o *Bacillus subtilis*. En Otro aspecto preferido, la célula de *Bacillus* es una *Bacillus alcalofílica*.

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

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

En un aspecto preferido, la célula huésped es una célula fúngica. “Fungi” como se utiliza en este caso incluye el filo 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 el Oomycota (como se cita en Hawksworth *et al.*, 1995, *supra*, página 171) y todos los hongos mitospóricos (Hawksworth *et al.*, 1995, *supra*).

En un aspecto más preferido, la célula huésped fúngica es una célula de levadura. “Levadura” como se utiliza en este caso incluye levadura ascoesporógena (Endomycetales), levadura basidiosporógena, 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 biología y actividades de levadura (Skinner, F.A., Passmore, S.M., y Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Serie No. 9, 1980).

En un aspecto incluso más preferido, la célula huésped de la levadura es una célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, o de *Yarrowia*.

En un aspecto más preferido, la célula huésped de levadura es una *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, o *Saccharomyces oviformis*. En otro aspecto más preferido, la célula huésped de levadura es una célula láctica de *Kluyveromyces*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.

En otro aspecto más preferido, 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 generalmente 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. En cambio, 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 un aspecto incluso más preferido, la célula huésped filamentosa fúngica es una célula de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromyces*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trametes*, o *Trichoderma*.

En un aspecto más preferido, la célula huésped filamentosa fúngica es una célula de *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otro aspecto más preferido, 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 reiculatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochromum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, o *Fusarium venenatum*. En otro aspecto más preferido, la célula huésped filamentosa fúngica es una célula de *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis subvermisporea*, *Coprinus cinereus*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma hazianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesel*, o *Trichoderma viride*.

Células fúngicas se pueden transformar por un proceso implicando formación de protoplasto, 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* y *Trichoderma* 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 descritas 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, Volume 194, pp 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, que comprende: (a) cultivo de una célula, que en su forma de tipo salvaje es capaz de producir el polipéptido, bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido. En un aspecto preferido, la célula es del género *Thielavia*. En un aspecto más preferido, la célula es *Thielavia terrestris*.

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

ES 2 358 243 T3

La presente invención también se refiere a métodos para producir un polipéptido de la presente invención, que comprende: (a) cultivo de una célula huésped bajo condiciones propicias para la producción del polipéptido, donde la célula huésped comprende una secuencia de nucleótidos mutante que comprende al menos una mutación en la secuencia codificante del polipéptido maduro de SEC ID n°: 1, donde la secuencia de nucleótidos mutante codifica un polipéptido que comprende o consiste en el polipéptido maduro de SEC ID n°: 2, y (b) recuperación del polipéptido.

En un aspecto preferido, el polipéptido maduro de SEC ID n°: 2 es los aminoácidos 18 a 481.

En los métodos de producción de la presente invención, las células se cultivan en un medio nutritivo adecuado para la producción del polipéptido usando métodos bien conocidos en la técnica. Por ejemplo, la célula se puede cultivar por cultivo en matraz de agitación, y fermentación a gran escala o a pequeña escala (incluyendo fermentaciones continuas, discontinuas, en flujo discontinuo, o en estado sólido) en fermentadores de laboratorio o industriales realizados 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. Medios adecuados están disponibles de proveedores comerciales o se pueden preparar según composiciones publicadas (p. ej., en catálogos de la Colección Americana de Cultivos Tipo). Si el polipéptido se segrega en el medio nutritivo, el polipéptido se puede recuperar directamente del medio. Si el polipéptido no es segregado en el medio, se puede recuperar 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 uso de anticuerpos específicos, formación de un producto enzimático, o desaparición de un sustrato enzimático. Por ejemplo, un ensayo enzimático se puede utilizar para determinar la actividad del polipéptido como se describe en este caso.

El polipéptido resultante puede ser recuperado usando 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 se pueden purificar por una variedad de procedimientos conocidos en la técnica incluyendo, pero no limitados a, cromatografía (p. ej., por intercambio iónico, por afinidad, hidrofóbica, por cromatoenfoco, y por exclusión de tamaño), procedimientos electroforéticos (p. ej., isoelectroenfoco preparatorio), solubilidad diferencial (p. ej., precipitación de sulfato amónico), SDS-PAGE, o extracción (véase, por ejemplo, Protein Purification, J.-C. Janson y Lars Riden, editores, VCH Publishers, Nueva York, 1989) para obtener polipéptidos substancialmente puros.

Plantas

La presente invención también se refiere a una planta transgénica, parte de planta, o célula vegetal que ha sido transformada con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de celobiohidrolasa de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido se puede recuperar de la planta o parte de la planta. Alternativamente, la planta o parte de la planta conteniendo el polipéptido recombinante se puede utilizar como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorando 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 monocotiledóneas son hierbas, tales como poa de prados (poa pratense, Poa), hierba forrajera tal como Festuca, Lolium, cespel templado, tal como Agrostis, y cereales, por ejemplo, trigo, avena, centeno, cebada, arroz, sorgo, y maíz (com).

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

Ejemplos de partes de la planta son tallo, callo, hojas, raíz, frutos, semillas, y tubérculos al igual que los tejidos individuales que comprenden estas partes, por ejemplo, epidermis, mesofilo, parenquima, tejidos vasculares, meristemas. Compartimentos de células vegetales específicas, tales como cloroplastos, apoplastos, mitocondria, vacuolas, peroxisomas y citoplasma son también considerados como una parte de planta. Además, cualquier célula vegetal, cualquiera que sea el origen del tejido, se considera como una parte de la planta. Asimismo, partes de la planta tales como tejidos específicos y células aisladas para facilitar la utilización de la invención son también consideradas partes de la planta, por ejemplo, embriones, endospermas, aleurona y revestimientos de semillas.

También se incluye dentro del campo de la presente invención la progenie de tales plantas, partes de planta, y células vegetales.

La planta transgénica o célula vegetal que expresa un polipéptido de la presente invención se puede construir conforme a métodos conocidos en la técnica. En resumen, la planta o célula vegetal se construye porque incorpora

uno o más constructos de expresión que codifican un polipéptido de la presente invención en el genoma huésped de la planta o genoma del cloroplasto y propagan la planta modificada resultante o célula vegetal en una planta transgénica o célula vegetal.

5 El constructo de expresión es convenientemente un constructo de ácidos nucleicos que comprende un polinucleótido que codifica un polipéptido de la presente invención operativamente enlazado con secuencias apropiadas reguladoras 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 en las que el constructo de expresión ha sido integrado y secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (esto depende del método de introducción de ADN que será usado).

15 La elección de secuencias reguladoras, tales como secuencias del promotor y del terminador y opcionalmente secuencias señal o de tránsito es determinada, 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, específica de la fase o tejido, y el producto genético puede ser dirigido a un tejido específico o parte de la planta tal como semillas o hojas. Secuencias reguladoras son, por ejemplo, descritas por Tague *et al.*, 1988, *Plant Physiology* 86: 506.

20 Para la expresión constitutiva, se puede usar 35S-CaMV, la ubiquitina 1 de maíz, y el promotor de actina 1 de arroz (Franck *et al.*, 1980, *Cell* 21: 285-294, Christensen *et al.*, 1992, *Plant Mo. Biol.* 18: 675-689; Zhang *et al.*, 1991, *Plant Cell* 3: 1155-1165)). Promotores específicos de los órganos pueden ser, por ejemplo, un promotor de tejidos sumidero de almacenamiento tal como semillas, tubérculos de patata, y frutos (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 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 desconocido de la proteína de semilla 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 napA de proteína de almacenamiento de *Brassica napus*, o cualquier otro promotor específico de semilla conocido en la técnica, por ejemplo, como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de la hoja tal como el promotor *rbcS* de arroz o tomate (Kyojuka *et al.*, 1993, *Plant Physiology* 102: 991-1000, el promotor del gen de metiltransferasa de adenina del virus de chlorella (Mittra and Higgins, 1994, *Plant Molecular Biology* 26: 85-93), o el promotor de gen *aldP* de arroz (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), o un promotor inducible dañado tal como el promotor *pin2* de la patata (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588). Asimismo, el promotor puede ser inducible por tratamientos abióticos tales como temperatura, sequía, o alteraciones en salinidad o inducidas por sustancias aplicadas exógenamente que activan el promotor, por ejemplo, etanol, estrógenos, hormonas de planta tales como etileno, ácido abscísico, y ácido giberélico, y metales pesados.

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

45 El gen marcador seleccionable y cualquier otra parte del constructo de expresión se pueden elegir de aquellos disponibles en la técnica.

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

55 Actualmente, la transferencia de genes mediada por *Agrobacterium tumefaciens* es el método de elección para generar dicotiledóneas transgénicas (para una revisión, ver Hooykas and Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38) y puede también ser usada para transformar monocotiledóneas, aunque otros métodos de transformación son frecuentemente usados para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas es el bombardeo de partículas (oro microscópico o partículas de tungsteno revestidas con ADN transformante) de callos embrionarios o embriones en desarrollo (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). Un método alternativo para la transformación de monocotiledóneas se basa en la transformación de protoplastos como se describe por Omirulleh *et al.*, 1993, *Plant Molecular Biology* 21: 415-428.

65 Después de la transformación, los transformantes que tienen incorporados el constructo de expresión se seleccionan y regeneran en plantas enteras según métodos bien conocidos en la técnica. Frecuentemente el procedimiento de transformación se diseña para la eliminación selectiva de genes de selección bien durante la regeneración o en las siguientes generaciones usando, por ejemplo, cotransformación con dos constructos de T-ADN separados o escisión específica del sitio del gen de selección por una recombinasa específica.

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

5

Eliminación o reducción de la actividad de la celobiohidrolasa

La presente invención también se refiere a métodos para producir un mutante de una célula madre, que comprende interrumpir o eliminar una secuencia polinucleótida, o una parte de la misma, codificar un polipéptido de la presente invención, que resulta en la célula mutante que produce menos cantidad de polipéptido que la célula madre cuando se cultiva bajo las mismas condiciones.

La célula mutante se puede construir reduciendo o eliminando la expresión de una secuencia de nucleótidos que codifica un polipéptido de la presente invención usando métodos bien conocidos en la técnica, por ejemplo, inserciones, interrupciones, reemplazos, o deleciones. En un aspecto preferido, la secuencia de nucleótidos es inactivada. La secuencia de nucleótidos para ser modificada o inactivada puede ser, por ejemplo, la región codificante o una parte de la misma esencial para la actividad, o un elemento regulador requerido para la expresión de la región codificante. Un ejemplo de tal regulador o secuencia de control puede ser una secuencia promotora o una parte funcional de la misma, es decir, una parte que es suficiente para afectar la expresión de la secuencia de nucleótidos. Otras secuencias de control para posible modificación incluyen, pero no de forma limitativa, una secuencia líder, secuencia de poliadenilación, secuencia de propéptido, secuencia de péptido señal, terminador de transcripción, y activador transcripcional.

La modificación o inactivación de la secuencia de nucleótidos puede ser realizada sometiendo la célula madre a mutagénesis y seleccionando para células mutantes en donde la expresión de la secuencia de nucleótidos ha sido reducida o eliminada. La mutagénesis, que puede ser aleatoria o específica, puede ser realizada, por ejemplo, usando un agente mutagenizante químico o físico adecuado, usando un oligonucleótido adecuado, o sometiendo la secuencia de ADN a mutagénesis generada por PCR. Además, la mutagénesis puede ser realizada usando cualquier combinación de estos agentes mutagenizantes.

30

Ejemplos de un agente físico o químico mutagenizante adecuado para este fin incluyen irradiación de (UV) ultravioleta, hidroxilamina, N-metil-N'-nitro-N-nitrosoguanidina (MNNG), O-metil hidroxilamina, ácido nitroso, etil metano sulfonato (EMS), bisulfito de sodio, ácido fórmico, y análogos de nucleótidos.

Cuando tales agentes son usados, la mutagénesis es típicamente realizada incubando la célula madre para ser mutagenizada en presencia del agente mutagenizante de elección bajo condiciones adecuadas, y cribando y/o seleccionando para células mutantes que presentan expresión reducida o ninguna expresión del gen.

Modificación o inactivación de la secuencia de nucleótidos se puede realizar por introducción, sustitución, o eliminación de uno o más nucleótidos en el gen o un elemento regulador requerido para la transcripción o traducción del mismo. Por ejemplo, nucleótidos pueden ser insertados o eliminados para resultar en la introducción de un codón de terminación, eliminación del codón de inicio, o un cambio en el marco de lectura. Tal modificación o inactivación se puede realizar por mutagénesis dirigida o mutagénesis generada de PCR conforme a métodos conocidos en la técnica. Aunque, en principio, la modificación puede ser realizada *in vivo*, es decir, directamente en la célula que expresa la secuencia de nucleótidos que debe ser modificada, se prefiere que la modificación sea realizada *in vitro* como se ejemplifica más abajo.

Un ejemplo de una manera conveniente de eliminar o reducir la expresión de una secuencia de nucleótidos por una célula se basa en técnicas de sustitución de genes, deleción de genes, o interrupción de genes. Por ejemplo, en el método de interrupción génica, una secuencia de ácidos nucleicos correspondiente a la secuencia de nucleótidos endógena es mutagenizada *in vitro* para producir una secuencia de ácidos nucleicos defectuosa que es luego transformada en la célula madre para producir un gen defectuoso. Por recombinación homóloga, la secuencia de ácidos nucleicos defectuosa reemplaza la secuencia de nucleótidos endógena. Puede ser deseable que la secuencia de nucleótidos defectuosa también codifique un marcador que se puede utilizar para la selección de transformantes en donde la secuencia de nucleótidos ha sido modificada o destruida. En un aspecto particularmente preferido, la secuencia de nucleótidos se interrumpe con un marcador seleccionable tal como se describe aquí.

Alternativamente, la modificación o inactivación de la secuencia de nucleótidos se puede realizar por técnicas antisentido o ARNi usando una secuencia complementaria a la secuencia de nucleótidos. Más específicamente, la expresión de la secuencia de nucleótidos por una célula se puede reducir o eliminar introduciendo una secuencia complementaria a la secuencia de nucleótidos del gen que puede ser transcrito en la célula y es capaz de hibridar al ARNm producido en la célula. Bajo condiciones que permiten a la secuencia de nucleótidos antisentido complementaria para hibridar al ARNm, es así reducida o eliminada la cantidad de proteína traducida.

La presente invención además se refiere a una célula mutante de una célula madre que comprende una rotura o deleción de una secuencia de nucleótidos que codifica el polipéptido o una secuencia de control de la misma, que resulta en la célula mutante produciendo menos del polipéptido o ningún polipéptido en comparación con la célula madre.

65

ES 2 358 243 T3

Las células deficitarias de polipéptido mutantes así creadas son particularmente útiles como células huéspedes para la expresión de polipéptidos homólogos y/o heterólogos. Por lo tanto, la presente invención además se refiere a métodos para producir un polipéptido homólogo o heterólogo que comprende: (a) cultivar la célula mutante bajo condiciones propicias para la producción del polipéptido; y (b) recuperar el polipéptido. El término “polipéptidos heterólogos” es definido aquí como polipéptidos que no son nativos a la célula huésped, una proteína nativa en la que las modificaciones se han hecho para alterar la secuencia nativa, o una proteína nativa cuya expresión es alterada de forma cuantitativa como resultado de una manipulación de la célula huésped por técnicas de ADN recombinante.

En otro aspecto, la presente invención se refiere a un método para la producción de un producto proteínico esencialmente libre de actividad de celobiohidrolasa por fermentación de una célula que produce tanto un polipéptido de la presente invención como el producto proteínico de interés al añadir una cantidad eficaz de un agente capaz de inhibir la actividad de la celobiohidrolasa al caldo de fermentación antes, durante, o después de que la fermentación ha sido completada, recuperando el producto de interés del caldo de fermentación, y opcionalmente sometiendo el producto recuperado a purificación adicional.

En otro aspecto, la presente invención se refiere a un método para la producción de un producto proteínico esencialmente libre de actividad de celobiohidrolasa al cultivar la célula bajo condiciones que permiten la expresión del producto, sometiendo el caldo de cultivo resultante a un tratamiento combinado del pH y de la temperatura para reducir la actividad de celobiohidrolasa sustancialmente, y recuperando el producto del caldo de cultivo. Alternativamente, el tratamiento combinado del pH y de la temperatura se puede realizar en una preparación enzimática recuperada del caldo de cultivo. El tratamiento combinado del pH y de la temperatura puede opcionalmente ser usado en combinación con un tratamiento con un inhibidor de celobiohidrolasa.

Conforme a este aspecto de la invención, es posible eliminar al menos 60%, preferiblemente al menos 75%, más preferiblemente al menos 85%, todavía más preferiblemente al menos 95%, y de la forma más preferible al menos 99% de la actividad de celobiohidrolasa. La eliminación completa de la actividad de celobiohidrolasa puede ser obtenida usando este método.

El tratamiento combinado del pH y de la temperatura se realiza preferiblemente a un pH en la gama de 2-4 o 9-11 y una temperatura en la gama de al menos 70-80°C durante un periodo temporal suficiente para lograr el efecto deseado, donde típicamente, 30 a 60 minutos es suficiente.

Los métodos usados para el cultivo y purificación del producto de interés se pueden realizar por métodos conocidos en la técnica.

Los métodos de la presente invención para producir un producto esencialmente libre de celobiohidrolasa es de interés particular en la producción de polipéptidos eucarióticos, en particular proteínas fúngicas tales como enzimas. La enzima puede ser seleccionada de, por ejemplo, una enzima amilolítica, enzima lipolítica, enzima proteolítica, enzima celulítica, oxidorreductasa, o enzima degradadora de la pared celular vegetal. Ejemplos de tales enzimas incluyen una aminopeptidasa, amilasa, amiloglucosidasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxirribonucleasa, esterasa, galactosidasa, beta galactosidasa, glucoamilasa, gluco-oxidasa, glucosidasa, haloperoxidasa, hemicelulasa, invertasa, isomerasa, lacasa, ligasa, lipasa, liasa, manosidasa, oxidasa, enzima pectinolítica, peroxidasa, fitasa, fenoloxidasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transferasa, transglutaminasa, o xilanas. Las células deficitarias de celobiohidrolasa pueden también ser usadas para expresar proteínas heterólogas de interés farmacéutico tales como hormonas, factores de crecimiento, receptores, y similares.

Será entendido que el término “polipéptidos eucarióticos” incluye no sólo polipéptidos nativos, sino también aquellos polipéptidos, por ejemplo, enzimas, que han sido modificados por sustituciones, deleciones o adiciones de aminoácidos, u otras modificaciones de este tipo para mejorar la actividad, termostabilidad, tolerancia al pH y similares.

En otro aspecto, la presente invención se refiere a un producto proteínico esencialmente libre de actividad de celobiohidrolasa que se produce por un método de la presente invención.

Composiciones

La presente invención también se refiere a composiciones que comprenden un polipéptido de la presente invención. Preferiblemente, las composiciones se enriquecen con tal polipéptido. El término “enriquecer” indica que la actividad de celobiohidrolasa de la composición ha sido aumentada, por ejemplo, con un factor de enriquecimiento de al menos 1.1.

La composición puede comprender un polipéptido de la presente invención como el componente enzimático más importante, por ejemplo, una composición monocomponente. Alternativamente, la composición puede comprender actividades enzimáticas múltiples, tales como una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica,

ribonucleasa, transglutaminasa, o xilanas. La(s) enzima(s) adicional(es) puede(n) ser producida(s), por ejemplo, por un microorganismo del género *Aspergillus*, preferiblemente *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, o *Aspergillus oryzae*; *Fusarium*, preferiblemente *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 sarcochromum*, *Fusarium sulphureum*, *Fusarium toruloseum*, *Fusarium trichothecioides*, o *Fusarium venenatum*; *Humicola*, preferiblemente *Humicola insolens* o *Humicola lanuginosa*; o *Trichoderma*, preferiblemente *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, o *Trichoderma viride*.

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

Ejemplos de usos preferidos de las composiciones de polipéptido de la invención se dan más abajo. 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.

Usos

La presente invención está también dirigida a los métodos siguientes para usar polipéptidos que tienen actividad de celobiohidrolasa, o composiciones de los mismos.

Degradación de biomasa a monosacáridos, disacáridos, y polisacáridos

Los polipéptidos que tienen actividad de celobiohidrolasa y células huéspedes de la presente invención se pueden utilizar en la producción de monosacáridos, disacáridos, y polisacáridos como materias primas químicas o de fermentación de la biomasa para la producción de etanol, plástico, u otros productos o productos intermedios. Los polipéptidos que tienen actividad de celobiohidrolasa pueden estar en forma de un caldo de fermentación crudo con o sin las células eliminadas o en forma de una preparación enzimática semi-purificada o purificada. Alternativamente, una célula huésped de la presente invención se puede utilizar como una fuente del polipéptido en un proceso de fermentación con la biomasa.

La biomasa puede incluir, pero no está limitada a, recursos de madera, residuos sólidos municipales, papel usado, y residuos de cosecha (véase, por ejemplo, Wiseloge *et al.*, 1995, en Handbook on Bioethanol (Charles E. Wyman, editor), págs.105-118, Taylor & Francis, Washington D.C.; Wyman, 1994, Bioresource Technology 50: 3-16; Lynd, 1990, Applied Biochemistry and Biotechnology 24/25: 695-719; Mosier *et al.*, 1999, Recent Progress in Bioconversion of Lignocellulosics, en Advances in Biochemical Engineering/Biotechnology, T. Scheper, managing editor, volumen 65, pp.23-40, Springer-Verlag, New York).

El polisacárido predominante en la pared celular primaria de biomasa es celulosa, el segundo más abundante es hemicelulosa, y el tercero es pectina. La pared celular secundaria, producida después de que la célula ha detenido su crecimiento, también contiene polisacáridos y se refuerza 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 generalmente es polimorfa, la celulosa se encuentra en tejido vegetal principalmente como una matriz cristalina insoluble de cadenas de glucano paralelas. Las hemicelulosas normalmente unen el hidrógeno a celulosa, al igual que a otras hemicelulosas, lo que ayuda a estabilizar la matriz de la pared celular.

Tres clases más importantes de glicohidrolasas se utilizan para descomponer la biomasa celulósica:

(1) 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 insolubles y solubles.

(2) 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 celobiohidrolasas (1,4-beta-D-glucano celobiohidrolasas, EC 3,2,1,91), que liberan D-celobiosa de 1,4-beta-glucanos.

(3) 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 glucósidos.

Estas tres clases de enzimas trabajan juntas sinérgicamente dando como resultado la decristalización eficaz e hidrólisis de celulosa nativa de biomasa para producir azúcares reductores.

Los polipéptidos que tienen actividad de celobiohidrolasa de la presente invención se pueden utilizar conjuntamente con las enzimas mencionadas anteriormente para además degradar el componente de celulosa del sustrato de biomasa, (véase, por ejemplo, Brigham *et al.*, 1995, en Handbook on Bioethanol (Charles E. Wyman, editor), págs.119-141, Taylor & Francis, Washington D.C.; Lee, 1997, Journal of Biotechnology 56: 1-24).

Etanol se puede producir por degradación enzimática de biomasa y conversión de los sacáridos liberados a etanol. Esta especie de etanol es frecuentemente referida como bioetanol o biocombustible. Se puede usar como un aditivo de combustible o suplemento en mezclas inferiores a 1% y hasta 100% (un sustituto de combustible).

Composiciones de detergente

Los polipéptidos que tienen actividad de celobiohidrolasa de la presente invención se pueden adicionar a y así convertirse en un componente de una composición de detergente.

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

En un aspecto específico, la presente invención proporciona un aditivo de detergente comprendiendo la enzima de la invención. El aditivo de detergente al igual que la composición de detergente puede comprender una o varias otras enzimas tales como una proteasa, lipasa, cutinasa, una amilasa, carbohidrasa, celulasa, pectinasa, mananasa, arabinasa, galactanasa, xilanasa, oxidasa, por ejemplo, una lacasa, y/o peroxidasa.

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

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

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.

Enzimas proteásicas preferidas disponibles comercialmente incluyen Alcalase™, Savinase™, Primase™, Duralase™, Esperase™, y Kannase™ (Novozymes A/S), Maxatase™, Maxacal™, Maxapem™, Properase™, Purafect™, Purafect OxP™, FN2™, y FN3™ (Genencor Internacional Inc.).

Lipasa: lipasas adecuadas incluyen aquellas de origen fúngico o bacteriano. Mutantes creados genéticamente por proteínas o modificados químicamente están incluidos. Ejemplos de lipasas útiles incluyen lipasas de *Humicola* (sinónimo de *Thermomyces*), por ejemplo, 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*, por ejemplo, de *P. alcaligenes* o *P. pseudoalcaligenes* (EP 218 272), *P. cepada* (EP 331 376), *P. stutzeri* (GB 1,372,034), *P. fluoresceins*, cepa de *Pseudomonas sp.* SD 705 (WO 95/06720 y WO 96/27002), *P. wisconsinensis* (WO 96/12012), una lipasa de *Bacillus*, por ejemplo, de *B. subtilis* (Dartois *et al.*, 1993, Biochemica et Biophysica Acta, 1131: 253-360), *B. stearothermophilus* (JP 64/744992) o *B. pumilus* (WO 91/16422).

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

Enzimas de lipasa preferidas disponibles comercialmente incluyen Lipolase™ y Lipolase Ultra™ (Novozymes A/S).

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

ES 2 358 243 T3

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

5 Amilasas disponibles comercialmente son DuramylTM, TermamylTM, FungamylTM y BANTM (Novozymes A/S), RapidaseTM y PurastarTM (Genencor Internacional Inc.).

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

15 Celulasas especialmente adecuadas son las celulasas neutras o alcalinas que tienen beneficios para el cuidado del color. Ejemplos de tales celulasas son celulasas descritas en EP 0 495 257, EP 0 531 372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasa tales como aquellos descritos en WO 94/07998, EP 0 531 315, patente estadounidense n.º. 5,457,046, patente EEUU n.º. 5,686,593, patente estadounidense n.º. 5,763,254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

20 Celulasas disponibles comercialmente incluyen CelluzymeTM, y CarezymeTM (Novozymes A/S), ClazinaseTM, y Puradax HATM (Genencor Internacional Inc.), y KAC-500(B)TM (Kao Corporation).

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

Una peroxidasa disponible comercialmente incluye GuardzymeTM (Novozymes A/S).

30 La(s) enzima(s) detergente(s) se puede(n) incluir en una composición de detergente añadiendo aditivos separados conteniendo una o más enzimas, o añadiendo un aditivo combinado comprendiendo todas estas enzimas. Un aditivo de detergente de la invención, es decir, un aditivo separado o un aditivo combinado, puede ser formulado, por ejemplo, como un granulado, líquido, suspensión acuosa, etc. Formulaciones de aditivo de detergente preferidas son granulados, en particular granulados no polvorientos, líquidos, en particular líquidos estabilizados, o suspensiones acuosas.

35 Granulados no polvorientos pueden ser producidos, por ejemplo, como se describe en las patentes estadounidenses Nos. 4,106,991 y 4,661,452 y pueden opcionalmente ser revestidas por métodos conocidos en la técnica. Ejemplos de materiales de recubrimiento cerosos 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 grasos etoxilados en los que el alcohol contiene de 12 a 20 átomos de carbono y en los cuales 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 de recubrimiento que forman películas adecuados para aplicación por técnicas de lecho fluidificado se dan en GB 1483591. Preparaciones enzimáticas líquidas pueden, por ejemplo, ser estabilizadas añadiendo un poliol tal como propilenglicol, un azúcar o alcohol de azúcar, ácido láctico o ácido bórico según métodos establecidos. Enzimas protegidas se pueden preparar según el método descrito en EP 238,216.

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

50 La composición de detergente comprende uno o más tensioactivos, que puede ser no iónico incluyendo zwitteriónico y/o catiónico y/o aniónico y/o semipolar. Los tensioactivos son típicamente presentes a un nivel de 0,1% a 60% en peso.

55 Cuando se incluyen aquí el detergente normalmente contiene de aproximadamente 1% a aproximadamente 40% de un tensioactivo aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, alquil sulfato (sulfato de alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, éster metílico de ácido alfa-sulfo graso, ácido alquil o alquenilsuccínico, o jabón.

60 Cuando se incluye en éste el detergente normalmente contiene de aproximadamente 0,2% a aproximadamente 40% de un tensioactivo no iónico tal como alcohol etoxilato, nonilfenol, etoxilato alquilpoliglicósido, óxido de alquildimetilamina, monoetanolamida de ácidos grasos etoxilados, monoetanolamida de ácidos grasos, alquil polihidroxi amida de ácidos grasos, o derivados de N-acilo N-alquilo de glucosamina ("glucamidas").

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

ES 2 358 243 T3

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

5 El detergente puede contener un sistema blanqueante que puede comprender una fuente de H₂O₂ tal como perborato o percarbonato que se puede combinar con un activador blanqueante formador de perácido tal como tetraacetiletile-nodiamina o nonanoiloxibencenosulfonato. Alternativamente, el sistema blanqueante puede comprender peroxiácidos de, por ejemplo, del tipo amida, imida, o sulfona.

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

15 El detergente puede también contener otros ingredientes de detergentes convencionales tales como, por ejemplo, acondicionadores de tejido incluyendo arcillas, reforzadores de espuma, supresores de espuma, agentes anticorrosivos, agentes de suspensión de suciedad, agentes de anti-reposición de suciedad, tintes, bactericidas, blanqueadores ópticos, hidrotropos, inhibidores de decoloración, o perfumes.

20 En las composiciones de detergente cualquier enzima, en particular la enzima de la invención, se puede adicionar en una cantidad correspondiente a 0,01-100 mg de proteína enzimática por litro de solución de lavado, preferiblemente 0,05-5 mg de proteína enzimática por litro de solución de lavado, en particular 0,1-1 mg de proteína enzimática por litro de solución de lavado.

25 La enzima de la invención puede adicionalmente ser incorporada en las formulaciones de detergente descritas en WO 97/07202, que está incorporada en la presente como referencia.

30 *Otros Usos*

Los polipéptidos que tienen actividad de celobiohidrolasa de la presente invención pueden también ser usados en combinación con otras glicohidrolasas y enzimas relacionadas, como se describe en este caso, en el tratamiento de tejidos como agentes de biopulido y para la reducción de pelusa, frisado, modificación de textura, y lavado a la piedra (N.K. Lange, en P. Suominen, T. Reinikainen (Eds.), *Trichoderma reesei Cellulases and Other Hydrolases*, Foundation for Biotechnical and Industrial Fermentation Research, Helsinki, 1993, pp. 263-272). Además, los polipéptidos descritos pueden también ser usados en combinación con otras glicohidrolasas y enzimas relacionadas, como se describe en este caso, en tratamiento de madera para biopulpaje o descortezado, fabricación de papel para modificación de fibra, blanqueamiento, y reducción de costes de energía de refinación, tratamiento de aguas rápidas, importante para aguas residuales, reciclaje de fibras de reciclaje lignocelulósico tal como desentintado y tratamiento de fibras secundarias, y utilización de residuo de madera (S.D. Mansfield y A.R. Esteghlalian en S.D. Mansfield y J.N. Saddler (Eds.), *Applications of Enzymes to Lignocellulosics*, ACS Symposium Series 855, Washington, D.C., 2003, pp. 2-29).

45 *Péptido señal*

La presente invención también se refiere a constructos de ácidos nucleicos que comprenden un gen que codifica una proteína operativamente enlazada a una secuencia de nucleótidos que codifica un péptido señal que comprende o que consiste en los aminoácidos 1 a 17 de SEC ID n°: 2, donde el gen es extranjero a la secuencia de nucleótidos. En un aspecto preferido, la secuencia de nucleótidos comprende nucleótidos 1 a 51 de SEC ID n°: 1. En otro aspecto preferido, la secuencia de nucleótidos consiste en los nucleótidos 1 a 51 de SEC ID n°: 1.

La presente invención también se refiere a vectores de expresión recombinantes y células huéspedes recombinantes que comprenden tales constructos de ácidos nucleicos.

55 La presente invención también se refiere a métodos para producir una proteína que comprende (a) cultivo de tal célula huésped recombinante bajo condiciones adecuadas para la producción de la proteína; y (b) recuperación de la proteína.

60 La proteína puede ser nativa o heteróloga a una célula huésped. El término "proteína" no pretende referirse aquí a una longitud específica del producto codificado y, por lo tanto, incluye péptidos, oligopéptidos, y proteínas. El término "proteína" también incluye dos o más polipéptidos combinados para formar el producto codificado. Las proteínas también incluyen polipéptidos híbridos que comprenden una combinación de secuencias polipeptídicas completas o parciales obtenidas de al menos dos proteínas diferentes donde una o más pueden ser heterólogas o nativas a la célula huésped. Las proteínas además incluyen variaciones creadas genéticamente y alélicas de origen natural de las proteínas anteriormente mencionadas y proteínas híbridas.

ES 2 358 243 T3

Preferiblemente, la proteína es una hormona o variante de la misma, enzima, receptor o parte de lo mismo, anticuerpo o parte del mismo, o indicador. En un aspecto más preferido, la proteína es una oxidoreductasa, transferasa, hidrolasa, liasa, isomerasa, o ligasa. En un aspecto incluso más preferido, la proteína es una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, ciclodextrina glicosiltransferasa, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, invertasa, lacasa, otra lipasa, manosidasa, mutanasa, oxidasa, enzima pectinolítica, peroxidasa, fitasa, polifenoloxidasa, enzima proteolítica, ribonucleasa, transglutaminasa o xilanasas.

El gen se puede obtener de cualquier fuente, procariótica, eucariótica, u otra fuente.

Ejemplos

Material

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

Geles de SDS-PAGE, tampón de carga, y tampón de desplazamiento fue de Invitrogen/Novex (Carlsbad, CA). Tripsina modificada de calidad de secuenciación fue de Princeton Separations (Aldelphia, NJ). Cepas de proteína BioSafe Commassie Blue G250 fueron de BioRad Laboratories (Hercules, CA).

Cepas

Cepa *Aspergillus oryzae* Jal250 (WO 99/61651) fue usada para la expresión del polipéptido de *Thielavia terrestris* que tiene actividad de celobiohidrolasa. La cepa de *Thielavia terrestris* NRRL 8126 fue usada como la fuente del gen para la proteína de la familia 6A.

Medios

Placas PDA fueron compuestas por litro de 39 gramos de agar de dextrosa de patata.

Medio NNCYP fue compuesto por litro de 5.0 g de $\text{NH}_4 \text{NO}_3$, 0.5 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.3 g de CaCl_2 , 2.5 g de ácido cítrico, 1.0 g de bacto-peptona, 5.0 g de extracto de levadura, 1 ml de solución de metales traza COVE, y suficiente K_2HPO_4 para conseguir un pH final de aproximadamente 5.4.

Medio NNCYPmod fue compuesto por litro de 1.0 g de NaCl , 5.0 g de NH_4NO_3 , 0.2 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.2 g de CaCl_2 , 2.0 g de ácido cítrico, 1.0 g de bacto-peptona, 5.0 g de extracto de levadura, 1 ml de solución de metales traza COVE, y suficiente K_2HPO_4 para conseguir el pH final de aproximadamente 5.4.

La solución de metales traza COVE fue compuesta por litro de 0.04 g de $\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$, 0.4 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 1.2 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0.7 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 0.8 g de $\text{Na}_2 \text{MoO}_4 \cdot 2\text{H}_2\text{O}$, y 10 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$.

Placas de LB fueron compuestas por litro de 10 g de triptona, 5 g de extracto de levadura, 5 g de cloruro sódico, y 15 g de Bacto Agar.

Medio MDU2BP fue compuesto por litro de 45 g de maltosa, 1 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g de NaCl , 2 g de $\text{K}_2\text{H}_2\text{PO}_4$, 12 g de KH_2PO_4 , 2 g de urea, y 500 μl de metales traza AMG; el pH fue ajustado a 5.0 y luego esterilizado por filtro con una unidad de filtrado de 0.22 μm .

Metales traza AMG fueron compuestos por litro de 14.3 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 2.5 g de $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.5 g de $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, 13.8 g de $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 8.5 g de $\text{MnSO}_4 \cdot 7\text{H}_2\text{O}$, y 3 g de ácido cítrico.

Medio SOC fue compuesto por 2% de triptona, 0.5% de extracto de levadura, 10 mM de NaCl , 2.5 mM de KCl , 10 mM de MgCl_2 , y 10 mM de MgSO_4 ; esterilizado por autoclave y luego se añadió glucosa esterilizada por filtro a 20 mM.

Medio de congelación fue compuesto por 60% SOC y 40% glicerol.

2X placas de medio YT fueron compuestas por litro de 16 g de triptona, 10 g de extracto de levadura, 5 g de NaCl , y 15 g de Bacto agar, esterilizadas por autoclave.

Medio YP fue compuesto por litro de 10 g de extracto de levadura y 20 g de bacto-peptona (Difco).

Medio Inductor de Celulasa (CIM) fue compuesto por litro de 20 g de celulosa, 10 g de sólidos de maíz fermentado, 1.45 g de $(\text{NH}_4)_2 \text{SO}_4$, 2.08 g de KH_2PO_4 , 0.28 g de CaCl_2 , 0.42 g de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, y 0.42 ml de solución de metales traza.

Solución de metales traza fue compuesta por litro de 216 g de $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$, 58 g de $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 27 g de $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 10 g de $\text{CuSO}_4 \cdot \text{SH}_2\text{O}$, 2.4 g de H_3BO_3 , y 336 g de ácido cítrico.

Ejemplo 1

Secuenciación de péptidos por espectrometría de masas en serie de un polipéptido que codifica una celobiohidrolasa de familia 6 de Thielavia terrestris NRRL 8126

5

Tres tapones de agarosa de una placa fresca de *Thielavia terrestris* NRRL 8126 crecida en PDA fueron inoculados en 100 ml de medio NNCYP suplementado con 1.5% de glucosa e incubado durante 25 horas a 42°C y 200 r.p.m. En un agitador orbital. Cincuenta ml de este cultivo fueron usados para inocular 1.8 litro de medio NNCYP suplementado con 0,4% de glucosa y 52 g de celulosa en polvo por litro y fue incubado a 42°C. El pH fue controlado a 5.0 por la adición de 15% de hidróxido amónico o 5 N de ácido fosfórico, según lo necesitado.

10

Las fermentaciones fueron realizadas a 42°C con oxígeno mínimo disuelto a 25% a una corriente de aire de 1.0 mM y una agitación de 1100 r.p.m. El medio de alimentación fue entregado en un vaso de fermentación de 2 litros a 0 horas con un índice de alimentación de 6.0-8.0 g/hora durante 120 horas. Los cultivos agrupados fueron centrifugados a 3000 x g durante 10 minutos y el sobrenadante fue filtrado a través de una unidad de filtración desechable con un prefiltro de fibra de vidrio (Nalgene, Rochester NY). El filtrado fue enfriado a 4°C para su almacenamiento.

15

Una alícuota de 0.3 ml del filtrado fue precipitada con 10% de ácido tricloroacético (ATC) -80% de acetona durante 20 minutos en hielo. La suspensión fue centrifugada durante 10 minutos a 13,000 x g. El sobrenadante fue eliminado y el granulado de proteína restante fue enjuagado con acetona fría. El granulado de proteína fue disuelto en 30 μ l de 1X dodecil sulfato de litio (LDS) tampón de carga de SDS-PAGE con 50 mM (DTT) de ditioneitol y calentado a 80°C durante 10 minutos. Una muestra de 15 μ l fue separada por SDS-PAGE usando un 4-12% gel de gradiente de SDS-PAGE Bis-Tris NuPAGE de 7 cm y tampón de desplazamiento de ácido 2-(N-morfolino)etanosulfónico (MES). El SDS-PAGE fue realizado bajo condiciones de reducción según el protocolo recomendado por el fabricante (Invitrogen, Carlsbad, CA). El gel fue eliminado del casete y enjuagado 3 veces con agua desionizada durante al menos 5 minutos cada vez y teñido con tinte de Bio-Safe Coomassie (BioRad Laboratories, Hercules, CA) durante 1 hora seguido de decoloración con agua doblemente destilado durante más de 30 minutos. Bandas proteicas observadas a aproximadamente 66 kDa y 73 kDa fueron cortadas y reducidas con 50 μ l de 10 mM DTT en 100 mM de bicarbonato de amonio durante 30 minutos. Después de la reducción, los pedazos de gel fueron alquilados con 50 μ l de 55 mM de yodoacetamida en 100 mM de bicarbonato de amonio durante 20 minutos. Los pedazos de gel secos fueron dejados rehidratarse en una solución de digestión de tripsina (6 ng/ μ l de tripsina de calidad de secuenciación en 50 mM bicarbonato de amonio) durante 30 minutos a temperatura ambiente, seguido de una digestión de 8 horas a 40°C. Cada uno de los pasos de reacción descritos fue seguido de numerosos lavados y prelavados con las soluciones deseadas. Cincuenta μ l de acetonitrilo fue usado para deshidratar los pedazos de gel entre reacciones y fueron secados por aire entre etapas. Péptidos fueron extraídos dos veces con 1% de ácido fórmico/2% de acetonitrilo en agua de calidad de HPLC durante 30 minutos. Soluciones de extracción peptídicas fueron transferidas a una placa de microtitulación de tipo PCR de 96 pocillos que fue enfriada a 10-15°C. Placas de microtitulación conteniendo las soluciones de péptido recuperadas fueron selladas para prevenir la evaporación y almacenadas a 4°C hasta que el análisis de espectrometría de masas pudiera ser realizado.

30

Para secuenciación peptídica de novo por espectrometría de masas en serie, un Q-Tof microTM, un espectrómetro de masas cuadrupolo de tiempo-de-vuelo híbrido ortogonal (Waters Micromass[®] MS Technologies, Milford, MA) fue usado para análisis LC/MS/MS. El Q-Tof microTM fue ajustado con un sistema UltimateTM capilar y de HPLC de nanoflujo que fue acoplado con un micro automuestreador FAMOS y un dispositivo de conmutación de columna Switchos II (LCPackings, San Francisco) para concentrar y desalar muestras. Las muestras fueron cargadas sobre una columna de seguridad (300 μ m ID x 5 cm, C18 pepmap) ajustadas en el bucle de inyección y lavadas con 0,1% de ácido fórmico en agua a 40 μ l/minuto durante 2 minutos usando la bomba Switchos II. Péptidos fueron separados en una columna capilar con nanoflujo de 75 μ m ID x 15 cm, C18, 3 μ m, 100 Å PepMapTM (LC Packings, San Francisco) a una velocidad de flujo de 175 nl/minuto de un flujo de división de 175 μ l/minuto usando un calibrador NAN-75 (Dionex, Sunnyvale, CA). Un gradiente de elución escalonado de 5% a 80% de acetonitrilo en 0,1% de ácido fórmico fue aplicado sobre un intervalo de 45 minutos. El eluyente de columna fue controlado a 215 nm e introducido en el Q-Tof microTM a través de una fuente de iones de electrospray ajustada con la interfaz de nanopulverización. El Q-Tof microTM es completamente controlado por microprocesador usando el software de MasslynxTM versión 3.5 (Waters Micromass[®] MS Technologies, Milford, MA). Los datos fueron adquiridos en modo de exploración y de una gama de masa de m/z 400 a 1990 con los criterios de conmutación para MS a MS/MS para incluir una intensidad iónica mayor que 10.0 cuentas por segundo y estados de carga de +2; +3, y +4. Espectros de análisis de hasta 4 especies de coelución con un tiempo de exploración de 1.9 segundos y tiempo de interexploración de 0.1 segundos pudieron ser obtenidos. Un voltaje de cono de 65 voltios fue típicamente usado y la energía de colisión fue programada para ser variada según la masa y estado de carga del péptido de elución y en la gama de 10-60 voltios. Los espectros adquiridos fueron combinados, alisados y centrados en una forma automatizada y una lista de valores máximos fue generada. Esta lista de valores máximos fue buscada contra las bases de datos privados y públicas seleccionadas usando software de ProteinLynxTM Global Server 1.1 (Waters Micromass[®] MS Technologies, Milford, MA). Resultados de las búsquedas con ProteinLynxTM fueron evaluados y proteínas no identificadas fueron analizadas además evaluando los espectros MS/MS de cada ión de interés y secuencia de novo fue determinada identificando las series de iones y y b y diferencias de masa correspondientes al aminoácido apropiado.

40

Secuencias peptídicas obtenidas de la secuenciación de novo por espectrometría de masas fueron obtenidas de varios iones cargados múltiplemente para la banda de gel de polipéptido de aproximadamente 73 kDa. Un ión peptídico

ES 2 358 243 T3

tríptico doblemente cargado de la secuencia 574.25 m/z fue determinado para ser Val-Pro-Ser-[Phe o Met oxidada]-Gln-Trp-[Ile o Leu]-Asp-Arg (aminoácidos 163 a 171 de SEC ID n°: 2). Un segundo ión peptídico tríptico doblemente cargado de 982.42 fue determinado para ser Gly-Ala-Asn-Pro-Pro-Tyr-Ala-Gly-[Ile o Leu]-[Phe o Met oxidada]-Val-Val-Tyr-Asp-[Leu o Ile]-pro-Asp-Arg (aminoácidos 193 a 210 de SEC ID n°: 2). Un enlace proteínico de 60 kDa fue también digerido en gel con tripsina y los péptidos resultantes fueron recuperados. Un ión peptídico doblemente cargado de 1159.02 fue secuenciado de novo determinado para ser [Ile o Leu]-[Ile o Leu]-[Phe o Met oxidada]-Val-[Ile o Leu]-Glu-pro-Asp-Ser-[Leu o Ile]-Ala-Asn-Met-Val-Thr-Asn-[Leu o Ile]-Asn-Val-Ala-Lys (aminoácidos 250 a 270 de SEC ID n°: 2).

10 Ejemplo 2

Construcción de biblioteca de ADNc con etiquetas de secuencias expresadas (EST)

15 Una alícuota de dos ml de un cultivo líquido de 24 horas (50 ml de NNCYPmod más 1% glucosa en un matraz de 250 ml, 45°C, 200 r.p.m.) de *Thielavia terrestris* NRRL 8126 fue usada para sembrar un matraz de 500 ml conteniendo 100 ml de medio NNCYPmod suplementado con 2% Sigmacell-20. El cultivo fue incubado a 45°C, 200 r.p.m. durante 3 días. Los micelios fueron cosechados por filtración a través de un embudo de Buchner con un prefiltro de fibra de vidrio (Nalgene, Rochester, NY) lavados dos veces con 10 mM de Tris-HCl-1 mM de EDTA pH 8 (TE), y congelados rápido en nitrógeno líquido.

20 ARN total fue aislado usando el método siguiente. Micelios congelados de *Thielavia terrestris* NRRL 8126 fueron molidos en un molinillo de café eléctrico. El material molido fue mezclado 1:1 v/v con 20 ml de Fenazol (Ambion, Inc., Austin, TX) en un tubo Falcon de 50 ml. Una vez que los micelios fueron suspendidos, ellos fueron extraídos con cloroformo y tres veces con una mezcla de fenol-cloroformo-isoamil alcohol 25:24:1 v/v/v. De la fase acuosa resultante, el ARN fue precipitado añadiendo 1/10 volúmenes de 3 M de acetato sódico pH 5.2 y 1.25 volúmenes de isopropanol. El ARN precipitado fue recuperado por centrifugado a 12,000 x g durante 30 minutos a 4°C. El granulado final fue lavado con 70% de etanol frío, secado al aire, y resuspendido en 500 ml de dietilpirocarbonato tratado con agua (agua DEPC).

30 La calidad y cantidad del ARN purificado fue evaluada con un Agilent Bioanalyzer 2100 (Agilent Technologies, Inc., Palo Alto, CA). ARNm poliadenilado fue aislado de 360 µg de ARN total con la ayuda de un equipo de Poly(A) Purist Magnetic (Ambion, Inc., Austin, TX) según las instrucciones del fabricante.

35 Para crear la biblioteca de ADNc, se empleó un equipo CloneMiner™ (Invitrogen, Carlsbad, CA) para construir una biblioteca direccional que no requiera el uso de clonación de enzima de restricción, así reduciendo el número de clones numéricos y y sego del tamaño.

40 Para asegurar la síntesis exitosa del ADNc, dos reacciones fueron realizadas en paralelo con dos concentraciones diferentes de ARNm (2.2 y 4.4 µg de poli(A)⁺ ARNm). Las muestras de ARNm fueron mezcladas con un cebador Biotin-attB2-Oligo(dt) (equipo de CloneMiner™, Invitrogen, Carlsbad, CA), 1X tampón de primera cadena (Invitrogen, Carlsbad, CA), 2 µl de 0.1 M (DTT) de ditiotreititol, 10 mM de mezcla de dATP, dTTP, dGTP, y dCTP, y agua a un volumen final de 18 y 16 µl, respectivamente.

45 Las mezclas de reacción fueron mezcladas cuidadosamente y luego 2 y 4 µl de transcriptasa inversa SuperScript™ fueron incubados y adicionados a 45°C durante 60 minutos para sintetizar la primera cadena complementaria. Para la síntesis de la segunda cadena, a cada reacción de la primera cadena se añadió 30 µl de 5X tampón de la segunda cadena (Invitrogen, Carlsbad, CA), 3 µl de una mezcla de 10 mM de dATP, dTTP, dGTP, y dCTP, 10 unidades de ADN-ligasa de *E. coli*, 40 unidades de ADN polimerasa I de *E. coli*, y 2 unidades de RNasa H de *E. coli* en un volumen total de 150 µl. Las mezclas fueron luego incubadas a 16°C durante dos horas. Después de la incubación de dos horas 2 µl de T4-ADN polimerasa fueron adicionados a cada reacción e incubados a 16°C durante 5 minutos para crear un ADNc de extremo romo. Las reacciones de ADNc fueron extraídas con una mezcla de fenol-cloroformo-isoamil alcohol 25:24:1 v/v/v y precipitadas en presencia de 20 µg de glicógeno, 120 µl de acetato amónico 5 M, y 660 µl de etanol. Después del centrifugado a 12,000 x g durante 30 minutos a 4°C los gránulos de ADNc fueron lavados con 70% de etanol frío, secados al vacío durante 2-3 minutos, y resuspendidos en 18 µl de DEPC-agua. A cada muestra de ADNc resuspendida se añadió 10 µl de 5X tampón adaptado, 10 µg de adaptador attB1 (provisto con el equipo; secuencia bicatenaria mostrada más abajo), 7 µl de 0.1 M de DTT, y 5 unidades de T4 ADN-ligasa.

5' -TCGTCGGGGACAACCTTTGTACAAAAAAGTTGG-3' (SEC ID n°: 3)

60 3' -CCCCTGTTGAAACATGTTTTTCAACCP-5' (SEC ID n°: 4)

65 Reacciones de ligadura fueron incubadas durante toda la noche a 16°C. Adaptadores en exceso fueron eliminados por cromatografía de exclusión de tamaño en 1 ml de resina Sphacryl™ S-500 HR (Amersham Biosciences, Piscataway, NJ). Fracciones de columna fueron recogidas según las instrucciones del equipo y las fracciones 3 a 14 fueron analizadas con un Agilent Bioanalyzer para determinar la fracción en la que los adaptadores attB1 comenzaron a eluir. Este análisis mostró que los adaptadores comenzaron a eluir alrededor de la fracción 10 u 11. Para la primera biblioteca, se agruparon las fracciones 6 a 11 y para la segunda biblioteca, se agruparon las fracciones 4-11.

ES 2 358 243 T3

La clonación del ADNc fue realizada por recombinación de ADN homóloga según el protocolo de Gateway System (Invitrogen, Carlsbad, CA) usando BP Clonase™ (Invitrogen, Carlsbad, CA) como la recombinasa. Cada reacción de recombinación de BP clonase™ contenía aproximadamente 70 ng de ADNc flanqueado por attB, 250 ng de pDONR™222, 2 µl de 5X Tampón de BP clonase™, 2 µl de TE, y 3 µl de BP Clonase™. Reacciones de recombinación fueron incubadas a 25°C durante toda la noche.

Reacciones de recombinación de BP inactivada por calor fueron luego divididas en 6 partes alícuotas y sometidas a electroporación en células electrocompetentes de ElectroMax™ DH10B (Invitrogen, Carlsbad, CA) en un BioRad Gene Pulser II (BioRad, Hercules, CA) con los siguientes parámetros: voltaje: 2.0 kV, resistencia: 200 capacidad 25 µF. Las células sometidas a electroporación fueron resuspendidas en 1 ml de medio SOC e incubadas a 37°C durante 60 minutos con agitación constante (200 r.p.m.). Después del periodo de incubación, las células transformadas fueron agrupadas y mezcladas 1:1 con medio de congelación. Una alícuota de 200 µl fue eliminada para titulación de biblioteca y luego el resto de cada biblioteca fue dividido en partes alícuotas en crioviales de 1.8 ml (Wheaton Science Products, Millville, NJ) y almacenada congelada a -80°C.

Cuatro diluciones en serie de cada biblioteca fueron preparadas: 1/100, 1/1000, 1/10⁴, 1/10⁵. De cada dilución 100 µl fue colocada en placas LB de 150 mm suplementadas con 50 µg de canamicina por ml e incubadas a 37°C durante toda la noche. El número de colonias en cada placa de dilución fue contado y usado para calcular el número total de transformantes en cada biblioteca.

La primera biblioteca mostró tener 5.4 millones de clones independientes y la segunda biblioteca mostró tener 9 millones de clones independientes.

Ejemplo 3

Preparación de molde y secuenciación nucleótida de clones de ADNc

Partes alícuotas de ambas bibliotecas fueron mezcladas y colocadas en placas sobre 25 x 25 cm de placas LB suplementadas con 50 µg de canamicina por ml. Colonias individuales fueron alineadas en placas de 96 pocillos conteniendo 100 µl de LB suplementado con 50 µg de canamicina por ml con la ayuda de un Robot Genetix QPix (Genetix Inc., Boston, MA). Cuarenta y cinco placas de 96 pocillos fueron obtenidas para un total de 4320 clones individuales. Las placas fueron incubadas durante toda la noche a 37°C con agitación a 200 r.p.m. Tras la incubación, 100 µl de 50% de glicerol estéril se añadió a cada pocillo. Los transformantes fueron replicados con la ayuda de una herramienta de 96 puntas (Boekel, Feasterville, PA) en placas secundarias de microcultivo de 96 pocillos de plato profundo (Advanced Genetic Technologies Corporation, Gaithersburg, MD) conteniendo 1 ml de Magnificent Broth™ (MacConnell Research, San Diego, CA) suplementado con 50 µg de canamicina por ml en cada pocillo. Las placas de microtitulación primarias fueron almacenadas congeladas a -80°C. Las placas de plato profundo secundarias fueron incubadas a 37°C durante toda la noche con agitación vigorosa (300 r.p.m.) en un agitador giratorio. Para prevenir vertido y contaminación cruzada, y para permitir aireación suficiente, cada placa de cultivo secundaria fue cubierta con un relleno de polipropileno (Advanced Genetic Technologies Corporation, Gaithersburg, MD) y una tapadera de plato de microtitulación de plástico. ADN plásmido fue preparado con un MWG Robot-Smart 384 (MWG Biotech Inc., High Point, NC) y Montage Plasmid Miniprep Kits (Millipore, Billerica, MA).

Reacciones de secuenciación fueron realizadas usando Big-Dye™ (Applied Biosystems, Inc., Foster City, CA) química de terminador (Giesecke *et al.*, 1992, Journal of Virology Methods 38: 47-60) y un cebador de secuenciación M 13 Directo (-20):

5'-GTAAAACGACGGCCAG-3'

(SEC ID n°: 5)

Las reacciones de secuenciación fueron realizadas en un formato de 384 pocillos con un Robot-Smart 384 (MWG Biotech Inc., High Point, NC) al igual que la eliminación de terminador con Millipore MultiScreen Seq384 Sequencing Clean-up Kits (Millipore, Billerica, MA). Las reacciones contenían 6 µl de ADN plásmido y 4 µl de mezcla maestra de secuenciación conteniendo 1 µl de 5X tampón de secuenciación (Millipore, Billerica, MA), 1 µl de Terminador de Byg-Dye™ (Applied Biosystems, Inc., Foster City, CA), 1.6 pmoles de cebador M13 Directo, y 1 µl de agua. La secuenciación del ADN single-pass fue realizada con un Secuenciador de ADN automatizado ABI PRISM Modelo 3700 (Applied Biosystems, Foster City, CA).

Ejemplo 4

Análisis de datos de secuencia de ADN de clones de ADNc

Llamada de base, atribución de valor de calidad, y ajuste de vector fueron realizados con la asistencia del software PHRED/PHRAP (Universidad de Washington, Seattle, WA). Análisis de agrupamiento del ESTs fue realizada con un Parcel Transcript Assembler v. 2.6.2. (Paracel, Inc., Pasadena, CA). Análisis de la agrupación EST indicó la presencia de 395 agrupaciones independientes.

ES 2 358 243 T3

Análisis de homología secuencial de las secuencias EST ensambladas contra las bases de datos PIR y ERDBP fue realizado con el programa Blastx (Altschul *et al.*, 1990, J. Mol. Biol. 215:403-410) en una agrupación Linux de 32 nodos (Paracel, Inc., Pasadena, CA) usando la matriz BLOSUM 62 (Henikoff, 1992, Proc. Natl. Acad. Sci. EEUU 89: 10915-10919) De estos, 246 tuvieron coincidencias con genes conocidos sea en las bases de datos de proteína privados o públicos y 149 no tuvieron ninguna coincidencia significativa contra estas bases de datos. Entre estos 246 genes, 13 tuvieron coincidencias contra homólogos bien caracterizados de genes de glicosil hidrolasa.

Ejemplo 5

Identificación de clones de ADNc que codifican una celobiohidrolasa de familia 6 (Cel6A)

Un clon de ADNc que codifica una celobiohidrolasa de familia 6 (Cel6A) fue inicialmente identificada por su homología a una proteína de familia 6A de *Humicola insolens* (NR 34811679). El análisis indicó que las dos proteínas fueron 50% idénticas al nivel de proteína sobre una extensión de 120 aminoácidos (360 pares de bases). Después este clon de identificación inicial, Tter11C9 fue recuperado de la placa de materia prima original congelada y alineada sobre una placa de LB suplementada con 50 µg de canamicina por ml. La placa fue incubada durante toda la noche a 37°C y al día siguiente una única colonia de la placa fue usada para inocular 3 ml de LB suplementado con 150 µg de canamicina por ml. El cultivo líquido fue incubado durante toda la noche a 37°C y ADN plásmido fue preparado con un Qiagen BioRobot 9600 (QIAGEN, Inc., Valencia, CA). ADN plásmido del clon Tter11C9 fue secuenciado otra vez con química de terminador Big-Dye™ como se ha descrito anteriormente, usando el cebador M13 directo y un cebador Poly-T mostrado debajo de la secuencia del extremo 3' del clon.

5'-TTTTTTTTTTTTTTTTTTTTTTTTTTVN-3'

(SEC ID nº: 6).

Donde V=G, A, C y N= G, A, C, T

El análisis de homología Blastx de la información de nueva secuencia indicó que el extremo 3 prima tuvo una identidad muy alta a la proteína de Familia 6 de *Talaromyces emersonii*. Estas proteínas fueron 97% idénticas sobre una extensión de 39 aminoácidos. También, el análisis del extremo 5 prima del clon Tter11C9 con el programa Interproscan (Zdobnov y Apweiler, 2001 Bioinformatics 17: 847-8.) mostró que el gen codificado por el clon Tter11C9 contenía la firma de secuencia de las proteínas de familia 6 de glicosil hidrolasa. Esta firma de secuencia conocida como el modelo PS00655 de Prosite (Sigrist *et al.*, 2002, Brief Bioinform. 3: 265-274) fue encontrada 202 aminoácidos desde el aminoácido de inicio metionina confirmando que el clon Tter11C9 codifica un gen de familia 6 de glicosil hidrolasa de *Thielavia terrestris*.

La secuencia de ADNc (SEC ID nº: 1) y la secuencia de aminoácidos deducida (SEC ID nº: 2) se muestran en las Figuras 1A y 1B. El clon de ADNc codifica un polipéptido de 481 aminoácidos. El contenido %G+C del clon de ADNc del gen es 66.9% y el de la región de codificación de proteína madura (nucleótidos 52 a 1443 de SEC ID nº: 1) es 66.7%. Usando el programa de software SignalP (Nielsen *et al.*, 1997, Protein Engineering 10: 1-6), un péptido señal de 17 residuos fue previsto. La proteína madura prevista contiene 464 aminoácidos con una masa molecular de 49.3 kDa.

Una alineación comparativa de secuencias de familia 6 de glicosil hidrolasa fue determinada usando el método de Clustal W (Higgins, 1989, *supra*) usando el Software de LASERGENE™ MEGALIGN™ (DNASTAR, Inc., Madison, WI) con una tabla de peso de residuo PAM250 y los siguientes parámetros de alineamiento múltiples: Penalización de gap de 10 y penalización de longitud de gap de 10. Los parámetros de alineación de parejas fueron Ktuple=1, penalización de espacio=3, ventanas=5, y diagonales=5. El alineamiento mostró que la secuencia de aminoácidos deducida gen de celobiohidrolasa Cel6A de *Thielavia terrestris* comparte un 74% de identidad a la secuencia de aminoácidos deducida del gen de celobiohidrolasa de *Chaetomium termofilum* (Geneseq ADP84824).

Una vez la identidad del clon Tter11C9 conteniendo pTter11C9 fue confirmada, una alícuota de 0.5 µl de ADN plásmido de este clon, que fue redesignado pTter6A, fue transferido en un frasco de células TOP10 de *E. coli* (Invitrogen, Carlsbad, CA), suavemente mezclado, y se incubó en hielo durante 10 minutos. Las células fueron luego sometidas a choque térmico a 42°C durante 30 segundos e incubadas nuevamente en hielo durante 2 minutos. Las células fueron resuspendidas en 250 µl de medio SOC e incubadas a 37°C durante 60 minutos con agitación constante (200 r.p.m.). Después del periodo de incubación, dos partes alícuotas de 30 µl fueron colocadas en placas sobre placas LB suplementadas con 50 µg de canamicina por ml e incubadas durante toda la noche a 37°C. Al día siguiente una única colonia fue escogida y alineada sobre tres crioviales de 1.8 ml conteniendo aproximadamente 1.5 mls de agarosa LB suplementada con 50 µg de canamicina por ml. Los viales fueron sellados con PetriSeal™ (Diversified Biotech, Boston MA) y depositados con la Colección de Patentes de Cultivos del Servicio de Investigación para la Agricultura, Centro de Investigación Regional del Norte, 1815 University Street, Peoria, Illinois, 61604, como NRRL B-30802, con fecha de depósito 17 de diciembre de 2004.

ES 2 358 243 T3

Ejemplo 6

Construcción del vector de expresión pAILo2

5 Vector de expresión pAILo2 fue construido modificando pBANe6 (Patente norteamericana n°. 6,461,837), que comprende un híbrido de los promotores de los genes para alfa-amilasa neutra de *Aspergillus niger* y triosa fosfato isomerasa de *Aspergillus oryzae* (promotor NA2-tpi), secuencia del terminador de amiloglicosidasa de *Aspergillus niger* (terminador AMG), y gen de acetamidasa de *Aspergillus nidulans* (amdS). Todos los pasos de mutagénesis fueron verificados por secuenciación usando Química de terminador de Big-Dye™ según se describe. La modificación de pBANe6 fue realizada eliminando primero tres sitios de restricción NcoI en las posiciones 2051, 2722, y 3397 pares de bases del marcador de selección amdS por mutagénesis sitio-dirigida. Todos los cambios fueron diseñados para ser “silenciosos” dejando la secuencia de proteína real del producto genético *amdS* invariada. La eliminación de estos tres sitios fue realizada simultáneamente con un GeneEditor™ *in vitro* Site-Directed Mutagenesis Kit (Promega, Madison, WI) según las instrucciones del fabricante usando los cebadores siguientes (nucleótido subrayado representa la base cambiada):

AMDS3NcoMut (2050):

5'-GTGCCCCATGATACGCCTCCGG-3' (SEC ID n°: 7)

AMDS2NcoMut (2721):

5'-GAGTCGTATTTCCAAGGCTCCTGACC-3' (SEC ID n°: 8)

AMDS1NcoMut (3396):

5'-GGAGGCCATGAAGTGGACCAACGG-3' (SEC ID n°: 9)

Un plásmido que comprende los tres cambios de secuencia previstos fue luego sometido a mutagénesis dirigida, usando un Equipo de mutagénesis dirigida de QuickChange™ (Stratagene, la Jolla, CA), para eliminar el sitio NcoI de restricción al final del terminador AMG en la posición 1643. Los siguientes cebadores (nucleótido subrayado representa la base cambiada) fueron usados para mutagénesis:

Cebador superior para mutagenizar la secuencia del terminador AMG:

5'-CACCGTGAAAGCCATGCTCTTTCCTTCGTGTAGAAGACCAGACAG-3' (SEC ID n°: 10)

Cebador inferior para mutagenizar la secuencia del terminador AMG:

5'-CTGGTCTTCTACACGAAGGAAAGAGCATGGCTTTCACGGTGTCTG-3' (SEC ID n°: 11)

El último paso en la modificación de pBANe6 fue la adición de un sitio de restricción NcoI nuevo al principio del poliligador usando un equipo de mutagénesis dirigida de quickChange™ y los siguientes cebadores (nucleótidos subrayados representan las bases cambiadas) para producir pAILo1 (figura 2).

Cebador superior para mutagenizar el promotor NA2-tpi:

5'-CTATATACACAACTGGATTTACCATGGGCCCGCGCCGCAGATC-3' (SEC ID n°: 12)

Cebador inferior para mutagenizar el promotor NA2-tpi:

5'-GATCTGCGGCCCGCGGGCCCATGGTAAATCCAGTTGTGTATATAG-3' (SEC ID n°: 13)

El gen *amdS* de pAILo2 fue intercambiado con el gen *pyrG* de *Aspergillus nidulans*. El plásmido pBANe10 (figura 3) fue usado como una fuente para el gen *pyrG* como un marcador de selección. Análisis de la secuencia de pBANe10 mostró que el marcador *pyrG* fue contenido dentro de un fragmento de restricción *NsiI* y no contiene sitios de restricción *NcoI* ni *PacI*. Ya que el *amdS* es también flanqueado por sitios de restricción *NsiI* la estrategia para intercambiar el marcador de selección fue un simple intercambio de fragmentos de restricción *NsiI*. ADN plásmido de pAILo2 y pBANe10 fueron digeridos con la enzima de restricción *NsiI* y los productos fueron purificados por electroforesis en gel de agarosa. El fragmento *NsiI* de pBANe10 conteniendo el gen *pyrG* fue ligado al esqueleto de pAILo2 para reem-

ES 2 358 243 T3

plazar el fragmento de ADN de NsiI original conteniendo el gen *amdS*. Clones recombinantes fueron analizados por el digirido de restricción para determinar que estos tienen el inserto correcto y también su orientación. Un clon con el gen *pyrG* transcrito en el sentido contrario a las agujas del reloj fue seleccionado. El plásmido nuevo fue designado pAILo2 (figura 4).

Ejemplo 7

Clonación del gen de celobiohidrolasa de la familia GH6A en un vector de expresión de Aspergillus oryzae

Dos cebadores oligonucleótidos sintéticos, mostrados abajo, fueron diseñados para amplificar por PCR el marco de lectura abierto en toda su longitud de *Thielavia terrestris* EST Tter11C9 que codifica una celobiohidrolasa de familia GH6A. Un In-Fusion Cloning Kit (BD Biosciences, Palo Alto, CA) fue usado para clonar el fragmento directamente en pAILo2.

Cebador In-fusion Directo:

5'-ACTGGATTACCATGGCTCAGAAGCTCCTTCT-3' (SEC ID n°: 14)

Cebador In-fusion inverso:

5'-TCACCTCTAGTTAATTAAGGGCGGGTTGGCGT-3' (SEC ID n°: 15)

Las letras en negrita representan la secuencia codificante. La secuencia restante contiene identidad de secuencia comparada con los sitios de inserción de pAILo2.

Cinuenta picomoles de cada uno de los cebadores anteriores fueron usados en una reacción PCR conteniendo 50 ng de pTter11C9 ADN, 1X Pfx tampón de amplificación (Invitrogen, Carlsbad, CA), 6 μ l de una mezcla de dATP 10 mM, dTTP, dGTP, y dCTP, 2.5 unidades de Platinum Pfx ADN-polimerasa (Invitrogen, Carlsbad, CA), 1 μ l de MgSO₄ 50 mM, y 5 μ l de 10X pCRx solución intensificadora (Invitrogen, Carlsbad, CA) en un volumen final de 50 μ l. Un Eppendorf Mastercycler 5333 (Eppendorf Scientific, Inc., Westbury, NY) fue usado para amplificar el fragmento programado para un ciclo a 98°C durante 2 minutos; y 35 ciclos cada uno a 94°C durante 30 segundos, 65°C durante 30 segundos, y 68°C durante 1.5 minutos. Después de los 35 ciclos, la reacción fue incubada a 68°C durante 10 minutos y luego se enfrió a 10°C hasta ulterior procesado. Un producto de reacción de PCR de 1.5 kb fue aislado en 0,8% de gel de agarosa GTG (Cambrex Bioproducts One Meadowlands Plaza East Rutherford, New Jersey 07073) usando 40 mM de base Tris -20 mM de acetato de sodio -1 mM tampón de EDTA disodio (TAE) y 0.1 μ g de bromuro de etidio por ml. La banda de ADN fue visualizada con la ayuda de un Dark Reader™ (Clare Chemical Research, Dolores, CO) para evitar mutaciones inducidas por UV. La banda de ADN de 1.5 kb fue cortada con una cuchilla de navaja desechable y purificada con un cubilete de centrifuga Ultrafree-DA (Millipore, Billerica, MA) según las instrucciones del fabricante.

El vector pAILo2 fue linealizado por digestión con *NcoI* y *PacI* (usando condiciones especificadas por el fabricante). El fragmento fue purificado por electroforesis en gel y ultrafiltración como se ha descrito anteriormente. La clonación del fragmento de PCR purificado en el vector pAILo2 linealizado y purificado fue realizada con un In-Fusion Cloning Kit (BD Biosciences, Palo Alto, CA). La reacción (20 μ l) contenía 1X tampón In-Fusion (BD Biosciences, Palo Alto, CA), 1X BSA (BD Biosciences, Palo Alto, CA), 1 μ l de enzima In-fusion (diluido 1:10) (BD Biosciences, Palo Alto, CA), 100 ng de pAILo2 digerido con *NcoI* y *PacI*, y 50 ng del producto de PCR purificado GH6A de *Thielavia terrestris*. La reacción fue incubada a temperatura ambiente durante 30 minutos. Una muestra de 2 μ l de la reacción fue usada para transformar células *E. coli* XL10 SoloPac® Gold (Stratagene, La Jolla, CA) según las instrucciones del fabricante. Después del periodo de recuperación, dos partes alícuotas de 100 μ l de la reacción de transformación fueron colocadas en placas sobre placas de medio 2X YT 150 mm suplementadas con 100 μ g de ampicilina por ml. Las placas fueron incubadas durante toda la noche a 37°C. Dos conjuntos de ocho clones putativos recombinantes fueron seleccionados al azar de las placas de selección y ADN plásmido fue obtenido de cada uno usando un Biorobot 9600 (QIAGEN, Inc., Valencia, CA). Los clones fueron analizados por digestión de restricción *NcoI*. Cuatro clones que tenían el modelo de digirido de restricción previsto fueron luego ordenados para confirmar que no había mutaciones en el inserto clonado. Clon #13 del segundo conjunto fue seleccionado y designado pAILo21 (figura 5).

Ejemplo 8

Expresión del gen de celobiohidrolasa de familia GH6A de Thielavia terrestris en Aspergillus oryzae JAL250

Protoplastos JAL250 (WO 99/61651) de *Aspergillus oryzae* fueron preparados según el método de Christensen *et al.*, 1988, Bio/Technology 6: 1419-1422. Cinco microgramos de pAILo21 fueron usados para transformar protoplastos JAL250 de *Aspergillus oryzae*. pAILo2 fue realizado como un control de vector.

ES 2 358 243 T3

La transformación de *Aspergillus oryzae* Jal250 con pAIl021 liberó aproximadamente 50 transformantes. Ocho transformantes fueron aislados en placas de PDA individuales e incubados durante cinco días a 34°C.

5 Placas de esporas confluyentes fueron lavadas con 5 ml de 0.01% Tween 80 y la suspensión de esporas fue usada para inocular 25 ml de medio MDU2BP en frascos de agitación de vidrio de 125 ml. Cultivos transformantes fueron incubados a 34°C con agitación constante a 200 r.p.m. En el quinto post-inoculación, los cultivos fueron centrifugados a 6000 x g y sus sobrenadantes fueron recogidos. Cinco micro-litros de cada sobrenadante fueron mezclados con un volumen igual de 2X tampón de carga (10% β -mercaptoetanol) y cargados sobre un gel Tris-glicina SDS-PAGE de 1.5 mm 8%-16% y teñidos con Simply Blue SafeStain (Invitrogen, Carlsbad, CA). Perfiles de SDS-PAGE de los caldos de cultivo mostraron que uno los transformantes (designado transformante 2) tuvo un enlace proteínico de aproximadamente 60 kDa.

15 Actividad bioquímica de la celobiohidrolasa Cel6A de *Thielavia terrestris* expresada en *Aspergillus oryzae* fue determinada por hidrólisis de celulosa de ácido fosfórico-hinchado (PASC) a glucosa en presencia de exceso de β -glucosidasa.

20 PASC fue obtenido de Avicel (Fluka, obtenido de Sigma-Aldrich, St. Luis, MO). Ácido ortofosfórico enfriado en hielo, 85% (VWR Internacional, Pittsburgh, PA) se añadió a Avicel, en la proporción de 30 ml de ácido a 1 g de Avicel, y se agitó en un baño de hielo durante una hora. Acetona fría (VWR Internacional, Pittsburgh, PA) fue añadida mientras se agitaba, usando 100 ml por g de Avicel. El compuesto acuoso fue transferido a un embudo de filtro de vidrio y lavado con acetona fría, usando tres lavados de 20 ml de acetona por g de Avicel. Finalmente, la celulosa fue lavada dos veces con 100 ml de agua por g de acetona. El PASC fue resuspendido en agua, a una concentración de 1% PASC, y almacenado a 4°C.

25 La actividad de celobiohidrolasa Cel6A de *Thielavia terrestris* fue comparada con celobiohidrolasa Cel6A de *Humicola insolens* usando 15.9 μ g de cada enzima Cel6 y 0.8 μ g de β -glucosidasa. Las enzimas fueron incubadas a 45°C en 1.182 ml de acetato sódico 50 mM, pH 5.0, 0.001% azida sódica, y 1.59 mg de PASC en placas de 96 pocillos profundos (Axygen Scientific, Union City, CA) selladas por un sellador de placa (ALPS-300, Abgene, Epsom, UK). Después de varios tiempos de incubación, las mezclas de reacción fueron centrifugadas y la concentración de glucosa del sobrenadante fue determinada con un analizador de glucosa (YSI, Inc., Yellow Springs, OH). Como se muestra en la figura 6, el curso de tiempo de hidrólisis y actividad de celobiohidrolasa Cel6A de *Thielavia terrestris* fue casi idéntica a aquella de celobiohidrolasa Cel6A de *Humicola insolens*.

35 Ejemplo 9

Análisis de huella genética de masa peptídica (PMF) tiempo de ionización/Desorción Láser Asistida por Matriz de espectrometría de masas de vuelo (MALDI-TOF MS) para verificación de proteína

40 La banda proteica de 60 kDa (ejemplo 8) fue digerida en gel con tripsina como se describe en el ejemplo 1. Los péptidos recuperados fueron analizados por análisis de huella genética de masa peptídica para verificación de proteína. Un espectrómetro de masas de tiempo de vuelo Maldi™-LR fue usado (Waters Micromass® MS Technologies, Milford, MA). Ácido alfa-ciano-4-hidroxicinámico recristalizado fue preparado mediante lavado de cantidades en miligramos del ácido alfa-ciano-4-hidroxicinámico (Sigma Chemical Co., St. Louis, MO) con 100% acetonitrilo (E.M. Science, Gibbstown, NJ) y mezclado íntegramente y centrifugado para formar un granulado matricial. La solución de acetonitrilo fue eliminada y descartada. Agua de calidad de HPLC (Fisher Chemicals, Fairlawn, NJ) fue añadida seguido de adición lenta de 28-30% de hidróxido amónico (J.T. Baker, Phillipsburg, NJ) hasta que casi todo el granulado fue disuelto. El granulado no disuelto fue descartado. Agua de HCl concentrada (Fisher Chemicals, Fairlawn, NJ) fue lentamente adicionada a la solución matricial hasta que una gran cantidad de matriz recristalizó. La matriz cristalizada fue eliminada por filtración y lavada varias veces con 0.1 M de HCl y dejada secar completamente. La solución de matriz final consistió en una solución de 10 mg/ml de ácido alfa-ciano-4-hidroxicinámico recristalizado en 50% acetonitrilo/50% acuoso 0.1% TFA. Un μ l de la solución de extracción peptídica obtenida de la digestión en gel de proteína fue mezclada con 1 μ l de la solución de matriz recristalizada y secada por manchas sobre una placa diana MALDI-TOF de acero inoxidable.

55 El espectrómetro de masas fue accionado en reflectrón y modo de ión positivo usando un voltaje de aceleración de +15 kV, voltaje de pulsación de 2535 voltios y voltaje de reflectrón de 2000 voltios. La gama de masa de toma de datos fue establecida de 640-3000 m/z. Un patrón de calibración de masa exacta que consiste en 1 μ l de 200 fmols/ μ l de ACTH (Clip de hormona Adenocorticotrófica 18-39 PM = 2,465,1989) (Sigma Chemical Co, St. Louis, MO) y 1 μ l de solución de matriz recristalizada fue usada para patrón interno y teñido en el pocillo diana de masa exacta adyacente. La toma de datos fue realizada usando una terminal de trabajo de microprocesador controlado por Windows NT usando software de espectrometría de masas Masslynx 4.0 (Waters Micromass® MS Technologies, Milford, MA). Los espectros adquiridos fueron combinados, alisados, y centrados, y una lista de valor máximo de masas de iones peptídicos generada. Esta lista de valor máximo fue buscado contra bases de datos usando software ProteinLynx™ Global Server 2.05 (Waters Micromass® MS Technologies, Milford, MA).

ES 2 358 243 T3

Los resultados del análisis de huella genética de masa peptídica indicó que las manchas de proteína de aproximadamente 60 kD es la proteína Cel6A de *Thielavia terrestris* basada en la proteína en la siguiente masa peptídica corresponde a celobiohidrolasa Cel6A de *Thielavia terrestris*:

Masa medida (m+H)	Masa teórica (m+H)	Secuencia	Posición en SEC ID NO:2
986.423	986.430	CANAESTYK	271-279
1147.792	1147.590	VPSFQWLDR	163-171
1193.856	1193.653	SLVIQYSDIR	240-249
1829.213	1828.990	NVTIDTLFAHTLSOIR	172-187
1964.197	1963.991	GANPPYAGIFVVYDLPDR	193-210
2301.317	2301.250	IIFVIEPDSLANMVTNLNVAK	250-270

Ejemplo 10

Expresión de celobiohidrolasa Cel6A de *Thielavia terrestris* por *Trichoderma reesei*

El promotor del gen de celobiohidrolasa I (CBHI) de *Trichoderma reesei* en plásmido de expresión pSMai155 (WO 05/074647) fue sustituido con un promotor de CBHII de *Trichoderma reesei* dando como resultado el plásmido pCW076. El promotor del gen de celobiohidrolasa II (CBHII) de *Trichoderma reesei* fue aislado del plásmido pEJG114 (WO 05/074647) por digestión con *SalI* y *NcoI*. La eliminación del promotor CBHI de pSMai155 fue realizada por digestión con *SalI* y *NcoI*. El fragmento de ADN de CBHII y plásmido pSMai155 linealizado, menos el promotor de CBHI, fueron visualizados con la ayuda de un Dark Reader™ (Clare Chemical Research, Dolores, CO). Ambos fragmentos fueron cortados con cuchillas de navaja desechables y purificadas con un QIAquick Gel Extraction Kit (QIAGEN Inc., Valencia, CA) según las instrucciones del fabricante. El promotor de CBHII fue luego ligado en pSMai155 linealizado utilizando los sitios *SalI* y *NcoI* obtenidos tras la eliminación del promotor CBHI usando un Rapid DNA Ligation Kit (Roche, Indianapolis, IN) según las instrucciones del fabricante. Las células competentes de grado subclonación XL1-Blue de *E. coli* (Stratagene, La Jolla, CA) fueron transformadas con el producto de ligadura. La identidad del constructo fue confirmada por secuenciación del ADN de la secuencia del promotor de CBHII de plásmidos purificados de *E. coli* transformado. Un clon conteniendo el plásmido recombinante fue designado pCW076 (figura 7).

Dos cebadores oligonucleótidos sintéticos mostrados abajo fueron diseñados para amplificar por PCR el gen de celobiohidrolasa Cel6A de *Thielavia terrestris* del plásmido pAILo21 (ejemplo 7). El cebador directo resulta en un extremo como 5' y el cebador inverso incorpora un sitio *PacI* en el extremo 3'. El fragmento Cel6A fue directamente clonado en pCW076 utilizando un sitio *NcoI* como en el extremo 5' y un sitio *PacI* en el extremo 3'.

Cebador directo: 5'-ATGGCTCAGAAGCTCCTTCTCGCCG-3' (SEC ID nº: 16)

Cebador inverso: 5'-CAGTCACCTCTAGTFAATTAATTAGAAGGGCGGG-3' (SEC ID nº: 17)

Cinuenta picomoles de cada uno de los cebadores arriba fueron usados en una reacción PCR que consiste en 50 ng de pAILo21, 1 µl de un 10 mM mezcla de dATP, dTTP, dGTP, y dCTP, 5 µl de 10X AmpliTaq® DNA Polymerase Buffer I (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA), y 5 unidades de AmpliTaq® DNA Polymerase (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA), en un volumen final de 50 µl. Un Eppendorf Mastercycler 5333 (Eppendorf Scientific, Inc., Westbury, NY) fue usado para amplificar el fragmento de ADN y fue programado para un ciclo a 95°C durante 3 minutos; y 30 ciclos cada uno a 95°C durante 45 segundos, 55°C durante 60 segundos, y 72°C durante 1 minuto 30 segundos. Después de los 30 ciclos, la reacción fue incubada a 72°C durante 10 minutos y luego enfriada a 4°C hasta ulterior tratamiento. El extremo 3' del fragmento Cel6A de PCR fue digerido usando *PacI*. El producto de digestión fue purificado usando un MinElute™ Reaction Cleanup Kit (QIAGEN, Valencia, CA) según las instrucciones del fabricante (QIAGEN Inc., Valencia, CA). El plásmido pCW076 fue digerido con *NcoI* y *PacI*. El sitio *NcoI* fue luego convertido en como usando una enzima Klenow para rellenar el sitio *NcoI* 5' cóncavo. La reacción de Klenow consistió en 20 µl de la mezcla de reacción de digestión pCW076 más 1 mM de dNTPs y 1 µl de enzima Klenow (Roche, Indianapolis, IN) que fue incubada brevemente a temperatura ambiente. El pCW076 linealizado fue purificado usando un MinElute™ Reaction Cleanup Kit (QIAGEN, Valencia, CA) según las instrucciones del fabricante (QIAGEN Inc., Valencia, CA). Estas reacciones resultaron en la creación de un extremo como 5' y un sitio *PacI* 3' compatible con el fragmento Cel6A generado. El fragmento Cel6A fue luego clonado en pCW076 usando

ES 2 358 243 T3

un Rapid DNA Ligation Kit (Roche, Indianapolis, IN) según las instrucciones del fabricante. Células Competentes de grado subclonación con XL1-Blue de *E. coli* (Stratagene, la Jolla, CA) fueron transformadas con el producto de ligadura. La identidad del constructo fue confirmada por secuenciación del ADN de la secuencia codificante de Cel6A de plásmidos purificados de *E. coli* transformado. Un clon conteniendo el plásmido recombinante fue designado pCW085 (figura 8).

Frascos de agitación conteniendo 25 ml de medio YP, 2% glucosa, y 10 mM uridina fueron inoculados con 5×10^7 esporas de cepa SaMe-FX16 de *Trichoderma reesei*. La incubación se efectuó durante 17 horas a 27°C con agitación a 90 r.p.m. Los micelios fueron luego lavados usando un "Vacuum Driven Disposable Filtration System" de 500 ml (Millipore) dos veces con aproximadamente 100 ml de agua desionizada. Los micelios fueron luego lavados dos veces con 1.2 M sorbitol y los protoplastos fueron generados suspendiendo los micelios en una solución filtrada estéril de 5 mg de GlucanexTM (Novozymes A/S, Bagsvaerd, DK) por ml y 1 mg de quitinasa (Sigma Chemical Co., St. Louis, MO) por ml en 20 ml de 1.2 M sorbitol. El digerido fue incubado a 34°C y suavemente agitado a 90 r.p.m. durante aproximadamente 20 minutos o hasta que los protoplastos fueron formados. Una vez que los protoplastos fueron generados la mezcla fue incubada en hielo hasta una digestión lenta. El digerido fue luego transferido a un Tubo Falcon de 50 ml y 30 ml de sorbitol helado 1.2 M se añadió al digerido. Los protoplastos fueron luego centrifugados durante 7 minutos a 200 r.p.m. Después del centrifugado el sobrenadante fue vertido y los protoplastos fueron lavados en 50 ml de sorbitol helado 1.2 M. Este paso fue repetido dos veces y después del segundo lavado los protoplastos fueron contados y resuspendidos en STC (1 M sorbitol, 10 mM CaCl₂, 10 mM Tris-HCl, pH 7.5) a una concentración de 1×10^8 protoplastos/ml. Los protoplastos fueron almacenados a -80°C hasta el uso.

Dos a cinco μ g de pCW085 linealizados con PmeI, 100 μ l de 1×10^8 protoplastos de la cepa SaMe-FX16 de *Trichoderma reesei*, y 250 μ l de tampón de PEG (50% PEG 4000, 10 mM CaCl₂, 10 mM Tris-HCl, pH 7.5) fueron suavemente mezclados juntos en un tubo Falcon 2059 de 12 ml y se incubó durante 30 minutos a temperatura ambiente. Después de la incubación, 3 ml de STC (1 M sorbitol, 10 mM CaCl₂, 10 mM Tris-HCl, pH 7.5) fueron adicionados a la mezcla reactiva. La reacción de transformación fue mezclada suavemente y luego la cantidad entera fue vertida sobre unas placas PDA conteniendo 100 ng de higromicina B (Sigma Chemical Co., St. Louis, MO) por ml e incubada a 28°C durante 5-7 días.

Veintiún transformantes fueron escogidos e inoculados en frascos de agitación de 250 ml conteniendo 25 ml de CIM. Los frascos de agitación fueron dejados crecer durante cinco días a 28°C mientras se agitaba a 200 r.p.m. Un transformante designado SaMe-D8 de *Trichoderma reesei* fue cultivado en una fermentación de 2 litros según el siguiente procedimiento.

Dos tapones de agarosa de una placa fresca de *Trichoderma reesei* SaMe-D8 crecida en placas PDA fueron inoculados en un matraz vibrante conteniendo 100 ml de medio compuesto por litro de 20 g de glucosa, 10 g de sólidos Corn Steep, 1.45 g de (NH₄)₂SO₄, 2.08 g de KH₂PO₄, 0.36 g de CaCl₂·2H₂O, 0.42 g de MgSO₄·7H₂O, y 0.2 ml de solución de metales traza. La solución de metales traza fue compuesta por litro de 216 g de FeCl₃·6H₂O, 58 g de ZnSO₄·7H₂O, 27 g de MnSO₄·H₂O, 10 g de CuSO₄·5H₂O, 2.4 g de H₃BO₃, y 336 g de ácido cítrico. El matraz vibrante fue incubado durante 2 días a 28°C y 200 r.p.m. En un agitador orbital. Cincuenta ml de este cultivo fueron usados para inocular 1.8 litros de medio compuesto por litro de 30 g de celulosa, 10 g de sólidos Corn Steep, 4 g de glucosa, 2.64 g de CaCl₂·2H₂O, 3.8 g de (NH₄)₂SO₄, 2.8 g de KH₂PO₄, 1.63 g de MgSO₄·7H₂O, 0.75 ml de solución de metales traza (igual que anteriormente), y 3 ml de ácido plurónico. Temperatura fue establecida a 28°C y el pH fue controlado a 4.75.

La fermentación fue ejecutada con un mínimo de oxígeno disuelto a 25% a un caudal de aire de 1.0 WM y una agitación de 1100 r.p.m. El medio de alimentación fue entregado en el vaso de fermentación de 2 litros según lo necesitado con un índice de alimento de 6.0-8.0 g/hora durante 7-8 días. El medio de alimentación fue compuesto por kg de 600 g de glucosa, 35.5 g de H₃PO₄, 20 g de celulosa, y 5 g de ácido plurónico.

El caldo de fermentación entero fue centrifugado a 3000 x g durante 10 minutos y el sobrenadante fue filtrado a través de una unidad de filtración desechable con un prefiltro de fibra de vidrio (Nalgene, Rochester NY). El filtrado fue enfriado a 4°C para almacenamiento.

El filtrado fue luego sometido a electroforesis en gel de poli(acrilamida) bidimensional para confirmar la expresión de la celobiohidrolasa Cel6A de *Thielavia terrestris* (véase ejemplo 11).

Ejemplo 11

Verificación de expresión de *Thielavia terrestris* Cel6A en *Trichoderma reesei* SaMe-D8

Electroforesis en gel de poli(acrilamida) bidimensional. Un ml de filtrado de la fermentación de 2 litros de *Trichoderma reesei* SaMe-D8 descrito en ejemplo 10 fue precipitado añadiendo 100 μ l de ácido tricloroacético saturado (ATC) a 4°C e incubándose durante 10 minutos en hielo seguido de adición de 9 ml de acetona enfriada en hielo y posterior incubación en hielo durante 20 minutos. La solución precipitada fue centrifugada a 10,000 x g durante 10 minutos a 4°C, el sobrenadante fue decantado, y el granulado fue enjuagado dos veces con acetona enfriada en hielo y secada al aire. El granulado seco fue disuelto en 0.2 ml de tampón de muestra por isoelectroenfoque (IEF) (9.0 M urea, 3.1%

ES 2 358 243 T3

(p/v) 3-[(3-colamidopropil) dimetil-amonio]-1-propanosulfonato (CHAPS, Pierce Chemical Co. Rockford, IL), 1% (v/v) pH 4-7 anfólitos, 50 mM (DTT) de ditioneitol, y 0.005% azul bromofenol en agua destilada). Solución stock de urea fue desionizada usando AG 501-X8 (D), malla de 20-5, resina de lecho mezclado de BioRad Laboratories (Hercules, CA). La solución desionizada fue almacenada a -20°C. La mezcla resultante fue dejada solubilizar durante varias horas con mezcla suave en un LabQuake™ Shaker (Lab Industries, Berkeley, CA). Doscientos μ l de cada mezcla de proteína de tampón de muestra IEF fue aplicado a una banda de IPG de 11 cm (BioRad Laboratories, Hercules, CA) en una bandeja de rehidratación de IPG (Amersham Biosciences, Piscataway, NJ). Una alícuota de fluido de 750 μ l de fluido de cobertura de banda seca (Amersham Biosciences, Piscataway, NJ) fue estratificada sobre las bandas IPG para prevenir la evaporación y permitió rehidratarse durante 12 horas mientras se aplicaba 30 voltios usando una unidad de isoelectroenfoque IPGPhor (Amersham Biosciences, Piscataway, NJ) a 20°C. La unidad IPGPhor fue programada para voltaje constante pero con una corriente máxima de 50 μ A por banda. Después de 12 horas de rehidratación, las condiciones de isoelectroenfoque fueron de la siguiente manera: 1 hora a 200 voltios, 1 hora a 500 voltios, y 1 hora a 1000 voltios. Luego un gradiente fue aplicado de 1000 voltios a 8000 voltios durante 30 minutos e isoelectroenfoque fue programado para ejecutarse a 8000 voltios y fue completado cuando >30,000 voltios por hora fue conseguido. Bandas de gel IPG fueron reducidas y alquiladas antes del segundo análisis de dimensión primero por reducción durante 15 minutos en 100 mg de ditioneitol por 10 ml de tampón de equilibrado de SDS (50 mM Tris HCl pH 8.8, 6.0 M urea, 2% (p/v) dodecilsulfato de sodio (SDS), 30% glicerol, y 0.002% (p/v) azul bromofenol) seguido de 15 minutos de alquilación en 250 mg de yodoacetamida por 10 ml de tampón de equilibrado a oscuras. Las bandas IPG fueron enjuagadas rápidamente en tampón de desplazamiento de SDS-PAGE (Invitrogen/Novex, Carlsbad, CA) y colocadas en un gel de Tris-Glicina SDS-PAGE de 11 cm, 1 pocillo 8-16% (BioRad Laboratories, Hercules, CA) y sometidas a electroforesis usando una unidad de electroforesis Criterion (BioRad Laboratories, Hercules, CA) a 50 voltios hasta que la muestra se introdujo en el gel y luego el voltaje fue aumentado a 200 voltios y dejado desplazarse hasta que el colorante azul de bromofenol alcanzó el fondo del gel.

Detección de polipéptido. El gel bidimensional fue eliminado del casete y enjuagado tres veces con agua durante al menos 5 minutos cada vez y teñido con tinte Bio-Safe™ Coomassie G250 (BioRad Laboratories, Hercules, CA) durante 1 hora seguido de decoloración con agua doblemente destilada durante más de 30 minutos. Manchas de gel de proteína observadas fueron cortadas usando un 2 mm Acu-Punch Biopsy Punch (Acuderm Inc., Ft. Lauderdale, FL) y se almacenaron en placas de noventa y seis pocillos que fueron prelavadas con 0.1% de ácido trifluoroacético (TFA) en 60% de acetonitrilo seguido de dos lavados adicionales con agua de calidad de HPLC. Las manchas de gel bidimensional teñido fueron almacenadas en 25-50 μ l de agua en las placas prelavadas a -20°C hasta ser digeridas. Una mancha de proteína correspondiente al PM previsto y punto teórico isoeléctrico (pI) del polipéptido Cel6A de *Thielavia terrestris* fue analizada por análisis de huella digital de masa peptídica MALDI-TOF MS como se describe en el ejemplo 9.

Los resultados del análisis de huella genética de masa peptídica indicaron que la mancha de proteína de aproximadamente 60 kD es la proteína de Cel6A de *Thielavia terrestris* basado en la siguiente masa peptídica corresponde a la celobiohidrolasa Cel6A de *Thielavia terrestris*:

Masa medida (m+H)	Masa teórica (m+H)	Secuencia	Posición en SEC ID NO:2
1043.124	1043.446	CANAESTYK	253-261
1147.379	1147.590	VPSFQWLDR	145-153
1193.478	1193.653	SLVIQYSDIR	222-231
1964.184	1963.991	GANPPYAGIFVVYDLPDR	175-192

Depósito de Material biológico

El siguiente material biológico ha sido depositado según las condiciones del tratado de Budapest con la Colección de Patentes de Cultivos del Servicio de Investigación para la Agricultura, Centro de investigación Regional del Norte, 1815 University Street, Peoria, Illinois, 61604, y se le dio el siguiente número de acceso:

Depósito	Número de acceso	Fecha de depósito
E. coli pTter6A	NRRL B-30802	17 de diciembre de 2004

ES 2 358 243 T3

La cepa ha sido depositada bajo condiciones que aseguran que acceso al cultivo estará disponible durante la pendency de esta solicitud de patente a una determinada por el Comisario de Patentes y Marcas para gozar de sus derechos bajo 37 C.F.R. §1.14 y 35 U.S.C. §122. El depósito representa un cultivo substancialmente puro de la cepa depositada. El depósito está disponible tal y como se requiere por las leyes de patente extranjeras en países donde duplicados de la solicitud en objeto, o de su progenie sean solicitados. No obstante, debe ser entendido que la disponibilidad de un depósito no constituye una licencia para poner en práctica la invención en objeto en derogación de los derechos sobre patentes concedidos por la acción gubernamental.

10

15

20

25

30

35

40

45

50

55

60

65

REIVINDICACIONES

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

- 5 (a) un polipéptido comprendiendo una secuencia de aminoácidos que tiene al menos 85%, incluso más preferiblemente al menos 90%, de la forma más preferible al menos 95%, e incluso de la forma más preferible al menos 97% de identidad con el polipéptido maduro de SEC ID n°: 2;
- 10 (b) un polipéptido que se codifica por un polinucleótido que se hibrida bajo al menos condiciones de astringencia altas con (i) la secuencia codificante del polipéptido maduro de SEC ID n°: 1, (ii) la secuencia de ADN genómico comprendiendo la secuencia codificante del polipéptido maduro de SEC ID n°: 1, o (iii) una cadena complementaria de (i) o (ii).

15 2. Polipéptido según la reivindicación 1, que comprende o que consiste en la secuencia de aminoácidos de SEC ID n°: 2.

3. Polipéptido según la reivindicación 1, que comprende o que consiste en el polipéptido maduro de SEC ID n°: 2.

20 4. Polipéptido según la reivindicación 1, que se codifica por el polinucleótido contenido en plásmido pTter6A que está contenido en *E. coli* NRRL B-30802.

25 5. Polipéptido de cualquiera de reivindicaciones 1-4, donde el polipéptido maduro es los aminoácidos 18 a 481 de SEC ID n°: 2. y la secuencia codificante del polipéptido maduro es los nucleótidos 52 a 1443 de SEC ID n°: 1.

6. Polinucleótido aislado que codifica el polipéptido según cualquiera de las reivindicaciones 1-5.

30 7. Constructo de ácidos nucleicos o un vector de expresión comprendiendo el polinucleótido según la reivindicación 6 operativamente enlazado a una o más secuencias de control que dirigen la producción del polipéptido en un huésped de expresión.

8. Célula huésped recombinante comprendiendo el constructo de ácidos nucleicos según la reivindicación 7.

35 9. Método para la producción del polipéptido según cualquiera de las reivindicaciones 1-5, comprendiendo: (a) cultivo de una célula, que en su forma de tipo salvaje sea capaz de producir el polipéptido, bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

40 10. Método para la producción del polipéptido según cualquiera de reivindicaciones 1-5, comprendiendo: (a) cultivo de una célula huésped comprendiendo un constructo de ácido nucleico comprendiendo un polinucleótido que codifica el polipéptido bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

45 11. Método para la producción del polipéptido según cualquiera de reivindicaciones 1-5, comprendiendo: (a) cultivo de una planta transgénica o una célula de planta comprendiendo un polinucleótido que codifica un polipéptido que tiene actividad de celobiohidrolasa bajo condiciones propicias para la producción del polipéptido; y (b) recuperación del polipéptido.

50 12. Planta transgénica, parte de planta o célula de planta, que ha sido transformada con un polinucleótido que codifica el polipéptido según cualquiera de las reivindicaciones 1-5.

13. Composición detergente comprendiendo un polipéptido que tiene actividad de celobiohidrolasa según cualquiera de reivindicaciones 1-5 y un tensioactivo.

55 14. Método para degradar o convertir la biomasa conteniendo celulosa y hemicelulosa, comprendiendo el tratamiento de la biomasa con una cantidad eficaz de un polipéptido que tiene actividad de celobiohidrolasa según cualquiera de las reivindicaciones 1-5.

60 15. Método según la reivindicación 14, comprendiendo además el tratamiento de la biomasa con una cantidad eficaz de endo-1,4-beta-glucanasa y beta-D-glucosidasa.

atggctcagaagctccttctcgccgcccgccttgccggccagcgccctcgctgctcccgtcgtcgaggagcgccag
M A Q K L L L A A A L A A S A L A A P V V E E R Q
aactgcggttccgtctggagccaatgcccggcattggctggctccggcgcgacctgctgcgcttcgggcaatacc
N C G S V W S Q C G G I G W S G A T C C A S G N T
tgcggtgagctgaaccgctactactcgcagtgccctgcccaacagccaggtgactacctcgaccagcaagaccacc
C V E L N P Y Y S Q C L P N S Q V T T S T S K T T
tccaccaccaccaggagcagcaccaccagccacagcagcggctccaccagcagcagcaccaccaccaccagcagt
S T T T R S S T T S H S S G P T S T S T T T T S S
cccgtggtcactaccccgcgagtaacctccatccccggcgggtgectcgtcaacggccagctggtccggcaaccgg
P V V T T P P S T S I P G G A S S T A S W S G N P
ttctcgggctgagatgtgggccaacgactactacgcctccgaggtctcgtcgctggccatccccagcatgacg
F S G V Q M W A N D Y Y A S E V S S L A I P S M T
ggcggccatggccaccaaggcggccgaggtggccaaggtgccagcttccagtggttgaccgcaacgtcaccatc
G A M A T K A A E V A K V P S F Q W L D R N V T I
gacacgctgttcgcccacacgctgtcgcagatccgcgcccgaaccagaaaggcggccaaccggccctacgcccggc
D T L F A H T L S Q I R A A N Q K G A N P P Y A G
atcttcgtggtctacgacctccggaccgcgactgcccggcggcggcgtccaacggcagttctccatcgccaac
I F V V Y D L P D R D C A A A A S N G E F S I A N
aacggggcggccaactacaagacgtacatcgacggatccggagcctcgtcatccagtactcagacatccgcac
N G A A N Y K T Y I D A I R S L V I Q Y S D I R I
atcttcgtcatcgagcccgactcgctggccaacatggtgaccaacctgaacgtggccaagtgcgccaacgcccgag
I F V I E P D S L A N M V T N L N V A K C A N A E
tcgacctacaaggagttgaccgtctacgcgctgcagcagctgaacctgcccaacgtggccatgtacctggacgcc
S T Y K E L T V Y A L Q Q L N L P N V A M Y L D A
ggccacgcccggctggctcggctggcccggccaacatccagccggcggccaacctcttcgcccagatctacacgagc
G H A G W L G W P A N I Q P A A N L F A E I Y T S
gccggcaagcccggcggcgtgcgcccctcgccaccaacgtggccaactacaacggctggagcctggccacgcccg
A G K P A A V R G L A T N V A N Y N G W S L A T P
ccctcgtacaccaggggcgaccccaactacgacgagagccactacgtccaggccctcgcccggctgctcaccgcc
P S Y T Q G D P N Y D E S H Y V Q A L A P L L T A
aacggcttccccggcacttcatcaccgacaccggcccgaacggcaagcagccgaccggacaacggcaatgggga
N G F P A H F I T D T G R N G K Q P T G Q R Q W G
gactggtgcaacggttatcggaactggcttcggcgtgcgcccagcagacaaacaccggcctcgacatcgaggacgcc
D W C N V I G T G F G V R P T T N T G L D I E D A

Fig. 1A

ES 2 358 243 T3

ttcgtctgggtcaagcccggcgagtgcgacggcagagcaacacgacctctccccgctacgactaccactgc
F V W V K P G G E C D G T S N T T S P R Y D Y H C
ggcctgtcggacgcgctgcagcctgctccggaggccggcacttggtccaggcctacttcgagcagctcctgacc
G L S D A L Q P A P E A G T W F Q A Y F E Q L L T
aacgccaacccgccctttaa
N A N P P F *

Fig. 1B

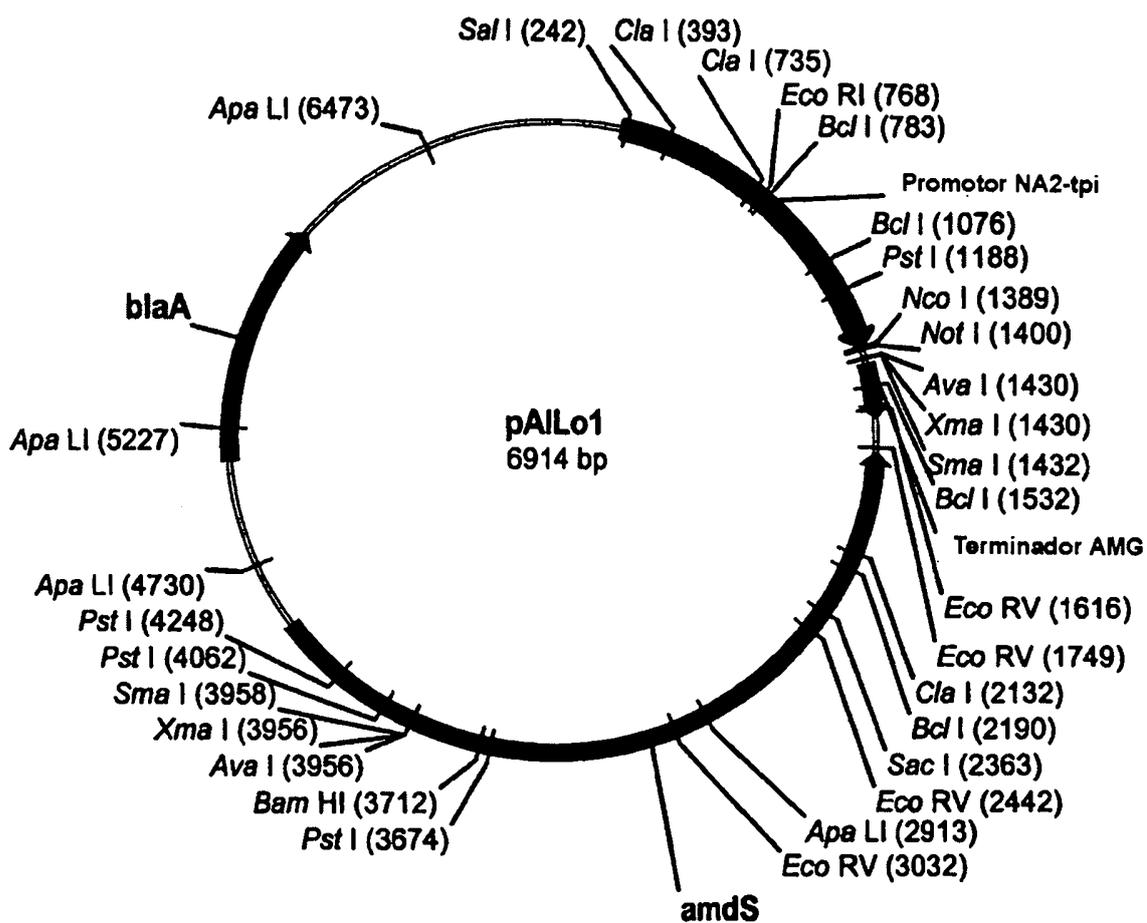


Fig. 2

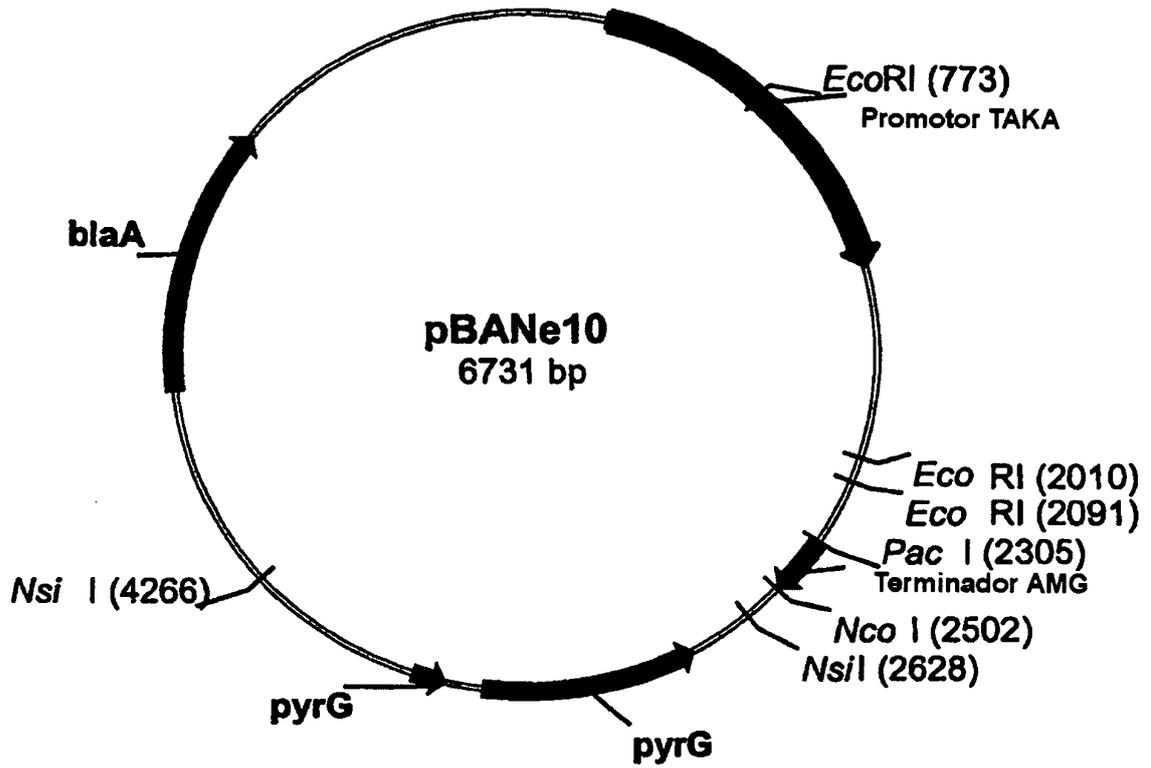


Fig. 3

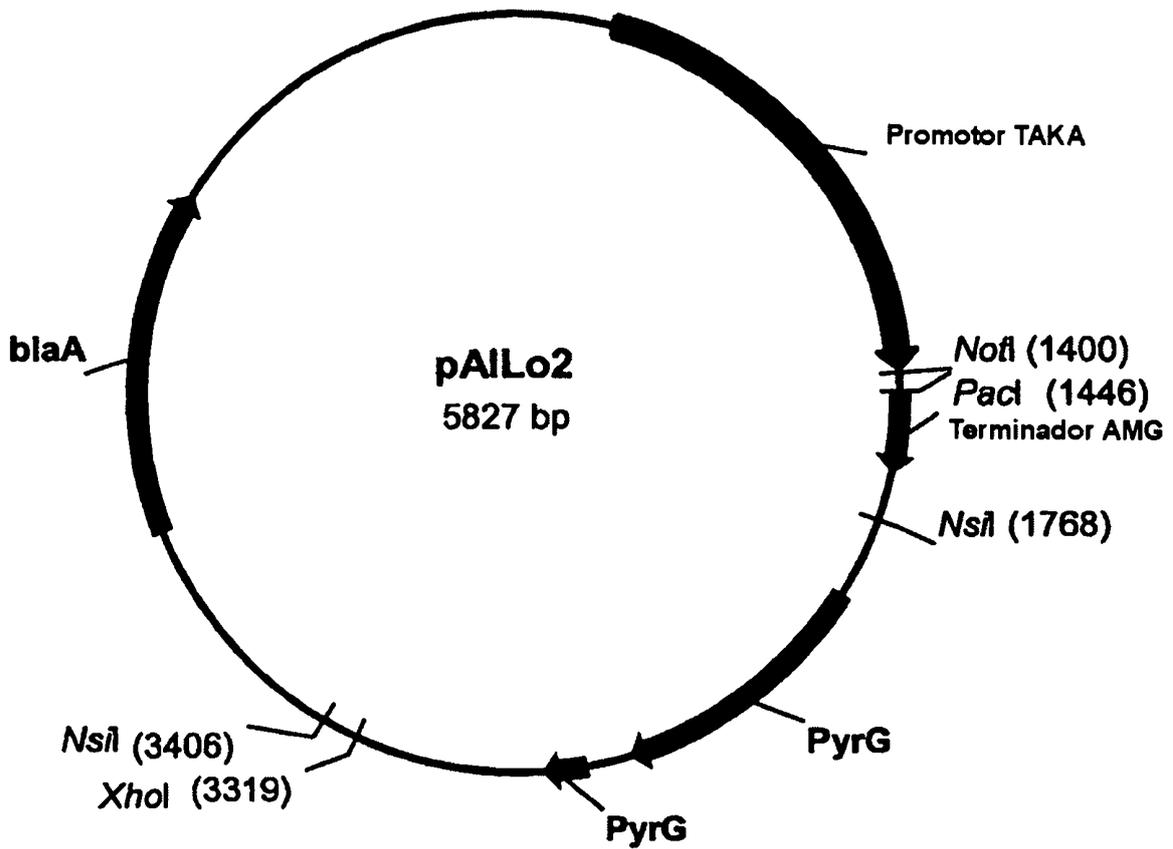


Fig. 4

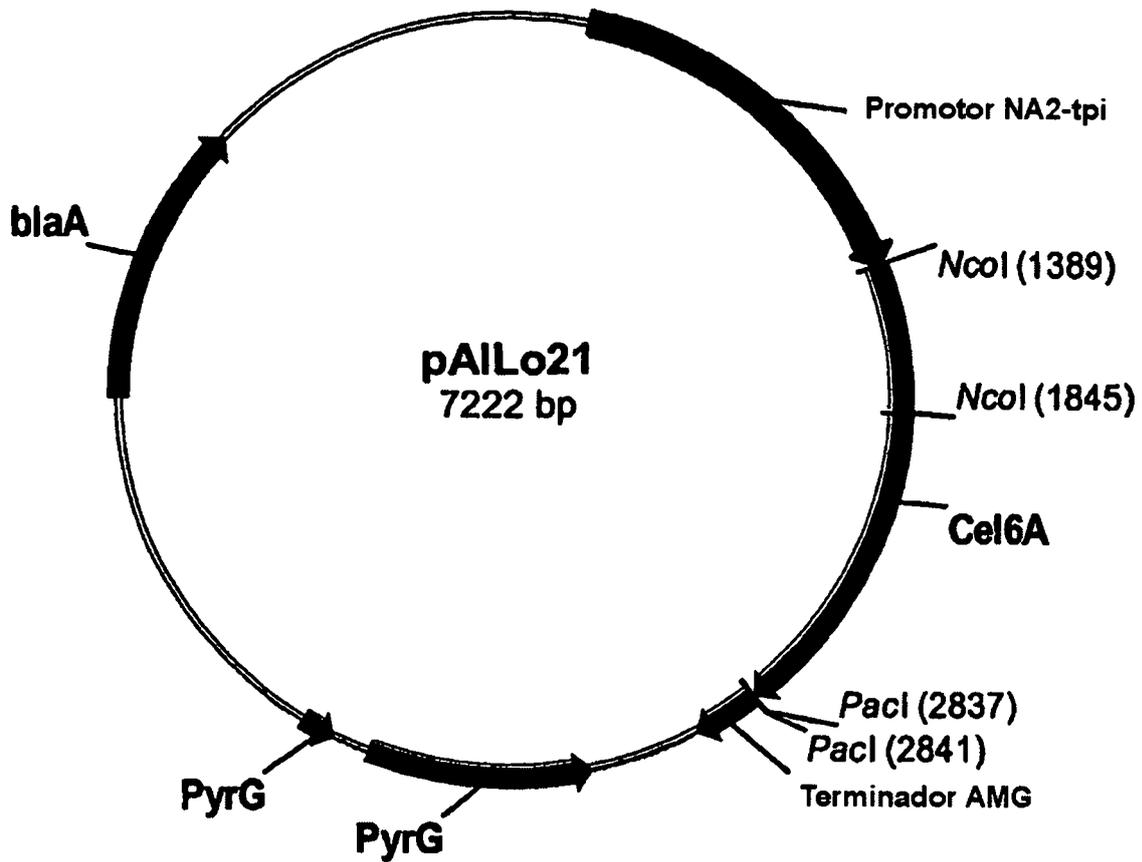


Fig. 5

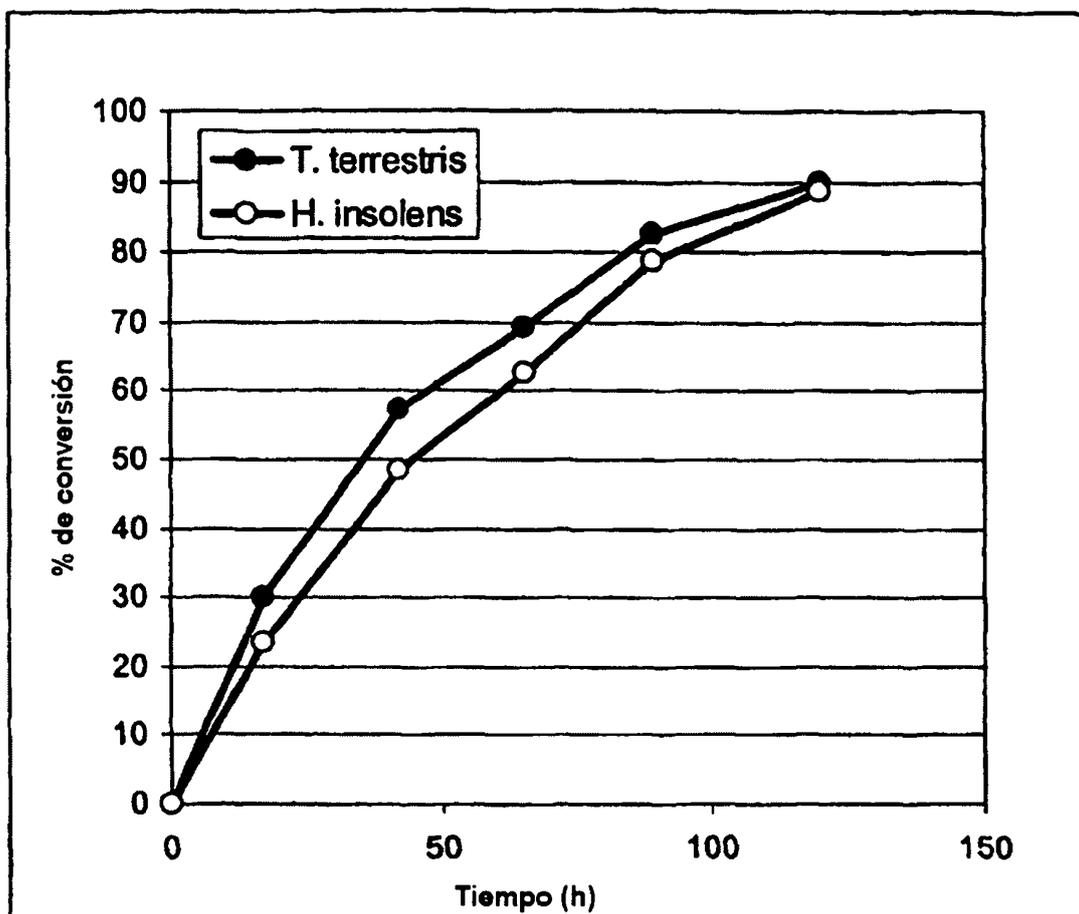


Fig. 6

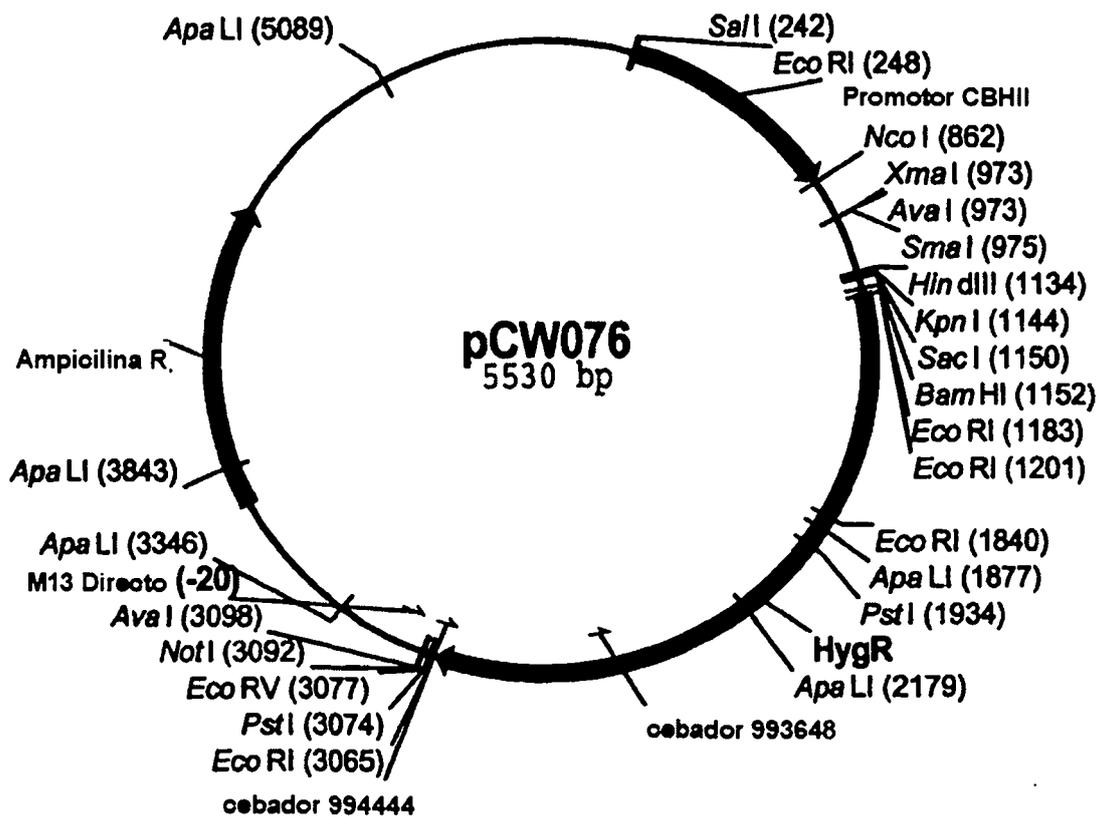


Fig. 7

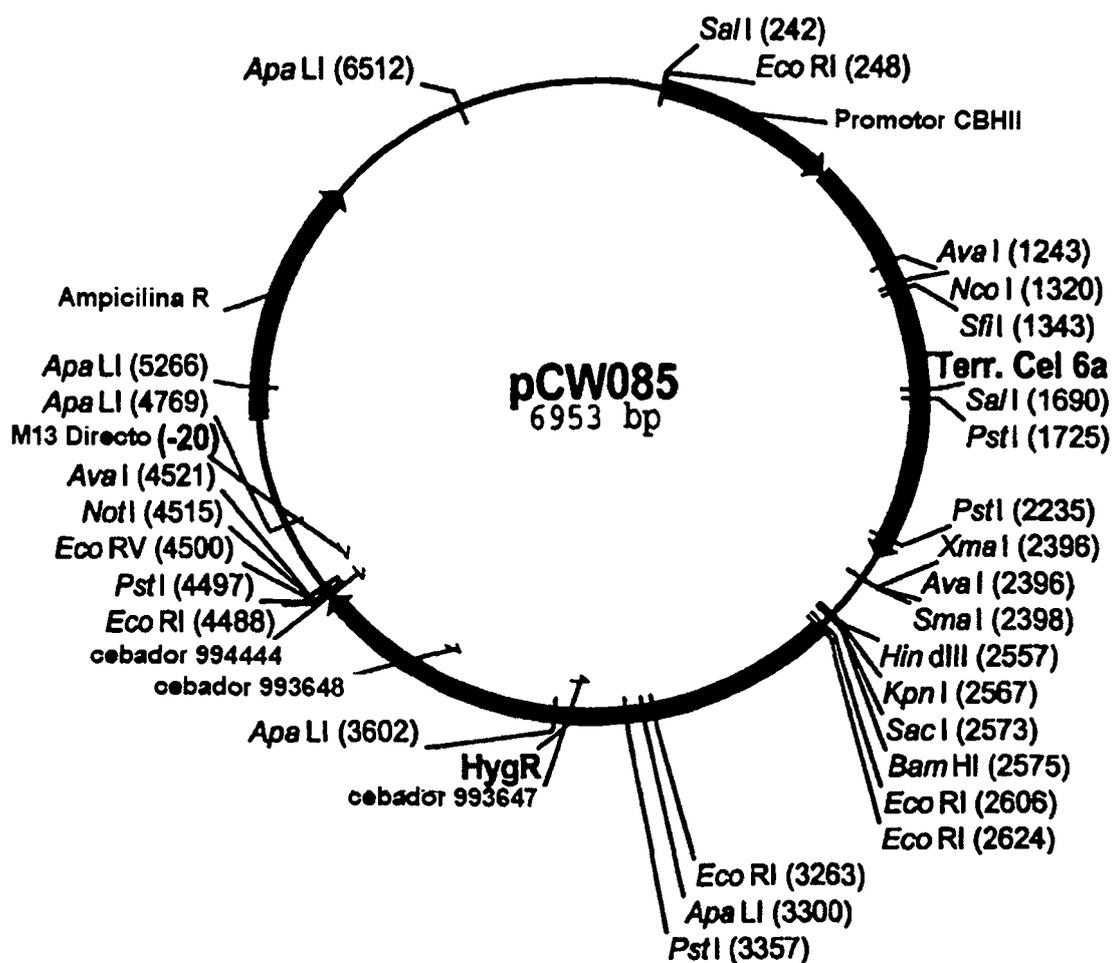


Fig. 8

ES 2 358 243 T3

LISTA DE SECUENCIAS

<110> Novozymes, Inc.

5 <120> Polipéptidos que tienen actividad de celobiohidrolasa y polinucleótidos que los codifican

<130> 10748.204-WO

10 <150> 60/642,274
<151> 2005-01-06

<160> 17

15 <170> PatentIn Version 3.2

<210> 1

20 <211> 1446
<212> ADN
<213> *Thielavia terrestris*

25 <400> 1

30 **atggctcaga agctccttct cgccgccc cttgcgcca gcgccctcgc tgctcccgtc 60**

gtcgaggagc gccagaactg cggttccgtc tggagccaat gcggcggcat tggctgggcc 120

ggcgcgacct gctgcgcttc gggcaatacc tgcggtgagc tgaaccctga ctactcgcag 180

35 **tgccctgcca acagccaggt gactacctcg accagcaaga ccacctccac caccaccagg 240**

agcagcacca ccagccacag cagcggctcc accagcacga gcaccaccac caccagcagt 300

cccgtggcca ctaccccgcc gagtacctcc atccccggcg gtgcctcgtc aacggccagc 360

40 **tggccggca acccgttctc gggcgtgcag atgtggcca acgactacta cgcctccgag 420**

gtctcgtcgc tggccatccc cagcatgacg ggcgccatgg ccaccaaggc ggccgaggtg 480

45 **gccaaggtgc ccagcttcca gtggcttgac cgcaacgtca ccatcgacac gctgttcgcc 540**

cacacgctgt cgcagatccg cgcggccaac cagaaaggcg ccaaccgcc ctacgcgggc 600

50 **atcttcgtgg tctacgacct tccggaccgc gactgcgccg ccgccgcgtc caacggcgag 660**

ttctccatcg cgaacaacgg ggcggccaac tacaagacgt acatcgacgc gatccggagc 720

ctcgtcatcc agtactcaga catccgcac atcttcgtca tcgagcccga ctcgctggcc 780

55 **aacatggtga ccaacctgaa cgtggccaag tgcgccaacg ccgagtcgac ctacaaggag 840**

ttgaccgtct acgcgctgca gcagctgaac ctgcccacg tggccatgta cctggacgcc 900

60 **ggccacgccg gctggctcgg ctggcccgcc aacatccagc cggccgcca cctcttcgcc 960**

gagatctaca cgagcgccg caagccggcc gccgtgcgcg gcctcgccac caacgtggcc 1020

aactacaacg gctggagcct ggccacgccg ccctcgtaca cccagggcga ccccaactac 1080

65 **gacgagagcc actacgtcca ggccctcgc ccgctgctca ccgccaacgg cttccccgcc 1140**

ES 2 358 243 T3

cacttcatca ccgacaccgg ccgcaacggc aagcagccga ccggacaacg gcaatgggga 1200
 gactggtgca acgttatcgg aactggcttc ggcgtgcgcc cgacgacaaa caccggcctc 1260
 5 gacatcgagg acgccttcgt ctgggtcaag cccggcggcg agtgcgacgg cacgagcaac 1320
 acgacctctc cccgctacga ctaccactgc ggctgtcgg acgcgctgca gcctgctccg 1380
 10 gaggccggca cttggttcca ggcctacttc gagcagctcc tgaccaacgc caaccgcccc 1440
 ttttaa 1446

15 <210> 2

<211> 481

<212> PRT

<213> *Thielavia terrestris*

20

<400> 2

25 Met Ala Gln Lys Leu Leu Leu Ala Ala Ala Leu Ala Ala Ser Ala Leu
 1 5 10 15

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

35 Gln Cys Gly Gly Ile Gly Trp Ser Gly Ala Thr Cys Cys Ala Ser Gly
 35 40 45

40 Asn Thr Cys Val Glu Leu Asn Pro Tyr Tyr Ser Gln Cys Leu Pro Asn
 50 55 60

45 Ser Gln Val Thr Thr Ser Thr Ser Lys Thr Thr Ser Thr Thr Thr Arg
 65 70 75 80

50 Ser Ser Thr Thr Ser His Ser Ser Gly Pro Thr Ser Thr Ser Thr Thr
 85 90 95

55 Thr Thr Ser Ser Pro Val Val Thr Thr Pro Pro Ser Thr Ser Ile Pro
 100 105 110

60 Gly Gly Ala Ser Ser Thr Ala Ser Trp Ser Gly Asn Pro Phe Ser Gly
 115 120 125

65 Val Gln Met Trp Ala Asn Asp Tyr Tyr Ala Ser Glu Val Ser Ser Leu
 130 135 140

Ala Ile Pro Ser Met Thr Gly Ala Met Ala Thr Lys Ala Ala Glu Val
 145 150 155 160

ES 2 358 243 T3

Ala Lys Val Pro Ser Phe Gln Trp Leu Asp Arg Asn Val Thr Ile Asp
 165 170 175

5

Thr Leu Phe Ala His Thr Leu Ser Gln Ile Arg Ala Ala Asn Gln Lys
 180 185 190

10

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

15

Asp Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe Ser Ile Ala
 210 215 220

20

Asn Asn Gly Ala Ala Asn Tyr Lys Thr Tyr Ile Asp Ala Ile Arg Ser
 225 230 235 240

25

Leu Val Ile Gln Tyr Ser Asp Ile Arg Ile Ile Phe Val Ile Glu Pro
 245 250 255

30

Asp Ser Leu Ala Asn Met Val Thr Asn Leu Asn Val Ala Lys Cys Ala
 260 265 270

35

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

40

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

45

Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Asn Leu Phe Ala
 305 310 315 320

50

Glu Ile Tyr Thr Ser Ala Gly Lys Pro Ala Ala Val Arg Gly Leu Ala
 325 330 335

55

Thr Asn Val Ala Asn Tyr Asn Gly Trp Ser Leu Ala Thr Pro Pro Ser
 340 345 350

60

Tyr Thr Gln Gly Asp Pro Asn Tyr Asp Glu Ser His Tyr Val Gln Ala
 355 360 365

65

Leu Ala Pro Leu Leu Thr Ala Asn Gly Phe Pro Ala His Phe Ile Thr
 370 375 380

Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Arg Gln Trp Gly
 385 390 395 400

ES 2 358 243 T3

Asp Trp Cys Asn Val Ile Gly Thr Gly Phe Gly Val Arg Pro Thr Thr
 405 410 415

5

Asn Thr Gly Leu Asp Ile Glu Asp Ala Phe Val Trp Val Lys Pro Gly
 420 425 430

10

Gly Glu Cys Asp Gly Thr Ser Asn Thr Thr Ser Pro Arg Tyr Asp Tyr
 435 440 445

15

His Cys Gly Leu Ser Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly Thr
 450 455 460

20

Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr Asn Ala Asn Pro Pro
 465 470 475 480

25

Phe

30

<210> 3

<211> 32

35

<212> ADN

<213> *Thielavia terrestris*

<400> 3

40

caactttgta tcgtcgggga caaaaaagt gg

32

<210> 4

45

<211> 27

<212> ADN

<213> *Thielavia terrestris*

50

<400> 4

ccctgttga aacatgttt tcaacc

27

55

<210> 5

<211> 16

<212> ADN

<213> *Escherichia coli*

60

<400> 5

gtaaacgac ggccag

16

65

<210> 6

<211> 25

ES 2 358 243 T3

	<212> ADN	
	<213> <i>Aspergillus nidulans</i>	
5	<220>	
	<221> misc característica	
	<222> (24)..(24)	
	<223> V=A, C, G	
10	<220>	
	<221> misc característica	
	<222> (25)..(25)	
15	<223> N=A, C, G, T	
	<400> 6	
20	tttttttt tttttttt tttvn	25
	<210> 7	
	<211> 22	
25	<212> ADN	
	<213> <i>Aspergillus nidulans</i>	
	<400> 7	
30	gtgccccatg atacgcctcc gg	22
	<210> 8	
35	<211> 26	
	<212> ADN	
	<213> <i>Aspergillus nidulans</i>	
40	<400> 8	
	gagtcgtatt tccaagctc ctgacc	26
45	<210> 9	
	<211> 24	
	<212> ADN	
50	<213> <i>Aspergillus niger</i>	
	<400> 9	
55	ggaggccatg aagtgaccca acgg	24
	<210> 10	
	<211> 45	
60	<212> ADN	
	<213> <i>Aspergillus niger</i>	
	<400> 10	
65	caccgtgaaa gccatgctct ttcttcgtg tagaagacca gacag	45

ES 2 358 243 T3

	<210> 11	
	<211> 45	
	<212> ADN	
5	<213> <i>Aspergillus niger</i>	
	<400> 11	
10	ctggtcttct acacgaagga aagagcatgg cttcacggt gtctg	45
	<210> 12	
	<211> 44	
15	<212> ADN	
	<213> <i>Aspergillus niger</i>	
	<400> 12	
20	ctatatacac aactggattt accatgggcc cgcgccgca gatc	44
	<210> 13	
25	<211> 44	
	<212> ADN	
	<213> <i>Aspergillus niger</i>	
30	<400> 13	
	gatctcggc cgcgggccca tggtaaatcc agtgtgtat atag	44
35	<210> 14	
	<211> 31	
	<212> ADN	
40	<213> <i>Thielavia terrestris</i>	
	<400> 14	
45	actggattac catggctcag aagctccttc t	31
	<210> 15	
	<211> 34	
50	<212> ADN	
	<213> <i>Thielavia terrestris</i>	
	<400> 15	
55	tcacctctag ttaattaa gggcgggtg gegt	34
	<210> 16	
60	<211> 25	
	<212> ADN	
	<213> <i>Thielavia terrestris</i>	
65	<400> 16	
	atggctcaga agctccttct cgccg	25

ES 2 358 243 T3

<210> 17

<211> 34

<212> ADN

5 <213> *Thielavia terrestris*

<400> 17

10 cagtcacctc tagttaatta affagaaggg cggg

34

15

20

25

30

35

40

45

50

55

60

65