

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 538 360**

51 Int. Cl.:

C12P 21/02 (2006.01)

C12N 9/42 (2006.01)

C07K 1/00 (2006.01)

C12N 15/80 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **20.07.2007 E 07868319 (0)**

97 Fecha y número de publicación de la concesión europea: **18.03.2015 EP 2046819**

54 Título: **Métodos para aumentar la secreción de polipéptidos que tienen actividad biológica**

30 Prioridad:

21.07.2006 US 832511 P

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

19.06.2015

73 Titular/es:

**NOVOZYMES, INC. (100.0%)
1445 DREW AVENUE
DAVIS, CA 95618, US**

72 Inventor/es:

MERINO, SANDRA

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 538 360 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

Métodos para aumentar la secreción de polipéptidos que tienen actividad biológica

5 Declaración sobre los derechos de las invenciones realizadas a continuación

Investigación y desarrollo sufragados con fondos federales

10 [0001] Esta invención se hizo con el apoyo del gobierno bajo el Subcontrato NREL nº 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

15 Campo de la invención

[0002] La presente invención se refiere a métodos de producción de polipéptidos segregados que tienen actividad biológica y a proteínas de fusión y polinucleótidos derivados.

20 Descripción de las técnicas relacionadas

[0003] La producción recombinante de un polipéptido heterólogo en una célula huésped fúngica, particularmente una célula fúngica filamentosa tal como *Aspergillus* o *Trichoderma* o una célula de levadura tal como *Saccharomyces*, puede proporcionar un vehículo más deseable para la producción del polipéptido en cantidades comercialmente importantes.

25

[0004] La producción recombinante de un polipéptido heterólogo segregado se realiza generalmente por construcción de un casete de expresión donde el ADN que codifica para el polipéptido está operativamente enlazado a un promotor adecuado para la célula huésped y una región codificante del péptido señal que codifica para una secuencia de aminoácidos enlazada en marco al amino terminal del polipéptido y dirige el polipéptido codificado en la vía secretora de la célula. El casete de expresión se introduce en la célula huésped, normalmente por transformación mediada por plásmido. La producción de la proteína heteróloga segregada se consigue entonces por cultivo de la célula huésped transformada bajo condiciones inductoras necesarias para el funcionamiento apropiado del promotor contenido en el casete de expresión.

30

[0005] Mientras que la expresión de un polipéptido heterólogo en una célula huésped se puede mejorar, a menudo se encuentra el obstáculo de que el polipéptido se segrega poco en el medio de cultivo. Un método para mejorar la secreción del polipéptido es reemplazar la secuencia codificante del péptido señal nativo con una región codificante del péptido señal foráneo para mejorar la secreción del polipéptido. No obstante, en algunos casos, tal sustitución no proporciona una mejora suficiente para producir el polipéptido en cantidades comercialmente relevantes. Otro método es fundir el polipéptido con otro polipéptido que sea altamente segregado por una célula huésped. El polipéptido altamente segregado funciona como un portador para transportar el polipéptido poco segregado o no segregado como una proteína de fusión a través de la vía secretora de la célula.

40

[0006] La WO 05/093050 divulga una proteína de fusión compuesta por un dominio catalítico de exo-celobiohidrolasa y un dominio catalítico de celulasa para aumentar el rendimiento de una enzima de celulasa. Gouka *et al.*, 1997, Applied and Environmental Microbiology Feb. 1997, p. 488-497, divulgan fusiones de gen de glucoamilasa que alivian las limitaciones para la producción de proteína en el *Aspergillus awamori*. Niissonen y Keranen, 1995, Current Genetics 28: 71-79, divulgan funciones múltiples de la celobiohidrolasa I en la producción mejorada de anticuerpos de fusión por *Trichoderma reesei*.

50

[0007] Es un objeto de la presente invención proporcionar métodos para aumentar la secreción de polipéptidos que tienen actividad biológica.

Resumen de la invención

55

[0008] La presente invención se refiere a métodos para producir un polipéptido segregado que tiene actividad biológica, que comprende:

60

(a) transformación de una célula huésped fúngica con un constructo de proteína de fusión que codifica una proteína de fusión, donde el constructo de proteína de fusión comprende:

- (i) un primer polinucleótido que comprende una secuencia de nucleótidos que codifica un péptido señal;
- (ii) un segundo polinucleótido que comprende una secuencia de nucleótidos que codifica al menos un dominio catalítico de una endoglucanasa o una parte de la misma; y
- (iii) un tercer polinucleótido que comprende una secuencia de nucleótidos que codifica al menos

65

un dominio catalítico de un polipéptido que tiene actividad biológica o una parte del mismo; donde la proteína de fusión es SEC ID nº: 74,

(b) cultivo de la célula huésped fúngica transformada bajo condiciones adecuadas para la producción de la proteína de fusión; y

(c) recuperación de la proteína de fusión, un componente de la misma, o una combinación de la misma, a partir del medio de cultivo, donde la proteína de fusión o el componente

Breve descripción de las figuras

[0009]

La figura 1 muestra un mapa de restricción de pMJ04.

La figura 2 muestra un mapa de restricción de pCaHj527.

La figura 3 muestra un mapa de restricción de pMT2188.

La figura 4 muestra un mapa de restricción de pCaHj568.

La figura 5 muestra un mapa de restricción de pMJ05.

La figura 6 muestra un mapa de restricción de pSMai130.

La figura 7 muestra la secuencia de ADN y la secuencia de aminoácidos de una secuencia de señal nativa de beta-glucosidasa de *Aspergillus oryzae* (SEC ID nº: 59 y 60).

La figura 8 muestra la secuencia de ADN y la secuencia de aminoácidos de una secuencia de señal de endoglucanasa V de *Humicola insolens* (SEC ID nº: 63 y 64).

La figura 9 muestra un mapa de restricción de pSMai135.

La figura 10 muestra un mapa de restricción de pSMai140.

La figura 11 muestra un mapa de restricción de pSaMe-F1.

La figura 12 muestra un mapa de restricción de pSaMe-FX.

La figura 13 muestra un mapa de restricción de pAlLo47.

Las figuras 14A, 14B, 14C y 14D muestran la secuencia de ADN y la secuencia de aminoácidos deducida de la proteína de fusión de la variante de beta-glucosidasa de *Aspergillus oryzae* (SEC ID nº: 73 y 74, respectivamente).

Las figuras 15A, 15B, 15C y 15D muestran la secuencia de ADN y la secuencia de aminoácidos deducida de la proteína de fusión de beta-glucosidasa de *Aspergillus oryzae* (SEC ID nº: 75 y 76, respectivamente).

Definiciones

[0010] Actividad de endoglucanasa: el término "actividad de endoglucanasa" se define aquí como una endo-1,4-beta-D-glucano 4-glucanohidrolasa (E.C. 3.2.1.4), que cataliza la endohidrólisis de los enlaces 1,4-beta-D-glicosídicos en la celulosa, derivados de la celulosa (tales como la carboximetilcelulosa y la hidroxietil celulosa), liquenina, enlaces beta-1,4 en glucanos mezclados beta-1,3 tales como beta-D-glucanos de cereal o xiloglucanos, y otro material vegetal que contenga componentes celulósicos. Para fines de la presente invención, la actividad de endoglucanasa se determina usando la hidrólisis de carboximetil celulosa (CMC) según el procedimiento de Ghose, 1987, Pure and Appl. Chem. 59: 257-268. Una unidad de actividad de endoglucanasa se define como 1,0 µmol de azúcares reductores producidos por minuto a 50°C, pH 4,8.

[0011] Actividad de beta-glucosidasa: el término "beta-glucosidasa" se define aquí como una beta-D-glucósido glucohidrolasa (E.C. 3.2.1.21), que cataliza la hidrólisis de residuos de beta-D-glucosa no reductora terminal con la liberación de beta-D-glucosa. Para fines de la presente invención, la actividad de beta-glucosidasa se determina según el procedimiento descrito por Venturi *et al.*, 2002, J. Basic Microbiol. 42: 55-66, excepto que se utilizan condiciones diferentes como se describe en este caso. Una unidad de actividad de beta-glucosidasa se define como 1,0 µmol de p-nitrofenol producido por minuto a 50°C, pH 5 de 4 mM de p-nitrofenil-beta-D-glucopiranosido como sustrato en 100 mM de citrato sódico, 0,01% de TWEEN® 20.

[0012] Polipéptido en toda su longitud: el término "polipéptido en toda su longitud" se define aquí como una forma precursora de un polipéptido que tiene actividad biológica, donde el precursor contiene una región de péptido señal y alternativamente también una región de propéptido, donde tras secreción de una célula, el péptido señal se divide y alternativamente también el propéptido se divide produciendo un polipéptido con actividad biológica.

[0013] Péptido señal: el término "péptido señal" se define aquí como un péptido enlazada en marco al amino terminal de un polipéptido y dirige el polipéptido codificado en una vía secretora de la célula.

[0014] Secuencia codificante de péptido señal: el término "secuencia codificante de péptido señal" se define aquí como una región codificante de péptido que codifica para una secuencia de aminoácido enlazada en marco al amino terminal de un polipéptido codificado y que dirige el polipéptido codificado en una vía secretora de la célula.

[0015] Propéptido: el término "propéptido" se define aquí como un péptido enlazado en marco al amino terminal de un polipéptido. El polipéptido resultante se conoce como una proenzima o un propolipéptido (o un zimógeno en algunos casos). Un propolipéptido es generalmente inactivo y se puede convertir en un polipéptido activo maduro por escisión

catalítica o autocatalítica del propéptido del propolipéptido. Cuando ambas regiones de péptido señal y de propéptido están presentes en el amino terminal de un polipéptido, la región de propéptido está enlazada en marco al amino terminal de un polipéptido y la región de péptido señal está enlazada en marco al amino terminal de la región de propéptido.

5

[0016] Secuencia codificante de propéptido: el término "secuencia codificante de propéptido" se define aquí como una región codificante de péptido que codifica para una secuencia de aminoácido enlazada en marco al amino terminal de un polipéptido que forma una proenzima o un propolipéptido (o un zimógeno en algunos casos).

10

[0017] Polipéptido maduro: el término "polipéptido maduro" se define aquí como un polipéptido que tiene actividad biológica que está en su forma final tras la traducción y cualquier modificación postraduccional, tal como tratamiento de N-terminal, truncamiento de C-terminal, glicosilación, fosforilación, etc.

15

[0018] Dominio catalítico: el término "dominio catalítico" se define aquí como una parte o región estructural de la secuencia de aminoácidos de una endoglucanasa o un polipéptido que tiene actividad biológica (por ejemplo, actividad de beta-glucosidasa), que posee la actividad catalítica de la endoglucanasa o el polipéptido que tiene actividad biológica (por ejemplo, beta-glucosidasa). El dominio catalítico también se denomina región "central" aquí.

20

[0019] Proteína de fusión: el término "proteína de fusión" se define aquí como un polipéptido que muestra actividad biológica y que comprende al menos tanto un dominio catalítico de endoglucanasa como un dominio catalítico de un polipéptido que tiene actividad biológica (por ejemplo, beta-glucosidasa).

25

[0020] Proteína de fusión de beta-glucosidasa: el término "proteína de fusión de beta-glucosidasa" se define aquí como un polipéptido que muestra actividad de beta-glucosidasa y que comprende al menos tanto un dominio catalítico de beta-glucosidasa como un dominio catalítico de endoglucanasa.

30

[0021] Componentes de una proteína de fusión: el término "componentes de una proteína de fusión" se define aquí como fragmentos individuales (divididos) de la proteína de fusión, donde cada fragmento tiene actividad biológica e incluye bien al menos el dominio catalítico de una endoglucanasa y al menos el dominio catalítico de un polipéptido que tiene actividad biológica. Por ejemplo, la presencia de un sitio de escisión, por ejemplo, el sitio Kex2, entre los componentes de al menos el dominio catalítico de una endoglucanasa y al menos el dominio catalítico de un polipéptido que tiene actividad biológica de la proteína de fusión puede dar como resultado la producción de un polipéptido que tiene actividad de endoglucanasa y otro polipéptido que tiene actividad biológica.

35

[0022] Componentes de una proteína de fusión de beta-glucosidasa: el término "componentes de una proteína de fusión de beta-glucosidasa" se define aquí como fragmentos individuales (divididos) de la proteína de fusión de beta-glucosidasa, donde cada fragmento tiene actividad de beta-glucosidasa y es bien al menos el dominio catalítico de la endoglucanasa y al menos el dominio catalítico de la beta-glucosidasa o al menos el dominio catalítico de la beta-glucosidasa. Por ejemplo, la presencia de un sitio de escisión, por ejemplo, el sitio Kex2, entre la endoglucanasa y componentes de beta-glucosidasa de la proteína de fusión pueden dar como resultado la producción de un polipéptido que tiene actividad de endoglucanasa y otro polipéptido que tiene actividad de beta-glucosidasa.

40

[0023] Módulo de unión de carbohidrato: el término "módulo de unión a carbohidratos (CBM)" se define aquí como una parte de la secuencia de aminoácidos de una endoglucanasa (celulasa) que está implicada en la unión de la endoglucanasa a celulosa (lignocelulosa). Los módulos de unión a carbohidratos generalmente funcionan por unión de manera no covalente de la endoglucanasa a la celulosa, un derivado de celulosa o un equivalente polisacárido de la misma. Los CBMs típicamente funcionan independientes del dominio catalítico.

45

[0024] Constructo de proteína de fusión: el término "constructo de proteína de fusión" se refiere a un constructo de ácido nucleico que está compuesto de diferentes genes o partes de los mismos en enlace operativo. Los componentes incluyen desde el final 5' una molécula de ADN que codifica al menos un dominio catalítico de endoglucanasa y una molécula de ADN que codifica al menos un dominio catalítico de un polipéptido que tiene actividad biológica.

50

[0025] Constructo de fusión de beta-glucosidasa: el término "constructo de fusión de beta-glucosidasa" se refiere a un constructo de ácido nucleico que está compuesto por diferentes genes o partes de los mismos en enlace operativo. Los componentes incluyen desde el final 5' una molécula de ADN que codifica al menos un dominio catalítico de endoglucanasa y una molécula de ADN que codifica al menos un dominio catalítico de beta-glucosidasa.

55

[0026] 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, de forma más preferible al menos 60% puro, incluso de forma más preferible al menos 80% puro, de la forma más preferible al menos 90% puro, e incluso de la forma más preferible al menos 95% puro, tal y como se determina por SDS-PAGE.

60

[0027] Polipéptido sustancialmente puro: el término "polipéptido sustancialmente puro" se refiere aquí a una preparación de polipéptido que contiene como mucho 10%, preferiblemente como mucho 8%, de forma más preferible como mucho

6%, de forma más preferible como mucho 5%, de forma más preferible como mucho 4%, de forma más preferible como mucho 3%, incluso de forma más preferible 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 de polipéptido con el que está originalmente o de forma recombinante asociado. Por lo tanto, se prefiere que el polipéptido sustancialmente puro sea al menos 92% puro, preferiblemente al menos 94% puro, de forma más preferible al menos 95% puro, de forma más preferible al menos 96% puro, de forma más preferible al menos 96% puro, de forma más preferible al menos 97% puro, de forma más preferible al menos 98% puro, incluso de forma más preferible al menos 99%, de la forma más preferible al menos 99,5% puro, e incluso de la forma más preferible 100% puro en peso del material de polipéptido total presente en la preparación. Los polipéptidos de la presente invención están preferiblemente en una forma sustancialmente pura, es decir, que la preparación de polipéptido está esencialmente libre de otro material de polipéptido con el cual está originalmente o de forma recombinante asociado. Esto se puede conseguir, por ejemplo, preparando el polipéptido mediante métodos recombinantes bien conocidos o por métodos de purificación tradicionales.

[0028] Identidad: la relación entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos se describe por el parámetro "identidad".

[0029] Para fines de la presente invención, el grado de identidad entre dos secuencias de aminoácidos se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, J. Mol. Biol. 48: 443-453) tal y como se implementa en el programa Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, Trends in Genetics 16: 276- 277), preferiblemente versión 3.0.0 o superior. Los parámetros opcionales usados son penalización por apertura de espacio de 10, penalización por extensión de espacio de 0,5 y la matriz de sustitución EBLOSUM62 (versión EMBOSS de BLOSUM62). El resultado de Needle marcado "identidad más larga" (obtenido utilizando la opción -nobrief) se usa como la identidad en porcentaje y se calculada de la siguiente manera:

$$\text{(Residuos Idénticos x 100)/(Longitud de Alineamiento – Número Total de Espacios en el Alineamiento)}$$

[0030] Para fines de la presente invención, el grado de identidad entre dos secuencias de desoxirribonucleótidos se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, *supra*) tal y como se implementa en el programa Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *supra*), preferiblemente versión 3.0.0 o superior. Los parámetros opcionales utilizados son penalización por apertura de espacio de 10, penalización por extensión de espacio de 0,5 y la matriz de sustitución EDNAFULL (versión EMBOSS de NCBI NUC4.4). El resultado de Needle marcado "identidad más larga" (obtenido utilizando la opción -nobrief) se usa como la identidad en porcentaje y se calcula de la siguiente manera:

$$\text{(Desoxirribonucleótidos Idénticos x 100)/(Longitud de Alineamiento – Número Total de Espacios en el Alineamiento)}$$

[0031] Fragmento de polipéptido: el término "fragmento de polipéptido" se define aquí como un polipéptido que tiene uno o más aminoácidos eliminados del amino y/o carboxilo terminal de una proteína de fusión (por ejemplo, proteína de fusión de beta-glucosidasa) o componentes de lo mismo, donde el fragmento tiene actividad biológica (por ejemplo, actividad de beta-glucosidasa).

[0032] Subsecuencia: el término "subsecuencia" se define aquí como una secuencia de nucleótido que tiene uno o más nucleótidos eliminados del extremo 5' y/o 3' de un polinucleótido, donde la subsecuencia codifica un fragmento de polipéptido que tiene actividad biológica, por ejemplo, actividad de beta-glucosidasa o actividad de endoglucanasa.

[0033] Polinucleótido aislado: el término "polinucleótido aislado" tal y como se usa aquí se refiere a un polinucleótido que es al menos 20% puro, preferiblemente al menos 40% puro, de forma más preferible al menos 60% puro, incluso de forma más preferible 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 se determina por electroforesis de agarosa.

[0034] Polinucleótido sustancialmente puro: el término "polinucleótido sustancialmente puro" como se usa aquí se refiere a una preparación polinucleótida libre de otros nucleótidos extraños o indeseados y en una forma adecuada para su uso dentro de sistemas de producción de proteína genéticamente modificada. Así, un polinucleótido sustancialmente puro contiene como mucho 10%, preferiblemente como mucho 8%, de forma más preferible como mucho 6%, de forma más preferible como mucho 5%, de forma más preferible como mucho 4%, de forma más preferible como mucho 3%, incluso de forma más preferible 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 o de forma recombinante asociado. Un polinucleótido sustancialmente puro puede, no obstante, incluir regiones no traducidas 5' y 3' de origen natural, tales como promotores y terminadores. Se prefiere que el polinucleótido sustancialmente puro sea al menos 90% puro, preferiblemente al menos 92% puro, de forma más preferible al menos 94% puro, de forma más preferible al menos 95% puro, de forma más preferible al menos 96% puro, de forma más preferible al menos 97% puro, incluso de forma más preferible 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 están preferiblemente en una forma sustancialmente pura, es decir, que la preparación polinucleótida está esencialmente libre de otro material polinucleótido con el cual está

originalmente o de forma recombinante asociado. Los polinucleótidos pueden ser de origen genómico, ADNc, ARN, semisintético sintético o cualquier combinación de los mismos.

[0035] ADNc: el término "ADNc" se define aquí como una molécula de ADN que se puede preparar por transcripción inversa de una molécula de ARNm, madura, empalmada, obtenida a partir de una célula eucariota. Al ADNc le faltan secuencias de intrones que están normalmente presentes en el ADN genómico correspondiente. La transcripción de ARN primario inicial es un precursor para ARNm que se procesa a través de una serie de pasos antes de aparecer como ARNm empalmado maduro. Estos pasos incluyen la eliminación de secuencias de intrones por un proceso denominado empalme. Al ADNc derivado ARNm le falta, por lo tanto, cualquier secuencias de intrones.

[0036] Constructo de ácidos nucleicos: el término "constructo de ácidos nucleicos" como se utiliza en este caso se refiere a una molécula de ácido nucleico, bien uni- o bicatenaria, que se aísla a partir de un gen de origen natural o que se modifica para contener segmentos de ácidos nucleicos de manera que de otro modo no existirían en la naturaleza. El término constructo de ácidos nucleicos es sinónimo del término "casete de expresión" cuando el constructo de ácidos nucleicos contiene las secuencias de control requeridas para la expresión de una secuencia codificante.

[0037] Secuencia de control: el término "secuencia de control" se define aquí para incluir todos los componentes que son necesarios o ventajosos para la expresión de un polinucleótido que codifica un polipéptido. Cada secuencia de control puede ser nativa o foránea de la secuencia de nucleótidos que codifica el polipéptido o nativas o foráneas entre sí. Tales secuencias de control incluyen, pero de forma no limitativa, una secuencia de poliadenilación, líder, secuencia de propéptido, promotor, secuencia de péptido señal y terminador de transcripción. Como mínimo, las secuencias de control incluyen un promotor y señales de parada transcripcionales y traduccionales. Las secuencias de control pueden estar provistas de enlaces para introducir sitios de restricción específicos que facilitan el ligamiento de las secuencias de control con la región de codificación de la secuencia de nucleótidos que codifica un polipéptido.

[0038] Operativamente enlazado: el término "operativamente enlazado" denota aquí una configuración en la que una secuencia de control se coloca en una posición apropiada con respecto a la secuencia de codificación de una secuencia polinucleótida de modo que la secuencia de control dirija la expresión de la secuencia codificante de un polipéptido. Además, el término "operativamente enlazado" también se refiere a dos polinucleótidos que están enlazados o fusionados, que se expresan juntos como una proteína de fusión o fusionada.

[0039] Secuencia codificante: cuando aquí se usa el término "secuencia codificante" se refiere a una secuencia de nucleótidos, que especifica directamente la secuencia de aminoácidos de su producto proteínico. Los límites de la secuencia codificante están generalmente determinados por un marco de lectura abierto, que empieza normalmente con el codón de inicio ATG o codones de inicio alternativos tales como GTG y TTG y termina 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.

[0040] 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.

[0041] Vector de expresión: el término "vector de expresión" se define aquí como una molécula de ADN lineal o circular que comprende un polinucleótido que codifica un polipéptido de la invención y está operativamente enlazada a nucleótidos adicionales que proporcionan su expresión.

[0042] Célula huésped: el término "célula huésped", como se utiliza en este caso, incluye cualquier tipo de célula que sea susceptible de transformación, transfección, transducción y similar con un constructo de ácidos nucleicos o vector de expresión que comprenda un polinucleótido.

[0043] Variante: cuando se usa aquí, el término "variante" se refiere a un polipéptido que tiene actividad biológica producida por un organismo que expresa una secuencia de nucleótidos modificada, por ejemplo, SEC ID n°: 25 o una secuencia homóloga de la misma, o la región codificante madura de la misma. La secuencia de nucleótidos modificada se obtiene por intervención humana mediante modificación de una secuencia de nucleótidos, por ejemplo, SEC ID n°: 23 o una secuencia homóloga de la misma, o la región codificante madura de la misma. La modificación puede ser una sustitución, una eliminación y/o una inserción de uno o varios aminoácidos, al igual que una sustitución de una o varias cadenas laterales de aminoácidos.

Descripción detallada de la invención

[0044] La presente invención se refiere a métodos para producir un polipéptido segregado que tiene actividad biológica, que comprende: (a) transformación de una célula huésped fúngica con un constructo de proteína de fusión que codifica una proteína de fusión, donde el constructo de proteína de fusión comprende: (i) un primer polinucleótido que comprende una secuencia de nucleótidos que codifica un péptido señal; (ii) un segundo polinucleótido que comprende una secuencia de nucleótidos que codifica al menos un dominio catalítico de una endoglucanasa o una parte de la misma; y (iii) un tercer polinucleótido que comprende una secuencia de nucleótidos que codifica al menos un dominio

catalítico de un polipéptido que tiene actividad biológica o una parte de la misma, donde la proteína de fusión es SEC ID nº: 74;

5 (b) cultivo de la célula huésped fúngica transformada bajo condiciones adecuadas para la producción de la proteína de fusión; y (c) recuperación de la proteína de fusión, un componente de la misma, o una combinación de los mismos, del medio de cultivo, donde la proteína de fusión o su componente tienen actividad biológica.

10 [0045] Una proteína de fusión se produce por fusión de una secuencia de nucleótidos que codifica un polipéptido que tiene actividad biológica o una parte del mismo con una secuencia de nucleótidos que codifica un polipéptido que tiene actividad de endoglucanasa o una parte del mismo y una secuencia de nucleótidos que codifica un péptido señal operativamente enlazada a la secuencia de nucleótidos que codifica el polipéptido que tiene actividad de endoglucanasa o una parte del mismo. Técnicas para producir proteínas de fusión se conocen en la técnica, e incluyen, por ejemplo, ligar las secuencias de codificación que codifican los polipéptidos de modo que éstas estén en marco y la expresión del polipéptido fusionado esté bajo el control del mismo promotor(es) y terminador. También se pueden construir proteínas de fusión utilizando tecnología de inteína donde se crean fusiones post-traduccionales (Cooper *et al.*, 1993, EMBO J. 12: 2575-2583 ; Dawson *et al.*, 1994, Science 266: 776- 779).

20 [0046] Una proteína de fusión que tiene actividad biológica que comprende un péptido señal, al menos el dominio catalítico de una endoglucanasa o una parte de la misma, y al menos el dominio catalítico de un polipéptido que tiene actividad biológica o una parte del mismo, aumenta la secreción de la proteína de fusión en comparación con la ausencia de al menos el dominio catalítico de la endoglucanasa o una parte de la misma. El aumento en la secreción de la proteína de fusión que tiene actividad biológica es al menos 5%, preferiblemente al menos 10%, de forma más preferible al menos 25%, incluso de forma más preferible al menos 50%, de forma más preferible al menos 100%, incluso de forma más preferible al menos 150%, incluso de forma más preferible al menos 200%, de la forma más preferible al menos 500%, e incluso de la forma más preferible al menos 1000% en comparación con la ausencia de al menos el dominio catalítico de la endoglucanasa.

30 [0047] En cada uno de los aspectos preferidos que aparecen más abajo, los componentes de un constructo de proteína de fusión (constructo de ácidos nucleicos) están operativamente enlazados desde el final 5' al final 3' del constructo.

[0048] En cada uno de los aspectos preferidos anteriores, los componentes de los constructos de proteína de fusión además comprenden una región promotora y/o una región terminadora.

35 Endoglucanasas y polinucleótidos de las mismas

[0049] Un polinucleótido que codifica un dominio catalítico, polipéptido maduro, o polipéptido en toda su longitud de una endoglucanasa, o partes de la misma, se pueden obtener a partir de cualquier organismo. Para fines de la presente invención, el término "polipéptido" se entenderá que incluye un polipéptido en toda su longitud, polipéptido maduro o dominio catalítico; o partes o fragmentos de los mismos que tienen actividad. El término "obtenido a partir de", como se utiliza en este caso en relación con una fuente dada, se refiere a que el polipéptido codificado por una secuencia de nucleótidos es producido por la fuente o por una cepa en la que la secuencia de nucleótidos de la fuente se ha insertado.

45 [0050] Muchas endoglucanasas tienen una estructura multidominio que consiste en un dominio catalítico separado de un dominio de unión a carbohidratos (CBM) por un péptido de enlace (Suurnakki *et al.*, 2000, Cellulose 7: 189-209). El dominio catalítico contiene el sitio activo, mientras que el CBM interactúa con la celulosa por unión de la enzima para ésta (van Tilbeurgh *et al.*, 1986, FEBS Letters 204: 223-227 ; Tomme *et al.*, 1988, European Journal of Biochemistry 170: 575-581).

50 [0051] Un polinucleótido que codifica un polipéptido que tiene actividad de endoglucanasa se puede obtener a partir de un gen que codifica un polipéptido bacteriano. Por ejemplo, el polipéptido puede ser un polipéptido bacteriano gram positivo incluyendo, pero no limitado a, un polipéptido de *Bacillus*, *Streptococcus*, *Streptomyces*, *Staphylococcus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus* u *Oceanobacillus*, por ejemplo, un polipéptido de *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis*, *Bacillus thuringiensis*, *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis* y *Streptococcus equi subsp. Zooepidemicus*, *Streptomyces lividans* o *Streptomyces murinus*; o un polipéptido bacteriano gram negativo incluyendo, pero no limitado a, un polipéptido de *E. coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Ilyobacter*, *Flavobacterium*, *Fusobacterium*, *Ilyobacter*, *Neisseria* o *Ureaplasma*.

60 [0052] Ejemplos de endoglucanasas bacterianas que se pueden usar como fuentes para los polinucleótidos en los métodos de la presente invención, incluyen, pero de forma no limitativa, una endoglucanasa de *Acidothermus cellulolyticus* (WO 91/05039; WO 93/15186; patente de EE.UU. nº 5.275.944; WO 96/02551; patente de EE.UU. nº 5.536.655, WO 00/70031, WO 05/093050); endoglucanasa III de *Thermobifida fusca* (WO 05/093050); y endoglucanasa V de *Thermobifida fusca* (WO 05/093050).

[0053] En otro aspecto preferido, un polinucleótido que codifica un polipéptido que tiene actividad de endoglucanasa se puede obtener a partir de un polipéptido de *Humicola insolens*.

[0054] Las técnicas usadas para aislar o clonar un polinucleótido que codifica un polipéptido que tiene actividad de endoglucanasa se conocen en la técnica e incluyen aislamiento de ADN genómico, preparación de ADNc o una combinación de los mismos. La clonación de los polinucleótidos a partir de tal ADN genómico se puede efectuar, por ejemplo, usando la reacción en cadena de polimerasa bien conocida (PCR) o la selección de anticuerpo de bibliotecas de expresión para detectar fragmentos de ADN clonado 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 ácido nucleico tales como reacción en cadena de la ligasa (LCR), transcripción activada ligada (LAT) y amplificación basada en secuencias de nucleótidos (NASBA) se pueden utilizar.

[0055] Se entenderá que para las especies anteriormente mencionadas la invención abarca tanto los estados perfectos como los imperfectos, y otros equivalentes taxonómicos, por ejemplo, anamorfos, independientemente del nombre de la especie por el que se conozcan. Expertos en la técnica reconocerán fácilmente la identidad de equivalentes apropiados.

[0056] Cepas de estas especies son de fácil acceso para el público en diferentes colecciones de cultivo, tales como la *American Type Culture Collection (ATCC)*, *Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM)*, *Centraalbureau Voor Schimmelcultures (CBS)* y *Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL)*.

[0057] En otro aspecto preferido, el polipéptido en toda su longitud, polipéptido maduro o dominio catalítico de la endoglucanasa, o una parte del mismo, es codificado por un polinucleótido obtenido a partir de un gen de endoglucanasa VI.

[0058] En otro aspecto preferido, el polipéptido en toda su longitud, polipéptido maduro o dominio catalítico de la endoglucanasa, o una parte del mismo, es codificado por un polinucleótido obtenido a partir de un gen de endoglucanasa de la familia 45.

Polipéptidos que tienen actividad biológica y polinucleótidos de los mismos

[0059] En un aspecto más preferido, el polipéptido es una endoglucanasa. En otro aspecto más preferido, el polipéptido es una celobiohidrolasa.

[0060] En otro aspecto preferido, un polinucleótido que codifica un polipéptido que tiene actividad de endoglucanasa se puede obtener a partir de *Humicola insolens*.

[0061] En otro aspecto preferido, el polipéptido en toda su longitud, polipéptido maduro o dominio catalítico de la beta-glucosidasa, o una parte del mismo, es codificado por un polinucleótido obtenido a partir de un gen de beta-glucosidasa de *Aspergillus oryzae*. En un aspecto más preferido, el polipéptido de longitud completo, polipéptido maduro o dominio catalítico de la beta-glucosidasa, o una parte del mismo, es codificado por un polinucleótido obtenido a partir de un gen de beta-glucosidasa de *Aspergillus oryzae* que comprende SEC ID n°: 23 que codifica el polipéptido de SEC ID n°: 24 o un gen mutante de beta-glucosidasa de *Aspergillus oryzae* que comprende SEC ID n°: 25 que codifica el polipéptido de SEC ID n°: 26.

[0062] En otro aspecto preferido, la beta-glucosidasa se segrega naturalmente. En otro aspecto preferido, la beta-glucosidasa no se segrega naturalmente.

Péptidos señal

[0063] El péptido señal puede ser cualquier péptido señal apropiado reconocido por una célula huésped para la secreción extracelular de una proteína de fusión de la presente invención. La secuencia señal es preferiblemente la que está naturalmente asociada al componente de endoglucanasa de la proteína de fusión que se va a expresar.

[0064] El final 5' de la secuencia codificante de la secuencia de nucleótidos que codifica un polipéptido puede contener intrínsecamente una región codificante del péptido señal naturalmente enlazado en el marco de lectura de traducción con el segmento de la región de codificación que codifica el polipéptido segregado. Alternativamente, el final 5' de la secuencia codificante puede contener una región codificante del péptido señal que sea foránea a la secuencia codificante. La región codificante del péptido señal foránea puede ser necesaria donde la secuencia codificante no contenga naturalmente una región codificante del péptido señal. Alternativamente, la región codificante del péptido señal foránea 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 dirija la proteína de fusión expresada en la vía secretora de una célula huésped de elección, es decir, segregada en un medio de cultivo, se puede utilizar en la presente invención.

[0065] Secuencias de codificación de péptido señal eficaces para células huésped fúngicas filamentosas son la secuencia codificante del péptido señal obtenida a partir de la endoglucanasa de *Humicola insolens*.

5 [0066] En un aspecto preferido, la secuencia codificante del péptido señal se obtiene a partir de un gen de *Humicola insolens* que codifica una endoglucanasa V. En un aspecto más preferido, la secuencia codificante de péptido señal codifica los aminoácidos 1 a 21 de SEC ID n°: 2. En otro aspecto más preferido, la secuencia codificante del péptido señal es los nucleótidos 1 a 63 de SEC ID n°: 1.

10 [0067] La proteína de fusión puede comprender además un segundo péptido señal que se una con el componente de beta-glucosidasa de la proteína de fusión. La secuencia codificante del péptido señal puede ser la secuencia codificante del péptido señal que esté naturalmente asociada a la secuencia codificante del polipéptido que tiene actividad biológica o puede ser una secuencia codificante del péptido señal diferente, tal como una de las descritas anteriormente.

15 [0068] En un aspecto preferido, la secuencia codificante del péptido señal es una secuencia naturalmente asociada a una secuencia codificante de beta-glucosidasa.

Enlaces

20 [0069] Tal y como se menciona *supra*, muchas endoglucanasas tienen una estructura multidominio que consiste en un dominio catalítico separado de uno o más módulos de unión a carbohidratos por un(os) péptido(s) enlazador(es). En los métodos de la presente invención, los constructos de proteína de fusión pueden además comprender una secuencia codificante de enlazador localizada 3' a la secuencia que comprende el dominio catalítico de endoglucanasa y 5' a la secuencia que comprende el dominio catalítico del polipéptido que tiene actividad biológica.

25 [0070] El enlazador se puede obtener a partir del mismo gen que el dominio catalítico de la endoglucanasa o a partir de un gen de endoglucanasa diferente. Por otro lado, el enlazador puede ser sintético en el origen.

30 [0071] Ejemplos de enlaces que se pueden usar en los métodos de la presente invención incluyen, pero de forma no limitativa, enlaces obtenidos a partir de los genes para la celobiohidrolasa I de *Trichoderma reesei* (Srisodsuk *et al.*, 1993, Journal of Biological Chemistry 268: 20765-20761); celobiohidrolasa Cel7A de *Hypocrea jecorina* (precedentemente *Trichoderma reesei*) (Mulakala *et al.*, 2005, Proteins 60: 598-605); endoglucanasa V de *Humicola insolens*; y endoglucanasa NRRL 8126 CEL7C *Thielavia terrestris*.

35 [0072] En un aspecto preferido, el enlazador se obtiene a partir de un gen de endoglucanasa de *Humicola insolens*. En otro aspecto preferido, el enlazador se obtiene a partir de un gen de endoglucanasa de *Trichoderma reesei*. En un aspecto más preferido, el enlazador se obtiene a partir de una endoglucanasa V del gen de *Humicola insolens* (eg5).

40 [0073] En otro aspecto preferido, el enlazador se obtiene a partir de un gen de endoglucanasa de *Thielavia terrestris*. En otro aspecto más preferido, el enlazador se obtiene a partir de un gen de endoglucanasa NRRL 8126 CEL7C de *Thielavia terrestris*.

45 [0074] En un aspecto preferido, el enlazador es al menos 5 residuos de aminoácidos. En un aspecto más preferido, el enlazador es al menos 15 residuos de aminoácidos. En un aspecto más preferido, el enlazador es al menos 25 residuos de aminoácidos.

50 [0075] En un aspecto preferido, el enlazador es entre aproximadamente 5 a aproximadamente 60 residuos de aminoácidos. En un aspecto más preferido, el enlazador es entre aproximadamente 15 a aproximadamente 50 residuos de aminoácidos. En un aspecto más preferido, el enlazador es entre aproximadamente 25 a aproximadamente 45 residuos de aminoácidos.

Módulos de unión a carbohidratos

55 [0076] Los módulos de unión a carbohidratos (CBMs) se definen como secuencias de aminoácidos contiguos con un pliegue discreto que tienen actividad de unión a carbohidratos y se encuentran comúnmente dentro de enzimas activas de carbohidratos. Varios tipos de CBMs se han descrito, y una mayoría de los mismos se enlazan a polisacáridos insolubles (véase Boraston *et al.*, 2004, Biochem J. 382: 769-781). Se han caracterizado módulos de unión a carbohidratos que median interacción con, por ejemplo, celulosa cristalina, celulosa no cristalina, quitina, beta-1,3-glucanos y glucanos de enlace beta-1,3-1,4-meclado, xilano, manano, galactano y almidón. Los módulos de unión a carbohidratos ocurren frecuentemente en celulasas multidominio. Mientras algunos CBMs confieren enlace específico a un subconjunto de estructuras de carbohidratos, otros son más generales en su capacidad para asociarse con varios polisacáridos. Los CBMs que confieren unión a celulosa a veces se denominan dominios de enlace a celulosa, o CBDs (Boraston *et al.*, 2004, Biochem J. 382: 769-781). Los CBMs se reagrupan por similitud de aminoácido; actualmente, hay descritas 48 familias de CBM.

65 [0077] Las glicosido hidrolasas pueden comprender más de un dominio catalítico y uno, dos, tres o más CBMs, y opcionalmente además comprenden una o varias regiones de secuencia de aminoácidos de polipéptido que conectan el

CBM(s) con el dominio(s) catalítico(s), una región del último tipo normalmente se denomina un "enlazador". Ejemplos de enzimas hidrolíticas que comprenden un CBM son celulasas, xilanasas, mananasas, arabinofuranosidasas, acetilsterasas y quitinasas (véase P. Tomme *et al.*, Cellulose-Binding Domains - Classification and Properties in Enzymatic Degradation of Insoluble Carbohydrates, John N. Saddler and Michael H. Penner (Eds.), ACS Symposium Series, No. 618, 1996).

[0078] Un CBM se puede localizar en el N o C terminal o en una posición interna de una proteína o polipéptido. La región de un polipéptido o proteína que constituye un CBM típicamente consiste en más de aproximadamente 30 y menos de aproximadamente 250 residuos de aminoácidos. Por ejemplo: aquellos CBMs enumerados y clasificados en la familia I consisten en 33-37 residuos de aminoácidos, aquellos enumerados y clasificados en la familia 2a consisten en 95-108 residuos de aminoácidos, y aquellos enumerados y clasificados en la familia 6 consisten en 85-92 residuos de aminoácidos. Por consiguiente, el peso molecular de una secuencia de aminoácidos que constituye un CBM estará típicamente en la gama de aproximadamente 4 kDa a aproximadamente 40 kDa, y normalmente por debajo de aproximadamente 35 kDa.

[0079] En los métodos de la presente invención, cualquier CBM puede ser utilizado. El CBM puede estar naturalmente asociado a la endoglucanasa o puede ser foráneo de la endoglucanasa.

[0080] En un aspecto preferido, un CBM se obtiene a partir de un gen de endoglucanasa (EG) de *Trichoderma reesei*. En un aspecto más preferido, un CBM se obtiene a partir de un gen de endoglucanasa EGI de *Trichoderma reesei*. En otro aspecto más preferido, un CBM se obtiene a partir de un gen de endoglucanasa EGII de *Trichoderma reesei*. En otro aspecto más preferido, un CBM se obtiene a partir de un EGV de endoglucanasa de *Trichoderma reesei*.

[0081] En otro aspecto preferido, un CBM se obtiene a partir de un gen de celobiohidrolasa (CBH) de *Trichoderma reesei*. En otro aspecto preferido, un CBM se obtiene a partir de un gen CBHI de *Trichoderma reesei* (Terri *et al.*, 1987, Gene 51: 42-52 ; Linder and Teeri, 1996, Biochemistry 93: 12251-12255). En otro aspecto preferido, un CBM se obtiene a partir de un gen CBHII de *Trichoderma reesei*.

[0082] En otro aspecto preferido, un CBM se obtiene a partir de un gen de endoglucanasa de *Thielavia terrestris*. En otro aspecto más preferido, un CBM se obtiene a partir de un gen de endoglucanasa NRRL 8126 CEL7C de *Thielavia terrestris*.

Sitio de escisión

[0083] En los métodos de la presente invención, los constructos de proteína de fusión pueden además comprender una secuencia de nucleótidos que codifica un sitio de escisión. El sitio de escisión está preferiblemente localizado entre la secuencia que comprende al menos el dominio catalítico de la endoglucanasa y la secuencia que comprende al menos el dominio catalítico del polipéptido que tiene actividad biológica. En la secreción de la proteína de fusión, el sitio se divide liberando el polipéptido que tiene actividad biológica a partir de la proteína de fusión.

[0084] Ejemplos de sitios de escisión incluyen, pero de forma no limitativa, un sitio Kex2 que codifica el dipéptido Lys-Arg (Martin *et al.*, 2003, J. Ind. Microbiol. Biotechnol. 3: 568-76 ; Svetina *et al.*, 2000, J. Biotechnol. 76: 245-251 ; Rasmussen-Wilson *et al.*, 1997, Appl. Environ. Microbiol. 63: 3488-3493 ; Ward *et al.*, 1995, Biotechnology 13: 498-503 ; and Contreras *et al.*, 1991, Biotechnology 9: 378-381), un sitio Ile-(Glu o Asp)-Gly-Arg, que es dividido por una proteasa de factor Xa después del residuo de arginina (Eaton *et al.*, 1986, Biochem. 25: 505-512); un sitio Asp-Asp-Asp-Asp-Ly, que es dividido por una enteroquinasa después de la lisina (Collins-Racie *et al.*, 1995, Biotechnology 13: 982-987); un sitio His- Tyr-Glu o un sitio His-Tyr-Asp, que es dividido por genenasa I (Carter *et al.*, 1989, Proteins: Structure, Function, and Genetics 6: 240-248); un sitio Leu-Val-Pro-Arg-Gly-Ser, que es dividido por trombina después del Arg (Stevens, 2003, Drug Discovery World 4: 35-48); un sitio Glu-Asn-Leu-Tyr-Phe-Gln-Gly, que es dividido por proteasa de TEV después del Gln (Stevens, 2003, *supra*); y un sitio Leu-Glu-Val- Leu-Phe-Gln-Gly-Pro, que es dividido por una forma genéticamente modificada de proteasa de rinovirus humano 3C después del Gln (Stevens, 2003, *supra*).

Proteínas de fusión

[0085] Una proteína de fusión que tiene actividad biológica de la presente invención que comprende un péptido señal, al menos el dominio catalítico de una endoglucanasa o una parte de la misma, y al menos el dominio catalítico de un polipéptido que tiene actividad biológica o una parte del mismo, aumenta la secreción de la proteína de fusión en comparación con la ausencia de al menos el dominio catalítico de la endoglucanasa o una parte del mismo. En cada uno de los aspectos preferidos a continuación, los componentes de una proteína de fusión se enlazan en marco desde el N terminal al C terminal de la proteína.

[0086] En un aspecto preferido, la proteína de fusión comprende un péptido señal, un polipéptido maduro de una endoglucanasa y un dominio catalítico de un polipéptido que tiene actividad biológica.

Promotores

[0087] La región promotora puede ser cualquier secuencia promotora apropiada reconocida por una célula huésped para la expresión de una proteína de fusión. La secuencia promotora contiene secuencias de control transcripcionales que median la expresión del polipéptido. El promotor puede ser cualquier secuencia de nucleótidos que muestre actividad transcripcional en la célula huésped de elección incluyendo promotores mutantes, truncados, tándem e híbridos, y se pueden obtener a partir de genes que codifican polipéptidos extracelulares o intracelulares bien nativos o foráneos (heterólogos) de la célula huésped. Promotores ejemplares incluyen tanto promotores constitutivos como promotores inducibles.

[0088] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de proteína de fusión 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 EE.UU. 75: 3727-3731), al igual que el promotor tac (DeBoer *et al.*, 1983, Proceedings of the National Academy of Sciences EE.UU. 80: 21-25). Otros promotores se describen en "Useful proteins from recombinant bacteria" in Scientific American, 1980, 242: 74-94; y en Sambrook *et al.*, 1989, *supra*.

[0089] Ejemplos de promotores adecuados para dirigir la transcripción de los constructos de proteína de fusión de la presente invención en una célula huésped fúngica filamentosa son promotores obtenidos a partir 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 *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), Daria de *Fusarium venenatum* (WO 00/56900), Quinn de *Fusarium venenatum* (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*, xilanasa I de *Trichoderma reesei*, xilanasa II de *Trichoderma reesei*, beta-xilosidasa de *Trichoderma reesei*, beta-tubulina de *Coprinus cinereus* y swollenin 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 promotores mutantes, truncados e híbridos de los mismos.

[0090] En un huésped de levadura, se obtienen promotores útiles a partir de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), galactocinasa de *Saccharomyces cerevisiae* (GAL1), alcohol dehidrogenasa/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 huésped de levadura se describen en Romanos *et al.*, 1992, Yeast 8: 423-488.

[0091] En un aspecto preferido, el promotor es un promotor de celobiohidrolasa. En un aspecto más preferido, el promotor es un promotor de celobiohidrolasa I (*cbh1*). En un aspecto más preferido incluso, el promotor es un promotor del gen de celobiohidrolasa I de *Trichoderma reesei* (*cbh1*). En un aspecto más preferido, el promotor es el promotor *cbh1* de *Trichoderma reesei* de los nucleótidos 505 a 1501 de SEC ID n°: 29 (número de registro D86235 del GENBANK™). En otro aspecto más preferido, el promotor es un promotor de celobiohidrolasa II (*cbh2*). En otro aspecto más preferido incluso, el promotor es un promotor del gen de celobiohidrolasa II de *Trichoderma reesei* (*cbh2*). En otro aspecto más preferido, el promotor es el promotor de *Trichoderma reesei* *cbh2* de los nucleótidos 1 a 582 de SEC ID n°: 30 (número de registro M55080 del GENBANK™).

[0092] En otro aspecto preferido, el promotor es el promotor NA2-tpi. En otro aspecto preferido, el promotor es un promotor de TAKA amilasa. En otro aspecto preferido, el promotor es un promotor de amiloglucosidasa de *Fusarium venenatum*. En otro aspecto preferido, el promotor es un promotor de proteasa de tipo tripsina de *Fusarium oxysporum*. En otro aspecto preferido, el promotor es un promotor de glucoamilasa de *Aspergillus niger* o *Aspergillus awamori* (*glaA*).

[0093] En un aspecto preferido, la región promotora dirige la expresión del primer, segundo y tercer polinucleótido, y alternativamente también el cuarto polinucleótido.

Terminadores

[0094] El terminador puede ser cualquier secuencia terminadora de transcripción adecuada reconocida por una célula huésped para terminar la transcripción. La secuencia terminadora está operativamente enlazada al 3' terminal de la secuencia de nucleótidos que codifica el polipéptido. Cualquier terminador que sea funcional en la célula huésped de elección se puede utilizar en la presente invención.

[0095] Terminadores preferidos para células huésped fúngicas filamentosas se obtienen a partir 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*.

5 [0096] Terminadores preferidos para células huésped de levadura se obtienen a partir 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 huésped de levadura son descritos por Romanos *et al.*, 1992, *supra*.

10 [0097] En un aspecto preferido, el terminador es un terminador del gen de celobiohidrolasa. En un aspecto más preferido, el terminador es un terminador del gen de celobiohidrolasa I (*cbh1*). En un aspecto más preferido incluso, el terminador es un terminador del gen de celobiohidrolasa I de *Trichoderma reesei* (*cbh1*). En un aspecto más preferido, el terminador es el terminador de *Trichoderma reesei cbh1* de SEC ID n°: 31. En otro aspecto más preferido, el terminador es un terminador del gen de celobiohidrolasa II (*cbh2*). En otro aspecto más preferido incluso, el terminador es un terminador del gen de celobiohidrolasa II de *Trichoderma reesei* (*cbh2*). En otro aspecto más preferido, el terminador es el terminador de *Trichoderma reesei cbh2* de SEC ID n°: 32.

[0098] En otro aspecto preferido, el terminador es un terminador de TAKA amilasa. En otro aspecto preferido, el promotor es un terminador de proteasa de tipo tripsina de *Fusarium oxysporum*. En otro aspecto preferido, el promotor es un terminador de glucoamilasa de *Aspergillus niger* o *Aspergillus awamori* (*glaA*).

Otras secuencias reguladoras

25 [0099] Los constructos de proteína de fusión pueden además comprender otros elementos reguladores tales como una secuencia líder de poliadenilación y otros elementos.

[0100] La secuencia reguladora también puede ser una secuencia líder adecuada, una región no traducida de un ARNm que sea importante para la traducción por la célula huésped. La secuencia líder está operativamente enlazada al 5' terminal de la secuencia de nucleótidos que codifica una proteína de fusión. Cualquier secuencia líder que sea funcional en la célula huésped de elección se puede utilizar en la presente invención.

[0101] Líderes que se prefieren para células huésped fúngicas filamentosas se obtienen a partir de los genes para TAKA amilasa de *Aspergillus oryzae* y triosa fosfato isomerasa de *Aspergillus nidulans*.

35 [0102] Líderes adecuados para células huésped de levadura se obtienen a partir de los genes para enolasa de *Saccharomyces cerevisiae* (ENO-1), 3-fosfoglicerato quinasa de *Saccharomyces cerevisiae*, alfa-factor de *Saccharomyces cerevisiae* y alcohol deshidrogenasa/gliceraldehído-3-fosfato deshidrogenasa de *Saccharomyces cerevisiae* (ADH2/GAP).

40 [0103] La secuencia reguladora también puede ser una secuencia de poliadenilación, una secuencia operativamente enlazada al 3' terminal de la secuencia de nucleótidos y, cuando se transcribe, es reconocida por la célula huésped como una señal para añadir residuos de poliadenosina al ARNm transcrito. Cualquier secuencia de poliadenilación que sea funcional en la célula huésped de elección se puede utilizar en la presente invención.

45 [0104] Secuencias de poliadenilación preferidas para células huésped fúngicas filamentosas se obtienen a partir 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*.

50 [0105] Secuencias de poliadenilación útiles para células huésped de levadura se describen en Guo y Sherman, 1995, Molecular Cellular Biology 15 5983-5990.

[0106] También puede ser deseable añadir secuencias reguladoras que permitan la regulación de la expresión del polipéptido con respecto al crecimiento de la célula huésped. Ejemplos de sistemas reguladores son aquellos que hacen que la expresión del gen se active (ON) o desactive (OFF) en respuesta a una sustancia química o estímulo físico, incluyendo la presencia de un compuesto regulador. Sistemas reguladores en sistemas procarióticos incluyen los sistemas operadores *lac*, *tac* y *trp*. En la levadura, el sistema ADH2 o sistema GAL1 se pueden utilizar. En hongos filamentosos, el promotor de TAKA alfa-amilasa, el promotor de glucoamilasa de *Aspergillus niger* y el promotor de glucoamilasa de *Aspergillus oryzae* se pueden utilizar como secuencias reguladoras. Otros ejemplos de secuencias reguladoras son aquellas que permiten amplificación génica. En sistemas eucarióticos, incluyen el gen de dihidrofolato reductasa, que se amplifica en presencia de metotrexato, y los genes de metalotioneína, que se amplifican con metales pesados. En estos casos, la secuencia de nucleótidos que codifica el polipéptido estaría operativamente enlazada con la secuencia reguladora.

65 [0107] Los constructos de proteína de fusión preferiblemente contienen uno o varios marcadores seleccionables que permiten una fácil selección de células transformadas, modificadas, transducida o similares. Un marcador seleccionable

es un gen cuyo producto proporciona resistencia biocida o vírica, resistencia a metales pesados, prototrofia a auxótrofos y similares.

[0108] Ejemplos de marcadores seleccionables bacterianos son los genes *dal* de *Bacillus subtilis* o *Bacillus licheniformis*, o marcadores que confieren resistencia antibiótica tales como resistencia a la ampicilina, canamicina, cloranfenicol o tetraciclina. Marcadores adecuados para células huésped de levadura son ADE2, HIS3, LEU2, LYS2, MET3, TRP1 y URA3. Marcadores seleccionables para su uso en una célula huésped fúngica filamentosa incluyen, pero de forma no 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. Se prefiere para su uso en una célula de *Aspergillus* los genes *amdS* y *pyrG* de *Aspergillus nidulans* o *Aspergillus oryzae* y el gen *bar* de *Streptomyces hygroscopicus*.

Vectores de expresión

[0109] La presente invención también se refiere a vectores de expresión recombinantes que comprenden un polinucleótido que codifica una proteína de fusión, un promotor y señales de parada transcripcional y traduccional. Los diferentes ácidos nucleicos y secuencias de control descritos aquí se pueden juntar juntos para producir un vector de expresión recombinante que puede incluir uno o varios sitios de restricción adecuados para permitir la inserción o la sustitución de la secuencia de nucleótidos que codifica el polipéptido en tales sitios. Alternativamente, un polinucleótido que codifica una proteína de fusión se puede expresar por inserción de la secuencia de nucleótidos o un constructo de proteína de fusión que comprende la secuencia en un vector apropiado para la expresión. Al crear el vector de expresión, la secuencia codificante se localiza en el vector de modo que la secuencia codificante esté operativamente enlazada a las secuencias de control apropiadas para la expresión.

[0110] El vector de expresión recombinante puede ser cualquier vector (por ejemplo, un plásmido o virus) que se puede someter convenientemente a procedimientos de ADN recombinante y puede provocar la expresión de la secuencia de nucleótidos. La elección del vector típicamente dependerá de la compatibilidad del vector con la célula huésped en la que el vector debe ser introducido. Los vectores pueden ser plásmidos circulares lineales o cerrados.

[0111] El vector puede ser un vector de replicación autónoma, es decir, un vector que exista como una entidad extracromosómica, cuya replicación sea 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, cuando se introduce en la célula huésped, se integre en el genoma y replique con el cromosoma(s) en el que se haya integrado. Además, un único vector o plásmido o dos o más vectores o plásmidos que juntos contengan el ADN total que se va a introducir en el genoma de la célula huésped, o un transposón se puede utilizar.

[0112] Los vectores de la presente invención preferiblemente contienen uno o varios marcadores seleccionables que permiten la fácil selección de células transformadas, modificadas, transducidas o similares. Ejemplos de marcadores seleccionables bacterianos, de levadura, fúngicos filamentosos se describen aquí.

[0113] Un vector de la presente invención preferiblemente contiene un elemento(s) que permite la integración del vector en el genoma de la célula huésped o la replicación autónoma del vector en la célula independiente del genoma.

[0114] Para la integración en el genoma de la célula huésped, el vector puede depender de la secuencia que codifica del polinucleótido el polipéptido o cualquier otro elemento del vector para la 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 ubicación(es) precisa en el cromosoma(s). Para aumentar la probabilidad de integración en una ubicación precisa, los elementos integracionales deberían contener preferiblemente un número suficiente de ácidos nucleicos, tal como de 100 a 10.000 pares de bases, preferiblemente de 400 a 10.000 pares de bases, y de la forma más preferible de 800 a 10.000 pares de bases, que tienen un grado alto de identidad con la secuencia objetivo correspondiente para mejorar la probabilidad de recombinación homóloga. Los elementos integracionales pueden ser cualquier secuencia que sea homóloga de la secuencia objetivo en el genoma de la célula huésped. Además, los elementos integracionales pueden ser secuencias de nucleótidos no codificantes o codificantes. Por otro lado, el vector se puede integrar en el genoma de la célula huésped por recombinación no homóloga.

[0115] Para la replicación autónoma, el vector puede comprender además un origen de replicación que permita al vector replicar de manera autónoma en la célula huésped en cuestión. El origen de replicación puede ser cualquier replicador plásmido mediando replicación autónoma que funcione en una célula. El término "origen de replicación" o "replicador plásmido" se define aquí como una secuencia de nucleótidos que permite a un plásmido o vector replicar in vivo.

[0116] Ejemplos de orígenes bacterianos de replicación son los orígenes de replicación de los plásmidos pBR322, pUC19, pACYC177 y pACYC184 permitiendo la replicación en el *E. coli*, y pUB110, pE194, pTA1060 y pAMB1 permitiendo la replicación en *Bacillus*.

[0117] Ejemplos de orígenes de replicación para su uso en una célula huésped de levadura son el origen de replicación de 2 micras, ARS1, ARS4, la combinación de ARS1 y CEN3 y la combinación de ARS4 y CEN6.

5 [0118] Ejemplos de orígenes de replicación útiles en una célula fúngica 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 que comprenden el gen se puede realizar según los métodos descritos en la WO 00/24883.

10 [0119] Más de una copia de un polinucleótido que codifica una proteína de fusión se puede insertar en la célula huésped para aumentar la producción del producto génico. 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 células que contienen copias amplificadas del gen marcador seleccionable, y por lo tanto copias adicionales del polinucleótido, se pueden seleccionar por cultivo de las
15 células en presencia del agente seleccionable apropiado.

[0120] 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 materia (véase, por ejemplo, Sambrook *et al.*, 1989, *supra*).

20 Células huésped

[0121] La presente invención también se refiere a células huésped fúngicas recombinantes, que comprenden un polinucleótido que codifica una proteína de fusión de la presente invención, que se usan ventajosamente en la
25 producción recombinante de la proteína. Un vector que comprende un polinucleótido de la presente invención se introduce en una célula huésped fúngica de modo que el vector se mantenga como un integrante cromosómico o como un vector extracromosomal que se duplica como se se ha descrito anteriormente. El término "célula huésped" abarca cualquier descendiente de una célula madre que no sea idéntico a la célula madre debido a mutaciones que se produzcan durante la replicación.

30 [0122] La célula huésped fúngica puede ser cualquier célula fúngica útil en la producción recombinante de un polipéptido de la presente invención.

[0123] "Hongo", como se utiliza en este caso, incluye los filos *Ascomycota*, *Basidiomycota*, *Chytridiomycota* y *Zygomycota* (como se definen por Hawkswort *et al.*, en, Ainsworth and Bisby's Dictionary of The Fungi, 8th edition, 35 1995, CAB International, University Press, Cambridge, Reino Unido) al igual que la *Oomycota* (como se cita en Hawkswort *et al.*, 1995, *supra*, página 171) y todos hongos mitospóricos (Hawkswort *et al.*, 1995, *supra*).

[0124] En un aspecto preferido, la célula huésped fúngica es una célula de levadura. "Levadura", como se utiliza en este caso, incluye levadura ascosporógena (*Endomycetales*), levadura basidioesporogénea y levadura de los *Fungi Imperfecti* (*Blastomycetes*). Dado que la clasificación de levadura puede cambiar en el futuro, para los fines de esta invención, la levadura debe ser definida como se describe en Biology and Activities of Yeast (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

45 [0125] En un aspecto más preferido, la célula huésped de levadura es una célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*.

[0126] En un aspecto más preferido, la célula huésped de levadura es una célula de *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*,
50 *Saccharomyces norbensis* o *Saccharomyces oviformis*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Kluyveromyces lactis*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.

[0127] En otro aspecto preferido, la célula huésped fúngica es una célula fúngica filamentosa. "Hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión Eumycota y Oomycota (como se define por Hawkswort *et al.*, 1995, *supra*). Los hongos filamentosos se caracterizan generalmente 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.

60 [0128] En un aspecto más preferido, la célula huésped fúngica filamentosa es una célula de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Chrysosporium*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filibasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromyces*, *Pleurotus*, *Schizophyllum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*,
65 *Trametes* o *Trichoderma*.

[0129] En un aspecto más preferido, la célula huésped fúngica filamentosa 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 fúngica filamentosa es una célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*,
 5 *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *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 fúngica filamentosa es una célula de *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis subvermispora*,
 10 *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium tropicum*, *Chrysosporium merdarium*, *Chrysosporium inops*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium zonatum*, *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 harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

[0130] Las células fúngicas se pueden transformar por un proceso que implica la formación de protoplasto, la transformación de los protoplastos y la regeneración de la pared celular de manera conocida de por sí. Procedimientos adecuados para la transformación de células huésped de *Aspergillus* y *Trichoderma* se describen en la EP 238 023 y en Yelton *et al.*, 1984, Proceedings of the National Academy of Sciences EE.UU. 81: 1470-1474. Métodos adecuados para transformar especies de *Fusarium* son descritos por Malardier *et al.*, 1989, Gene 78: 147-156, and WO 96/00787. La levadura se puede transformar utilizando 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; and Hinnen *et al.*, 1978, Proceedings of the National Academy of Sciences EE.UU. 75: 1920.
 20
 25

Producción y recuperación

[0131] En los métodos de la presente invención, la célula huésped fúngica se cultiva en un medio nutritivo adecuado para la producción de un polipéptido que tiene actividad biológica utilizando métodos bien conocidos en la técnica. Por ejemplo, la célula se puede cultivar por cultivo en matraz de agitación y fermentación a pequeña o gran escala (incluyendo fermentación continua, de lote, de lote alimentado o de estado sólido) en laboratorio o fermentadores industriales realizadas en un medio adecuado y bajo condiciones que permitan que el polipéptido se exprese y/o aisle. El cultivo tiene lugar en un medio nutritivo adecuado que comprende fuentes de nitrógeno y carbono y sales inorgánicas, utilizando procedimientos conocidos en la técnica. Medios adecuados están disponibles de proveedores comerciales o se pueden preparar según composiciones publicadas (por ejemplo, en catálogos de la *American Type Culture Collection*).
 30
 35

[0132] En los métodos de la presente invención, el polipéptido que tiene actividad biológica se selecciona del grupo que consiste en una proteína de fusión, componentes de la proteína de fusión y una combinación de la proteína de fusión y sus componentes.
 40

[0133] En un aspecto preferido, el polipéptido que tiene actividad biológica es una proteína de fusión.

[0134] En otro aspecto preferido, el polipéptido que tiene actividad biológica es un componente(s) de una proteína de fusión.
 45

[0135] En otro aspecto preferido, el polipéptido que tiene actividad biológica es una combinación de una proteína de fusión y componentes de la misma.
 50

[0136] Los polipéptidos que tienen actividad biológica se pueden detectar usando métodos conocidos en la técnica que son específicos para los polipéptidos. Estos métodos de detección pueden incluir el uso de anticuerpos específicos, la formación de un producto enzimático o la 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, que puede incluir tanto actividad de endoglucanasa como una actividad biológica específica.
 55

[0137] El polipéptido resultante que tiene actividad biológica, por ejemplo, proteína de fusión de beta-glucosidasa o un componente del mismo, se puede recuperar utilizando métodos conocidos en la técnica. Por ejemplo, el polipéptido se puede recuperar del medio nutritivo por procedimientos convencionales incluyendo, pero no limitado a, centrifugado, filtración, extracción, secado por pulverización, evaporación o precipitación.
 60

[0138] Los polipéptidos de la presente invención se puede purificar por una variedad de procedimientos conocidos en la técnica incluyendo, pero no limitado a, cromatografía (por ejemplo, de intercambio iónico, afinidad, hidrofóbica, cromatoenfoco y exclusión de tamaño), procedimientos electroforéticos (por ejemplo, isoelectroenfoco preparativo), solubilidad diferencial (por ejemplo, precipitación de sulfato amónico), SDS-PAGE o extracción (véase, por ejemplo,
 65

Protein Purification, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989) para obtener polipéptidos sustancialmente puros.

Composiciones

[0139] La presente invención también se refiere a composiciones que comprenden un polipéptido que tiene actividad biológica de la presente invención. Preferiblemente, las composiciones se enriquecen con tal polipéptido. El término "enriquecido" indica que la actividad biológica de la composición ha sido aumentada, por ejemplo, con un factor de enriquecimiento de al menos 1.1.

[0140] La composición puede comprender un polipéptido de la presente invención como el componente enzimático mayor, por ejemplo, una composición monocomponente. Alternativamente, la composición puede comprender actividades enzimáticas múltiples, tal como una aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, celulasa, quitinasa, cutinasa, glicosiltransferasa de ciclodextrina, desoxiribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, haloperoxidasa, invertasa, lacasa, lipasa, manosidasa, oxidasa, enzima pectinolítica, peptidoglutaminasa, peroxidasa, fitasa, polifenoloxidasas, enzima proteolítica, ribonucleasa, transglutaminasa o xilanasas. La enzima(s) adicional(es) se puede producir, 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 sarcochroum*, *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*.

[0141] Las composiciones de polipéptido se pueden preparar conforme a métodos conocidos en la técnica y pueden ser en forma de un líquido o una composición seca. Por ejemplo, la composición de polipéptido puede ser en forma de un granulado o un microgranulado. El polipéptido que se va a incluir en la composición se puede estabilizar conforme a métodos conocidos en la técnica.

[0142] Más adelante se dan ejemplos de usos preferidos de las composiciones de polipéptido de la invención. La dosificación de la composición de polipéptido de la invención y otras condiciones bajo las que la composición se usa se pueden determinar basándose en métodos conocidos en la técnica.

Usos

[0143] La presente invención se dirige también a métodos para el uso de las proteínas de fusión o componentes de las mismas o composiciones de las mismas.

Métodos de tratamiento de material celulósico

[0144] Los métodos de la presente invención son particularmente útiles para mejorar la secreción de polipéptidos que tienen actividad celulolítica o hemicelulolítica en cantidades comercialmente importantes, que se pueden usar para degradar o convertir material lignocelulósico. Tales polipéptidos incluyen, pero de forma no limitativa, endoglucanasas, celobiohidrolasas, beta-glucosidasas, xilanasas, beta-xilosidasas, arabinofuranosidasas, acetil xilano estererasas y ácido ferúlico estererasas. Consecuentemente, la presente invención también se refiere a métodos para degradar o convertir un material celulósico, que comprende: tratamiento del material celulósico con una cantidad eficaz de una composición de enzima celulolítica en presencia de una cantidad eficaz de una proteína de fusión o un componente de la misma que tiene actividad celulolítica o actividad hemicelulolítica obtenida según los métodos instantáneos. Para fines de ilustración, un polipéptido que tiene actividad de beta-glucosidasa obtenido según los métodos de la presente invención, por ejemplo, una proteína de fusión de beta-glucosidasa o un componente de la misma, se usa para fines ilustrativos.

[0145] La biomasa celulósica puede incluir, pero no está limitada a, recursos madereros, residuos sólidos municipales, papel usado, plantas de cultivo y residuos de planta de cultivo (véase, por ejemplo, Wiseloge *et al.*, 1995, in Handbook on Bioethanol (Charles E. Wyman, editor), pp. 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, in Advances in Biochemical Engineering/Biotechnology, T. Scheper, managing editor, Volume 65, pp.23-40, Springer-Verlag, New York).

[0146] El polisacárido predominante en la pared celular primaria de la biomasa es la celulosa, el segundo más abundante es la hemicelulosa y el tercero es la pectina. La pared celular secundaria, producida después de que la célula haya dejado de crecer, también contiene polisacáridos y está reforzada por lignina polimérica reticulada de manera covalente en hemicelulosa. La celulosa es un homopolímero de anhidrocelobiosa y así un beta-(1-4)-D-glucano lineal, mientras que las hemicelulosas incluyen una variedad de compuestos, tales como xilanos, xiloglucanos, arabinoxilanos y mananos en estructuras complejas ramificadas con un espectro de sustituyentes. Aunque generalmente polimorfa, la

celulosa se encuentra en tejido vegetal principalmente como una matriz cristalina insoluble de cadenas de glucanos paralelos. Las hemicelulosas normalmente el hidrógeno se enlazan con la celulosa, al igual que con otras hemicelulosas, lo que ayuda a estabilizar la matriz de la pared celular. Se entiende aquí que el término "material celulósico" o "celulosa" también abarca lignocelulosa.

[0147] En los métodos de la presente invención, la composición enzimática celulolítica puede comprender cualquier proteína implicada en el procesamiento del material celulósico en glucosa, o hemicelulosa en xilosa, manosa, galactosa y arabinosa, sus polímeros, o productos derivados de éstos como se describe abajo. La composición enzimática celulolítica puede ser una preparación monocomponente, por ejemplo, una endoglucanasa, una preparación multicomponente, por ejemplo, endoglucanasa, celobiohidrolasa, beta-glucosidasa o una combinación de preparaciones de proteína multi y monocomponente. Las proteínas celulolíticas pueden tener actividad, es decir, hidrolizar celulosa, bien en el rango de pH ácido, neutro o alcalino. La composición enzimática celulolítica puede además comprender un polipéptido que tenga actividad aumentada celulolítica según la WO 2005/074647, WO 2005/074656 y la solicitud publicada EE.UU. 2007/0077630.

[0148] La proteína celulolítica puede ser de origen fúngico o bacteriano, que se puede obtener o aislar y purificar a partir de microorganismos que se conozcan por ser capaces de producir enzimas celulolíticas, por ejemplo, especies de *Bacillus*, *Pseudomonas*, *Humicola*, *Coprinus*, *Thielavia*, *Fusarium*, *Myceliophthora*, *Acremonium*, *Cephalosporium*, *Scytalidium*, *Penicillium* o *Aspergillus* (véase, por ejemplo, EP 458162), especialmente aquellos producidos por una cepa seleccionada a partir de *Humicola insolens* (reclasificada como *Scytalidium thermophilum*, véase, por ejemplo, la patente de EE.UU. n° 4.435.307), *Coprinus cinereus*, *Fusarium oxysporum*, *Myceliophthora thermophila*, *Meripilus giganteus*, *Thielavia terrestris*, *Acremonium sp.*, *Acremonium persicinum*, *Acremonium acremonium*, *Acremonium brachyphenium*, *Acremonium dichromosporum*, *Acremonium obclavatum*, *Acremonium pinkertoniae*, *Acremonium roseogriseum*, *Acremonium incoloratum* y *Acremonium furatum*; preferiblemente de DSM 1800 de *Humicola insolens*, DSM 2672 de *Fusarium oxysporum*, CBS 117.65 de *Myceliophthora thermophila*, RYM-202 de *Cephalosporium sp.*, CBS 478.94 de *Acremonium sp.*, CBS 265.95 de *Acremonium sp.*, CBS 169.65 de *Acremonium persicinum*, AHU 9519 de *Acremonium acremonium*, CBS 535.71 de *Cephalosporium sp.*, CBS 866.73 de *Acremonium brachyphenium*, CBS 683.73 de *Acremonium dichromosporum*, CBS 311.74 de *Acremonium obclavatum*, CBS 157.70 de *Acremonium pinkertoniae*, CBS 134.56 de *Acremonium roseogriseum*, CBS 146.62 de *Acremonium incoloratum* y CBS 299.70H de *Acremonium furatum*. Proteínas celulolíticas también se pueden obtener de *Trichoderma* (particularmente *Trichoderma viride*, *Trichoderma reesei* y *Trichoderma koningii*), *Bacillus alcalophilus* (véase, por ejemplo, la patente de EE.UU. n° 3.844.890 y la EP 458162) y *Streptomyces* (véase, por ejemplo, la EP 458162). Mutantes químicamente modificados o creados genéticamente a partir de proteínas de proteínas celulolíticas también se pueden usar.

[0149] Proteínas celulolíticas especialmente adecuadas son las celulasas descritas en la EP 495.257, EP 531.372, WO 96/11262, WO 96/29397, WO 98/08940. Otros ejemplos son variantes de celulasas tales como los descritos en la WO 94/07998, EP 531.315, patente EE.UU. n° 4.435.307, patente de EE.UU. n° 5.457.046, patente de EE.UU. n° 5.648.263, patente de EE.UU. n° 5.686.593, patente de EE.UU. n° 5.691.178, patente de EE.UU. n° 5.763.254, patente de EE.UU. n° 5.776.757, WO 89/09259, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

[0150] Como se ha mencionado anteriormente, las proteínas celulolíticas usadas en los métodos de la presente invención pueden ser preparaciones monocomponente, es decir, un componente esencialmente libre de otros componentes celulolíticos. El componente único puede ser un componente recombinante, es decir, producido por clonación de una secuencia de ADN que codifica el componente único y la célula posterior transformada con la secuencia de ADN y expresada en un huésped (véase, por ejemplo, WO 91/17243 y WO 91/17244). El huésped es preferiblemente un huésped heterólogo (la enzima es foránea del huésped), pero el huésped puede ser también bajo ciertas condiciones un huésped homólogo (la enzima es nativa del huésped). Proteínas celulolíticas monocomponentes también se puede preparar por purificación de tal proteína a partir de un caldo de fermentación.

[0151] Las proteínas celulolíticas usadas en los métodos de la presente invención se pueden producir por fermentación de las cepas microbianas mencionadas anteriormente en un medio nutritivo que contiene fuentes de nitrógeno y carbono adecuadas y sales inorgánicas, utilizando procedimientos conocidos en la técnica (véase, por ejemplo, Bennett, J.W. y LaSure, L. (eds.), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991). Medios adecuados están disponibles de proveedores comerciales o se pueden preparar según composiciones publicadas (por ejemplo, en catálogos de la American Type Culture Collection). Rangos de temperatura y otras condiciones adecuadas para el crecimiento y la producción de proteína celulolítica se conocen en la técnica (véase, por ejemplo, Bailey, J.E., and Ollis, D.F., *Biochemical Engineering Fundamentals*, McGraw-Hill Book Company, NY, 1986).

[0152] La fermentación puede ser cualquier método de cultivo de una célula que de como resultado la expresión o el aislamiento de una proteína celulolítica. La fermentación puede, por lo tanto, entenderse como que comprende el cultivo en matraz de agitación, o fermentación a pequeña o gran escala (incluyendo, fermentación continua, de lote, de lote alimentado o de estado sólido) en laboratorio o fermentadores industriales realizadas en un medio adecuado y bajo condiciones que permitan que la proteína celulolítica se exprese o aisle. Las proteínas celulolíticas resultantes producidas por los métodos anteriormente descritos se pueden recuperar del medio de fermentación y purificar por procedimientos convencionales como se describe en este caso.

[0153] Ejemplos de preparaciones enzimáticas celulolíticas comerciales adecuados para el uso en la presente invención incluyen, por ejemplo, CELLUCLAST™ (disponible de Novozymes A/S) y NOVOZYM™ 188 (disponible de Novozymes A/S). Otras preparaciones disponibles comercialmente que comprenden celulosa que se pueden utilizar incluyen CELLUZYME™, CEREFLO™ y ULTRAFLO™ (Novozymes A/S), LAMINEX™ y SPEZYME™ CP (Genencor Int.), ROHAMENT™ 7069 W (Röhm GmbH) y FIBREZYME® LDI, FIBREZYME® LBR o VISCOSTAR® 150L (Dyadic International, Inc., Jupiter, FL, EE.UU.). Las enzimas de celulosa se agregan en cantidades efectivas desde aproximadamente 0,001% a aproximadamente 5,0 % en peso de sólidos, de forma más preferible de aproximadamente 0,025% a aproximadamente 4,0% en peso de sólidos, y de la forma más preferible de aproximadamente 0,005% a aproximadamente 2,0% en peso de sólidos.

[0154] Las proteínas celulolíticas resultantes o proteínas de beta-glucosidasa producidas por los métodos anteriormente descritos se pueden recuperar del medio de fermentación por procedimientos convencionales incluyendo, pero no limitado a, centrifugado, filtración, secado por pulverización, evaporación o precipitación. La proteína recuperada puede luego purificarse más por una variedad de procedimientos cromatográficos, por ejemplo, cromatografía de intercambio de iones, cromatografía de filtración en gel, cromatografía de afinidad o similares.

[0155] La actividad de una proteína celulolítica se puede determinar utilizando cualquier método conocido en la técnica.

[0156] Ejemplos de preparaciones celulolíticas adecuados para su uso en la presente invención incluyen, por ejemplo, CELLUCLAST™ (disponible de Novozymes A/S) y NOVOZYM™ 188 (disponible de Novozymes A/S). Otras preparaciones disponibles comercialmente que comprenden celulosa que se pueden utilizar incluyen CELLUZYME™, CEREFLO™ y ULTRAFLO™ (Novozymes A/S), LAMINEX™ y SPEZYME™ CP (Genencor Int.) y ROHAMENT™ 7069 W (Röhm GmbH). Las enzimas de celulosa se agregan en cantidades efectivas de aproximadamente 0,001% a aproximadamente 5,0 % en peso de sólidos, de forma más preferible de aproximadamente 0,025% a aproximadamente 4,0% en peso de sólidos, y de la forma más preferible de aproximadamente 0,005% a aproximadamente 2,0% en peso de sólidos.

[0157] Como se ha mencionado anteriormente, las proteínas celulolíticas usadas en los métodos de la presente invención pueden ser preparaciones monocomponentes, es decir, un componente esencialmente libre de otros componentes celulolíticos. El componente único puede ser un componente recombinante, es decir, producido por clonación de una secuencia de ADN que codifica el componente único y la posterior célula transformada con la secuencia de ADN y expresada en un huésped (véase, por ejemplo, WO 91/17243 y WO 91/17244). Otros ejemplos de proteínas celulolíticas monocomponente incluyen, pero de forma no limitativa, aquellos descritos en la JP-07203960-A y la WO-9206209. El huésped es preferiblemente un huésped heterólogo (la enzima es foránea del huésped), pero el huésped puede ser también bajo ciertas condiciones un huésped homólogo (la enzima es nativa del huésped). Las proteínas celulolíticas monocomponentes también se pueden preparar por purificación de tal proteína a partir de un caldo de fermentación.

[0158] Ejemplos de proteínas celulolíticas monocomponente útiles en la práctica de los métodos de la presente invención incluyen, pero de forma no limitativa, endoglucanasa, celobiohidrolasa y otras enzimas útiles en la biomasa celulósica de degradación.

[0159] El término "endoglucanasa" ya se ha definido aquí. El término "celobiohidrolasa" se define aquí como una 1,4-beta-D-glucano celobiohidrolasa (E.C. 3.2.1.91), que cataliza la hidrólisis de enlaces 1,4-beta-D-glucosídicos en la celulosa, celooligosacáridos o cualquier beta-1,4- glucosa enlazada que contenga celobiosa liberadora de polímero desde los extremos reductores o no reductores de la cadena. Para fines de la presente invención, la actividad de celobiohidrolasa se puede determinar según los procedimientos descritos por Lever *et al.*, 1972, Anal. Biochem. 47: 273-279 y por van Tilbeurgh *et al.*, 1982, FEBS Letters 149: 152-156 ; van Tilbeurgh y Claeysens, 1985, FEBS Letters 187: 283-288. En la presente invención, el método de Lever *et al.* se empleó para valorar la hidrólisis de celulosa en residuos de maíz.

[0160] Los polipéptidos de la presente invención se usan conjuntamente con proteínas celulolíticas para degradar el componente celulósico del sustrato de biomasa, (véase, por ejemplo, Brigham *et al.*, 1995, in Handbook on Bioethanol (Charles E. Wyman, editor), pp.119-141, Taylor & Francis, Washington D.C. ; Lee, 1997, Journal of Biotechnology 56: 1-24).

[0161] Las cantidades óptimas de un polipéptido que tiene actividad de beta-glucosidasa y de proteínas celulolíticas depende de diferentes factores que incluyen, pero sin estar limitados a, la mezcla de proteínas celulolíticas de componente, el sustrato celulósico, la concentración de sustrato celulósico, el pretratamiento(s) del sustrato celulósico, la temperatura, el tiempo, el pH y la inclusión del organismo fermentador (por ejemplo, levadura para sacarificación y fermentación simultánea). El término "proteínas celulolíticas" se define aquí como aquellas proteínas o mezclas de proteínas que pueden hidrolizar o convertir o degradar celulosa bajo las condiciones evaluadas. Sus cantidades se miden normalmente por un ensayo común tal como BCA (ácido bicinonínico, P.K. herrero *et al.*, 1985, v. Biochem. 150: 76), y la cantidad preferida se añade en proporción a la cantidad de biomasa hidrolizada.

5 [0162] En un aspecto preferido, la cantidad de polipéptido que tiene actividad de beta-glucosidasa por g de material celulósico es de aproximadamente 0,01 a aproximadamente 2,0 mg, preferiblemente de aproximadamente 0,025 a aproximadamente 1,5 mg, de forma más preferible de aproximadamente 0,05 a aproximadamente 1,25 mg, de forma más preferible de aproximadamente 0,075 a aproximadamente 1,25 mg, de forma más preferible de aproximadamente 0,1 a aproximadamente 1,25 mg, incluso de forma más preferible de aproximadamente 0,15 a aproximadamente 1,25 mg, y de la forma más preferible de aproximadamente 0,25 a aproximadamente 1,0 mg por g de material celulósico.

10 [0163] En otro aspecto preferido, la cantidad de proteínas celulolíticas por g de material celulósico es de aproximadamente 0,5 a aproximadamente 50 mg, preferiblemente de aproximadamente 0,5 a aproximadamente 40 mg, de forma más preferible de aproximadamente 0,5 a aproximadamente 25 mg, de forma más preferible de aproximadamente 0,75 a aproximadamente 20 mg, de forma más preferible de aproximadamente 0,75 a aproximadamente 15 mg, incluso de forma más preferible de aproximadamente 0,5 a aproximadamente 10 mg, y de la forma más preferible de aproximadamente 2,5 a aproximadamente 10 mg por g de material celulósico.

15 [0164] Los métodos de la presente invención pueden utilizarse para degradar o convertir un material celulósico, por ejemplo, lignocelulosa, en muchas sustancias útiles, por ejemplo, productos químicos y combustibles. Además de etanol, algún artículo de consumo y productos químicos de especialidad que se pueden producir a partir de celulosa incluyen xilosa, acetona, acetato, glicina, lisina, ácidos orgánicos (por ejemplo, ácido láctico), 1,3-propanodiol, butanodiol, glicerol, etilenglicol, furfural, polihidroxialcanoatos, y ácido cis,cis-mucónico (Lynd, L. R., Wyman, C. E., and Gerngross, T. U., 1999, Biocommodity Engineering, Biotechnol. Prog., 15: 777-793 ; Philippidis, G. P., 1996, Cellulose bioconversion technology, in Handbook on Bioethanol: Production and Utilization, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212 ; and Ryu, D. D. Y., and Mandels, M., 1980, Cellulases: biosynthesis and applications, Enz. Microb. Technol., 2: 91-102). Los beneficios de coproducción potenciales se extienden más allá de la síntesis de múltiples productos orgánicos a partir de carbohidrato fermentable. Residuos ricos en lignina que permanecen después de tratamiento biológico se pueden convertir en productos químicos derivados de la lignina, o se pueden usar para la producción de energía.

25 [0165] Métodos convencionales usados para procesar el material celulósico conforme a los métodos de la presente invención son bien entendidas a los expertos en el art. los métodos de la presente invención se pueden implementar utilizando cualquier equipo de tratamiento de biomasa convencional configuró para operar conforme a la invención.

30 [0166] Tal equipo puede incluir un reactor agitado de lote alimentado, un reactor agitado por lotes, un reactor agitado de flujo continuo con ultrafiltración, un reactor de columna de flujo de pistón continuo (Fernanda de Castilhos Corazza, Flávio Faria de Moraes, Gisella Maria Zanin e Ivo Neitzel, 2003, Optimal control in fed-batch reactor for the cellobiose hydrolysis, Acta Scientiarum. Technology 25(1): 33-38 ; Gusakov, A.V., and Sinitsyn, A.P., 1985, Kinetics of the enzymatic hydrolysis of cellulose: 1. A mathematical model for a batch reactor process, Enz. Microb. Technol. 7: 346-352), un reactor de fricción (Ryu, S.K., y Lee, J.M., 1983, Bioconversion of waste cellulose by using an attrition bioreactor, Biotechnol. Bioeng. 25: 53-65), o un reactor con agitación intensiva inducida por un campo electromagnético (Gusakov, A.V., Sinitsyn, A.P., Davydkin, I.Y., Davydkin, V.Y., Protas, O.V., 1996, Enhancement of enzymatic cellulose hydrolysis using a novel type of bioreactor with intensive stirring induced by electromagnetic field, Appl. Biochem. Biotechnol. 56: 141-153).

35 [0167] Los métodos convencionales incluyen, pero de forma no limitativa, sacarificación, fermentación, hidrólisis y fermentación separada (SHF), sacarificación y fermentación simultánea (SSF), sacarificación y cofermentación simultánea (SSCF), hidrólisis y fermentación híbrida (HHF) y conversión microbiana directa (DMC).

40 [0168] La SHF utiliza pasos de proceso separados para primero hidrolizar enzimáticamente celulosa en azúcares fermentables y luego, en un paso posterior, azúcares de fermentos en etanol. En la SSF, la hidrólisis enzimática de celulosa y la fermentación de azúcar fermentable en etanol se combinan en un paso (Philippidis, G. P., 1996, Cellulose bioconversion technology, in Handbook on Bioethanol: Production and Utilization, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212). La SSCF incluye la cofermentación de azúcares múltiples (Sheehan, J., and Himmel, M., 1999, Enzymes, energy and the environment: A strategic perspective on the U.S. Department of Energy's research and development activities for bioethanol, Biotechnol. Prog. 15: 817-827). La HHF incluye dos pasos separados realizados en el mismo reactor pero a temperaturas diferentes, es decir, sacarificación enzimática a alta temperatura seguida de SSF a una temperatura inferior que la cepa de fermentación pueda tolerar. La DMC combina los tres procesos (producción de celulasa, hidrólisis de celulosa y fermentación) en un paso (Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S., 2002, Microbial cellulose utilization: Fundamentals and biotechnology, Microbiol. Mol. Biol. Reviews 66: 506-577).

45 [0169] "Fermentación" o "proceso de fermentación" se refiere a cualquier proceso de fermentación o cualquier proceso que comprenda un paso de fermentación. Un proceso de fermentación incluye, sin limitación, procesos de fermentación usados para producir productos de fermentación incluyendo alcoholes (por ejemplo, arabinitol, butanol, etanol, glicerol, metanol, 1,3-propanodiol, sorbitol, y xilitol); ácidos orgánicos (por ejemplo, ácido acético, ácido acetónico, ácido adípico, ácido ascórbico, ácido cítrico, ácido 2,5-diketo-D-glucónico, ácido fórmico, ácido fumárico, ácido glucárico, ácido glucónico, ácido glucurónico, ácido glutárico, ácido 3-hidroxipropiónico, ácido itacónico, ácido láctico, ácido málico, ácido malónico, ácido oxálico, ácido propiónico, ácido succínico y ácido xilónico); cetonas (por ejemplo, acetona); aminoácidos

(por ejemplo, ácido aspártico, ácido glutámico, glicina, lisina, serina y treonina); gases (por ejemplo, metano, hidrógeno (H₂), dióxido de carbono (CO₂) y monóxido de carbono (CO)). Procesos de fermentación también incluyen procesos de fermentación usados en la industria de alcohol consumible (por ejemplo, cerveza y vino), industria lechera (por ejemplo, productos lácteos fermentados), industria del cuero e industria del tabaco.

[0170] Una proteína de fusión o un componentes de la misma que tiene actividad celulolítica o actividad hemicelulolítica obtenida según los métodos de la presente invención, por ejemplo, una proteína de fusión de beta-glucosidasa o un componentes de la misma, y células huésped de la misma, se pueden usar en la producción de monosacáridos, disacáridos y polisacáridos como materias primas químicas o de fermentación a partir de biomasa para la producción de etanol, plásticos, otros productos o productos intermedios. En particular, los polipéptidos y células huésped se pueden utilizar para aumentar el valor del procesamiento de residuos (grano de destiladoras secados, granos utilizados de elaboración de cerveza, bagazo de caña de azúcar, etc.) por solubilización parcial o completa de celulosa o hemicelulosa. Al impulsar el procesamiento de material celulósico por la preparación de la enzima celulolítica en glucosa, xilosa, manosa, galactosa y arabinosa, sus polímeros o productos derivados de éstos como se describe abajo, los polipéptidos pueden ser en forma de un caldo de fermentación crudo con o sin las células o en forma de una preparación enzimática semi-purificada o purificada. El polipéptido puede ser una preparación monocomponente, una preparación de proteína de varios componentes o una combinación de preparaciones de proteína de varios componentes y monocomponente. Alternativamente, una célula huésped se puede utilizar como una fuente de tal polipéptido en un proceso de fermentación con la biomasa. La célula huésped también puede contener genes nativos o heterólogos que codifiquen proteína celulolítica al igual que otras enzimas útiles en el procesamiento de biomasa.

[0171] La presente invención además se refiere a métodos para producir una sustancia orgánica, que comprende: (a) sacarificación de un material celulósico con una cantidad eficaz de una composición de enzima celulolítica i en presencia de una cantidad eficaz de una proteína de fusión o un componentes de la misma que tiene actividad celulolítica o actividad hemicelulolítica obtenida según los métodos instantáneos; (b) fermentación del material celulósico sacarificado del paso (a) con uno o más microorganismos fermentativos; y (c) recuperación de la sustancia orgánica de la fermentación. Como se ha indicado anteriormente, para fines de ilustración, un polipéptido que tiene actividad de beta-glucosidasa obtenido según los métodos de la presente invención, por ejemplo, una proteína de fusión de beta-glucosidasa o un componente de la misma, se usa para fines ilustrativos. El polipéptido que tiene actividad de beta-glucosidasa puede ser en forma de un caldo de fermentación crudo con o sin las células o en forma de una preparación semi-purificada o purificada enzimática. La proteína de beta-glucosidasa puede ser una preparación monocomponente, una preparación de proteína de varios componentes, o una combinación de preparaciones de proteína de varios componentes y monocomponente.

[0172] La sustancia puede ser cualquier sustancia derivada de la fermentación. En un aspecto preferido, la sustancia es un alcohol. Se entiende que el término "alcohol" abarca una sustancia que contiene una o más fracciones de hidroxilo. En un aspecto más preferido, el alcohol es arabinitol. En otro aspecto más preferido, el alcohol es butanol. En otro aspecto más preferido, el alcohol es etanol. En otro aspecto más preferido, el alcohol es glicerol. En otro aspecto más preferido, el alcohol es metanol. En otro aspecto más preferido, el alcohol es 1,3-propanodiol. En otro aspecto más preferido, el alcohol es sorbitol. En otro aspecto más preferido, el alcohol es xilitol. Véase, por ejemplo, Gong, C. S., Cao, N. J., Du, J., and Tsao, G. T., 1999, Ethanol production from renewable resources, en *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207- 241 ; Silveira, M. M., y Jonas, R., 2002, The biotechnological production of sorbitol, *Appl. Microbiol. Biotechnol.* 59: 400-408 ; Nigam, P., y Singh, D., 1995, Processes for fermentative production of xylitol - a sugar substitute, *Process Biochemistry* 30 (2): 117-124 ; Ezeji, T. C., Qureshi, N. and Blaschek, H. P., 2003, Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and in situ recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19 (6): 595-603.

[0173] En otro aspecto preferido, la sustancia es un ácido orgánico. En otro aspecto más preferido, el ácido orgánico es ácido acético. En otro aspecto más preferido, el ácido orgánico es ácido acetónico. En otro aspecto más preferido, el ácido orgánico es ácido adípico. En otro aspecto más preferido, el ácido orgánico es ácido ascórbico. En otro aspecto más preferido, el ácido orgánico es ácido cítrico. En otro aspecto más preferido, el ácido orgánico es ácido 2,5-diketo-D-gluónico. En otro aspecto más preferido, el ácido orgánico es ácido fórmico. En otro aspecto más preferido, el ácido orgánico es ácido fumárico. En otro aspecto más preferido, el ácido orgánico es ácido glucárico. En otro aspecto más preferido, el ácido orgánico es ácido glucónico. En otro aspecto más preferido, el ácido orgánico es ácido glucurónico. En otro aspecto más preferido, el ácido orgánico es ácido glutárico. En otro aspecto preferido, el ácido orgánico es ácido 3-hidroxipropiónico. En otro aspecto más preferido, el ácido orgánico es ácido itacónico. En otro aspecto más preferido, el ácido orgánico es ácido láctico. En otro aspecto más preferido, el ácido orgánico es ácido málico. En otro aspecto más preferido, el ácido orgánico es ácido malónico. En otro aspecto más preferido, el ácido orgánico es ácido oxálico. En otro aspecto más preferido, el ácido orgánico es ácido propiónico. En otro aspecto más preferido, el ácido orgánico es ácido succínico. En otro aspecto más preferido, el ácido orgánico es ácido de xilónico. Véase, por ejemplo, Chen, R., y Lee, Y. Y., 1997, Membrane-mediated extractive fermentation for lactic acid production from cellulosic biomass, *Appl. Biochem. Biotechnol.* 63-65: 435-448.

[0174] En otro aspecto preferido, la sustancia es una cetona. Se entiende que el término "cetona" abarca una sustancia que contiene una o más fracciones de cetona. En otro aspecto más preferido, la cetona es acetona. Véase, por ejemplo, Qureshi y Blaschek, 2003, *supra*.

5 [0175] En otro aspecto preferido, la sustancia es un aminoácido. En otro aspecto más preferido, el ácido orgánico es ácido aspártico. En otro aspecto más preferido, el aminoácido es ácido glutámico. En otro aspecto más preferido, el aminoácido es glicina. En otro aspecto más preferido, el aminoácido es lisina. En otro aspecto más preferido, el aminoácido es serina. En otro aspecto más preferido, el aminoácido es treonina. Véase, por ejemplo, Richard, A., y Margaritis, A., 2004, Empirical modeling of batch fermentation kinetics for poly(glutamic acid) production and other microbial biopolymers, *Biotechnology and Bioengineering* 87 (4): 501-515.

10 [0176] En otro aspecto preferido, la sustancia es un gas. En otro aspecto más preferido, el gas es metano. En otro aspecto más preferido, el gas es H₂. En otro aspecto más preferido, el gas es CO₂. En otro aspecto más preferido, el gas es CO. Véase, por ejemplo, Kataoka, N., A. Miya, y K. Kiriya, 1997, Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria, *Water Science and Technology* 36 (6-7): 41-47 ; and Gunaseelan V.N. in *Biomass and Bioenergy*, Vol. 13 (1-2), pp. 83-114, 1997 , Anaerobic digestion of biomass for methane production: A review.

15 [0177] La producción de una sustancia de material celulósico requiere típicamente cuatro pasos principales. Estos cuatro pasos son pretratamiento, hidrólisis enzimática, fermentación y recuperación. A continuación se ejemplifica un proceso para producir etanol, pero se entiende que se pueden utilizar procesos similares para producir otras sustancias, por ejemplo, las sustancias anteriormente descritas.

20 [0178] Pretratamiento. En el paso pretratamiento o prehidrólisis, el material celulósico se calienta para descomponer la lignina y la estructura de carbohidrato, solubilizar la mayor parte de la hemicelulosa y hacer accesible la fracción de celulosa a las enzimas celulolíticas. El calentamiento se realiza bien directamente con vapor o en lodo, donde un catalizador también se puede añadir al material para acelerar las reacciones. Catalizadores incluyen ácidos fuertes, tales como ácido sulfúrico y SO₂, o álcali, tal como hidróxido sódico. El propósito de la fase de pretratamiento es facilitar la penetración de las enzimas y microorganismos. La biomasa celulósica también puede someterse a un pretratamiento de explosión de vapor hidrotermal (véase solicitud de patente EE.UU. n° 20020164730). No obstante, se entiende que al practicar los métodos de la presente invención, cualquier pretratamiento se puede usar utilizando pretratamiento térmico, químico y/o mecánico.

25 [0179] Sacarificación. En el paso de la hidrólisis enzimática, también conocido como sacarificación, se añaden enzimas como las descritas en este caso al material pretratado para convertir la fracción de celulosa en glucosa y/o otros azúcares. La sacarificación se realiza generalmente en los reactores de tanque agitado o fermentadores bajo condiciones de pH, temperatura y mezcla controladas. Un paso de sacarificación puede durar hasta 200 horas. La sacarificación se puede realizar a temperaturas de aproximadamente 30°C a aproximadamente 65°C, en particular alrededor de 50°C, y a un pH en la gama entre aproximadamente 4 a aproximadamente 5, especialmente alrededor de pH 4,5. Para producir glucosa que se pueda metabolizar por levadura, la hidrólisis se realiza típicamente en presencia de una beta-glucosidasa.

30 [0180] Fermentación. En el paso de fermentación, azúcares, liberados del material celulósico como resultado de los pasos de pretratamiento y de hidrólisis enzimática, son fermentados en etanol por un organismo fermentador, tal como levadura. La fermentación puede también efectuarse simultáneamente con la hidrólisis enzimática en el mismo vaso, nuevamente bajo condiciones de pH, temperatura y de mezcla controladas. Cuando la sacarificación y la fermentación se realizan simultáneamente en el mismo vaso, el proceso se denomina generalmente sacarificación y fermentación simultánea o SSF.

35 [0181] Cualquier material celulósico adecuado o materia prima se puede utilizar en el paso de fermentación practicando la presente invención. El material se selecciona generalmente basado en el producto de fermentación deseada, es decir, la sustancia que se va a obtener de la fermentación, y el proceso empleado, como bien se conoce en la técnica. Ejemplos de sustratos adecuados para su uso en los métodos de la presente invención, incluyen materiales con celulosa, tales como madera o residuos vegetales o azúcares de bajo peso molecular DP1-3 obtenidos de material celulósico procesado que se puede metabolizar por el microorganismo fermentador, y que se pueden suministrar por adición directa al medio de fermentación.

40 [0182] El término "medio de fermentación" se entenderá que se refiere a un medio antes de que el microorganismo(s) de fermentación se añada, tal como, un medio resultante de un proceso de sacarificación, al igual que un medio usado en un proceso de sacarificación y fermentación simultánea (SSF).

45 [0183] "Microorganismo de fermentación" se refiere a cualquier microorganismo adecuado para su uso en un proceso de fermentación deseado. Microorganismos fermentativos adecuados pueden fermentar, es decir, convertir, azúcares, tales como glucosa, xilosa, arabinosa, manosa, galactosa u oligosacáridos directa o indirectamente en el producto de fermentación deseado. Ejemplos de microorganismos fermentativos incluyen organismos fúngicos, tales como levadura. La levadura preferida incluye cepas del *Saccharomyces spp.*, y en particular, *Saccharomyces cerevisiae*. La levadura

disponible comercialmente incluye, por ejemplo, RED STAR®/Lesaffre Ethanol Red (disponible de RED STAR®/Lesaffre, EE.UU.) FALI (disponible de Fleischmann's Yeast, una división de Burns Philp Food Inc., EE.UU.), SUPERSTART (disponible de Alltech), GERT STRAND (disponible de Gert Strand AB, Suecia) y FERMIOL (disponibles de DSM Specialties).

5

[0184] En un aspecto preferido, la levadura es un *Saccharomyce spp.* En un aspecto más preferido, la levadura es *Saccharomyces cerevisiae*. En otro aspecto más preferido, la levadura es *Saccharomyces distaticus*. En otro aspecto más preferido, la levadura es *Saccharomyces uvarum*. En otro aspecto preferido, la levadura es un *Kluyveromyces*. En otro aspecto más preferido, la levadura es *Kluyveromyces marxianus*. En otro aspecto más preferido, la levadura es *Kluyveromyces fragilis*. En otro aspecto preferido, la levadura es una *Candida*. En otro aspecto más preferido, la levadura es *Candida pseudotropicalis*. En otro aspecto más preferido, la levadura es *Candida brassicae*. En otro aspecto preferido, la levadura es una *Clavispora*. En otro aspecto más preferido, la levadura es *Clavispora lusitaniae*. En otro aspecto más preferido, la levadura es *Clavispora opuntiae*. En otro aspecto preferido, la levadura es un *Pachysolen*. En otro aspecto más preferido, la levadura es *Pachysolen tannophilus*. En otro aspecto preferido, la levadura es un *Bretannomyces*. En otro aspecto más preferido, la levadura es *Bretannomyces clausenii* (Philippidis, G. P., 1996, Cellulose bioconversion technology, in Handbook on Bioethanol: Production and Utilization, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212).

10

15

20

[0185] Bacterias que pueden fermentar eficazmente glucosa en etanol incluyen, por ejemplo, *Zymomonas mobilis* y *Clostridium thermocellum* (Philippidis, 1996, *supra*).

[0186] Se conoce en la técnica que los organismos anteriormente descritos también pueden usarse para producir otras sustancias, como se describe en este caso.

25

[0187] La clonación de genes heterólogos en *Saccharomyces cerevisiae* (Chen, Z., Ho, N. W. Y., 1993, Cloning and improving the expression of Pichia stipitis xylose reductase gene in *Saccharomyces cerevisiae*, Appl. Biochem. Biotechnol. 39-40: 135-147 ; Ho, N. W. Y., Chen, Z, Brainard, A. P., 1998, Genetically engineered *Saccharomyces* yeast capable of effectively cofermenting glucose and xylose, Appl. Environ. Microbiol. 64: 1852-1859) o en bacterias tales como *Escherichia coli* (Beall, D. S., Ohta, K., Ingram, L. O., 1991, Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*, Biotech. Bioeng. 38: 296- 303), *Klebsiella oxytoca* (Ingram, L. O., Gomes, P. F., Lai, X., Moniruzzaman, M., Wood, B. E., Yomano, L. P., York, S. W., 1998, Metabolic engineering of bacteria for ethanol production, Biotechnol. Bioeng. 58: 204-214), y *Zymomonas mobilis* (Zhang, M., Eddy, C., Deanda, K., Finkelstein, M., and Picataggio, S., 1995, Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*, Science 267: 240-243 ; Deanda, K., Zhang, M., Eddy, C., and Picataggio, S., 1996, Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering, Appl. Environ. Microbiol. 62: 4465-4470) ha llevado a la construcción de organismos capaces de convertir hexosas y pentosas en etanol (cofermentación).

30

35

40

[0188] La levadura u otro microorganismo típicamente se añade a la celulosa degradada o hidrolizada y la fermentación se realiza durante aproximadamente 24 a aproximadamente 96 horas, tal como de aproximadamente 35 a aproximadamente 60 horas. La temperatura está típicamente entre aproximadamente 26°C a aproximadamente 40°C, en particular en aproximadamente de 32°C, y en aproximadamente pH 3 a aproximadamente pH 6, en particular aproximadamente pH 4-5.

45

[0189] En un aspecto preferido, la levadura u otro microorganismo se aplica a la celulosa degradada o hidrolizada y la fermentación se realiza durante aproximadamente 24 a aproximadamente 96 horas, tal como típicamente 35-60 horas. En un aspecto preferido, la temperatura está generalmente entre aproximadamente 26 a aproximadamente 40°C, en particular aproximadamente 32°C, y el pH está generalmente de aproximadamente pH 3 a aproximadamente pH 6, preferiblemente aproximadamente pH 4-5. La levadura u otro microorganismo se aplica preferiblemente en cantidades de aproximadamente 10^5 a 10^{12} , preferiblemente de aproximadamente 10^7 a 10^{10} , especialmente aproximadamente 5×10^7 de concentración de células viables por ml de caldo de fermentación. Durante una fase de producción de etanol la concentración de células de levadura debería preferiblemente estar en la gama de aproximadamente 10^7 a 10^{10} , especialmente alrededor de aproximadamente 2×10^8 . Más información sobre el uso de levadura para la fermentación se puede encontrar en, por ejemplo, "The Alcohol Textbook" (Editors K. Jacques, T.P. Lyons and D.R. Kelsall, Nottingham University Press, United Kingdom 1999), que es incorporado por la presente por referencia.

50

55

[0190] El proceso más ampliamente usado en la técnica es el proceso de sacarificación y fermentación simultánea (SSF) donde no hay fase de retención para la sacarificación, lo que significa que la levadura y la enzima se agregan juntas.

60

[0191] Para la producción de etanol, después de la fermentación la trituración se destila para extraer el etanol. El etanol obtenido según el proceso de la invención se puede utilizar como, por ejemplo, etanol de combustible, etanol potable, es decir, licores neutros potables o etanol industrial.

65

[0192] Un estimulador de fermentación se puede utilizar en combinación con cualquiera de los procesos enzimáticos descritos aquí para mejorar más el proceso de fermentación, y en particular, el rendimiento del microorganismo

- 5 fermentador, tal como, mejora del índice y rendimiento de etanol. Un "estimulador de fermentación" se refiere a estimuladores para el crecimiento de los microorganismos fermentativos, en particular, levadura. Estimuladores de fermentación preferidos para el crecimiento incluyen vitaminas y minerales. Ejemplos de vitaminas incluyen multivitaminas, biotina, pantotenato, ácido nicotínico, meso-inositol, tiamina, piridoxina, ácido para-aminobenzóico, ácido fólico, riboflavina y vitaminas A, B, C, D y E. Véase, por ejemplo, Alfenore *et al.*, Improving ethanol production and viability of *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag (2002), que se incorpora en la presente por referencia. Ejemplos de minerales incluyen minerales y sales minerales que pueden suministrar nutrientes que comprenden P, K, Mg, S, Ca, Fe, Zn, Mn y Cu.
- 10 [0193] Recuperación. El alcohol se separa del material celulósico fermentado y purificado por métodos convencionales de destilación. Se puede obtener etanol con una pureza de hasta aproximadamente 96 vol.%, que se puede usar como, por ejemplo, etanol de combustible, etanol potable, es decir, licores neutros potables, o etanol industrial.
- 15 [0194] Para otras sustancias, cualquier método conocido en la técnica se puede usar incluyendo, pero no limitado a, cromatografía (por ejemplo, intercambio iónico, afinidad, hidrofóbica, cromatografía de exclusión por tamaño), procedimientos electroforéticos (por ejemplo, isoelectroenfoque preparativo), solubilidad diferencial (por ejemplo, precipitación de sulfato amónico), SDS-PAGE, destilación o extracción.
- 20 [0195] En los métodos de la presente invención, además de la beta-glucosidasa, la preparación enzimática celulolítica y polipéptido(s) de mejora celulolítica pueden ser suplementados con una o varias actividades enzimáticas adicionales para mejorar la degradación del material celulósico. Enzimas adicionales preferidas son hemicelulasas, esterases (por ejemplo, lipasas, fosfolipasas y/o cutinasas), proteasas, lacasas, peroxidasas o mezclas derivadas.
- 25 [0196] En los métodos de la presente invención, la enzima(s) adicional(es) se pueden adicionar antes o durante la fermentación, incluyendo durante o después de la propagación del microorganismo(s) de fermentación.
- 30 [0197] Las enzimas referenciadas aquí se puede derivar u obtener a partir de cualquier origen adecuado, incluyendo bacteriano, fúngico, levadura u origen mamífero. El término "obtenido" se refiere aquí a que la enzima se puede haber aislado de un organismo que produce naturalmente la enzima como una enzima nativa. El término "obtenido" se refiere también aquí a que la enzima puede haber sido producida a partir de recombinación en un organismo huésped, donde la enzima producida de recombinación es nativa o foránea del organismo huésped o tiene una secuencia de aminoácidos modificada, por ejemplo, tiene uno o varios aminoácidos que están eliminados, insertados y/o sustituidos, es decir, una enzima producida de forma recombinante que es una secuencia de aminoácidos nativa mutante y/o un fragmento de ella o una enzima producida por procesos de redistribución de ácidos nucleicos conocido en la técnica.
- 35 Dentro del significado de una enzima nativa están las variantes naturales y en el significado de una enzima foránea están variantes obtenidas de forma recombinante, tal como por mutagénesis dirigida al sitio o redistribución.
- 40 [0198] Las enzimas también puede ser purificadas. El término "purificado" como se utiliza en este caso cubre enzimas libres de otros componentes del organismo del que se derivan. El término "purificado" cubre también enzimas libres de componentes del organismo nativo del que se obtienen. Las enzimas pueden ser purificadas, con sólo cantidades menores presentes de otras proteínas. La expresión "otras proteínas" se refiere en particular a otras enzimas. El término "purificado" como se utiliza en este caso también se refiere a la eliminación de otros componentes, particularmente otras proteínas y más particularmente otras enzimas presentes en la célula de origen de la enzima de la invención. La enzima puede ser "polipéptido sustancialmente puro", es decir, libre de otros componentes del organismo donde se produce, es decir, por ejemplo, un organismo huésped para enzimas producidas de forma recombinante.
- 45 [0199] Las enzimas usadas en la presente invención pueden estar en cualquier forma adecuada para su uso en los procesos descritos aquí, tal como, por ejemplo, un caldo de fermentación crudo con o sin células, un polvo seco o granulado, un granulado no polvoriento, un líquido, un líquido estabilizado o una enzima protegida. Los granulados se pueden producir, por ejemplo, como se describe en las patentes de EE.UU. nº 4.106.991 y 4.661.452, y pueden opcionalmente estar recubiertos por procesos conocidos en la técnica. Las preparaciones enzimáticas líquidas pueden, por ejemplo, estabilizarse añadiendo estabilizadores tales como un azúcar, un alcohol de azúcar u otro poliol, y/o ácido láctico u otro ácido orgánico según el proceso establecido. Enzimas protegidas se pueden preparar según el proceso descrito en la EP 238.216.
- 50
- 55 Composiciones detergentes
- 60 [0200] Los métodos de la presente invención son particularmente útiles para mejorar la secreción de polipéptidos en cantidades comercialmente importantes para su uso en composiciones detergentes. Tales polipéptidos incluyen, pero de forma no limitativa, proteasas, enzimas celulolíticas, amilasas y peroxidasas, o cualquier otra enzima o proteína biológica útil para la industria de los detergentes.
- 65 [0201] La composición de detergente de la presente invención se puede formular, por ejemplo, como una composición de detergente para lavado de ropa a mano o a máquina que incluye una composición de aditivo de lavandería adecuada para el pretratamiento de tejidos manchados y una composición de suavizante para la ropa añadida en el aclarado, o se

puede formular como una composición de detergente para su uso en operaciones de limpieza de superficies duras del hogar en general, o se puede formular para operaciones de lavado de vajilla a mano o a máquina.

5 [0202] En un aspecto específico, la presente invención proporciona un aditivo de detergente que comprende un polipéptido que tiene actividad biológica (por ejemplo, proteína de fusión, un componente de la misma o combinaciones de los mismos) obtenido según la presente invención. El aditivo de detergente al igual que la composición de detergente puede comprender unas o más enzimas más tales como una proteasa, lipasa, cutinasa, una amilasa, carbohidrasa, celulasa, pectinasa, mananasa, arabinasa, galactanasa, xilanasas, oxidasas, por ejemplo, una lacasa y/o peroxidasa.

10 [0203] En general, las propiedades de los componentes enzimáticos deberían ser compatibles con el detergente seleccionado (es decir, pH óptimo, compatibilidad con otros ingredientes enzimáticos y no enzimáticos, etc.), y los componentes enzimáticos deberían estar presentes en cantidades eficaces.

15 [0204] Proteasas: proteasas adecuadas incluyen aquellas de origen animal, vegetal o microbiano. Se prefiere el origen microbiano. Mutantes químicamente modificados o creados genéticamente de proteína están incluidos. La proteasa puede ser una serina proteasa o una metaloproteasa, preferiblemente una proteasa microbiana alcalina 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 la WO 89/06279). Ejemplos de proteasas de tipo tripsina son tripsina (por ejemplo, de origen porcino o bovino) y la proteasa de *Fusarium* descrita en la WO 89/06270 y la WO 94/25583.

20 [0205] Ejemplos de proteasas útiles son las variantes descritas en la WO 92/19729, WO 98/20115, WO 98/20116 y WO 98/34946, especialmente las variantes con sustituciones en una o varias de las siguientes posiciones: 27, 36, 57, 76, 87, 97, 101, 104, 120, 123, 167, 170, 194, 206, 218, 222, 224, 235 y 274.

25 [0206] Enzimas proteásicas disponibles comercialmente preferidas incluyen ALCALASE™, SAVINASE™, PRIMASE™, DURALASE™, ESPERASE™, AND KANNASE™ (NOVOZYMES A/S), MAXATASE™, MAXACAL™, MAXAPEM™, PROPERASE™, PURAFECT™, PURAFECT OXP™, FN2™ y FN3™ (Genencor International Inc.).

30 [0207] Lipasas: lipasas adecuadas incluyen aquellas de origen bacteriano o fúngico. Mutantes químicamente modificados o creados genéticamente de proteína están incluidos. Ejemplos de lipasas útiles incluyen lipasas de *Humicola* (sinónimo de *Thermomyces*), por ejemplo, de *H. lanuginosa* (*T. lanuginosus*) como se describe en la EP 258 068 y la EP 305 216 o de *H. insolens* como se describe en la WO 96/13580, una lipasa de *Pseudomonas*, por ejemplo, de *P. alcaligenes* o *P. pseudoalcaligenes* (EP 218 272), *P. cepacia* (EP 331 376), *P. stutzeri* (GB 1.372.034), *P. fluorescens*, *Pseudomonas* sp. cepa 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).

35 [0208] Otros ejemplos son variantes de lipasa tales como los descritos en la 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.

40 [0209] Lipasas disponibles comercialmente preferidas incluyen LIPOLASE™, LIPEX™ y Lipolase ULTRA™ (Novozymes A/S).

45 [0210] Amilasas: amilasas adecuadas (α y/o β) incluyen aquellas de origen bacteriano o fúngico. Mutantes químicamente modificados o creados genéticamente de proteína están incluidos. Las amilasas incluyen, por ejemplo, α -amilasas obtenidas de *Bacillus*, por ejemplo, una cepa especial de *Bacillus licheniformis*, descrita con más detalle en la GB 1.296.839.

50 [0211] Ejemplos de amilasas útiles son las variantes descritas en la WO 94/02597, WO 94/18314, WO 96/23873 y WO 97/43424, especialmente las variantes con sustituciones en una o varias 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.

55 [0212] Amilasas disponibles comercialmente son DURAMYL™, TERMAMYL™, FUNGAMYL™ y BAN™ (Novozymes A/S), Rapidase™ y Purastar™ (de Genencor International Inc.).

60 [0213] Celulasas: celulasas adecuadas incluyen aquellas de origen bacteriano o fúngico. Mutantes químicamente modificados o creados genéticamente de proteína están incluidos. Celulasas adecuadas incluyen celulasas del género *Bacillus*, *Pseudomonas*, *Humicola*, *Fusarium*, *Thielavia*, *Acremonium* o *Trichoderma* por ejemplo, las celulasas fúngicas producidas a partir de *Humicola insolens*, *Myceliophthora thermophila* y *Fusarium oxysporum* descritas en la patente de EE.UU. n° 4.435.307, patente de EE.UU. n° 5.648.263, patente de EE.UU. n° 5.691.178, patente de EE.UU. n° 5.776.757 y WO 89/09259.

65 [0214] Celulasas especialmente adecuadas son las celulasas alcalinas o neutras que tienen beneficios de cuidado del color. Ejemplos de tales celulasas son las celulasas descritas en la 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 los descritos en la WO 94/07998, EP 0 531 315, patente de EE.UU. n° 5.457.046, patente de EE.UU. n° 5.686.593, patente de EE.UU. n° 5.763.254, WO 95/24471, WO 98/12307 y PCT/DK98/00299.

5 [0215] Celulasas disponibles comercialmente incluyen CELLUCLAST®, CELLUZYME™ y CAREZYME™ (Novozymes A/S), Clazinase™ y Puradax HA™ (Genencor International Inc.) y KAC-500(B)™ (Kao Corporation).

10 [0216] Peroxidasas/oxidasas: peroxidasas/oxidasas adecuadas incluyen aquellas de origen vegetal, bacteriano o fúngico. Mutantes químicamente modificados o creados genéticamente de proteína están incluidos. Ejemplos de peroxidasas útiles incluyen peroxidasas de Coprinus, por ejemplo, de C. cinereus, y variantes de las misma como aquellas descritas en la WO 93/24618, WO 95/10602 y WO 98/15257.

[0217] Peroxidasas disponibles comercialmente incluyen GUARDZYME™ (Novozymes A/S).

15 [0218] El componente(s) enzimático se puede incluir en una composición de detergente añadiendo aditivos separados que contengan una o varias enzimas o añadiendo un aditivo combinado que comprenda todas estas enzimas. Un aditivo de detergente de la presente invención, es decir, un aditivo separado o un aditivo combinado, se puede formular, por ejemplo, como un granulado, líquido, lodo, etc. Formulaciones de aditivo de detergente preferidas son granulados, en particular granulados no polvorientos, líquidos, en particular líquidos estabilizados, o lodos.

20 [0219] Granulados no polvorientos se pueden producir, por ejemplo, como se describen en la patente de EE.UU. N° 4.106.991 y 4.661.452 y pueden opcionalmente ser recubiertos 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 donde el alcohólico contiene de 12 a 20 átomos de carbono y en el cual hay de 15 a 80 unidades de óxido de etileno; alcoholes grasos; ácidos grasos; y mono-, di- y triglicéridos de ácidos grasos. Ejemplos de materiales de recubrimiento que forman película adecuados para la aplicación por técnicas de lecho fluidificado se dan en la GB 1483591. Preparaciones enzimáticas líquidas se pueden estabilizar, por ejemplo, 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 la EP 238.216.

25 [0220] La composición de detergente de la presente invención puede tener cualquier forma conveniente, por ejemplo, una barra, una tableta, un polvo, un gránulo, una pasta o un líquido. Un detergente líquido puede ser acuoso, contener típicamente hasta 70% agua y 0-30% de solvente orgánico, o no acuoso.

30 [0221] La composición de detergente comprende uno o varios surfactantes, que pueden ser no iónicos incluyendo semipolar y/o aniónico y/o catiónico y/o zwitteriónico. Los surfactantes están típicamente presentes en un nivel de 0,1 % a 60% en peso.

35 [0222] Cuando se incluye en esto, el detergente contendrá normalmente de aproximadamente 1% a aproximadamente 40% de un surfactante aniónico tal como alquilbencenosulfonato lineal, alfa-olefinsulfonato, sulfato de alquilo (sulfato de alcohol graso), etoxisulfato alcohólico, alcanosulfonato secundario, alfa-sulfo éster metílico de ácido graso, ácido o jabón de alquil- o alquenilsuccínico.

40 [0223] Cuando se incluye en el mismo, el detergente contendrá normalmente de aproximadamente 0,2% a aproximadamente 40% de un surfactante no iónico tal como alcohol etoxilato, nonilfenol etoxilato, alquilpoliglicósido, alquildimetilaminaóxido, monoetanolamida de ácido graso etoxilado, monoetanolamida de ácido graso, amida de ácido graso de alquilo de polihidroxi o N-acil-N-alquilo derivados de glucosamina ("glucamidas").

45 [0224] El detergente puede contener 0-65 % de un constructor de detergente o agente complejante tal como zeolita, difosfato, trifosfato, fosfonato, carbonato, citrato, ácido nitrilotriacético, ácido etilenodiaminatetraacético, ácido dietileno-triaminopentaacético, ácido alquil- o alquenilsuccínico, silicatos solubles o silicatos estratificados (por ejemplo, SKS-6 de Hoechst).

50 [0225] El detergente puede comprender uno o varios polímeros. Ejemplos son carboximetilcelulosa, poli(vinilpirrolidona), poli(etileno glicol), poli(vinil alcohol), poli(vinilpiridina-N-óxido), poli(vinilimidazola), policarboxilatos tales como poliacrilatos, copolímeros de ácido maléico/acrílico y copolímeros de ácido lauril metacrilato/acrílico.

55 [0226] 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 de formación de perácido tal como tetraacetiltilenodiamina o nonanoiloxibencenosulfonato. Alternativamente, el sistema blanqueante puede comprender peroxiácidos de, por ejemplo, amida, imida o tipo sulfona.

60 [0227] El componente(s) enzimático de la composición de detergente de la presente invención se puede estabilizar usando agentes estabilizantes convencionales, por ejemplo, un poliol tal como propilenglicol o glicerol, un azúcar o alcohol de azúcar, ácido láctico, ácido bórico o un derivado de ácido bórico, por ejemplo, un éster de borato aromático, o

un derivado de ácido fenil borónico tal como ácido 4-formilfenilo borónico y la composición se puede formular como se describe en, por ejemplo, WO 92/19709 y WO 92/19708.

5 [0228] El detergente también puede contener otros ingredientes de detergentes convencionales tales como suavizantes de ropa incluyendo arcillas, potenciadores de espuma, supresores de espuma, agentes anticorrosión, agentes suspensores de suciedad, agentes antiredeposición de suciedad, tintes, bactericidas, blanqueadores ópticos, hidrotropos, inhibidores de decoloración o perfumes.

10 [0229] En las composiciones de detergente se puede añadir cualquier componente enzimático, en particular los polipéptidos que tienen actividad biológica de la presente invención, en una cantidad que corresponde 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.

15 [0230] Los polipéptidos que tienen actividad biológica de la presente invención se pueden incorporar adicionalmente a las formulaciones de detergente descritas en la WO 97/07202, que se incorpora en la presente como referencia.

Otros usos

20 [0231] Polipéptidos que tienen actividad biológica (por ejemplo, proteína de fusión, un componente de las mismas o combinaciones de las mismas) obtenidos según la presente invención o también se puede usar 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 reducir la pelusa, la formación de bolas, la modificación de textura y lavado a la piedra (N.K. Lange, in 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 también se puede usar en combinación con otras glicohidrolasas y enzimas relacionadas, como se describe en este caso, en el tratamiento de madera para biopulpaje o descortezado, fabricación de papel para modificación de fibra, blanqueo y reducción de costes de energía de refinación, tratamiento de aguas rápidas, importante para reciclaje de aguas residuales, reciclaje de fibra lignocelulósica tal como destintado y tratamiento de fibra secundaria, y utilización de residuo de madera (S.D, Mansfield and A.R. Esteghlalian in S.D, Mansfield and J.N. Saddler (Eds.), *Applications of Enzymes to Lignocellulosics*, ACS Symposium Series 855, Washington, D.C., 2003, pp. 2-29).

[0232] La presente invención se describe más en detalle mediante los siguientes ejemplos que no deberían interpretarse como limitación del ámbito de la invención.

35 Ejemplos

[0233] Los productos químicos usados como tampones y sustratos fueron productos comerciales de al menos calidad reactiva.

40 Secuenciación del ADN

[0234] La secuenciación del ADN se realizó utilizando un analizador genético modelo 3130X de Applied Biosystems (Applied Biosystems, Foster City, CA, EE.UU.) using dye terminator chemistry (Giesecke *et al.*, 1992, *Journal of Virol. Methods* 38: 47-60). Las secuencias fueron ensambladas utilizando phred/phrap/consed (University of Washington, Seattle, WA, EE.UU.) con cebadores específicos de secuencia.

Cepa

50 [0235] *Trichoderma reesei* RutC30 (ATCC 56765; Montencourt and Eveleigh, 1979, *Adv. Chem. Ser.* 181: 289-301) fue derivada de *Trichoderma reesei* Qm6A (ATCC 13631; Mandels and Reese, 1957, *J. Bacteriol.* 73: 269-278). La cepa de *Trichoderma reesei* RutC30 y de *Aspergillus oryzae* Jai355 (WO 02/062973) se usaron para la expresión de la proteína de fusión de beta-glicosidasa.

Medios y soluciones

55 [0236] El medio YP estaba compuesto por litro de 10 g de extracto de levadura y 20 g de bacto triptona.

60 [0237] El medio inductor de celulasa estaba compuesto por litro de 20 g de celulosa, 10 g de sólidos de maíz remojados, 1,45 g de (NH₄)₂SO₄, 2,08 g de KH₂PO₄, 0,28 g de CaCl₂, 0,42 g de MgSO₄·7H₂O y 0,42 ml de solución de metales traza.

[0238] La solución de metales traza estaba 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.

65 [0239] El STC estaba compuesto por 1 M sorbitol, 10 mM CaCl₂ y 10 mM de Tris-HCl, pH 7,5.

- [0240] Las placas de COVE estaban compuestas por litro de 342 g de sacarosa, 10 ml de solución de sales de COVE, 10 ml de 1 M de acetamida, 10 ml de 1,5 M de CsCl y 25 g de agar noble.
- 5 [0241] La solución de sales de COVE estana compuesta por litro de 26 g de KCl, 26 g de MgSO₄, 76 g de KH₂PO₄ y 50 ml de solución de metales traza de COVE.
- [0242] La solución de metales traza de COVE estaba compuesta por litro de 0,04 g de Na₂B₄O₇·10H₂O, 0,4 g de CuSO₄·5H₂O, 1,2 g de FeSO₄·7H₂O, 0,7 g de MnSO₄·H₂O, 0,8 g de Na₂MoO₂·H₂O y 10 g de ZnSO₄·7H₂O.
- 10 [0243] Las placas de COVE2 estaban compuestas por litro de 30 g de sacarosa, 20 ml de solución de sales de COVE, 25 g de agar noble y 10 ml de 1 M de acetamida.
- [0244] Las placas PDA estaban compuestas por litro de 39 gramos de agar de dextrosa de patata.
- 15 [0245] El medio LB estaba compuesto por litro de 10 g de triptona, 5 g de extracto de levadura, 5 g de cloruro sódico.
- [0246] Las placas 2X YT estaban 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.
- 20 [0247] El medio MDU2BP estaba compuesto por litro de 45 g de maltosa, 1 g de MgSO₄·7H₂O, 1 g de NaCl, 2 g de K₂HSO₄, 12 g de KH₂PO₄, 2 g de urea y 500 µl de solución de metales traza AMG, el pH se ajustó a 5,0 y luego se esterilizó por filtro con una unidad de filtración de 0,22 µm.
- 25 [0248] La solución de metales traza AMG estaba compuesta por litro de 14,3 g de ZnSO₄·7H₂O, 2,5 g de CuSO₄·5H₂O, 0,5 g de NiCl₂·6H₂O, 13,8 g de FeSO₄·7H₂O, 8,5 g de MnSO₄·7H₂O y 3 g de ácido cítrico.
- [0249] Las placas de medio mínimo estaban compuestas por litro de 6 g de NaNO₃, 0,52 de KCl, 1,52 g de KH₂PO₄, 1 ml de solución de metales traza de COVE, 20 g de agar noble, 20 ml de 50% de glucosa, 2,5 ml de 20% de MgSO₄·7H₂O y 20 ml de solución madre de biotina.
- 30 [0250] La solución madre de biotina estaba compuesta por litro de 0,2 g de biotina.
- [0251] El medio SOC estaba compuesto por 2% de triptona, 0,5% de extracto de levadura, 10 mM de NaCl, 2,5 mM de KCl, 10 mM de MgCl₂ y 10 mM de MgSO₄, seguido de adición de glucosa esterilizada por filtro a 20 mM después del autoclave.
- 35

Ejemplo 1: construcción de vector de expresión pMJ04

- 40 [0252] El vector de expresión pMJ04 se construyó por amplificador de PCR del terminador del gen de celobiohidrolasa 1 de *Trichoderma reesei* (cbh1, CEL7A) del ADN genómico de *Trichoderma reesei* RutC30 utilizando los cebadores 993429 (antisentido) y 993428 (sentido) mostrados a continuación. El cebador antisentido fue diseñado para tener un sitio *Pac* I en el final 5' y un sitio *Spe* I en el final 3' del cebador sentido.
- 45 Cebador 993429 (antisentido):
- 5'-AACGTTAATTAAGGAATCGTTTTGTGTTT-3' (SEC ID n°: 33)
- Cebador 993428 (sentido):
- 50 5'-AGTACTAGTAGCTCCGTGGCGAAAGCCTG-3' (SEC ID n°: 34)
- [0253] El ADN genómico de *Trichoderma reesei* RutC30 se aisló utilizando un kit de DNEASY® Plant Maxi (QIAGEN Inc., Valencia, CA, EE.UU.).
- 55 [0254] Las reacciones de amplificación (50 µl) estuvieron compuestas por 1X tampón de reacción ThermoPol (New England Biolabs, Beverly, MA, EE.UU.), 0,3 mM de dNTPs, 100 ng de ADN genómico de *Trichoderma reesei* RutC30, 0,3 µM de cebador 993429, 0,3 µM de cebador 993428 y 2 unidades de polimerasa de ADN Veny (New England Biolabs, Beverly, MA, EE.UU.). Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® (Eppendorf Scientific, Inc., Westbury, NY, EE.UU.) programado durante 5 ciclos, cada uno durante 30 segundos a 94°C, 30 segundos a 50°C y 60 segundos a 72°C, seguidos de 25 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 65°C y 120 segundos a 72°C (extensión final de 5 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando 40 mM de Tris base-20 mM de sodio acetate-1 mM de tampón de EDTA disodio (TAE) donde una banda de producto de 229 pares de bases fue cortado del gel y purificado utilizando un equipo de extracción de gel de QIAQUICK® (QIAGEN Inc., Valencia, CA, EE.UU.) según las instrucciones del fabricante.
- 60
- 65

[0255] El fragmento de PCR resultante fue digerido con *Pac* I y *Spe* I y ligado en pAILo1 (WO 05/067531) digerido con las mismas enzimas de restricción utilizando un equipo de ligamiento rápido (Roche, Indianapolis, IN, EE.UU.), para generar pMJ04 (figura 1).

5 Ejemplo 2: construcción de pCaHj568

[0256] El plásmido pCaHj568 fue construido a partir de pCaHj170 (patente de EE.UU. n° 5.763.254) y pMT2188. El plásmido pCaHj170 comprende la región de codificación de longitud completa de endoglucanasa V de *Humicola insolens* (CEL45A) (SEC ID n°: 1, que codifica la secuencia de aminoácidos de SEC ID n°: 2). La construcción de pMT2188 fue iniciada por amplificación por PCR del origen de replicación pUC19 de pCaHj483 (WO 98/00529) usando los cebadores 142779 y 142780 mostrados a continuación. El cebador 142780 introduce un sitio *Bbu* I en el fragmento de PCR.

142779:

5'-TTGAATTGAAAATAGATTGATTTAAACTTC-3' (SEC ID n°: 35)

142780:

5'-TTGCATGCGTAATCATGGTCATAGC-3' (SEC ID n°: 36)

[0257] Un sistema de PCR EXPAND® (Roche Molecular Biochemicals, Basel, Suiza) se usó después siguiendo las instrucciones del fabricante para esta amplificación. Los productos de PCR fueron separados en un gel de agarosa y un fragmento de 1160 pares de bases fue aislado y purificado utilizando un kit de giro de extracción de gel Jetquick (Genomed, Wielandstr, Alemania).

[0258] El gen URA3 fue amplificado a partir del vector de clonación de *Saccharomyces cerevisiae* general pYES2 (nvitrogen, Carlsbad, CA, EE.UU.) utilizando los cebadores 140288 y 142778 mostrados a continuación utilizando un sistema de PCR EXPAND®.

El cebador 140288 introdujo un sitio Eco RI en el fragmento de PCR.

5'-TTGAATTCATGGGTAATAACTGATAT-3' (SEC ID n°: 37)

142778:

5'-AAATCAATCTATTTTCAATTCAATTCATCATT-3' (SEC ID n°: 38)

[0259] Los productos de PCR fueron separados en un gel de agarosa y un fragmento de 1126 pares de bases fue aislado y purificado utilizando un equipo de giro de extracción de gel Jetquick.

[0260] Los dos fragmentos de la PCR fueron fusionados mediante mezcla y amplificados usando los cebadores 142780 y 140288 arriba mostrados por el método de empalme de solapamiento (Horton *et al.*, 1989, Gene 77: 61-68). Los productos de PCR fueron separados en un gel de agarosa y un fragmento de 2263 pares de bases fue aislado y purificado utilizando un equipo de giro de extracción de gel Jetquick.

[0261] El fragmento resultante fue digerido con *Eco* RI y *Bbu* I y ligado utilizando protocolos estándar al fragmento mayor de pCaHj483 digerido con las mismas enzimas de restricción. La mezcla de ligamiento fue transformada en la cepa de *E. coli* pirF-negativa DB6507 (ATCC 35673) hacha competente por el método de Mandel e Higa, 1970, J. Mol. Biol. 45: 154. Los transformantes se seleccionaron en el medio sólido M9 (Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, 2nd edition, Cold Spring Harbor Laboratory Press) suplementado por litro con 1 g de ácidos de casamino, 500 µg de tiamina y 10 mg de canamicina. Un plásmido de un transformante fue aislado y designado pCaHj527 (figura 2).

[0262] El promotor NA2-tpi presente en pCaHj527 fue sometido a mutagénesis dirigida al sitio por un simple método de PCR utilizando un sistema de PCR EXPAND® según las instrucciones del fabricante. Los nucleótidos 134-144 fueron convertidos de GTACTAAAACC (SEC ID n°: 39) a CCGTTAAATTT (SEC ID n°: 40) usando el cebador mutagénico 141223 mostrado a continuación.

Cebador 141223:

5'-GGATGCTGTTGACTCCGAAATTTAACGGTTTGGTCTTGCATCCC-3' (SEC ID n°: 41)

Los nucleótidos 423-436 fueron convertidos de ATGCAATTTAAACT (SEC ID n°: 42) a CGGCAATTTAACGG (SEC ID n°: 43) usando el cebador mutagénico 141222 mostrado a continuación.

Cebador 141222:

5'-GGTATTGTCCTGCAGACGGCAATTTAACGGCTTCTGCGAATCGC-3' (SEC ID n°: 44)

El plásmido resultante fue designado pMT2188 (figura 3).

5 [0263] La región codificante de endoglucanasa V de *Humicola insolens* fue transferida de pCaHj170 como un fragmento de *Bam* HI-*Sal* I en pMT2188 digerido con *Bam* HI y *Xho* I para generar pCaHj568 (figura 4). El plásmido pCaHj568 comprende un promotor NA2-*tpi* mutado operativamente enlazado a la secuencia codificante en toda su longitud de endoglucanasa V de *Humicola insolens*.

10 **Ejemplo 3: construcción de pMJ05**

[0264] El plásmido pMJ05 fue construido por amplificación de PCR de la región de codificación en toda su longitud de endoglucanasa V de *Humicola insolens* de 915 pares de bases de pCaHj568 usando los cebadores HiEGV-F y HiEGV-R mostrados a continuación.

15 HiEGV-F (sentido):

5'-AAGCTTAAGCATGCGTTCCTCCCCCTCC-3' (SEC ID n°: 45)

20 HiEGV-R (antisentido):

5'-CTGCAGAATTCTACAGGCACTGATGGTACCAG-3' (SEC ID n°: 46)

25 [0265] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de reacción ThermoPol (New England Biolabs, Beverly, MA, EE.UU.), 0,3 mM de dNTPs, 10 ng/µl de pCaHj568, 0,3 µM de cebador HiEGV-F, 0,3 µM de cebador HiEGV-R y 2 unidades de polimerasa de ADN Vent (New England Biolabs, Beverly, MA, EE.UU.). Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado 5 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 50°C y 60 segundos a 72°C, seguidos de 25 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 65°C y 120 segundos a 72°C (extensión final de 5 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1.0% usando el tampón de TAE donde una banda de producto de 937 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

35 [0266] El fragmento purificado de 937 pares de bases se usó como plantilla de ADN para amplificaciones posteriores utilizando los siguientes cebadores:

HiEGV-R (antisentido):

5'-CTGCAGAATTCTACAGGCACTGATGGTACCAG-3' (SEC ID n°: 47)

40

HiEGV-F-solapamiento (sentido):

5'-ACCGCGGACTGCGCATCATGCGTTCCTCCCCCTCC-3' (SEC ID n°: 48)

45

Las secuencias de cebador en cursiva son homólogas a 17 pares de bases del promotor del gen de celobiohidrolasa I de *Trichoderma reesei* (*cbh1*) y las secuencias de cebador subrayadas son homólogas a 29 pares de bases de la región de codificación de endoglucanasa V de *Humicola insolens*. Un solapamiento de 36 pares de bases entre el promotor y la secuencia codificante permitió una fusión precisa de un fragmento de 994 pares de bases que comprende el promotor *cbh1* de *Trichoderma reesei* con el fragmento de 918 pares de bases que comprende la región de codificación de endoglucanasa V de *Humicola insolens*.

50

[0267] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de reacción ThermoPol, 0,3 mM de dNTPs, 1 µl del fragmento de PCR purificado de 937 pares de bases, 0,3 µM del cebador HiEGV-F-solapamiento, 0,3 µM de cebador HiEGV-R y 2 unidades de polimerasa de DNA Vent. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 5 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 50°C y 60 segundos a 72°C, seguidos de 25 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 65°C y 120 segundos a 72°C (extensión final de 5 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando el tampón de TAE donde una banda de producto de 945 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

55

60

[0268] Se realizó una PCR separada para amplificar la secuencia promotora de *Trichoderma reesei cbh1* que se extiende desde 994 pares de bases arriba del codón de inicio ATG del gen del ADN genómico de *Trichoderma reesei* RutC30 utilizando los cebadores mostrados debajo (el cebador sentido se diseñó para tener un sitio de restricción *Sal* I en el final 5'). El ADN genómico de *Trichoderma reesei* RutC30 se aisló utilizando un equipo Plant Maxi de UNEASY®.

65

TrCBHlpro-F (sentido):

5'-AAACGTCGACCGAATGTAGGATTGTTATC-3' (SEC ID n°: 49)

5 TrCBHlpro-R (antisentido):

5'-GATGCGCAGTCCGCGGT-3' (SEC ID n°: 50)

10 [0269] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de reacción ThermoPol, 0,3 mM de dNTPs, 100 ng/µl de ADN genómico de *Trichoderma reesei* RutC30, 0,3 µM de cebador TrCBHlpro-F, 0,3 µM de cebador TrCBHlpro-R y 2 unidades de polimerasa de DNA Vent. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 55°C y 120 segundos a 72°C (extensión final de 5 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón de TAE donde una banda de producto de 998 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

15 [0270] El fragmento de PCR purificado de 998 pares de bases se usó como plantilla de ADN para amplificaciones posteriores utilizando los cebadores mostrados a continuación.

20 TrCBHlpro-F:

5'-AAACGTCGACCGAATGTAGGATTGTTATC-3' (SEC ID n°: 51)

25 TrCBHlpro-R-solapamiento:

5'-GGAGGGGGGAGGAACGCATGATGCGCAGTCCGCGGT-3' (SEC ID n°: 52)

30 [0271] Las secuencias en cursiva son homólogas por 17 pares de bases del promotor de *Trichoderma reesei cbh1* y las secuencias subrayadas son homólogas por 29 pares de bases de la región de codificación de endoglucanasa V de *Humicola insolens*. Un solapamiento de 36 pares de bases entre el promotor y la secuencia codificante permitió la fusión del fragmento de 994 pares de bases que comprende el promotor de *Trichoderma reesei cbh1* al fragmento de 918 pares de bases que comprende la región de codificación en toda su longitud de endoglucanasa V de *Humicola insolens*.

35 [0272] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de reacción ThermoPol, 0,3 mM de dNTPs, 1 µl del fragmento de PCR purificado de 998 pares de bases, 0,3 µM de cebador TrCBH1pro-F, 0,3 µM de cebador TrCBH1pro-R-solapamiento y 2 unidades de polimerasa de ADN Vent. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 5 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 50°C y 60 segundos a 72°C, seguidos de 25 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 65°C y 120 segundos a 72°C (extensión final de 5 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón de TAE donde una banda de producto de 1017 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

45 [0273] El fragmento de PCR promotor de 1017 pares de bases de *Trichoderma reesei cbh1* y el fragmento de PCR de 945 pares de bases de endoglucanasa V de *Humicola insolens* se usaron como plantilla de ADN para amplificación posterior utilizando los siguientes cebadores para fundir precisamente el promotor de 994 pares de bases *cbh1* a la región de codificación en toda su longitud de endoglucanasa V de 918 pares de bases utilizando la PCR de superposición.

50 TrCBHlpro-F:

5'-AAACGTCGACCGAATGTAGGATTGTTATC-3' (SEC ID n°: 53)

55 HiEGV-R:

5'-CTGCAGAATTCTACAGGCACTGATGGTACCAG-3' (SEC ID n°: 54)

60 [0274] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de reacción de ThermoPol, 0,3 mM dNTPs, 0,3 µM de cebador TrCBH1pro-F, 0,3 µM de cebador HiEGV-R y 2 unidades de polimerasa de ADN Vent. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 5 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 50°C y 60 segundos a 72°C, seguidos de 25 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 65°C y 120 segundos a 72°C (extensión final de 5 minutos). Los productos reactivos fueron aislados en un gel de agarosa al 1.0% usando tampón TAE donde una banda de producto de 1926 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

65

[0275] El fragmento resultante de 1926 pares de bases fue clonado en un vector pCR®-Blunt-II-TOPO® (Invitrogen, Carlsbad, CA, EE.UU.) utilizando un equipo de clonación por PCR TOPO® de ZEROBLUNT® (Invitrogen, Carlsbad, CA, EE.UU.) siguiendo el protocolo del fabricante.

5 El plásmido resultante fue digerido con *Not* I y *Sal* I y el fragmento de 1926 pares de bases se purificó con gel utilizando un equipo de extracción de gel QIAQUICK® y se ligó usando T4 ADN-ligasa (Roche, Indianapolis, IN, EE.UU.) en pMJ04, que fue también digerido con las mismas dos enzimas de restricción, para generar pMJ05 (figura 5). El plásmido pMJ05 comprende el promotor y el terminador de la celobiohidrolasa I de *Trichoderma reesei* operativamente enlazado a la secuencia codificante en toda su longitud de endoglucanasa V de *Humicola insolens*.

10 **Ejemplo 4: construcción del vector de expresión pSMai130**

[0276] Un fragmento de ADN de 2586 pares de bases que abarca desde el codón de inicio ATG al codón de terminación TAA de la secuencia codificante en toda su longitud de beta-glucosidasa de *Aspergillus oryzae* (SEC ID n°: 21 para secuencia de ADNc y SEC ID n°: 22 para la secuencia de aminoácidos deducida; *E. coli* DSM 14240) fue amplificado por PCR a partir de pJal660 (WO 2002/095014) como plantilla con los cebadores 993467 (sentido) y 993456 (antisentido) mostrados a continuación. Un sitio *Spe* I fue diseñado en el final 5' del cebador antisentido para facilitar ligamiento. Las secuencias de cebador en cursiva son homólogas por 24 pares de bases del promotor de *Trichoderma reesei cbh1* y las secuencias subrayadas son homólogas por 22 pares de bases de la región de codificación de beta-glucosidasa de *Aspergillus oryzae*.

20 Cebador 993467:

5'-*ATAGTCAACCGCGGACTGCGCATCATGAAGCTTGGTTGGATCGAGG*-3' (SEC ID n°: 55)

25 Cebador 993456:

5'-*ACTAGTTTACTGGGCCTTAGGCAGCG*-3' (SEC ID n°: 56)

30 [0277] Las reacciones de amplificación (50 µl) estaban compuestas por tampón de amplificación *Pfx* (Invitrogen, Carlsbad, CA, EE.UU.), 0,25 mM de dNTPs, 10 ng de pJal660, 6,4 µM de cebador 993467, 3,2 µM de cebador 993456, 1 mM de MgCl₂ y 2,5 unidades de *Pfx* ADN polimerasa (Invitrogen, Carlsbad, CA, EE.UU.). Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 1 minuto a 94°C, 1 minuto a 55°C y 3 minutos a 72°C (extensión final de 15 minutos). Los productos reactivos fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 2586 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0278] Un PCR separado se realizó para amplificar la secuencia promotora de *Trichoderma reesei cbh1* que se extiende desde 1000 pares de bases arriba del codón de inicio ATG del gen, usando el cebador 993453 (sentido) y el cebador 993463 (antisentido) mostrado debajo para generar un fragmento de PCR de 1000 pares de bases.

40 Cebador 993453:

5'-*GTCGACTCGAAGCCCGAATGTAGGAT*-3' (SEC ID n°: 57)

45 Cebador 993463:

5'-*CCTCGATCCAACCAAGCTTCATGATGCGCAGTCCGCGGTTGACTA*-3' (SEC ID n°: 58)

50 Las secuencias de cebador en cursiva son homólogas a los 24 pares de bases del promotor de *Trichoderma reesei cbh1* y las secuencias de cebador subrayadas son homólogas a los 22 pares de bases de la región de codificación en toda su longitud de beta-glucosidasa de *Aspergillus oryzae*. El solapamiento de 46 pares de bases entre el promotor y la secuencia codificante permitió la fusión del fragmento de 1000 pares de bases que comprende el promotor de *Trichoderma reesei cbh1* con el fragmento de 2586 pares de bases que comprende la región de codificación de beta-glucosidasa de *Aspergillus oryzae*.

55 [0279] Las reacciones de amplificación (50 µl) estaban compuestas por el tampón de amplificación *Pfx*, 0,25 mM de dNTPs, 100 ng de ADN genómico de *Trichoderma reesei* RutC30, 6,4 µM de cebador 993453, 3,2 µM de cebador 993463, 1 mM de MgCl₂ y 2,5 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 1 minuto a 94°C, 1 minuto a 55°C y 3 minutos a 72°C (extensión final de 15 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 1000 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

65 [0280] Los fragmentos purificados fueron usados como plantilla de ADN para amplificación posterior por PCR de superposición usando el cebador 993453 (sentido) y el cebador 993456 (antisentido) arriba mostrados para fusionar precisamente el fragmento de 1000 pares de bases que comprende el promotor de *Trichoderma reesei cbh1* con el

fragmento de 2586 pares de bases que comprende la región de codificación en toda su longitud de beta-glucosidasa de *Aspergillus oryzae*.

[0281] Las reacciones de amplificación (50 µl) estaban compuestas por tampón de amplificación *Pfx*, 0,25 mM de dNTPs, 6,4 µM de cebador 99353, 3,2 µM de cebador 993456, 1 mM de MgCl₂ y 2,5 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 1 minuto a 94°C, 1 minuto a 60°C y 4 minutos a 72°C (extensión final de 15 minutos).

[0282] El fragmento resultante de 3586 pares de bases fue digerido con *Spe* I y *Sal* I y ligado en el pMJ04, digerido con las mismas dos enzimas de restricción, para generar pSMai130 (figura 6). El plásmido pSMai130 comprende el promotor de gen de celobiohidrolasa I de *Trichoderma reesei* y el terminador operativamente enlazado a la secuencia señal y la secuencia codificante de beta-glucosidasa nativa de *Aspergillus oryzae* (es decir, la secuencia de codificación de beta-glucosidasa de *Aspergillus oryzae* en toda su longitud).

Ejemplo 5: construcción de pSMai135

[0283] La región de codificación madura de beta-glucosidasa de *Aspergillus oryzae* (menos la secuencia señal nativa, véase figura 7; SEC ID n°: 59 y 60) de Lys-20 al codón de parada TAA fue amplificada por PCR a partir de pJaL660 como plantilla con el cebador 993728 (sentido) y el cebador 993727 (antisentido) mostrado a continuación.

Cebador 993728:

5'-TGCCGGTGTGGCCCTTGCCAAGGATGATCTCGCGTACTCCC-3' (SEC ID n°: 61)

Cebador 993727:

5'-GACTAGTCTTACTGGGCCTTAGGCAGCG-3' (SEC ID n°: 62)

Las secuencias en cursiva son homólogas a 20 pares de bases de la secuencia de señal de endoglucanasa V de *Humicola insolens* y las secuencias subrayadas son homólogas a 22 pares de bases de la región de codificación de beta-glucosidasa de *Aspergillus oryzae*. Un sitio *Spe* I fue diseñado en el final 5' del cebador antisentido.

[0284] Las reacciones de amplificación (50 µl) estaban compuestas por tampón de amplificación *Pfx*, 0,25 mM de dNTPs, 10 ng/µl de pJaL660, 6,4 µM de cebador 993728, 3,2 µM de cebador 993727, 1 mM de MgCl₂ y 2,5 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 1 minuto a 94°C, 1 minuto a 55°C y 3 minutos a 72°C (extensión final de 15 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 2523 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0285] Una amplificación por PCR separada se realizó para amplificar 1000 pares de bases del promotor de *Trichoderma reesei cbh1* y 63 pares de bases de la secuencia de señal de endoglucanasa V de *Humicola insolens* (codón de inicio ATG para Ala-21, Figura 8, SEC ID n°: 63 y 64), usando el cebador 993724 (sentido) y el cebador 993729 (antisentido) mostrado a continuación.

Cebador 993724:

5'-ACGCGTCGACCGAATGTAGGATTGTTATCC-3' (SEC ID n°: 65)

Cebador 993729:

5'-GGGAGTACGCGAGATCATCCTTGGCAAGGGCCAACACCGGCA-3' (SEC ID n°: 66)

[0286] Las secuencias de cebador en cursiva son homólogas a 20 pares de bases de la secuencia de señal de endoglucanasa V de *Humicola insolens* y las secuencias de cebador subrayadas son homólogas a 22 pares de bases de la región de codificación de beta-glucosidasa de *Aspergillus oryzae*.

[0287] El plásmido pMJ05, que comprende la región de codificación de endoglucanasa V de *Humicola insolens* bajo el control del promotor *cbh1*, se usó como modelo para generar un fragmento de 1063 pares de bases que comprende el promotor de *Trichoderma reesei cbh1* y el fragmento de secuencia de señal de endoglucanasa V de *Humicola insolens*. Un solapamiento de 42 pares de bases fue compartido entre el promotor de *Trichoderma reesei cbh1* y la secuencia señal de endoglucanasa V de *Humicola insolens* y la secuencia codificante madura de beta-glucosidasa de *Aspergillus oryzae* para proporcionar un enlace perfecto entre el promotor y el codón de inicio ATG de la región de codificación de beta-glucosidasa de *Aspergillus oryzae* de 2523 pares de bases.

[0288] Las reacciones de amplificación (50 µl) estaban compuestas por tampón de amplificación *Pfx*, 0,25 mM de dNTPs, 10 ng/µl de pMJ05, 6,4 µM de cebador 993728, 3,2 µM de cebador 993727, 1 mM de MgCl₂ y 2,5 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 1 minuto a 94°C, 1 minuto a 60°C y 4 minutos a 72°C (extensión final de 15 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 1063 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0289] Los fragmentos de solapamiento purificados fueron usados como modelos para amplificación usando el cebador 993724 (sentido) y el cebador 993727 (antisentido) anteriormente descritos para fusionar de forma precisa el fragmento de 1063 pares de bases que comprende el promotor de *Trichoderma reesei cbh1* y la secuencia señal de endoglucanasa V de *Humicola insolens* con el fragmento de 2523 pares de bases que comprende el marco de la región de codificación madura de beta-glucosidasa de *Aspergillus oryzae* por PCR de superposición.

[0290] Las reacciones de amplificación (50 µl) estaban compuestas por tampón de amplificación *Pfx*, 0,25 mM de dNTPs, 6,4 µM de cebador 993724, 3,2 µM de cebador 993727, 1 mM de MgCl₂ y 2,5 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 1 minuto a 94°C, 1 minuto a 60°C y 4 minutos a 72°C (extensión final de 15 minutos). Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 3591 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0291] El fragmento resultante de 3591 pares de bases fue digerido con *Spe* I y *Sal* I y ligado en pMJ04 digerido con las mismas enzimas de restricción para generar pSMai135 (figura 9). El plásmido pSMai135 comprende el promotor del gen de celobiohidrolasa I de *Trichoderma reesei* y el terminador operativamente enlazado a la secuencia señal de endoglucanasa V de *Humicola insolens* y la secuencia codificante madura de beta-glucosidasa de *Aspergillus oryzae*.

Ejemplo 6: expresión de beta-glucosidasa de *Aspergillus oryzae* con la señal de secreción de endoglucanasa V de *Humicola insolens*

[0292] El plásmido pSMai135 que codifica la beta-glucosidasa de *Aspergillus oryzae* madura enlazado a la señal de secreción de endoglucanasa V de *Humicola insolens* (figura 8) se introdujo en el RutC30 de *Trichoderma reesei* por transformación mediada por PEG (Penttila *et al.*, 1987, Gene 61 155-164). El plásmido contenía el gen *amdS* de *Aspergillus nidulans* para permitir que los transformantes crecieran en acetamida como única fuente de nitrógeno.

[0293] El RutC30 de *Trichoderma reesei* fue cultivado a 27°C y 90 r.p.m. en 25 ml de medio YP suplementado con 2% (p/v) de glucosa y 10 mM de uridina durante 17 horas. Se recogieron micelios por filtración utilizando un sistema de filtración desechable impulsado por vacío (Millipore, Bedford, MA, EE.UU.) y se lavaron dos veces con agua desionizada y dos veces con 1,2 M de sorbitol. Se generaron protoplastos por suspensión de los micelios lavados en 20 ml de 1,2 M de sorbitol que contenían 15 mg de GLUCANEX® (Novozymes A/S, Bagsværd, Dinamarca) por ml y 0,36 unidades de quitinasa (Sigma Chemical Co., St. Louis, MO, EE.UU.) por ml e incubación durante 15-25 minutos a 34°C con agitación suave a 90 r.p.m. Se recogieron protoplastos por centrifugado durante 7 minutos a 400 x g y se lavaron dos veces con frío 1,2 M de sorbitol. Los protoplastos se contaron utilizando un hemocitómetro y se resuspendieron en STC a una concentración final de 1 X 10⁸ de protoplastos por ml. Los protoplastos de exceso se almacenaron en un contenedor de congelación 1°C Cryo (Nalgene, Rochester, NY, EE.UU.) a 80°C.

[0294] Aproximadamente 7 µg de pSMai135 digerido con *Pme* I se añadió a 100 µl de solución de protoplasto y se mezcló suavemente, seguido de 260 µl de tampón de PEG, mezclado e incubado a temperatura ambiente durante 30 minutos. Luego se añadió STC (3 ml) y se mezcló y la solución de transformación se colocó en placas sobre placas de COVE usando selección *amdS* de *Aspergillus nidulans*. Las placas fueron incubadas a 28°C durante 5-7 días. Los transformantes se subcultivaron sobre placas COVE2 y se dejaron crecer a 28°C.

[0295] Sesenta y siete transformantes designados SMA135 obtenidos con pSMai135 fueron subcultivados en placas frescas que contenían acetamida se dejaron esporular durante 7 días a 28°C.

[0296] Los 67 transformantes SMA135 de *Trichoderma reesei* fueron cultivados en matraces de agitación con deflectores de 125 ml que contenían 25 ml de medios de inducción de celulasa a pH 6,0 inoculados con esporas de los transformantes e incubados a 28°C y 200 r.p.m. durante 7 días. RutC30 de *Trichoderma reesei* se usó como control. Las muestras de caldo de cultivo se retiraron el día 7. Un ml de cada caldo de cultivo fue centrifugado a 15.700 x g durante 5 minutos en un microcentrifugador y los sobrenadantes transferidos a nuevos tubos. Las muestras fueron almacenadas a 4°C hasta ensayo enzimático. Los sobrenadantes fueron evaluados para actividad de beta-glucosidasa utilizando p-nitrofenil-beta-D-glucopiranosido como sustrato, como se describe a continuación.

[0297] La actividad de beta-glucosidasa fue determinada a temperatura ambiente usando 25 µl de partes alícuotas de sobrenadantes de cultivo, diluidas 1:10 en 50 mM de succinato pH 5,0, en 200 de 0,5 mg/ml p-nitrofenilo- beta-D-glucopiranosido como sustrato en 50 mM de succinato pH 5,0. Después de 15 minutos de incubación la reacción se

detuvo añadiendo 100 µl de 1 M Tris-HCl pH 8,0 y la absorbancia se leyó espectrofotométricamente en 405 nm. Una unidad de actividad de beta-glucosidasa correspondía a la producción de 1 µmol de p-nitrofenilo por minuto por litro a pH 5,0, temperatura ambiente. La beta-glucosidasa de *Aspergillus niger* (NOVOZYM™ 188, Novozymes A/S, Bagsvaerd, Dinamarca) se usó como un estándar enzimático.

5

[0298] Varios transformantes SMA135 mostraron actividades de beta-glucosidasa varias veces superiores que la segregada por RutC30 de *Trichoderma reesei*. De los transformantes SMA135 seleccionados, el transformante SMA135-04 produjo la máxima actividad de beta-glucosidasa.

10

[0299] SDS-PAGE se efectuó utilizando geles (5% resolución) Tris-HCl de CRITERION® (Bio-Rad, Hercules, CA, EE.UU.) con el Sistema CRITERION®, (Bio-Rad, Hercules, CA, EE.UU.). Cinco µl de los sobrenadantes del día 7 (véase arriba) fueron suspendidos en 2X concentración de tampón de muestra de Laemmli, (Bio-Rad, Hercules, CA, EE.UU.) y hervidos en presencia de 5% de beta-mercaptoetanol durante 3 minutos. Las muestras de sobrenadantes fueron cargadas en un gel de poliacrilamida y sometidas a electroforesis con 1X Tris/Glicina/SDS como tampón de ejecución, (Bio-Rad, Hercules, CA, EE.UU.). El gel resultante fue manchado con mancha Coomassie de BIO-SAFE®, (Bio-Rad, Hercules, CA, EE.UU.).

15

20

[0300] De los 38 transformantes SMA135 de *Trichoderma reesei* analizados por SDS-PAGE, 26 produjeron una proteína de aproximadamente 110 kDa que no era visible en el RutC30 de *Trichoderma reesei* como control. El transformante SMA135-04 de *Trichoderma reesei* produjo el máximo nivel de beta-glucosidasa como se destaca por abundancia de la banda de 110 kDa vista por SDS-PAGE.

Ejemplo 7: construcción del vector de expresión pSMai140

25

[0301] El vector de expresión pSMai140 fue construido digiriendo el plásmido pSATE111BG41 (WO 04/099228), que porta la región de codificación en toda su longitud de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 (SEC ID n°: 23 que codifica la secuencia de aminoácidos de SEC ID n°: 24), con *Nco* I. El fragmento resultante de 1243 pares de bases fue aislado en un gel de agarosa al 1,0% usando tampón TAE y purificado utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

30

35

[0302] El vector de expresión pSMai135 fue digerido con *Nco* I y un fragmento de 8286 pares de bases fue aislado en un gel de agarosa al 1,0% usando tampón TAE y purificado utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante. El fragmento BG41 de la variante de beta-glucosidasa de *Aspergillus oryzae* digerido por *Nco* I de 1243 pares de bases fue luego ligado al vector de 8286 pares de bases, utilizando T4 ADN-ligasa (Roche, Indianapolis, IN, EE.UU.) según protocolo del fabricante, para crear el vector de expresión pSMai140 (figura 10). El plásmido pSMai140 comprende el promotor de gen de celobiohidrolasa I de *Trichoderma reesei* (CEL7A) y el terminador operativamente enlazado a la secuencia de señal de endoglucanasa V de *Humicola insolens* y la secuencia codificante madura de la variante de beta-glucosidasa de *Aspergillus oryzae*.

40

Ejemplo 8: transformación de *Trichoderma reesei* RutC30 con pSMai140

45

[0303] El plásmido pSMai140 fue linealizado con *Pme* I y transformado en la cepa RutC30 de *Trichoderma reesei* como se describe en el ejemplo 6. Un total de 100 transformantes se obtuvieron de cuatro experimentos de transformación independientes, todos fueron cultivados en matraces de agitación en medio inducido por celulosa, y la actividad de beta-glucosidasa fue medida del medio de cultivo de los transformantes como se describe en el ejemplo 6. Un número de transformantes SMA140 de *Trichoderma reesei* mostró actividades de beta-glucosidasa varias veces superiores que la de *Trichoderma reesei* RutC30.

50

[0304] La presencia de la proteína de la variante BG41 de la beta-glucosidasa de *Aspergillus oryzae* en el medio de cultivo fue detectada por electroforesis en gel de SDS-poliacrilamida como se describe en el ejemplo 6 y coloración de Coomassie de los mismos 13 sobrenadantes de cultivo cuya actividad enzimática fue analizada. Los trece transformantes que tenían alta actividad de β-glucosidasa, también expresaron los aproximadamente 110 KDa de la variante BG41 de beta-glucosidasa de *Aspergillus oryzae*, a rendimientos variables.

55

[0305] El más alto transformante de expresión de variante de beta-glucosidasa, como evaluado por ensayo de actividad de beta-glucosidasa y electroforesis en gel de SDS-poliacrilamida, fue designado *Trichoderma reesei* SMA140-43.

Ejemplo 9: construcción del vector de expresión pSaMe-F1

60

[0306] Un fragmento de ADN que contenía 209 pares de bases del promotor del gen de celobiohidrolasa I de *Trichoderma reesei* y la región central (nucleótidos 1 a 702 de SEC ID n°: 1 que codifican los aminoácidos 1 a 234 de SEC ID n°: 2; WO 91/17243) del gen de endoglucanasa V de *Humicola insolens* fue amplificado por PCR usando pMJ05 como modelo utilizando los cebadores mostrados por debajo.

65

995103:

5'-cccaagcttagccaagaaca-3' (SEC ID n°: 67)

995137:

5 5'-gggggaggaacgcatgggatctggacggc-3' (SEC ID n°: 68)

[0307] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de amplificación *Pfx*, 10 mM de dNTPs, 50 mM de MgSO₄, 10 ng/µl de pMJ05, 50 picomoles de cebador 995103, 50 picomoles de cebador 995137 y 2 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 55°C y 60 segundos a 72°C (extensión final de 3 minutos).

[0308] Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 911 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0309] Un fragmento de ADN que contenía 806 pares de bases del gen BG41 de la variante de beta-glucosidasa de *Aspergillus oryzae* se amplificó por PCR utilizando pSMai140 como modelo y los cebadores mostrados por debajo.

20 995133:

5'-gccgtccagatccccatgcttctctcccc-3' (SEC ID n°: 69)

995111:

25 5'-ccaagcttgctcagagtttc-3' (SEC ID n°: 70)

[0310] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de amplificación *Pfx*, 10 mM de dNTPs, 50 mM de MgSO₄, 100 ng de pSMai140, 50 picomoles de cebador 995133, 50 picomoles de cebador 995111 y 2 unidades de *Pfx* ADN polimerasa. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado durante 30 ciclos cada uno durante 30 segundos a 94°C, 30 segundos a 55°C y 120 segundos a 72°C (extensión final de 3 minutos).

[0311] Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 806 pares de bases fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0312] Los dos fragmentos de la PCR anterior fueron luego sometidos a PCR de superposición. Los fragmentos de superposición purificados fueron usados como modelos para amplificación utilizando el cebador 995103 (sentido) y el cebador 995111 (antisentido) anteriormente descritos para fusionar de forma precisa el fragmento de 702 pares de bases que comprende 209 pares de bases del promotor del gen de celobiohidrolasa I de *Trichoderma reesei* y la secuencia central de endoglucanasa V de *Humicola insolens* con el fragmento de 806 pares de bases que comprende una parte de la región de codificación de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 por PCR de superposición.

[0313] Las reacciones de amplificación (50 µl) estaban compuestas por 1X tampón de amplificación *Pfx*, 10 mM de dNTPs, 50 mM de MgSO₄, 2,5 µl de cada fragmento (20 ng/µl), 50 picomoles de cebador 995103, 50 picomoles de cebador 995111 y 2 unidades de *Pfx* ADN polimerasa de alta fidelidad. Las reacciones fueron incubadas en un MASTERCYCLER® 5333 de EPPENDORF® programado para una desnaturalización inicial de 3 minutos a 95°C seguida de 30 ciclos cada uno durante 1 minuto de desnaturalización, 1 minuto de recocido a 60°C y una extensión de 3 minutos a 72°C.

[0314] Los productos de reacción fueron aislados en un gel de agarosa al 1,0% usando tampón TAE donde una banda de producto de 1,7 kb fue cortada del gel y purificada utilizando un equipo de extracción de gel QIAQUICK® según las instrucciones del fabricante.

[0315] El fragmento 1,7 kb fue ligado en un vector Blunt pCR®4 (Invitrogen, Carlsbad, CA, EE.UU.) según las instrucciones del fabricante. El constructo se transformó luego en células de *E. coli* químicamente competentes TOP10 de ONE SHOT® (Invitrogen, Carlsbad, CA, EE.UU.) según el procedimiento de transformación química rápida del fabricante. Se seleccionaron colonias y se analizaron por aislamiento de plásmido y digestión con *Hind* III para liberar el fragmento de PCR de superposición de 1,7 kb.

[0316] El plásmido pSMai140 fue también digerido con *Hind* III para linearizar el plásmido. Ambos fragmentos digeridos fueron combinados en una reacción de ligamiento utilizando un equipo de ligamiento de ADN rápido siguiendo las instrucciones del fabricante para producir pSaMe-F1 (figura 11).

[0317] Células competentes de grado subclonación de *E. coli* XL1-Blue (Stratagene, La Jolla, CA, EE.UU.) se transformaron con el producto de ligamiento. La identidad del constructo fue confirmada por secuenciación del ADN del promotor del gen de celobiohidrolasa I de *Trichoderma reesei*, la secuencia de señal de endoglucanasa V de *Humicola insolens*, en núcleo de endoglucanasa V de *Humicola insolens*, la secuencia señal de endoglucanasa V de *Humicola insolens*, la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 y la secuencia terminadora del gen de celobiohidrolasa I de *Trichoderma reesei* a partir de los plásmidos purificados de *E. coli* transformado. Un clon que contenía el plásmido recombinante fue designado pSaMe-F1. El plásmido pSaMe-F1 comprende el promotor y el terminador del gen de celobiohidrolasa I de *Trichoderma reesei* y la secuencia de péptido señal de endoglucanasa V de *Humicola insolens* enlazada directamente al polipéptido de núcleo de endoglucanasa V de *Humicola insolens* que se fusionan directamente con el péptido señal de endoglucanasa V de *Humicola insolens* que está enlazado directamente a la secuencia codificante madura de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41. La secuencia de ADN y secuencia de aminoácidos deducida de la proteína de fusión de variante de beta-glucosidasa de *Aspergillus oryzae* BG se muestra en SEC ID n°: 73 y 74, respectivamente (véase figuras 14A, 14B, 14C y 14D).

15 **Ejemplo 10: transformación de *Trichoderma reesei* RutC30 con pSaMe-F1**

[0318] Matrices de agitación que contenían 25 ml de medio YP suplementado con 2% de glucosa y 10 mM de uridina fueron inoculados con 5×10^7 de esporas de *Trichoderma reesei* RutC30. Tras la incubación toda la noche durante aproximadamente 16 horas a 27°C, 90 r.p.m., los micelios fueron recogidos utilizando un sistema de filtración desechable impulsado por vacío. Los micelios fueron lavados dos veces en 100 ml de agua desionizada y dos veces en 1,2 M de sorbitol. Se generaron protoplastos como se describe en el ejemplo 6.

[0319] Dos microgramos de ADN pSaMe-F1 linealizado con *Pme* I, 100 µl de protoplastos de RutC30 de *Trichoderma reesei* y 50% de PEG (4000) fueron mezclados e incubados durante 30 minutos a temperatura ambiente. Luego, 3 ml de STC se añadió y el contenido se vertió sobre una placa de COVE suplementada con 10 mM de uridina. La placa se incubó luego a 28°C. Transformantes empezaron a aparecer el día 6 y fueron escogidos para placas COVE2 para crecimiento a 28°C y 6 días. Veintidós transformantes de *Trichoderma reesei* fueron recuperados.

[0320] Los transformantes fueron cultivados en frascos de agitación en el medio de inducción de celulosa y la actividad de beta-glucosidasa se midió como se describe en el ejemplo 6. Varios transformantes pSaMe-F1 produjeron actividad de beta-glucosidasa. Un transformante, denominado SaMeF1-9 de *Trichoderma reesei*, produjo la máxima cantidad de beta-glucosidasa, y tuvo dos veces la actividad de una cepa que expresa la variante de beta-glucosidasa de *Aspergillus oryzae* (ejemplo 9).

[0321] La actividad de endoglucanasa se evaluó utilizando un ensayo de superposición de carboximetilcelulosa (CMC) según Beguin, 1983, Analytical Biochem. 131(2): 333-336. Cinco µg de la proteína de cinco de las muestras de caldo (las que tuvieron la máxima actividad de beta-glucosidasa) fueron diluidos en el tampón de muestra nativa, (Bio-Rad, Hercules, CA, EE.UU.) y se ejecutaron en un gel Tris-HCl al 8-16% de CRITERION®, (Bio-Rad, Hercules, CA, EE.UU.) usando 10X tampón de ejecución Tris/glicina, (Bio-Rad, Hercules, CA, EE.UU.) y luego el gel se dejó sobre una placa que contenía 1% de carboximetilcelulosa (CMC). Después de 1 hora de incubación a 37°C, el gel fue manchado con 0,1% de rojo Congo durante 20 minutos. La placa se desmanchó luego usando 1 M de NaCl para identificar regiones de indicativo de aclaramiento de actividad de endoglucanasa. Dos zonas de aclaramiento fueron visibles, una zona superior en aproximadamente 110 kDa y una zona inferior en aproximadamente 25 kDa. El tamaño de proteína previsto de la endoglucanasa V de *Humicola insolens* y la fusión de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 es 118 kDa si las dos proteínas no se dividen y permanecen como un único polipéptido; la glicosilación del dominio de núcleo de la endoglucanasa V individual y la beta-glucosidasa lleva a migración de las proteínas individuales a PM más altos que los previstos de la secuencia primaria. Si las dos proteínas son divididas luego los tamaños predichos para el dominio de núcleo de la endoglucanasa V de *Humicola insolens* es 24 kDa y 94 kDa para la variante de beta-glucosidasa de *Aspergillus oryzae* BG41. Debido a que hubo una zona de aclaramiento en aproximadamente 110 kDa, este resultado indicó que mínimamente una población de proteína de fusión de endoglucanasa y beta-glucosidasa permanece intacta como una única proteína grande. La zona de aclaramiento inferior muy probablemente representa una actividad de endoglucanasa endógena, y posiblemente adicionalmente se origina de la escisión parcial del dominio de núcleo de la endoglucanasa V de *Humicola insolens* a partir de la β-glucosidasa de *Aspergillus oryzae*.

[0322] Los resultados demostraron que el núcleo de endoglucanasa V de *Humicola insolens* estaba activo incluso mientras estaba fusionado a la beta-glucosidasa de *Aspergillus oryzae*. Además, el aumento de la actividad de beta-glucosidasa pareció originarse de la secreción aumentada de proteína con respecto a la eficiencia de secreción de la beta-glucosidasa de no fusión. Al enlazar la secuencia de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 con el núcleo de la endoglucanasa V de *Humicola insolens*, se segregó más beta-glucosidasa.

60 **Ejemplo 11: construcción del vector pSaMe-FX**

[0323] El plásmido pSaMe-FX fue construido modificando pSaMe-F1. El plásmido pSaMe-F1 fue digerido con Bst Z17 y Eco RI para generar un fragmento de 1 kb que contenía la secuencia codificante de la variante de beta-glucosidasa BG41 y un fragmento de 9,2 kb que contenía el resto del plásmido. Los fragmentos fueron separados en un gel de agarosa 1,0% usando tampón TAE y el fragmento de 9.2 kb fue cortado y purificado utilizando un equipo de extracción

de gel QIAQUICK® según las instrucciones del fabricante. El plásmido pSMai135 fue también digerido con Bst Z17 y Eco RI para generar un fragmento de 1 kb que contenía bases homólogas a la secuencia codificante de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 y un fragmento de 8,5 kb con el resto del plásmido. El fragmento de 1 kb fue aislado y purificado como por encima.

[0324] Los fragmentos de 9,2 kb y 1 kb fueron combinados en una reacción de ligamiento utilizando un equipo de ligamiento de ADN rápido siguiendo las instrucciones del fabricante para producir pSaMe-FX, que es idéntico a pSaMe-F1 excepto en que éste contenía la secuencia codificante madura de beta-glucosidasa de tipo salvaje en lugar de la secuencia codificante madura de la variante.

[0325] Células competentes de *E. coli* SURE® (Stratagene, La Jolla, CA, EE.UU.) fueron transformadas con el producto de ligamiento. La identidad del constructo fue confirmada por secuenciación del ADN del promotor de gen de celobiohidrolasa I de *Trichoderma reesei*, secuencia de señal de endoglucanasa V de *Humicola insolens*, secuencia núcleo de endoglucanasa V de *Humicola insolens*, secuencia señal de endoglucanasa V de *Humicola insolens*, secuencia codificante madura de beta-glucosidasa de *Aspergillus oryzae* y la secuencia terminadora de gen de celobiohidrolasa I de *Trichoderma reesei* a partir de plásmidos purificados de *E. coli* transformado. Un clon que contenía el plásmido recombinante fue designado pSaMe-FX (figura 12). La secuencia de ADN y la secuencia de aminoácidos deducida de la proteína de fusión de beta-glucosidasa de *Aspergillus oryzae* se muestra en SEC ID n°: 75 y 76, respectivamente (véanse figuras 15A, 15B, 15C y 15D).

Ejemplo 12: transformación y expresión de transformantes de *Trichoderma*

[0326] El constructo pSaMe-FX fue linealizado con *Pme* I y transformado en la cepa RutC30 de *Trichoderma reesei* como se describe en el ejemplo 10. Un total de 63 transformantes fueron obtenidos de una única transformación. Se cultivaron transformantes en matraces de agitación en medio de inoculación de celulosa y la actividad de beta-glucosidasa se midió como se describe en el ejemplo 6. Varios transformantes de pSaMe-FX produjeron actividad de beta-glucosidasa. Un transformante designado SaMe-FX16 produjo dos veces la cantidad de actividad de beta-glucosidasa en comparación con SaMeF1-9 de *Trichoderma reesei* (ejemplo 10).

Ejemplo 13: análisis de transformantes de *Trichoderma reesei*

[0327] Una proteína de fusión fue construida como se describe en el ejemplo 9 por fundición del núcleo de la endoglucanasa V de *Humicola insolens* (que contenía su propia secuencia señal nativa) con la secuencia codificante madura de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 enlazada a la secuencia señal de endoglucanasa V de *Humicola insolens*. Este constructo de fusión dio como resultado un aumento dos veces la actividad de beta-glucosidasa segregada en comparación con la secuencia codificante madura de la variante de beta-glucosidasa de *Aspergillus oryzae* BG41 enlazada a la secuencia señal de endoglucanasa V de *Humicola insolens*. Un segundo constructo de fusión se hizo como se describe en el ejemplo 11 que consistió en el núcleo de endoglucanasa V de *Humicola insolens* (que contenía su propia secuencia señal) fusionado con la secuencia codificante de beta-glucosidasa de tipo salvaje de *Aspergillus oryzae* enlazada a la secuencia señal de endoglucanasa V de *Humicola insolens*, y esto llevó a una mejora incluso mayor de la actividad de beta-glucosidasa. La cepa transformada con la fusión de tipo salvaje tuvo dos veces la actividad de beta-glucosidasa segregada con respecto a la cepa transformada con la fusión de la variante de beta-glucosidasa BG41.

Ejemplo 14: clonación de la secuencia codificante de proteína de fusión de beta-glucosidasa en un vector de expresión de *Aspergillus oryzae*

[0328] Dos cebadores oligonucleótidos sintéticos, mostrados por debajo, se diseñaron para amplificación por PCR del marco de lectura abierto en toda su longitud de pSaMeFX que codifica la proteína de fusión de beta-glucosidasa.

Cebador sentido PCR:

5'-GGACTGCGCAGCATGCGTTC-3' (SEC ID n°: 71)

Cebador antisentido PCR:

5'-AGTTAATTAATTACTGGGCCTTAGGCAGCG-3' (SEC ID n°: 72)

Las letras en negrita representan la secuencia codificante. La "G" subrayada en el cebador sentido representa un cambio de base introducido para crear un sitio de restricción *Sph* I. La secuencia restante contiene identidad de secuencia comparada con los sitios de inserción de pSaMeFX. La secuencia subrayada en el cebador antisentido representa un sitio de restricción *Pac* I añadido para facilitar la clonación de este gen en el vector de expresión pAIlO2 (WO 04/099228).

[0329] Cincuenta picomoles de cada uno de los cebadores anteriores fueron usados en una reacción por PCR que contenía 50 ng de ADN de pSaMeFX, 1X de tampón de amplificación *Pfx*, 6 µl de 10 mM de mezcla de dATP, dTTP,

dGTP y dCTP, 2,5 unidades de *Pfx* ADN polimerasa de PLATINUM®, y 1 µl de 50 mM de MgSO₄ en un volumen final de 50 µl. Un MASTERCYCLER® 5333 de EPPENDORF® se usó para amplificar el fragmento programado durante 1 ciclo a 98°C durante 2 minutos; y 35 ciclos cada uno a 96°C durante 30 segundos, 61 °C durante 30 segundos y 68°C durante 3 minutos. Después de los 35 ciclos, la reacción fue incubada a 68°C durante 10 minutos y luego enfriada a 10°C hasta más procesado. Un producto de reacción por PCR de 3,3 kb fue aislado en un gel de GTG-agarosa al 0,8% (Cambrex Bioproducts One Meadowlands Plaza East Rutherford, NJ, EE.UU.) usando tampón de TAE y 0,1 µg de bromuro de etidio por ml. El ADN fue visualizado con la ayuda de un DARK READER™ (Clare Chemical Research, Dolores, CO, EE.UU.) para evitar mutaciones inducidas por UV. Una banda de ADN de 3,3 kb fue cortada con una cuchilla desechable y purificada con una cápsula de giro ULTRAFREE®-DA (Millipore, Billerica, MA, EE.UU.) según las instrucciones del fabricante.

[0330] El producto de PCR purificado de 3,3 kb fue clonado en un vector pCR®4Blunt-TOPO® (Invitrogen, Carlsbad, CA, EE.UU.). Cuatro microlitros del producto de PCR purificado fueron mezclados con 1 µl de una solución de cloruro sódico de 2 M y 1 µl del vector TOPO®. La reacción fue incubada a temperatura ambiente durante 15 minutos y luego 2 µl de la reacción se usó para transformar células de *E. coli* químicamente competentes TOP10 de One Shot® según las instrucciones del fabricante. Tres partes alícuotas de 83 µl cada una de la reacción de transformación se extendieron sobre tres placas de 2X YT de 150 mm suplementadas con 100 µg de ampicilina por ml y se incubaron durante toda la noche a 37°C.

[0331] Ocho colonias recombinantes se usaron para inocular cultivos líquidos que contenían 3 ml de medio LB suplementado con 100 µg de ampicilina por ml. ADN plásmido se preparó a partir de estos cultivos utilizando un BIOROBOT® 9600 (QIAGEN Inc., Valencia, CA, EE.UU.). Se analizaron clones por digestión de enzima de restricción con *Pac* I. El ADN plásmido de cada clon fue digerido con *Pac* I y analizado por electroforesis en gel de agarosa al 1,0% usando tampón TAE. Los ocho clones tuvieron el modelo de digestión de restricción previsto y los clones 5, 6, 7 y 8 fueron seleccionados para ser secuenciados para confirmar que no había ninguna mutación en el inserto clonado. El análisis de secuencias de sus finales 5' y 3' indicó que los 4 clones tenían la secuencia correcta. Los clones 5 y 7 fueron seleccionados para más secuenciación. Ambos clones fueron secuenciados para valores Phred Q mayores que 40 para asegurar que no había ningún error inducido por PCR. Los clones 5 y 7 mostraron que tenían la secuencia prevista y clon 5 fue seleccionado para reclonación en pAILo2.

[0332] El ADN plásmido del clon 5 fue linealizado por digestión con *Sph* I. El clon linealizado se hizo luego con los extremos redondeados añadiendo 1,2 µl de un 10 mM de mezcla de dATP, dTTP, dGTP y dCTP y 6 unidades de T4 ADN polimerasa (New England Biolabs, Inc., Ipswich, MA, EE.UU.). La mezcla fue incubada a 12°C durante 20 minutos y luego la reacción fue detenida añadiendo 1 µl de 0,5 M de EDTA y calentamiento a 75°C durante 20 minutos para inactivar la enzima. Un fragmento de 3,3 kb que codifica la proteína de fusión de beta-glucosidasa fue purificado por electroforesis en gel y ultrafiltración como se ha descrito anteriormente.

[0333] El vector pAILo2 fue linealizado por digestión con *Nco* I. El vector linealizado se hizo luego con los extremos redondeados añadiendo 0,5 µl de una mezcla de 10 mM de dATP, dTTP, dGTP y dCTP y una unidad de ADN polimerasa I. La mezcla se incubó a 25°C durante 15 minutos y luego la reacción fue detenida añadiendo 1 µl de 0,5M de EDTA y calentamiento a 75°C durante 15 minutos para inactivar las enzimas. Luego el vector fue digerido con *Pac* I. El vector con los extremos redondeados fue purificado por electroforesis en gel y ultrafiltración como se ha descrito anteriormente.

[0334] La clonación del fragmento de 3,3 kb que codifica la proteína de fusión de beta-glucosidasa en el vector pAILo2 linealizado y purificado se realizó con un equipo de ligamiento rápido. Una muestra de 1 µl de la reacción se usó para transformar células Gold de *E. coli* XL10 SOLOPACK® (Stratagene, La Jolla, CA, EE.UU.) según las instrucciones del fabricante. Después del periodo de recuperación, dos partes alícuotas de 100 µl de la reacción de transformación se colocaron en placas sobre dos placas 2X YT de 150 mm suplementadas con 100 µg de ampicilina por ml e incubadas durante toda la noche a 37°C. Un conjunto de ocho clones recombinantes putativos fue seleccionado al azar de las placas de selección y se obtuvo ADN plásmido a partir de cada uno usando un BIOROBOT® 9600. Los clones 1-4 fueron seleccionados para secuenciación con los cebadores específicos de pAILo2 para confirmar que el vector de unión/inserto tenía la secuencia correcta. El clon 3 tenía una unión de vector/inserto perfecto y fue denominado pAILo47 (figura 13).

[0335] Para crear una cepa de expresión sin marcador, una digestión de endonucleasa de restricción se realizó para separar el gen *blaA* que confiere resistencia a la ampicilina antibiótica del resto del constructo de expresión. Treinta microgramos de pAILo47 fueron digeridos con *Pme* I. El ADN digerido fue luego purificado por electroforesis en gel de agarosa como se ha descrito anteriormente. Una banda de ADN de 6,4 kb que contenía el constructo de expresión pero a la que le faltaba el gen *blaA* fue cortada con una hoja de afeitar y purificada con un equipo de extracción de gel QIAQUICK®.

Ejemplo 15: expresión de la proteína de fusión de beta-glucosidasa en el *Aspergillus oryzae* JaL355

65

[0336] Se prepararon protoplastos de *Aspergillus oryzae* JaL355 (WO 00/240694) según el método de Christensen *et al.*, 1988, Bio/Technology 6: 1419-1422. Diez microlitros del constructo de expresión purificado del ejemplo 14 se usaron para transformar protoplastos de JaL355 de *Aspergillus oryzae*. La transformación de *Aspergillus oryzae* JaL355 liberó aproximadamente 90 transformantes. Cincuenta transformantes fueron aislados en placas PDA individuales e incubados durante cinco días a 34°C.

[0337] Cuarenta y ocho placas de esporas confluyentes fueron lavadas con 3 ml de 0,01% TWEEN® 80 y la suspensión de esporas fue usada para inocular 25 ml de medio MDU2BP en matraces de agitación de vidrio de 125 ml. Los cultivos de transformantes fueron incubados a 34°C con agitación constante a 200 r.p.m. Después de 5 días, 1 ml de partes alícuotas de cada cultivo fue centrifugado a 12.000 x g y sus sobrenadantes recogidos. Cinco µl de cada sobrenadante fue mezclado con un volumen igual de 2X tampón de carga (10% de beta-mercaptoetanol) y cargado sobre un 1,5 mm 8 %-16 % de gel de SDS-PAGE tris-glicina y manchado con mancha de proteína azul Coomassie G250 de BIO-SAFE® (Bio-Rad, Hercules, CA, EE.UU.). Los perfiles de SDS-PAGE de los caldos de cultivo mostraron que 33 de los 48 transformantes fueron capaces de expresar una proteína nueva con un peso molecular aparente muy cerca del previsto 118 kDa. El transformante 21 produjo el mejor rendimiento y fue seleccionado para otros estudios.

Ejemplo 16: aislamiento de espóra única del transformante 21 de *Aspergillus oryzae* JaL355

[0338] Las esporas del transformante 21 de *Aspergillus oryzae* JaL355 se extendieron sobre una placa PDA y se incubaron durante cinco días a 34°C. Un área pequeña de la placa de esporas confluyente se lavó con 0,5 ml de 0,01% de TWEEN® 80 para resuspender las esporas. Una parte alícuota de 100 µl de la suspensión de esporas fue diluida a un volumen final de 5 ml con 0,01% de TWEEN® 80. Con la ayuda de un hemocitómetro, la concentración de esporas fue determinada y diluida a una concentración final de 0,1 esporas por microlitros. Una parte alícuota de 200 µl de la dilución de esporas se extendió sobre 150 mm de placas de medio mínimo y se incubó durante 2-3 días a 34°C. Colonias emergentes fueron cortadas de las placas y transferidas a placas PDA e incubadas durante 3 días a 34°C. Luego, las esporas se extendieron a través de las placas y se incubaron otra vez durante 5 días a 34°C.

[0339] Las placas de esporas confluyentes fueron lavadas con 3 ml de 0,01% de TWEEN® 80 y la suspensión de esporas se usó para inocular 25 ml de medio MDU2BP en matraces de agitación de vidrio de 125 ml. Cultivos monospóricos fueron incubados a 34°C con agitación constante a 200 r.p.m. Después de 5 días, una parte alícuota de 1 ml de cada cultivo fue centrifugada a 12.000 x g y sus sobrenadantes recogidos. Cinco de cada sobrenadante se mezcló con un volumen igual de 2X tampón de carga (10% beta-mercaptoetanol) y se cargó sobre un gel de SDS-PAGE tris-glicina 1,5 mm 8%-16% y se manchó con mancha de proteína G250 azul Coomassie de BIO-SAFE®. Los perfiles de SDS-PAGE de los caldos de cultivo mostraron que los ocho transformantes eran capaces de expresar la proteína de fusión de beta-glucosidasa a niveles altísimos y uno de los cultivos denominados *Aspergillus oryzae* JaL355A1Lo47 produjo el mejor rendimiento.

[0340] La invención descrita y reivindicada aquí no está limitada en su alcance por los aspectos específicos aquí descritos, dado que estos aspectos son ilustraciones de diferentes aspectos de la invención. Cualquier aspecto equivalente está dentro del campo de esta invención. De hecho, serán aparentes para los expertos en la técnica diferentes modificaciones de la invención además de las mostrada y descrita aquí a partir de la descripción precedente. Tales modificaciones están también destinadas a incluirse dentro del campo de las reivindicaciones anexas. En caso de conflicto, la presente divulgación, incluyendo definiciones, prevalecerá.

[0341] Se citan aquí varias referencias, las descripciones se incorporan por referencia en su totalidad.
Listado de secuencias

[0342]

<110> Novozymes, Inc.

<120> Métodos de aumento de secreción de polipéptidos que tienen actividad biológica

<130> 10916.204-WO

<150> US 60/832,511

<151> 2006-07-21

<160> 76

<170> Versión de patentIn 3.4

<210> 1

<211> 923

<212> ADN

<213> *Humicola insolens*

ES 2 538 360 T3

<400> 1

atgcggtcct cccccctcct ccgctccgcc gttgtggccg ccctgccggt gttggccctt 60
 gccgctgatg gcaggtccac ccgctactgg gactgctgca agccttcgtg cggctggggcc 120
 aagaaggctc ccgtgaacca gcctgtcttt tcctgcaacg ccaacttcca gcgtatcacg 180
 gacttcgacg ccaagtccgg ctgcgagccg ggcgggtgctg cctactcgtg cgccgaccag 240
 accccatggg ctgtgaacga cgacttcgcg ctcggttttg ctgccacctc tattgccggc 300
 agcaatgagg cgggctggtg ctgcgccctgc tacgagctca ccttcacatc cggtcctggt 360
 gctggcaaga agatggtcgt ccagtccacc agcactggcg gtgatcttgg cagcaaccac 420
 ttogatctca acatccccgg cggcggcgtc ggcattctcg acggatgcac tccccagttc 480
 ggcggctctgc ccggccagcg ctacggcggc atctcgtccc gcaacgagtg cgatcggttc 540
 cccgacgccc tcaagcccgg ctgctactgg cgcttcgact ggttcaagaa cgccgacaat 600
 ccgagcttca gcttccgtca ggtccagtgc ccagccgagc tcgtcgctcg caccggatgc 660
 cgccgcaacg acgacggcaa cttccctgcc gtccagatcc cctccagcag caccagctct 720
 ccggtcaacc agcctaccag caccagcacc acgtccacct ccaccacctc gagcccgcca 780
 gtccagccta cgactcccag cggctgcact gctgagaggt gggctcagtg cggcgggcaat 840
 ggctggagcg gctgcaccac ctgcgctcgt ggcagcactt gcacgaagat taatgactgg 900
 taccatcagt gcctgtagaa ttc 923

5

<210> 2

<211> 305

<212> PRT

10

<213> Humicola insolens

<400> 2

15

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro

ES 2 538 360 T3

Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu
260 265 270

Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys
275 280 285

Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys
290 295 300

Leu
305

<210> 3

<211> 1188

5 <212> ADN

<213> Myceliophthora thermophila

<400> 3

ES 2 538 360 T3

cgacttgaaa cgccccaaat gaagtcctcc atcctcgcca gcgtcttcgc cacgggcgcc 60
 gtggctcaaa gtggtcctgt gcagcaatgt ggtggcatcg gatggcaagg atcgaccgac 120
 tgtgtgtcgg gctaccactg cgtctaccag aacgattggt acagccagtg cgtgcctggc 180
 gcggcgtcga caacgctgca gacatcgacc acgtccaggc ccaccgccac cagcaccgcc 240
 cctccgtcgt ccaccacctc gcctagcaag ggcaagctga agtggctcgg cagcaacgag 300
 tcgggcgccg agttcgggga gggcaattac cccggcctct ggggcaagca cttcatcttc 360
 ccgtcgactt cggcgattca gacgctcatc aatgatggat acaacatctt ccggatcgac 420
 ttctcgatgg agcgtctggt gcccaaccag ttgacgtcgt ccttcgacca gggttacctc 480
 cgcaacctga ccgaggtggt caacttcgtg acgaacgcgg gcaagtacgc cgtcctggac 540
 ccgcacaact acggccggta ctacggcaac atcatcacgg acacgaacgc gttccggacc 600
 ttctggacca acctggccaa gcagttcgcc tccaactcgc tcgtcatctt cgacaccaac 660
 aacgagtaca acacgatgga ccagaccctg gtgctcaacc tcaaccaggc cgccatcgac 720
 ggcacccggg ccgccggcgc gacctcgcag tacatcttcg tcgagggcaa cgcgtggagc 780
 ggggcctgga gctggaacac gaccaacacc aacatggccg ccctgacgga cccgcagaac 840
 aagatcgtgt acgagatgca ccagtacctc gactcggaca gctcgggcac ccacgccgag 900
 tgcgtcagca gcaccatcgg cgcacagcgc gtcgtcggag ccaccagtg gctccgcgcc 960
 aacggcaagc tcggcgtcct cggcgagttc gccggcggcg ccaacgccgt ctgccagcag 1020
 gccgtcaccg gcctcctcga ccacctccag gacaacagcg acgtctggct gggtgccctc 1080
 tgggtgggccg ccggtccctg gtggggcgac tacatgtact cgttcgagcc tccttcgggc 1140
 accggctatg tcaactacaa ctcgatcttg aagaagtact tgccgtaa 1188

<210> 4

<211> 389

<212> PRT

<213> Myceliophthora thermophila

<400> 4

5

Met Lys Ser Ser Ile Leu Ala Ser Val Phe Ala Thr Gly Ala Val Ala
1 5 10 15

Gln Ser Gly Pro Trp Gln Gln Cys Gly Gly Ile Gly Trp Gln Gly Ser
20 25 30

Thr Asp Cys Val Ser Gly Tyr His Cys Val Tyr Gln Asn Asp Trp Tyr
35 40 45

Ser Gln Cys Val Pro Gly Ala Ala Ser Thr Thr Leu Gln Thr Ser Thr
50 55 60

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

Ser Pro Ser Lys Gly Lys Leu Lys Trp Leu Gly Ser Asn Glu Ser Gly
85 90 95

Ala Glu Phe Gly Glu Gly Asn Tyr Pro Gly Leu Trp Gly Lys His Phe
100 105 110

Ile Phe Pro Ser Thr Ser Ala Ile Gln Thr Leu Ile Asn Asp Gly Tyr
115 120 125

Asn Ile Phe Arg Ile Asp Phe Ser Met Glu Arg Leu Val Pro Asn Gln
130 135 140

Leu Thr Ser Ser Phe Asp Gln Gly Tyr Leu Arg Asn Leu Thr Glu Val
145 150 155 160

Val Asn Phe Val Thr Asn Ala Gly Lys Tyr Ala Val Leu Asp Pro His
165 170 175

Asn Tyr Gly Arg Tyr Tyr Gly Asn Ile Ile Thr Asp Thr Asn Ala Phe
180 185 190

Arg Thr Phe Trp Thr Asn Leu Ala Lys Gln Phe Ala Ser Asn Ser Leu
195 200 205

Val Ile Phe Asp Thr Asn Asn Glu Tyr Asn Thr Met Asp Gln Thr Leu
210 215 220

Val Leu Asn Leu Asn Gln Ala Ala Ile Asp Gly Ile Arg Ala Ala Gly
 225 230 235 240

Ala Thr Ser Gln Tyr Ile Phe Val Glu Gly Asn Ala Trp Ser Gly Ala
 245 250 255

Trp Ser Trp Asn Thr Thr Asn Thr Asn Met Ala Ala Leu Thr Asp Pro
 260 265 270

Gln Asn Lys Ile Val Tyr Glu Met His Gln Tyr Leu Asp Ser Asp Ser
 275 280 285

Ser Gly Thr His Ala Glu Cys Val Ser Ser Thr Ile Gly Ala Gln Arg
 290 295 300

Val Val Gly Ala Thr Gln Trp Leu Arg Ala Asn Gly Lys Leu Gly Val
 305 310 315 320

Leu Gly Glu Phe Ala Gly Gly Ala Asn Ala Val Cys Gln Gln Ala Val
 325 330 335

Thr Gly Leu Leu Asp His Leu Gln Asp Asn Ser Asp Val Trp Leu Gly
 340 345 350

Ala Leu Trp Trp Ala Ala Gly Pro Trp Trp Gly Asp Tyr Met Tyr Ser
 355 360 365

Phe Glu Pro Pro Ser Gly Thr Gly Tyr Val Asn Tyr Asn Ser Ile Leu
 370 375 380

Lys Lys Tyr Leu Pro
 385

<210> 5

<211> 1232

<212> ADN

<213> Myceliophthora thermophila

<400> 5

5

ES 2 538 360 T3

ggatccactt agtaacggcc gccagtggtc tggaaagcat gaagtctctc ttccctgtcac 60
 ttgtagcgac cgtcgcgctc agctcgccag tattctctgt cgcagctctgg gggcaatgcg 120
 gcggcattgg cttcagcggg agcaccgtct gtgatgcagg cgccggctgt gtgaagctca 180
 acgactatta ctctcaatgc caaccggcg ctcccactgc tacatccgcg gcgccaagta 240
 gcaacgcacc gtccggcact tcgacggcct cggccccctc ctccagcctt tgctctggca 300
 gccgcacgcc gttccagttc ttcggtgtca acgaatccgg cgcggagttc ggcaacctga 360
 acatcccggg tgttctgggc accgactaca cctggccgtc gccatccagc attgacttct 420
 tcatgggcaa gggaatgaat accttccgta ttccgttctc catggagcgt cttgtcccc 480
 ctgccactgg catcacagga cctctcgacc agacgtactt gggcggcctg cagacgattg 540
 tcaactacat caccggcaaa ggcggctttg ctctcattga cccgcacaac tttatgatct 600
 acaatggcca gacgatctcc agtaccagcg acttccagaa gttctggcag aacctcgcag 660
 gagtgtttaa atcgaacagt cacgtcatct tcgatgttat gaacgagcct cacgatattc 720
 ccgcccagac cgtgttccaa ctgaaccaag ccgctgtcaa tggcatccgt gcgagcggtg 780
 cgacgtcgca gctcattctg gtcgagggca caagctggac tggagcctgg acctggacga 840
 cctctggcaa cagcgatgca ttcggtgcca ttaaggatcc caacaacaac gtcgcatcc 900
 agatgcatca gtacctggat agcgatggct ctggcacttc gcagacctgc gtgtctccca 960
 ccatcgggtc cgagcggttg caggctgcga ctcaatggtt gaagcagaac aacctcaagg 1020
 gcttcctggg cgagatcggc gccggctcta actccgcttg catcagcgtc gtgcagggtg 1080
 cgttgtgttc gatgcagcaa tctggtgtgt ggctcggcgc tctctggtgg gctgcgggcc 1140
 cgtggtgggg cgactactac cagtccatcg agccgcctc tggcccggcg gtgtccgcga 1200
 tcctcccga ggcctgctg ccgttcgcgt aa 1232

5 <210> 6
 <211> 397
 <212> PRT
 <213> Myceliophthora thermophila

10 <400> 6

ES 2 538 360 T3

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

ES 2 538 360 T3

Leu Gly Thr Asp Tyr Thr Trp Pro Ser Pro Ser Ser Ile Asp Phe Phe
 115 120 125
 Met Gly Lys Gly Met Asn Thr Phe Arg Ile Pro Phe Leu Met Glu Arg
 130 135 140
 Leu Val Pro Pro Ala Thr Gly Ile Thr Gly Pro Leu Asp Gln Thr Tyr
 145 150 155 160
 Leu Gly Gly Leu Gln Thr Ile Val Asn Tyr Ile Thr Gly Lys Gly Gly
 165 170 175
 Phe Ala Leu Ile Asp Pro His Asn Phe Met Ile Tyr Asn Gly Gln Thr
 180 185 190
 Ile Ser Ser Thr Ser Asp Phe Gln Lys Phe Trp Gln Asn Leu Ala Gly
 195 200 205
 Val Phe Lys Ser Asn Ser His Val Ile Phe Asp Val Met Asn Glu Pro
 210 215 220
 His Asp Ile Pro Ala Gln Thr Val Phe Gln Leu Asn Gln Ala Ala Val
 225 230 235 240
 Asn Gly Ile Arg Ala Ser Gly Ala Thr Ser Gln Leu Ile Leu Val Glu
 245 250 255
 Gly Thr Ser Trp Thr Gly Ala Trp Thr Trp Thr Thr Ser Gly Asn Ser
 260 265 270
 Asp Ala Phe Gly Ala Ile Lys Asp Pro Asn Asn Asn Val Ala Ile Gln
 275 280 285
 Met His Gln Tyr Leu Asp Ser Asp Gly Ser Gly Thr Ser Gln Thr Cys
 290 295 300
 Val Ser Pro Thr Ile Gly Ala Glu Arg Leu Gln Ala Ala Thr Gln Trp
 305 310 315 320
 Leu Lys Gln Asn Asn Leu Lys Gly Phe Leu Gly Glu Ile Gly Ala Gly
 325 330 335
 Ser Asn Ser Ala Cys Ile Ser Ala Val Gln Gly Ala Leu Cys Ser Met
 340 345 350
 Gln Gln Ser Gly Val Trp Leu Gly Ala Leu Trp Trp Ala Ala Gly Pro

ES 2 538 360 T3

355

360

365

Trp Trp Gly Asp Tyr Tyr Gln Ser Ile Glu Pro Pro Ser Gly Pro Ala
370 375 380

Val Ser Ala Ile Leu Pro Gln Ala Leu Leu Pro Phe Ala
385 390 395

<210> 7

<211> 1303

<212> ADN

<213> Basidiomiceto CBS 495.95

<400> 7

5

ES 2 538 360 T3

ggaaagcgtc agtatggtga aatttgcgct tgtggcaact gtcggcgcaa tcttgagcgc 60
 ttctgcggcc aatgcggctt ctatctacca gcaatgtgga ggcattggat ggtctgggtc 120
 cactgtttgc gacgccggtc tcgcttgcgt tatcctcaat gcgtactact ttcagtgtt 180
 gacgcccgcc gcgggccaga caacgacggg ctcgggcgca ccggcgtcaa catcaacctc 240
 tcaactcaacg gtcactacgg ggagctcaca ctcaacaacc gggacgacgg cgacgaaaac 300
 aactaccact ccgtcgacca ccacgacct acccgccatc tctgtgtctg gtcgcgtctg 360
 ctctggctcc aggacgaagt tcaagttctt cgggtggaat gaaagcggcg ccgaattcgg 420
 gaacactgct tggccagggc agctcgggaa agactataca tggccttcgc ctagcagcgt 480
 ggactacttc atgggggctg gattcaatac attccgtatc accttcttga tggagcgtat 540
 gagccctccg gctaccggac tcaactggccc attcaaccag acgtacctgt cgggcctcac 600
 caccattgtc gactacatca cgaacaaagg aggatacgct cttattgacc cccacaactt 660
 catgcgttac aacaacggca taatcagcag cacatctgac ttcgcgactt ggtggagcaa 720
 tttggccact gtattcaaat ccacgaagaa cgccatcttc gacatccaga acgagccgta 780
 cggaatcgat ggcgagaccg tatacgaact gaatcaagct gccatcaatt cgatccgcgc 840
 cgctggcgct acgtcacagt tgattctggt tgaaggaacg tcatacactg gagcttggac 900
 gtgggtctcg tccggaaacg gagctgcttt cgcggccggt acggatcctt acaacaacac 960
 ggcaattgaa atgcaccaat acctcgacag cgacggttct gggacaaacg aagactgtgt 1020
 ctctccacc attgggtcgc aacgtctcca agctgccact gcgtggctgc acaaacagg 1080
 actcaagga ttctcggag agacgggtgc tgggtcgaat tcccagtgca tcgacgccgt 1140
 gttcgatgaa ctttgctata tgcaacagca aggcggctcc tggatcggtg cactctggtg 1200
 ggctgcgggt ccctggtggg gcacgtacat ttactcgatt gaacctccga gcggtgccgc 1260
 tatcccagaa gtccttcttc agggtctcgc tccattcctc tag 1303

<210> 8

<211> 429

<212> PRT

<213> Basidiomiceto CBS 495.95

<400> 8

5

ES 2 538 360 T3

Met Val Lys Phe Ala Leu Val Ala Thr Val Gly Ala Ile Leu Ser Ala
1 5 10 15

Ser Ala Ala Asn Ala Ala Ser Ile Tyr Gln Gln Cys Gly Gly Ile Gly
20 25 30

Trp Ser Gly Ser Thr Val Cys Asp Ala Gly Leu Ala Cys Val Ile Leu
35 40 45

Asn Ala Tyr Tyr Phe Gln Cys Leu Thr Pro Ala Ala Gly Gln Thr Thr
50 55 60

Thr Gly Ser Gly Ala Pro Ala Ser Thr Ser Thr Ser His Ser Thr Val
65 70 75 80

Thr Thr Gly Ser Ser His Ser Thr Thr Gly Thr Thr Ala Thr Lys Thr
85 90 95

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

Gly Arg Val Cys Ser Gly Ser Arg Thr Lys Phe Lys Phe Phe Gly Val
115 120 125

Asn Glu Ser Gly Ala Glu Phe Gly Asn Thr Ala Trp Pro Gly Gln Leu
130 135 140

Gly Lys Asp Tyr Thr Trp Pro Ser Pro Ser Ser Val Asp Tyr Phe Met
145 150 155 160

Gly Ala Gly Phe Asn Thr Phe Arg Ile Thr Phe Leu Met Glu Arg Met
165 170 175

Ser Pro Pro Ala Thr Gly Leu Thr Gly Pro Phe Asn Gln Thr Tyr Leu
180 185 190

Ser Gly Leu Thr Thr Ile Val Asp Tyr Ile Thr Asn Lys Gly Gly Tyr
195 200 205

Ala Leu Ile Asp Pro His Asn Phe Met Arg Tyr Asn Asn Gly Ile Ile
210 215 220

Ser Ser Thr Ser Asp Phe Ala Thr Trp Trp Ser Asn Leu Ala Thr Val

ES 2 538 360 T3

agccccccgt tcaggcacac ttggcatcag atcagcttag cagcgcctgc acagcatgaa 60
 gctctcgcag tgggccgcgc tggcggcact caccgcgacg gcgctcgccg cccctcgcgc 120
 cagcagcccg caggcgcgga ggcaggett c agccggctgc tcgtctgagg tcacgctcga 180
 cgccagcacc aacgtttgga agaagtacac gctgcacccc aacagctact accgcaagga 240
 ggttgaggcc gcggtggcgc agatctcgga cccggacctc gccgcccaagg ccaagaaggt 300
 ggccgacgtc ggcaccttcc tgtggctcga ctcgatcgag aacatcggca agctggagcc 360
 ggcgatccag gacgtgccct gcgagaacat cctgggcctg gtcactctag acctgccggg 420
 ccgcgactgc gcggccaagg cgtccaacgg cgagctcaag gtcggcgaga tcgaccgcta 480
 caagaccgag tacatcgaca gtgagtgtg cccccgggt tcgagaagag cgtgggggaa 540
 agggaaaggg ttgactgact gacacggcgc actgcagaga tcgtgtcgat cctcaaggca 600
 caccccaaca cggcgttcgc gctggtcatc gagccggact cgctgcccaa cctggtgacc 660
 aacagcaact tggacacgtg ctcgagcagc gcgtcgggct accgcgaagg cgtggcttac 720
 gccctcaaga acctcaacct gcccaacgtg atcatgtacc tcgacgccgg ccacggcggc 780
 tggctcggct gggacgcaa cctgcagccc ggcgcgcagg agctagccaa ggcgtacaag 840
 aacgccggct cgccaagca gctccgcggc ttctcgacca acgtggccgg ctggaactcc 900
 tggtgagctt ttttccattc catttcttct tcctcttctc tcttctctcc cactctgcag 960
 cccccctcc cccaagcacc cactggcgtt ccggcttgct gactcggcct ccctttcccc 1020
 gggcaccagg gatcaatcgc ccggcgaatt ctcccaggcg tccgacgcca agtacaacaa 1080
 gtgccagaac gagaagatct acgtcagcac cttcggctcc gcgctccagt cggccggcat 1140
 gcccaaccac gccatcgtcg acacgggccg caacggcgtc accggcctgc gcaaggagtg 1200
 gggtgactgg tgcaacgtca acggtgcagg ttcgttgtct tctttttctc ctcttttgtt 1260
 tgcacgtcgt ggtccttttc aagcagccgt gtttggttg gggagatgga ctccggctga 1320
 tgttctgctt cctctctagg cttcggcgtg cgcccgaaga gcaacacggg cctcagactg 1380
 gccgacgcgt tcgtgtgggt caagcccggc ggcgagtcgg acggcaccag cgacagctcg 1440
 tcgccgcgct acgacagctt ctgcggcaag gacgacgcct tcaagccctc gcccgaggcc 1500
 ggcacctgga acgaggccta cttcgagatg ctgctcaaga acgccgtgcc gtcgttctaa 1560
 gacggtccag catcatccgg 1580

5 <210> 10
 <211> 396
 <212> PRT
 <213> Thielavia terrestris
 <400> 10

ES 2 538 360 T3

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

Leu Ala Ala Pro Ser Pro Thr Thr Pro Gln Ala Pro Arg Gln Ala Ser
 20 25 30

ES 2 538 360 T3

Ala Gly Cys Ser Ser Ala Val Thr Leu Asp Ala Ser Thr Asn Val Trp
35 40 45

Lys Lys Tyr Thr Leu His Pro Asn Ser Tyr Tyr Arg Lys Glu Val Glu
50 55 60

Ala Ala Val Ala Gln Ile Ser Asp Pro Asp Leu Ala Ala Lys Ala Lys
65 70 75 80

Lys Val Ala Asp Val Gly Thr Phe Leu Trp Leu Asp Ser Ile Glu Asn
85 90 95

Ile Gly Lys Leu Glu Pro Ala Ile Gln Asp Val Pro Cys Glu Asn Ile
100 105 110

Leu Gly Leu Val Ile Tyr Asp Leu Pro Gly Arg Asp Cys Ala Ala Lys
115 120 125

Ala Ser Asn Gly Glu Leu Lys Val Gly Glu Ile Asp Arg Tyr Lys Thr
130 135 140

Glu Tyr Ile Asp Lys Ile Val Ser Ile Leu Lys Ala His Pro Asn Thr
145 150 155 160

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

Asn Ser Asn Leu Asp Thr Cys Ser Ser Ser Ala Ser Gly Tyr Arg Glu
180 185 190

Gly Val Ala Tyr Ala Leu Lys Asn Leu Asn Leu Pro Asn Val Ile Met
195 200 205

Tyr Leu Asp Ala Gly His Gly Gly Trp Leu Gly Trp Asp Ala Asn Leu
210 215 220

Gln Pro Gly Ala Gln Glu Leu Ala Lys Ala Tyr Lys Asn Ala Gly Ser
225 230 235 240

Pro Lys Gln Leu Arg Gly Phe Ser Thr Asn Val Ala Gly Trp Asn Ser
245 250 255

Trp Asp Gln Ser Pro Gly Glu Phe Ser Gln Ala Ser Asp Ala Lys Tyr
260 265 270

Asn Lys Cys Gln Asn Glu Lys Ile Tyr Val Ser Thr Phe Gly Ser Ala
275 280 285

ES 2 538 360 T3

Leu Gln Ser Ala Gly Met Pro Asn His Ala Ile Val Asp Thr Gly Arg
 290 295 300

Asn Gly Val Thr Gly Leu Arg Lys Glu Trp Gly Asp Trp Cys Asn Val
 305 310 315 320

Asn Gly Ala Gly Phe Gly Val Arg Pro Thr Ser Asn Thr Gly Leu Glu
 325 330 335

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

Thr Ser Asp Ser Ser Ser Pro Arg Tyr Asp Ser Phe Cys Gly Lys Asp
 355 360 365

Asp Ala Phe Lys Pro Ser Pro Glu Ala Gly Thr Trp Asn Glu Ala Tyr
 370 375 380

Phe Glu Met Leu Leu Lys Asn Ala Val Pro Ser Phe
 385 390 395

- <210> 11
- <211> 1203
- <212> ADN
- <213> Thielavia terrestris

5

<400> 11

ES 2 538 360 T3

atgaagtacc tcaacctcct cgcagctctc ctgcgcgtcg ctctctctc cctcgctgca 60
 cccagcatcg aggccagaca gtcgaacgtc aaccataca tcggcaagag cccgctcgtt 120
 attaggtcgt acgcccmeta gcttgaggag accgtcagga ccttccagca acgtggcgac 180
 cagctcaacg ctgcgaggac acggacggtg cagaacggtg cgactttcgc ctggatctcg 240
 gataccaatg gtattggagc cattcgacct ctcatccaag atgctctcgc ccagcaggct 300
 cgcactggac agaaggtcat cgtccaaatc gtcgtctaca acctcccaga tcgcgactgc 360
 tctgccaacg cctcgactgg agagttcacc gtaggaaacg acggtctcaa ccgatacaag 420
 aactttgtca acaccatcgc ccgcgagctc tcgactgctg acgctgacaa gctccacttt 480
 gccctcctcc tcgaaccgga cgcacttgcc aacctcgtca ccaacgcgaa tgccccagg 540
 tgccgaatcg ccgctcccgc ttacaaggag ggtatcgctt acaccctcgc caccttgtcc 600
 aagcccaacg tcgacgtcta catcgacgcc gccaacggtg gctggctcgg ctggaacgac 660
 aacctccgcc ccttcgccga actcttcaag gaagtctacg acctcgcccg ccgcatcaac 720
 cccaacgcca aggtccgcgg cgtccccgtc aacgtctcca actacaacca gtaccgcgct 780
 gaagtccgcg agcccttcac cgagtggaag gacgcctggg acgagagccg ctacgtcaac 840
 gtctcacc cgcacctcaa cgcgctcggc ttctccgcgc acttcatcgt tgaccaggga 900
 cgcggtggca agggcggtat caggacggag tggggccagt ggtgcaacgt taggaacgct 960
 gggttcggta tcaggcctac tgcggatcag ggcgtgctcc agaaccgaa tgtggatgcg 1020
 attgtgtggg ttaagccggg tggagagtcg gatggcacga gtgatttgaa ctogaacagg 1080
 tatgatccta cgtgcaggag tccggtggcg catgttcccg ctctgagggc tggccagtgg 1140
 ttcaacgagt atgttgtaa cctcgttttg aacgctaacc cccctcttga gcctacctgg 1200
 taa 1203

5 <210> 12
 <211> 400
 <212> PRT
 <213> Thielavia terrestris

10 <400> 12

ES 2 538 360 T3

Met Lys Tyr Leu Asn Leu Leu Ala Ala Leu Leu Ala Val Ala Pro Leu
1 5 10 15

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

Tyr Ile Gly Lys Ser Pro Leu Val Ile Arg Ser Tyr Ala Gln Lys Leu
35 40 45

Glu Glu Thr Val Arg Thr Phe Gln Gln Arg Gly Asp Gln Leu Asn Ala
50 55 60

Ala Arg Thr Arg Thr Val Gln Asn Val Ala Thr Phe Ala Trp Ile Ser
65 70 75 80

Asp Thr Asn Gly Ile Gly Ala Ile Arg Pro Leu Ile Gln Asp Ala Leu
85 90 95

Ala Gln Gln Ala Arg Thr Gly Gln Lys Val Ile Val Gln Ile Val Val
100 105 110

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

Phe Thr Val Gly Asn Asp Gly Leu Asn Arg Tyr Lys Asn Phe Val Asn
130 135 140

Thr Ile Ala Arg Glu Leu Ser Thr Ala Asp Ala Asp Lys Leu His Phe
145 150 155 160

ES 2 538 360 T3

Ala Leu Leu Leu Glu Pro Asp Ala Leu Ala Asn Leu Val Thr Asn Ala
 165 170 175

Asn Ala Pro Arg Cys Arg Ile Ala Ala Pro Ala Tyr Lys Glu Gly Ile
 180 185 190

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

Asp Ala Ala Asn Gly Gly Trp Leu Gly Trp Asn Asp Asn Leu Arg Pro
 210 215 220

Phe Ala Glu Leu Phe Lys Glu Val Tyr Asp Leu Ala Arg Arg Ile Asn
 225 230 235 240

Pro Asn Ala Lys Val Arg Gly Val Pro Val Asn Val Ser Asn Tyr Asn
 245 250 255

Gln Tyr Arg Ala Glu Val Arg Glu Pro Phe Thr Glu Trp Lys Asp Ala
 260 265 270

Trp Asp Glu Ser Arg Tyr Val Asn Val Leu Thr Pro His Leu Asn Ala
 275 280 285

Val Gly Phe Ser Ala His Phe Ile Val Asp Gln Gly Arg Gly Gly Lys
 290 295 300

Gly Gly Ile Arg Thr Glu Trp Gly Gln Trp Cys Asn Val Arg Asn Ala
 305 310 315 320

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

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

Thr Ser Asp Leu Asn Ser Asn Arg Tyr Asp Pro Thr Cys Arg Ser Pro
 355 360 365

Val Ala His Val Pro Ala Pro Glu Ala Gly Gln Trp Phe Asn Glu Tyr
 370 375 380

Val Val Asn Leu Val Leu Asn Ala Asn Pro Pro Leu Glu Pro Thr Trp
 385 390 395 400

ES 2 538 360 T3

<210> 13
 <211> 1501
 <212> ADN
 <213> Thielavia terrestris

5

<400> 13

```

gccgttgtca agatgggcca gaagacgctg cacggattcg ccgccacggc tttggccggt      60
ctcccccttg tgaaggctca gcagcccggc aacttcacgc cggagggtgca cccgcaactg     120
ccaacgtgga agtgcacgac cgccggcggc tgcgttcagc aggacacttc ggtggtgctc     180
gactggaact accgttggat ccacaatgcc gacggcaccg cctcgtgcac gacgtccagc     240
ggggtcgacc acacgctgtg tccagatgag gcgacctgcg cgaagaactg cttcgtggaa     300
ggcgtcaact acacgagcag cgggtgtcacc acatccggca gttcgtctgac gatgaggcag     360
tatttcaagg ggagcaacgg gcagaccaac agcgtttcgc ctcgtctcta cctgctcggc     420
tcggatggaa actacgtaat gctcaagctg ctcggccagg agctgagctt cgatgtcgat     480
ctctccacgc tcccctgcgg cgagaacggc gcgctgtacc tgtccgagat ggacgcgacc     540
ggtggcagga accagtataa caccggcggg gccaaactacg gctcgggcta ctgtgacgcc     600
cagtgtcccg tgcagacgtg gatgaacggc acgctgaaca ccaacgggca gggctactgc     660
tgcaacgaga tggacatcct cgaggccaac tcccgcgcca acgcgatgac acctcacccc     720
tgcgccaacg gcagctgcga caagagcggg tgcggactca acccctacgc cgagggctac     780
aagagctact acggaccggg cctcacgggt gacacgtoga agcccttcac catcattacc     840
cgcttcatca ccgacgacgg cacgaccagc ggcaccctca accagatcca gcggatctat     900
gtgcagaatg gcaagacggg cgcgtcggct gcgtccggag gcgacatcat cacggcatcc     960
ggctgcacct cggcccaggc gttcggcggg ctggccaaca tgggcgcggc gcttggacgg    1020
ggcatggtgc tgaccttcag catctggaac gacgctgggg gctacatgaa ctggctcgac    1080
agcggcaaca acggcccgtg cagcagcacc gagggcaacc cgtccaacat cctggccaac    1140
taccgggaca cccacgtggt cttctccaac atccgctggg gagacatcgg ctcgacggtc    1200
caggtctcgg gaggcggcaa cggcggctcg accaccacca cgtcgaccac cacgctgagg    1260
acctcgacca cgaccaccac caccgccccg acggccactg ccacgcactg gggacaatgc    1320
ggcggaatcg gggtagtca accgcctcct gcattctggt gaggaagtta actaacgtgg    1380
cctacgcagt ggactggacc gaccgtctgc gaatcgccgt acgcatgcaa ggagctgaac    1440
ccctggtact accagtgcct ctaaagtatt gcagtgaagc catactccgt gctcggcatg    1500
g                                                                                   1501
    
```

<210> 14
<211> 464
<212> PRT
<213> Thielavia terrestris
<400> 14

5

ES 2 538 360 T3

Met Gly Gln Lys Thr Leu His Gly Phe Ala Ala Thr Ala Leu Ala Val
1 5 10 15

Leu Pro Phe Val Lys Ala Gln Gln Pro Gly Asn Phe Thr Pro Glu Val
20 25 30

His Pro Gln Leu Pro Thr Trp Lys Cys Thr Thr Ala Gly Gly Cys Val
35 40 45

Gln Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Ile His
50 55 60

Asn Ala Asp Gly Thr Ala Ser Cys Thr Thr Ser Ser Gly Val Asp His
65 70 75 80

Thr Leu Cys Pro Asp Glu Ala Thr Cys Ala Lys Asn Cys Phe Val Glu
85 90 95

Gly Val Asn Tyr Thr Ser Ser Gly Val Thr Thr Ser Gly Ser Ser Leu
100 105 110

Thr Met Arg Gln Tyr Phe Lys Gly Ser Asn Gly Gln Thr Asn Ser Val
115 120 125

Ser Pro Arg Leu Tyr Leu Leu Gly Ser Asp Gly Asn Tyr Val Met Leu
130 135 140

Lys Leu Leu Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Thr Leu
145 150 155 160

Pro Cys Gly Glu Asn Gly Ala Leu Tyr Leu Ser Glu Met Asp Ala Thr
165 170 175

Gly Gly Arg Asn Gln Tyr Asn Thr Gly Gly Ala Asn Tyr Gly Ser Gly
180 185 190

Tyr Cys Asp Ala Gln Cys Pro Val Gln Thr Trp Met Asn Gly Thr Leu
195 200 205

Asn Thr Asn Gly Gln Gly Tyr Cys Cys Asn Glu Met Asp Ile Leu Glu
210 215 220

Ala Asn Ser Arg Ala Asn Ala Met Thr Pro His Pro Cys Ala Asn Gly
225 230 235 240

Ser Cys Asp Lys Ser Gly Cys Gly Leu Asn Pro Tyr Ala Glu Gly Tyr
245 250 255

ES 2 538 360 T3

Lys Ser Tyr Tyr Gly Pro Gly Leu Thr Val Asp Thr Ser Lys Pro Phe
260 265 270

Thr Ile Ile Thr Arg Phe Ile Thr Asp Asp Gly Thr Thr Ser Gly Thr
275 280 285

Leu Asn Gln Ile Gln Arg Ile Tyr Val Gln Asn Gly Lys Thr Val Ala
290 295 300

Ser Ala Ala Ser Gly Gly Asp Ile Ile Thr Ala Ser Gly Cys Thr Ser
305 310 315 320

Ala Gln Ala Phe Gly Gly Leu Ala Asn Met Gly Ala Ala Leu Gly Arg
325 330 335

Gly Met Val Leu Thr Phe Ser Ile Trp Asn Asp Ala Gly Gly Tyr Met
340 345 350

Asn Trp Leu Asp Ser Gly Asn Asn Gly Pro Cys Ser Ser Thr Glu Gly
355 360 365

Asn Pro Ser Asn Ile Leu Ala Asn Tyr Pro Asp Thr His Val Val Phe
370 375 380

Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser Thr Val Gln Val Ser Gly
385 390 395 400

Gly Gly Asn Gly Gly Ser Thr Thr Thr Thr Ser Thr Thr Thr Leu Arg
405 410 415

Thr Ser Thr Thr Thr Thr Thr Thr Ala Pro Thr Ala Thr Ala Thr His
420 425 430

Trp Gly Gln Cys Gly Gly Ile Gly Trp Thr Gly Pro Thr Val Cys Glu
435 440 445

Ser Pro Tyr Ala Cys Lys Glu Leu Asn Pro Trp Tyr Tyr Gln Cys Leu
450 455 460

<210> 15

<211> 1368

<212> ADN

<213> Thielavia terrestris

5

<400> 15

accgatccgc tcgaagatgg cgcccaagtc tacagttctg gccgcctggc tgctctcctc 60

gctggccgcg gccagcaga tcggcaaagc cgtgcccagag gtccacccca aactgacaac 120

ES 2 538 360 T3

gcagaagtgc actctccgcg gcgggtgcaa gcctgtccgc acctcgggtcg tgctcgactc 180
gtccgcgcgc tcgctgcaca aggtcgggga cccaacacc agctgcagcg tcggcgggga 240
cctgtgctcg gacgcgaagt cgtgcggcaa gaactgcgcg ctcgagggcg tcgactacgc 300
ggccccacggc gtggcgacca agggcgacgc cctcacgctg caccagtggc tcaagggggc 360
cgacggcacc tacaggaccg tctcgccgcg cgtatacctc ctgggcgagc acgggaagaa 420
ctacgaggac ttcaagctgc tcaacgccga gctcagcttc gacgtcgacg tgtcccagct 480
cgtctgcggc atgaacggcg ccctgtactt ctccgagatg gagatggacg gcggccgcag 540
cccgtgaac ccggcgggcg ccacgtacgg cacgggctac tgcgacgcg agtgccccaa 600
gttggacttt atcaacggcg aggtatctt tctctcttct gtttttcttt tccatcgctt 660
tttctgaccg gaatccgcc tcttagctca acaccaacca cacgtacggg gcgtgctgca 720
acgagatgga catctgggag gccaacgcgc tggcgcaggc gctcacgccg caccctgca 780
acgcgacgcg ggtgtacaag tgcgacacgg cggacgagtg cgggcagccg gtgggcgtgt 840
gcgacgaatg ggggtgctcg tacaaccctt ccaacttcgg ggtcaaggac tactacgggc 900
gcaacctgac ggtggacacg aaccgcaagt tcacgggtgac gacgcagttc gtgacgtcca 960
acgggcgggc ggacggcgag ctgaccgaga tccggcggct gtacgtgcag gacggcgtgg 1020
tgatccagaa ccacgcggtc acggcgggcg gggcgcagta cgacagcatc acggacggct 1080
tctgcaacgc gacggccacc tggacgcagc agcggggcgg gctcgcgcgc atgggcgagg 1140
ccatcggccg cggcatggtg ctcatcttca gcctgtgggt tgacaacggc ggcttcatga 1200
actggctcga cagcggcaac gccgggccct gcaacgccac cgagggcgac ccggccctga 1260
tcctgcagca gcacccggac gccagcgtca ctttctcaa catccgatgg ggcgagatcg 1320
gcagcacgta caagagcgag tgcagccact agagtagagc ttgtaatt 1368

<210> 16

<211> 423

<212> PRT

<213> Thielavia terrestris

<400> 16

5

ES 2 538 360 T3

Met Ala Pro Lys Ser Thr Val Leu Ala Ala Trp Leu Leu Ser Ser Leu
1 5 10 15

Ala Ala Ala Gln Gln Ile Gly Lys Ala Val Pro Glu Val His Pro Lys
20 25 30

Leu Thr Thr Gln Lys Cys Thr Leu Arg Gly Gly Cys Lys Pro Val Arg
35 40 45

Thr Ser Val Val Leu Asp Ser Ser Ala Arg Ser Leu His Lys Val Gly

ES 2 538 360 T3

50						55										60
Asp	Pro	Asn	Thr	Ser	Cys	Ser	Val	Gly	Gly	Asp	Leu	Cys	Ser	Asp	Ala	
65					70					75					80	
Lys	Ser	Cys	Gly	Lys	Asn	Cys	Ala	Leu	Glu	Gly	Val	Asp	Tyr	Ala	Ala	
				85					90					95		
His	Gly	Val	Ala	Thr	Lys	Gly	Asp	Ala	Leu	Thr	Leu	His	Gln	Trp	Leu	
			100					105					110			
Lys	Gly	Ala	Asp	Gly	Thr	Tyr	Arg	Thr	Val	Ser	Pro	Arg	Val	Tyr	Leu	
		115					120					125				
Leu	Gly	Glu	Asp	Gly	Lys	Asn	Tyr	Glu	Asp	Phe	Lys	Leu	Leu	Asn	Ala	
	130					135					140					
Glu	Leu	Ser	Phe	Asp	Val	Asp	Val	Ser	Gln	Leu	Val	Cys	Gly	Met	Asn	
145					150					155					160	
Gly	Ala	Leu	Tyr	Phe	Ser	Glu	Met	Glu	Met	Asp	Gly	Gly	Arg	Ser	Pro	
				165					170					175		
Leu	Asn	Pro	Ala	Gly	Ala	Thr	Tyr	Gly	Thr	Gly	Tyr	Cys	Asp	Ala	Gln	
			180					185					190			
Cys	Pro	Lys	Leu	Asp	Phe	Ile	Asn	Gly	Glu	Leu	Asn	Thr	Asn	His	Thr	
		195					200					205				
Tyr	Gly	Ala	Cys	Cys	Asn	Glu	Met	Asp	Ile	Trp	Glu	Ala	Asn	Ala	Leu	
	210					215					220					
Ala	Gln	Ala	Leu	Thr	Pro	His	Pro	Cys	Asn	Ala	Thr	Arg	Val	Tyr	Lys	
225					230					235					240	
Cys	Asp	Thr	Ala	Asp	Glu	Cys	Gly	Gln	Pro	Val	Gly	Val	Cys	Asp	Glu	
				245					250					255		
Trp	Gly	Cys	Ser	Tyr	Asn	Pro	Ser	Asn	Phe	Gly	Val	Lys	Asp	Tyr	Tyr	
			260					265					270			
Gly	Arg	Asn	Leu	Thr	Val	Asp	Thr	Asn	Arg	Lys	Phe	Thr	Val	Thr	Thr	
		275					280					285				
Gln	Phe	Val	Thr	Ser	Asn	Gly	Arg	Ala	Asp	Gly	Glu	Leu	Thr	Glu	Ile	
	290					295					300					

ES 2 538 360 T3

Arg Arg Leu Tyr Val Gln Asp Gly Val Val Ile Gln Asn His Ala Val
 305 310 315 320

Thr Ala Gly Gly Ala Thr Tyr Asp Ser Ile Thr Asp Gly Phe Cys Asn
 325 330 335

Ala Thr Ala Thr Trp Thr Gln Gln Arg Gly Gly Leu Ala Arg Met Gly
 340 345 350

Glu Ala Ile Gly Arg Gly Met Val Leu Ile Phe Ser Leu Trp Val Asp
 355 360 365

Asn Gly Gly Phe Met Asn Trp Leu Asp Ser Gly Asn Ala Gly Pro Cys
 370 375 380

Asn Ala Thr Glu Gly Asp Pro Ala Leu Ile Leu Gln Gln His Pro Asp
 385 390 395 400

Ala Ser Val Thr Phe Ser Asn Ile Arg Trp Gly Glu Ile Gly Ser Thr
 405 410 415

Tyr Lys Ser Glu Cys Ser His
 420

- <210> 17
- <211> 1011
- <212> ADN
- <213> Thielavia terrestris

5

<400> 17

ES 2 538 360 T3

atgaccctac ggctccctgt catcagcctg ctggcctcgc tggcagcagg cgccgtcgtc 60
 gtcccacggg cggagtttca cccccctctc ccgacttggga aatgcacgac ctccggggggc 120
 tgogtgcagc agaacaccag cgtcgtcctg gaccgtgact cgaagtacgc cgcacacagc 180
 gccggctcgc ggacggaatc ggattacgcg gcaatgggag tgtccacttc gggcaatgcc 240
 gtgacgctgt accactacgt caagaccaac ggcaccctcg tccccgcttc gccgcgcatac 300
 tacctcctgg gcgcggacgg caagtacgtg cttatggacc tcctcaacca ggagctgtcg 360
 gtggacgtcg acttctcggc gctgccgtgc ggcgagaacg gggccttcta cctgtccgag 420
 atggcggcgg acgggcgggg cgacgcgggg gcggggcagc ggtactgcga cgcgcagtgc 480
 cagggctact gctgcaacga gatggacatc ctcgaggcca actcgatggc gacggccatg 540
 acgccgcacc cgtgcaaggg caacaactgc gaccgcagcg gctgcgggcta caaccctac 600
 gccagcggcc agcgcggctt ctacgggccc ggcaagacgg tcgacacgag caagcccttc 660
 accgtcgtca cgcagttcgc cgccagcggc ggcaagctga cccagatcac ccgcaagtac 720
 atccagaacg gccgggagat cggcggcggc ggcaccatct ccagctgcgg ctccgagtct 780
 tcgacgggcg gcctgaccgg catgggagcag gcgctggggc gcggaatggt gctggccatg 840
 agcatctgga acgacgcggc ccaggagatg gcatggctcg atgccggcaa caacggccct 900
 tgcgccagtg gccagggcag cccgtccgtc attcagtcgc agcatcccga caccacgctc 960
 gtcttctcca acatcaggtg gggcgacatc gggctacca cgaagaacta g 1011

5 <210> 18
 <211> 336
 <212> PRT
 <213> Thielavia terrestris

10 <400> 18

Met Thr Leu Arg Leu Pro Val Ile Ser Leu Leu Ala Ser Leu Ala Ala
 1 5 10 15

Gly Ala Val Val Val Pro Arg Ala Glu Phe His Pro Pro Leu Pro Thr
 20 25 30

Trp Lys Cys Thr Thr Ser Gly Gly Cys Val Gln Gln Asn Thr Ser Val
 35 40 45

Val Leu Asp Arg Asp Ser Lys Tyr Ala Ala His Ser Ala Gly Ser Arg
 50 55 60

Thr Glu Ser Asp Tyr Ala Ala Met Gly Val Ser Thr Ser Gly Asn Ala
 65 70 75 80

Val Thr Leu Tyr His Tyr Val Lys Thr Asn Gly Thr Leu Val Pro Ala
 85 90 95

Ser Pro Arg Ile Tyr Leu Leu Gly Ala Asp Gly Lys Tyr Val Leu Met
 100 105 110

Asp Leu Leu Asn Gln Glu Leu Ser Val Asp Val Asp Phe Ser Ala Leu
 115 120 125

Pro Cys Gly Glu Asn Gly Ala Phe Tyr Leu Ser Glu Met Ala Ala Asp
 130 135 140

Gly Arg Gly Asp Ala Gly Ala Gly Asp Gly Tyr Cys Asp Ala Gln Cys
 145 150 155 160

Gln Gly Tyr Cys Cys Asn Glu Met Asp Ile Leu Glu Ala Asn Ser Met
 165 170 175

Ala Thr Ala Met Thr Pro His Pro Cys Lys Gly Asn Asn Cys Asp Arg
 180 185 190

Ser Gly Cys Gly Tyr Asn Pro Tyr Ala Ser Gly Gln Arg Gly Phe Tyr
 195 200 205

Gly Pro Gly Lys Thr Val Asp Thr Ser Lys Pro Phe Thr Val Val Thr
 210 215 220

Gln Phe Ala Ala Ser Gly Gly Lys Leu Thr Gln Ile Thr Arg Lys Tyr
 225 230 240

Ile Gln Asn Gly Arg Glu Ile Gly Gly Gly Gly Thr Ile Ser Ser Cys
 245 250 255

Gly Ser Glu Ser Ser Thr Gly Gly Leu Thr Gly Met Gly Glu Ala Leu
 260 265 270

Gly Arg Gly Met Val Leu Ala Met Ser Ile Trp Asn Asp Ala Ala Gln
 275 280 285

Glu Met Ala Trp Leu Asp Ala Gly Asn Asn Gly Pro Cys Ala Ser Gly
 290 295 300

Gln Gly Ser Pro Ser Val Ile Gln Ser Gln His Pro Asp Thr His Val
 305 310 315 320

Val Phe Ser Asn Ile Arg Trp Gly Asp Ile Gly Ser Thr Thr Lys Asn
 325 330 335

<210> 19

<211> 1480

<212> ADN

<213> Cladorrhinum foecundissimum

<400> 19

5

ES 2 538 360 T3

gatccgaatt cctcctctcg ttctttagtc acagaccaga catctgccca cgatggttca 60
caagttcgcc ctccctcaccg gcctcgccgc ctccctcgca tctgccccagc agatcggcac 120
cgtcgtcccc gagtctcacc ccaagcttcc caccaagcgc tgcactctcg ccggtggtg 180
ccagaccgtc gacacctcca tcgtcatcga cgccttccag cgtcccctcc acaagatcgg 240
cgacccttcc actccttgcg tcgtcggcgg ccctctctgc cccgacgcca agtcctgcgc 300
tgagaactgc gcgctcgagg gtgtcgacta tgcctcctgg ggcatacaaga ccgagggcga 360
cgccctaact ctcaaccagt ggatgccga cccggcgaac cctggccagt acaagacgac 420
tactccccgt acttaccttg ttgctgagga cggcaagaac tacgaggatg tgaagctcct 480
ggctaaggag atctcgtttg atgccgatgt cagcaacctt ccctgcggca tgaacggtgc 540
tttctacttg tctgagatgt tgatggatgg tggacgtggc gacctcaacc ctgctggtgc 600
cgagtatggt accggttact gtgatgcgca gtgcttcaag ttggatttca tcaacggcga 660
ggccaacatc gaccaaaagc acggcgcctg ctgcaacgaa atggacattt tcgaatccaa 720
ctcgcgcgcc aagaccttcg tccccaccc ctgcaacatc acgcaggtct acaagtgcga 780
aggcgaagac gagtgccggcc agcccgtcgg cgtgtgcgac aagtgggggt gcggcttcaa 840
cgagtacaaa tggggcgtcg agtccttcta cggccggggc tcgcagttcg ccatcgactc 900
ctccaagaag ttcaccgtca ccacgcagtt cctgaccgac aacggcaagg aggacggcgt 960
cctcgtcgag atccgccgct tgtggcacca ggatggcaag ctgatcaaga acaccgctat 1020
ccaggttgag gagaactaca gcacggactc ggtgagcacc gagttctgcg agaagactgc 1080
ttctttcacc atgcagcgcg gtggtctcaa ggcgatgggc gaggctatcg gtcgtggtat 1140
ggtgctggtt ttcagcatct gggcggatga ttcgggtttt atgaactggt tggatgcgga 1200
gggtaatggc ccttgcagcg cgactgaggg cgatccgaag gagattgtca agaataagcc 1260
ggatgctagg gttacgttct caaacattag gattggtgag gttggtagca cgtatgctcc 1320
gggtgggaag tgcggtgta agagcagggt tgctaggggg cttactgctt cttaaggggg 1380
gtgtgaagag aggaggaggt gttggtgggg gttggagatg ataattgggc gagatggtgt 1440
agagcggggt ggttgatata gaatacgttg aattggatgt 1480

5 <210> 20
<211> 440
<212> PRT
<213> Cladorrhinum foecundissimum

10 <400> 20

ES 2 538 360 T3

Met Val His Lys Phe Ala Leu Leu Thr Gly Leu Ala Ala Ser Leu Ala
 1 5 10 15

Ser Ala Gln Gln Ile Gly Thr Val Val Pro Glu Ser His Pro Lys Leu
 20 25 30

Pro Thr Lys Arg Cys Thr Leu Ala Gly Gly Cys Gln Thr Val Asp Thr
 35 40 45

Ser Ile Val Ile Asp Ala Phe Gln Arg Pro Leu His Lys Ile Gly Asp
 50 55 60

Pro Ser Thr Pro Cys Val Val Gly Gly Pro Leu Cys Pro Asp Ala Lys
 65 70 75 80

Ser Cys Ala Glu Asn Cys Ala Leu Glu Gly Val Asp Tyr Ala Ser Trp
 85 90 95

ES 2 538 360 T3

Gly Ile Lys Thr Glu Gly Asp Ala Leu Thr Leu Asn Gln Trp Met Pro
 100 105 110

Asp Pro Ala Asn Pro Gly Gln Tyr Lys Thr Thr Thr Pro Arg Thr Tyr
 115 120 125

Leu Val Ala Glu Asp Gly Lys Asn Tyr Glu Asp Val Lys Leu Leu Ala
 130 135 140

Lys Glu Ile Ser Phe Asp Ala Asp Val Ser Asn Leu Pro Cys Gly Met
 145 150 155 160

Asn Gly Ala Phe Tyr Leu Ser Glu Met Leu Met Asp Gly Gly Arg Gly
 165 170 175

Asp Leu Asn Pro Ala Gly Ala Glu Tyr Gly Thr Gly Tyr Cys Asp Ala
 180 185 190

Gln Cys Phe Lys Leu Asp Phe Ile Asn Gly Glu Ala Asn Ile Asp Gln
 195 200 205

Lys His Gly Ala Cys Cys Asn Glu Met Asp Ile Phe Glu Ser Asn Ser
 210 215 220

Arg Ala Lys Thr Phe Val Pro His Pro Cys Asn Ile Thr Gln Val Tyr
 225 230 235 240

Lys Cys Glu Gly Glu Asp Glu Cys Gly Gln Pro Val Gly Val Cys Asp
 245 250 255

Lys Trp Gly Cys Gly Phe Asn Glu Tyr Lys Trp Gly Val Glu Ser Phe
 260 265 270

Tyr Gly Arg Gly Ser Gln Phe Ala Ile Asp Ser Ser Lys Lys Phe Thr
 275 280 285

Val Thr Thr Gln Phe Leu Thr Asp Asn Gly Lys Glu Asp Gly Val Leu
 290 295 300

Val Glu Ile Arg Arg Leu Trp His Gln Asp Gly Lys Leu Ile Lys Asn
 305 310 315 320

Thr Ala Ile Gln Val Glu Glu Asn Tyr Ser Thr Asp Ser Val Ser Thr
 325 330 335

Glu Phe Cys Glu Lys Thr Ala Ser Phe Thr Met Gln Arg Gly Gly Leu
 340 345 350

Lys Ala Met Gly Glu Ala Ile Gly Arg Gly Met Val Leu Val Phe Ser
 355 360 365

Ile Trp Ala Asp Asp Ser Gly Phe Met Asn Trp Leu Asp Ala Glu Gly
 370 375 380

Asn Gly Pro Cys Ser Ala Thr Glu Gly Asp Pro Lys Glu Ile Val Lys
 385 390 395 400

Asn Lys Pro Asp Ala Arg Val Thr Phe Ser Asn Ile Arg Ile Gly Glu
 405 410 415

Val Gly Ser Thr Tyr Ala Pro Gly Gly Lys Cys Gly Val Lys Ser Arg
 420 425 430

Val Ala Arg Gly Leu Thr Ala Ser
 435 440

- <210> 21
- <211> 1380
- <212> ADN
- <213> Trichoderma reesei
- <400> 21

5

ES 2 538 360 T3

atggcgcct cagttacact gccgttgacc acggccatcc tggccattgc ccggctcgtc 60
 gccgcccagc aaccgggtac cagcaccccc gaggtccatc ccaagttgac aacctacaag 120
 tgtacaaagt ccgggggggtg cgtggcccag gacacctcgg tggtccttga ctggaactac 180
 cgctggatgc acgacgcaaa ctacaactcg tgcaccgtca acggcggcgt caacaccacg 240
 ctctgccctg acgaggcgac ctgtggcaag aactgcttca tcgagggcgt cgactacgcc 300
 gcctcggggcgc tcacgacctc gggcagcagc ctcacatga accagtacat gcccagcagc 360
 tctggcggct acagcagcgt ctctcctcgg ctgtatctcc tggactctga cggtgagtac 420
 gtgatgctga agctcaacgg ccaggagctg agcttcgacg tcgacctctc tgctctgccg 480
 tgtggagaga acggctcgtc ctacctgtct cagatggacg agaacggggg cgccaaccag 540
 tataacacgg ccggtgccaa ctacgggagc ggctactgcg atgctcagtg ccccgctccag 600
 acatggagga acggcaccct caacactagc caccagggct tctgctgcaa cgagatggat 660
 atcctggagg gcaactcgag ggcgaatgcc ttgaccctc actcttgac ggccacggcc 720
 tgcgactctg ccggttgccg cttcaacccc tatggcagcg gctacaaaag ctactacggc 780
 cccggagata ccggtgacac ctccaagacc ttcacatca tcaccagtt caacacggac 840
 aacggctcgc cctcgggcaa ccttgtgagc atcaccgcga agtaccagca aaacggcgtc 900
 gacatcccca ggcgccagcc cggcggcgac accatctcgt cctgcccgtc cgcctcagcc 960
 tacggcggcc tcgccacatc gggcaaggcc ctgagcagcg gcatgggtgct cgtgttcagc 1020
 atttggaacg acaacagcca gtacatgaac tggctcgaca gcggcaacgc cggcccctgc 1080
 agcagcaccg agggcaacct atccaacatc ctggccaaca accccaacac gcacgtcgtc 1140
 ttctccaaca tccgctgggg agacattggg tctactacga actcgactgc gccccgccc 1200
 ccgcctgcgt ccagcacgac gttttcgact acacggagga gctcgacgac ttcgagcagc 1260
 ccgagctgca cgcagactca ctgggggcag tgcggtggca ttgggtacag cgggtgcaag 1320
 acgtgcacgt cgggcactac gtgccagtat agcaacgact actactcgca atgcctttag 1380

5 <210> 22
 <211> 459
 <212> PRT
 <213> Trichoderma reesei

10 <400> 22

Met Ala Pro Ser Val Thr Leu Pro Leu Thr Thr Ala Ile Leu Ala Ile
 1 5 10 15

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

His Pro Lys Leu Thr Thr Tyr Lys Cys Thr Lys Ser Gly Gly Cys Val
 35 40 45

Ala Gln Asp Thr Ser Val Val Leu Asp Trp Asn Tyr Arg Trp Met His
 50 55 60

Asp Ala Asn Tyr Asn Ser Cys Thr Val Asn Gly Gly Val Asn Thr Thr
 65 70 75 80

Leu Cys Pro Asp Glu Ala Thr Cys Gly Lys Asn Cys Phe Ile Glu Gly
 85 90 95

Val Asp Tyr Ala Ala Ser Gly Val Thr Thr Ser Gly Ser Ser Leu Thr
 100 105 110

Met Asn Gln Tyr Met Pro Ser Ser Ser Gly Gly Tyr Ser Ser Val Ser
 115 120 125

Pro Arg Leu Tyr Leu Leu Asp Ser Asp Gly Glu Tyr Val Met Leu Lys
 130 135 140

Leu Asn Gly Gln Glu Leu Ser Phe Asp Val Asp Leu Ser Ala Leu Pro
 145 150 155 160

Cys Gly Glu Asn Gly Ser Leu Tyr Leu Ser Gln Met Asp Glu Asn Gly

ES 2 538 360 T3

Thr Ser Ser Ser Pro Ser Cys Thr Gln Thr His Trp Gly Gln Cys Gly
420 425 430

Gly Ile Gly Tyr Ser Gly Cys Lys Thr Cys Thr Ser Gly Thr Thr Cys
435 440 445

Gln Tyr Ser Asn Asp Tyr Tyr Ser Gln Cys Leu
450 455

<210> 23

<211> 2586

<212> ADN

<213> Aspergillus oryzae

<400> 23

5

ES 2 538 360 T3

atgaagcttg gttggatcga ggtggccgca ttggcggctg cctcagtagt cagtgccaaag 60
 gatgatctcg cgtactcccc tcctttctac ccttccccat gggcagatgg tcagggtgaa 120
 tgggcggaag tatacaaacg cgctgtagac atagtttccc agatgacgtt gacagagaaa 180
 gtcaacttaa cgactggaac aggatggcaa ctagagaggt gtgttggaaca aactggcagt 240
 gttcccagac tcaacatccc cagcttgtgt ttgcaggata gtcctcttgg tattcgtttc 300
 tcggactaca attcagcttt ccctgicgggt gttaatgtcg ctgccacctg ggacaagacg 360
 ctcgcctacc ttcgtggtca ggcaatgggt gaggagttca gtgataaggg tattgacgtt 420
 cagctgggtc ctgctgctgg ccctctcgggt gctcatccgg atggcggtag aaactgggaa 480
 ggtttctcac cagatccagc cctcaccgggt gtactttttg cggagacgat taagggtatt 540
 caagatgctg gtgtcattgc gacagctaag cattatatca tgaacgaaca agagcatttc 600
 cgccaacaac ccgaggctgc gggttacgga ttcaacgtaa gcgacagttt gagttccaac 660
 gttgatgaca agactatgca tgaattgtac ctctggccct tcgcggatgc agtacgcgct 720
 ggagtcggtg ctgtcatgtg ctcttacaac caaatcaaca acagctacgg ttgicgagaat 780
 agcgaactc tgaacaagct tttgaaggcg gagcttggtt tccaaggctt cgtcatgagt 840
 gattggaccg ctcatcacag cggcgtaggc gctgctttag caggtctgga tatgtcgatg 900
 cccggtgatg ttaccttcga tagtggtagc tctttctggg gtgcaaactt gacggtcggt 960
 gtccttaacg gtacaatccc ccaatggcgt gttgatgaca tggctgtccg tatcatggcc 1020
 gcttattaca aggttggccg cgacaccaa tacaccctc ccaacttcag ctcggtggacc 1080
 agggacgaat atggtttcgc gcataacat gtttcggaag gtgcttacga gaggtcaac 1140
 gaattcgtgg acgtgcaacg cgatcatgcc gacctaacc gtcgcatcgg cgcgcagagc 1200
 actgttctgc tgaagaacaa ggggtgccttg cccttgagcc gcaaggaaaa gctggtcgcc 1260
 cttctgggag aggatgcggg ttccaactcg tggggcgcta acggctgtga tgaccgtggt 1320
 tgcgataacg gtacccttgc catggcctgg ggtagcggta ctgcgaattt cccatacctc 1380

ES 2 538 360 T3

gtgacaccag agcagggcat tcagaacgaa gttcttcagg gccgtggtaa tgtcttcgcc 1440
 gtgaccgaca gttgggctct cgacaagatc gctgcggtg cccgccaggc cagcgtatct 1500
 ctctgtttcg tcaactccga ctcaggagaa ggctatctta gtgtggatgg aaatgagggc 1560
 gatcgttaaca acatcactct gtggaagaac ggcgacaatg tggtaaacac cgcagcgaat 1620
 aactgtaaca acaccgttgt catcatccac tccgtcggac cagttttgat cgatgaatgg 1680
 tatgaccacc ccaatgtcac tggattctc tgggctggc tgccaggcca ggagtctggc 1740
 aactccattg ccgatgtgct gtacggctgt gtcaaccctg gcgccaaagc tcctttcact 1800
 tggggcaaga cccgggagtc gtatggttct cccttggca aggatgcca caatggcaac 1860
 ggagcgcgcc agtctgattt caccagggt gttttcatcg attacgcca tttcgataag 1920
 ttcaatgaga cccctatcta cgagtttggc tacggcttga gctacaccac cttcgagctc 1980
 tccgacctcc atgttcagcc cctgaacgcg tcccgataca ctcccaccag tggcatgact 2040
 gaagctgcaa agaactttgg tgaaattggc gatgcgtcgg agtacgtgta tccggagggg 2100
 ctggaaagga tccatgagtt tatctatccc tggatcaact ctaccgacct gaaggcacgc 2160
 tctgacgatt ctaactacgg ctgggaagac tccaagtata ttcccgaagg cgccacggat 2220
 gggctctgcc agccccgttt gcccgctagt ggtggtgccg gaggaaacc cggctctgtac 2280
 gaggatcttt tccgcgtctc tgtgaaggtc aagaacacgg gcaatgtcgc cggatgatgaa 2340
 gttcctcagc tgtacgtttc cctaggcggc ccgaatgagc ccaaggtggt actgcgcaag 2400
 tttgagcgta ttcacttggc cccttcgcag gaggccgtgt ggacaacgac ccttaccctg 2460
 cgtgacctg caaactggga cgtttcggct caggactgga ccgtcactcc ttaccccaag 2520
 acgatctacg ttggaaactc ctcacggaaa ctgccgtcc aggcctcgtc gcctaaggcc 2580
 cagtaa 2586

<210> 24

<211> 861

<212> PRT

<213> *Aspergillus oryzae*

<400> 24

5

ES 2 538 360 T3

Met Lys Leu Gly Trp Ile Glu Val Ala Ala Leu Ala Ala Ala Ser Val
1 5 10 15

Val Ser Ala Lys Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser
20 25 30

Pro Trp Ala Asp Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala
35 40 45

ES 2 538 360 T3

Val Asp Ile Val Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr
50 55 60

Thr Gly Thr Gly Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser
65 70 75 80

Val Pro Arg Leu Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu
85 90 95

Gly Ile Arg Phe Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn
100 105 110

Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Gln Ala
115 120 125

Met Gly Glu Glu Phe Ser Asp Lys Gly Ile Asp Val Gln Leu Gly Pro
130 135 140

Ala Ala Gly Pro Leu Gly Ala His Pro Asp Gly Gly Arg Asn Trp Glu
145 150 155 160

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

Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
180 185 190

Ile Met Asn Glu Gln Glu His Phe Arg Gln Gln Pro Glu Ala Ala Gly
195 200 205

Tyr Gly Phe Asn Val Ser Asp Ser Leu Ser Ser Asn Val Asp Asp Lys
210 215 220

Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
225 230 235 240

Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
245 250 255

Gly Cys Glu Asn Ser Glu Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
260 265 270

Gly Phe Gln Gly Phe Val Met Ser Asp Trp Thr Ala His His Ser Gly
275 280 285

Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Val
290 295 300

ES 2 538 360 T3

Thr Phe Asp Ser Gly Thr Ser Phe Trp Gly Ala Asn Leu Thr Val Gly
305 310 315 320

Val Leu Asn Gly Thr Ile Pro Gln Trp Arg Val Asp Asp Met Ala Val
325 330 335

Arg Ile Met Ala Ala Tyr Tyr Lys Val Gly Arg Asp Thr Lys Tyr Thr
340 345 350

Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Phe Ala His
355 360 365

Asn His Val Ser Glu Gly Ala Tyr Glu Arg Val Asn Glu Phe Val Asp
370 375 380

Val Gln Arg Asp His Ala Asp Leu Ile Arg Arg Ile Gly Ala Gln Ser
385 390 395 400

Thr Val Leu Leu Lys Asn Lys Gly Ala Leu Pro Leu Ser Arg Lys Glu
405 410 415

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

Ala Asn Gly Cys Asp Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
435 440 445

Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu
450 455 460

Gln Ala Ile Gln Asn Glu Val Leu Gln Gly Arg Gly Asn Val Phe Ala
465 470 475 480

Val Thr Asp Ser Trp Ala Leu Asp Lys Ile Ala Ala Ala Ala Arg Gln
485 490 495

Ala Ser Val Ser Leu Val Phe Val Asn Ser Asp Ser Gly Glu Gly Tyr
500 505 510

Leu Ser Val Asp Gly Asn Glu Gly Asp Arg Asn Asn Ile Thr Leu Trp
515 520 525

Lys Asn Gly Asp Asn Val Val Lys Thr Ala Ala Asn Asn Cys Asn Asn
530 535 540

Thr Val Val Ile Ile His Ser Val Gly Pro Val Leu Ile Asp Glu Trp
545 550 555 560

Tyr Asp His Pro Asn Val Thr Gly Ile Leu Trp Ala Gly Leu Pro Gly
 565 570 575

Gln Glu Ser Gly Asn Ser Ile Ala Asp Val Leu Tyr Gly Arg Val Asn
 580 585 590

Pro Gly Ala Lys Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
 595 600 605

Gly Ser Pro Leu Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln
 610 615 620

Ser Asp Phe Thr Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys
 625 630 635 640

Phe Asn Glu Thr Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr
 645 650 655

Thr Phe Glu Leu Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg
 660 665 670

Tyr Thr Pro Thr Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu
 675 680 685

Ile Gly Asp Ala Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile
 690 695 700

His Glu Phe Ile Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser
 705 710 715 720

Ser Asp Asp Ser Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu
 725 730 735

Gly Ala Thr Asp Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly
 740 745 750

Ala Gly Gly Asn Pro Gly Leu Tyr Glu Asp Leu Phe Arg Val Ser Val
 755 760 765

Lys Val Lys Asn Thr Gly Asn Val Ala Gly Asp Glu Val Pro Gln Leu
 770 775 780

Tyr Val Ser Leu Gly Gly Pro Asn Glu Pro Lys Val Val Leu Arg Lys
 785 790 795 800

Phe Glu Arg Ile His Leu Ala Pro Ser Gln Glu Ala Val Trp Thr Thr

ES 2 538 360 T3

805

810

815

Thr Leu Thr Arg Arg Asp Leu Ala Asn Trp Asp Val Ser Ala Gln Asp
820 825 830

Trp Thr Val Thr Pro Tyr Pro Lys Thr Ile Tyr Val Gly Asn Ser Ser
835 840 845

Arg Lys Leu Pro Leu Gln Ala Ser Leu Pro Lys Ala Gln
850 855 860

<210> 25

<211> 2586

5 <212> ADN

<213> *Aspergillus oryzae*

<400> 25

ES 2 538 360 T3

atgaagcttg gttggatcga ggtggccgca ttggcggctg cctcagtagt cagtgccaaag 60
 gatgatctcg cgtactcccc tcctttctac ccttccccat gggcagatgg tcagggtgaa 120
 tgggcggaag tatacaaacg cgctgtagac atagtttccc agatgacgtt gacagagaaa 180
 gtcaacttaa cgactggaac aggatggcaa ctagagaggt gtgttggaca aactggcagt 240
 gttcccagac tcaacatccc cagcttgtgt ttgcaggata gtccctcttg tattcgtttc 300
 tcggactaca attcagcttt ccctgcggtt gttaatgtcg ctgccacctg ggacaagacg 360
 ctcgcctacc ttcgtggtca ggcaatgggt gaggagttca gtgataaggg tattgacgtt 420
 cagctgggtc ctgctgctgg ccctctcggg gctcatccgg atggcggtag aaactgggaa 480
 agtttctcac cagatccagc cctcaccggt gtactttttg cggagacgat taagggtatt 540
 caagatgctg gtgtcattgc gacagctaag cattatatca tgaacgaaca agagcatttc 600
 cgccaacaac ccgaggctgc gggttacgga ttcaacgtaa gcgacagttt gagttccaac 660
 gttgatgaca agactatgca tgaattgtac ctctggcoct tcgcggatgc agtacgcgct 720
 ggagtcggtg ctgttatgtg ctcttacaac caaatcaaca acagctacgg ttgcgagaat 780
 agcgaaactc tgaacaagct tttgaaggcg gagcttggtt tccaaggctt cgtcatgagt 840
 gattggaccg ctcaacacag cggcgtaggc gctgctttag caggctctgga tatgtcgatg 900
 cccggtgatg ttaccttcca tagtggtacg tctttctggg gtgcaaactt gacggtcggg 960
 gtccttaacg gtacaatccc ccaatggcgt gttgatgaca tggctgtccg tatcatggcc 1020
 gcttattaca aggttggccg cgacaccaa tacaccctc ccaacttcag ctcggtggacc 1080
 agggacgaat atggtttcgc gcataacat gtttcggaag gtgcttacga gagggccaac 1140
 gaattcgtgg acgtgcaacg cgatcatgcc gacctaatec gtcgcatcgg cgcgcagagc 1200
 actgttctgc tgaagaacaa ggggtgcctt cccttgagcc gcaaggaaaa gctggtcgcc 1260

ES 2 538 360 T3

cttctgggag aggatgcggg ttccaactcg tggggcgcta acggctgtga tgaccgtggt 1320
 tgcgataacg gtacccttgc catggcctgg ggtagcggta ctgcgaattt cccatacctc 1380
 gtgacaccag agcaggcgat tcagaacgaa gttcttcagg gccgtggtaa tgtcttcgcc 1440
 gtgaccgaca gttgggcgct cgacaagatc gctgcggctg cccgccaggc cagcgtatct 1500
 ctcgtgttcg tcaactccga ctcaggagaa ggctatctta gtgtggatgg aatgagggc 1560
 gatcgtaaca acatcactct gtggaagaac ggcgacaatg tgggcaagac cgcagcgaat 1620
 aactgtaaca acaccgttgt catcatccac tccgtcggac cagttttgat cgatgaatgg 1680
 tatgaccacc ccaatgtcac tggattctc tgggctggtc tgccaggcca ggagtctggt 1740
 aactccattg ccgatgtgct gtacggtcgt gtcaaccctg gcgccaagtc tcctttcact 1800
 tggggcaaga cccgggagtc gtatggttct cccttggca aggatgcca caatggcaac 1860
 ggagcgcccc agtctgattt caccaggggt gttttcatcg attaccgcca tttcgataag 1920
 ttcaatgaga ccctatcta cgagtttggc tacggcttga gctacaccac cttcgagctc 1980
 tccgacctcc atgttcagcc cctgaacgcg tcccataca ctcccaccag tggcatgact 2040
 gaagctgcaa agaactttgg tgaaattggc gatgcgtcgg agtacgtgta tccggagggg 2100
 ctggaaagga tccatgagtt tatctatccc tggatcaact ctaccgacct gaaggcatcg 2160
 tctgacgatt ctaactacgg ctgggaagac tccaagtata ttcccgaagg cgccacggat 2220
 gggctctgcc agccccgttt gcccgctagt ggtggtgccg gaggaaacct cggctctgtac 2280
 gaggatcttt tccgcgtctc tgtgaaggtc aagaacacgg gcaatgtcgc cggatgatgaa 2340
 gttcctcagc tgtacgtttc cctaggcggc ccgaatgagc ccaagggtgg actgcgcaag 2400
 tttgagcgta ttcacttggc cccttcgcag gaggccgtgt ggacaacgac ccttaccctg 2460
 cgtgaccttg caaactggga cgtttcggct caggactgga ccgtcactcc ttaccccaag 2520
 acgatctacg ttggaaactc ctcaaggaaa ctgccgctcc aggcctcgct gcctaaggcc 2580
 cagtaa 2586

<210> 26
 <211> 861
 <212> PRT
 <213> *Aspergillus oryzae*

<400> 26

5

ES 2 538 360 T3

Met Lys Leu Gly Trp Ile Glu Val Ala Ala Leu Ala Ala Ala Ser Val
1 5 10 15

Val Ser Ala Lys Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser
20 25 30

Pro Trp Ala Asp Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala

ES 2 538 360 T3

	35		40		45														
Val	Asp	Ile	Val	Ser	Gln	Met	Thr	Leu	Thr	Glu	Lys	Val	Asn	Leu	Thr				
	50					55					60								
Thr	Gly	Thr	Gly	Trp	Gln	Leu	Glu	Arg	Cys	Val	Gly	Gln	Thr	Gly	Ser				
	65				70					75					80				
Val	Pro	Arg	Leu	Asn	Ile	Pro	Ser	Leu	Cys	Leu	Gln	Asp	Ser	Pro	Leu				
				85					90					95					
Gly	Ile	Arg	Phe	Ser	Asp	Tyr	Asn	Ser	Ala	Phe	Pro	Ala	Gly	Val	Asn				
			100					105						110					
Val	Ala	Ala	Thr	Trp	Asp	Lys	Thr	Leu	Ala	Tyr	Leu	Arg	Gly	Gln	Ala				
		115					120							125					
Met	Gly	Glu	Glu	Phe	Ser	Asp	Lys	Gly	Ile	Asp	Val	Gln	Leu	Gly	Pro				
	130					135					140								
Ala	Ala	Gly	Pro	Leu	Gly	Ala	His	Pro	Asp	Gly	Gly	Arg	Asn	Trp	Glu				
	145				150					155					160				
Ser	Phe	Ser	Pro	Asp	Pro	Ala	Leu	Thr	Gly	Val	Leu	Phe	Ala	Glu	Thr				
				165					170					175					
Ile	Lys	Gly	Ile	Gln	Asp	Ala	Gly	Val	Ile	Ala	Thr	Ala	Lys	His	Tyr				
			180					185						190					
Ile	Met	Asn	Glu	Gln	Glu	His	Phe	Arg	Gln	Gln	Pro	Glu	Ala	Ala	Gly				
		195					200					205							
Tyr	Gly	Phe	Asn	Val	Ser	Asp	Ser	Leu	Ser	Ser	Asn	Val	Asp	Asp	Lys				
	210					215					220								
Thr	Met	His	Glu	Leu	Tyr	Leu	Trp	Pro	Phe	Ala	Asp	Ala	Val	Arg	Ala				
	225				230					235					240				
Gly	Val	Gly	Ala	Val	Met	Cys	Ser	Tyr	Asn	Gln	Ile	Asn	Asn	Ser	Tyr				
				245					250					255					
Gly	Cys	Glu	Asn	Ser	Glu	Thr	Leu	Asn	Lys	Leu	Leu	Lys	Ala	Glu	Leu				
			260					265					270						
Gly	Phe	Gln	Gly	Phe	Val	Met	Ser	Asp	Trp	Thr	Ala	Gln	His	Ser	Gly				
		275					280					285							

ES 2 538 360 T3

Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Val
 290 295 300

Thr Phe Asp Ser Gly Thr Ser Phe Trp Gly Ala Asn Leu Thr Val Gly
 305 310 315 320

Val Leu Asn Gly Thr Ile Pro Gln Trp Arg Val Asp Asp Met Ala Val
 325 330 335

Arg Ile Met Ala Ala Tyr Tyr Lys Val Gly Arg Asp Thr Lys Tyr Thr
 340 345 350

Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Phe Ala His
 355 360 365

Asn His Val Ser Glu Gly Ala Tyr Glu Arg Val Asn Glu Phe Val Asp
 370 375 380

Val Gln Arg Asp His Ala Asp Leu Ile Arg Arg Ile Gly Ala Gln Ser
 385 390 395 400

Thr Val Leu Leu Lys Asn Lys Gly Ala Leu Pro Leu Ser Arg Lys Glu
 405 410 415

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

Ala Asn Gly Cys Asp Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
 435 440 445

Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu
 450 455 460

Gln Ala Ile Gln Asn Glu Val Leu Gln Gly Arg Gly Asn Val Phe Ala
 465 470 475 480

Val Thr Asp Ser Trp Ala Leu Asp Lys Ile Ala Ala Ala Ala Arg Gln
 485 490 495

Ala Ser Val Ser Leu Val Phe Val Asn Ser Asp Ser Gly Glu Gly Tyr
 500 505 510

Leu Ser Val Asp Gly Asn Glu Gly Asp Arg Asn Asn Ile Thr Leu Trp
 515 520 525

Lys Asn Gly Asp Asn Val Val Lys Thr Ala Ala Asn Asn Cys Asn Asn
 530 535 540

ES 2 538 360 T3

Thr Val Val Ile Ile His Ser Val Gly Pro Val Leu Ile Asp Glu Trp
545 550 555 560

Tyr Asp His Pro Asn Val Thr Gly Ile Leu Trp Ala Gly Leu Pro Gly
565 570 575

Gln Glu Ser Gly Asn Ser Ile Ala Asp Val Leu Tyr Gly Arg Val Asn
580 585 590

Pro Gly Ala Lys Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
595 600 605

Gly Ser Pro Leu Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln
610 615 620

Ser Asp Phe Thr Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys
625 630 635 640

Phe Asn Glu Thr Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr
645 650 655

Thr Phe Glu Leu Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg
660 665 670

Tyr Thr Pro Thr Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu
675 680 685

Ile Gly Asp Ala Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile
690 695 700

His Glu Phe Ile Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser
705 710 715 720

Ser Asp Asp Ser Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu
725 730 735

Gly Ala Thr Asp Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly
740 745 750

Ala Gly Gly Asn Pro Gly Leu Tyr Glu Asp Leu Phe Arg Val Ser Val
755 760 765

Lys Val Lys Asn Thr Gly Asn Val Ala Gly Asp Glu Val Pro Gln Leu
770 775 780

Tyr Val Ser Leu Gly Gly Pro Asn Glu Pro Lys Val Val Leu Arg Lys
785 790 795 800

ES 2 538 360 T3

Phe Glu Arg Ile His Leu Ala Pro Ser Gln Glu Ala Val Trp Thr Thr
805 810 815

Thr Leu Thr Arg Arg Asp Leu Ala Asn Trp Asp Val Ser Ala Gln Asp
820 825 830

Trp Thr Val Thr Pro Tyr Pro Lys Thr Ile Tyr Val Gly Asn Ser Ser
835 840 845

Arg Lys Leu Pro Leu Gln Ala Ser Leu Pro Lys Ala Gln
850 855 860

<210> 27

<211> 3060

<212> ADN

<213> Aspergillus fumigatus

<400> 27

5

ES 2 538 360 T3

atgagattcg gttggctcga ggtggccgct ctgacggccg cttctgtagc caatgccag 60
 gtttgtgatg ctttcccgtc attgttccgg atatagttga caatagtcac ggaaataatc 120
 aggaattggc tttctctcca ccattctacc cttcgccttg ggctgatggc cagggagagt 180
 gggcagatgc ccatcgacgc gccgtcgaga tcgtttctca gatgacactg gcggagaagg 240
 ttaaccttac aacgggtact ggggtgggtg cgactttttt gttgacagtg agctttcttc 300
 actgaccatc tacacagatg ggaaatggac cgatgcgtcg gtcaaaccgg cagcgttccc 360
 aggtaagctt gcaattctgc aacaacgtgc aagtgtagtt gctaaaacgc ggtggtgcag 420
 acttggtatc aactggggtc tttgtggcca ggattcccct ttgggtatcc gtttctgtga 480
 gctatacccg cggagtcttt cagtccttgt attatgtgct gatgattgtc tctgtatagc 540
 tgacctcaac tccgccttcc ctgctggtac taatgtcgcc gcgacatggg acaagacact 600
 cgcctacctt cgtggcaagg ccatgggtga ggaattcaac gacaagggcg tggacatctt 660
 gctggggcct gctgctggtc ctctcggcaa ataccggac ggcggcagaa tctgggaagg 720
 cttctctcct gatccggttc tcaactggtg acttttcgcc gaaactatca agggtatcca 780
 agacgcgggt gtgattgcta ctgccaagca ttacattctg aatgaacagg agcatttccg 840
 acaggttggc gaggcccagg gatatggtta caacatcacg gagacgatca gctccaacgt 900
 ggatgacaag accatgcacg agttgtacct ttggtgagta gttgacactg caaatgagga 960
 ccttgattga tttgactgac ctggaatgca ggccctttgc agatgctgtg cgcggtaaga 1020
 ttttccgtag acttgacctc gcgacgaaga aatcgctgac gaaccatcgt agctggcgtt 1080
 ggcgctgtca tgtgttccta caatcaaac aacaacagct acggttgtca aaacagtcaa 1140
 actctcaaca agctcctcaa ggctgagctg ggcttccaag gcttcgtcat gagtgactgg 1200

ES 2 538 360 T3

agcgctcacc acagcgggtgt cggcgctgcc ctcgctgggt tggatatgtc gatgcctgga 1260
 gacatttcct tcgacgacgg actctccttc tggggcacga acctaactgt cagtgttctt 1320
 aacggcaccg ttccagcctg gcgtgtcgat gacatggctg ttcgtatcat gaccgcgtac 1380
 tacaaggttg gtcgtgaccg tcttcgtatt cccctaact tcagctcctg gaccgggat 1440
 gagtacggct gggagcattc tgctgtctcc gagggagcct ggaccaaggt gaacgacttc 1500
 gtcaatgtgc agcgcagtca ctctcagatc atccgtgaga ttggtgccgc tagtacagtg 1560
 ctcttgaaga acacgggtgc tcttcctttg accggcaagg aggttaaagt ggggtgttctc 1620
 ggtgaagacg ctggttccaa cccgtgggggt gctaacggct gccccgaccg cggctgtgat 1680
 aacggcactc ttgctatggc ctggggtagt ggtactgcca acttcctta ccttgtcacc 1740
 cccgagcagg ctatccagcg agaggtcac agcaacggcg gcaatgtctt tgctgtgact 1800
 gataacgggg ctctcagcca gatggcagat gttgcatctc aatccagggt agtgcgggct 1860
 cttagaaaaa gaacgttctc tgaatgaagt ttttaacca ttgcgaacag cgtgtctttg 1920
 gtgtttgtca acgccgactc tggagaggggt ttcacagtg tcgacggcaa cgagggtgac 1980
 cgcaaaaatc tactctgtg gaagaacggc gaggccgtca ttgacactgt tgtcagccac 2040
 tgcaacaaca cgatttggt tattcacagt gttgggcccg tcttgatcga ccggtggtat 2100
 gataacccca acgtcactgc catcatctgg gccggcttgc ccggtcagga gagtggcaac 2160
 tccttggtcg acgtgctcta tggccgcgtc aaccacagcg ccaagacccc gttcacctgg 2220
 ggcaagactc gggagtctta cggggctccc ttgctcaccg agcctaaca tggcaatggt 2280
 gctccccagg atgatttcaa cgagggcgtc ttcattgact accgtcactt tgacaagcgc 2340
 aatgagaccc ccatttatga gtttggccat ggcttgagct acaccacctt tggttactct 2400
 caccttcggg ttcaggccct caatagttcg agttcggcat atgtcccgc tagcggagag 2460
 accaagcctg cgccaaccta tggtgagatc gtagtgccg ccgactacct gtatcccag 2520
 ggtctcaaaa gaattaccaa gtttatttac ccttggtca actcgaccga cctcgaggat 2580
 tcttctgacg acccgaacta cggctgggag gactcggagt acattcccga aggcgctagg 2640
 gatgggtctc ctcaaccct cctgaaggct ggcggcgctc ctggtggtaa ccctaccctt 2700
 tatcaggatc ttgtaggggt gtcggccacc ataaccaaca ctggtaacgt cgccggttat 2760
 gaagtccctc aattggtgag tgaccgcgat gttccttgcg ttgcaatttg gctaactcgc 2820
 ttctagtatg tttcactggg cggaccgaac gagcctcggg tcgttctgcg caagttcgac 2880
 cgaatcttcc tggctcctgg ggagcaaaaag gtttgacca cgactctta ccgtcgtgat 2940
 ctgcaccaatt gggatgtgga ggctcaggac tgggtcatca caaagtaccc caagaaagtg 3000
 cacgtcggca gtcctcgcg taagctgcct ctgagagcgc ctctgccccg tgtctactag 3060

<210> 28
<211> 863
<212> PRT
5 <213> *Aspergillus fumigatus*

<400> 28

Met Arg Phe Gly Trp Leu Glu Val Ala Ala Leu Thr Ala Ala Ser Val
1 5 10 15

Ala Asn Ala Gln Glu Leu Ala Phe Ser Pro Pro Phe Tyr Pro Ser Pro
20 25 30

Trp Ala Asp Gly Gln Gly Glu Trp Ala Asp Ala His Arg Arg Ala Val
35 40 45

Glu Ile Val Ser Gln Met Thr Leu Ala Glu Lys Val Asn Leu Thr Thr
50 55 60

Gly Thr Gly Trp Glu Met Asp Arg Cys Val Gly Gln Thr Gly Ser Val
65 70 75 80

Pro Arg Leu Gly Ile Asn Trp Gly Leu Cys Gly Gln Asp Ser Pro Leu
85 90 95

Gly Ile Arg Phe Ser Asp Leu Asn Ser Ala Phe Pro Ala Gly Thr Asn
100 105 110

Val Ala Ala Thr Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Lys Ala
115 120 125

Met Gly Glu Glu Phe Asn Asp Lys Gly Val Asp Ile Leu Leu Gly Pro
130 135 140

Ala Ala Gly Pro Leu Gly Lys Tyr Pro Asp Gly Gly Arg Ile Trp Glu
145 150 155 160

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

Ile Lys Gly Ile Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr
180 185 190

Ile Leu Asn Glu Gln Glu His Phe Arg Gln Val Gly Glu Ala Gln Gly
195 200 205

Tyr Gly Tyr Asn Ile Thr Glu Thr Ile Ser Ser Asn Val Asp Asp Lys
210 215 220

ES 2 538 360 T3

Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala
 225 230 235 240

Gly Val Gly Ala Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr
 245 250 255

Gly Cys Gln Asn Ser Gln Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu
 260 265 270

Gly Phe Gln Gly Phe Val Met Ser Asp Trp Ser Ala His His Ser Gly
 275 280 285

Val Gly Ala Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Ile
 290 295 300

Ser Phe Asp Asp Gly Leu Ser Phe Trp Gly Thr Asn Leu Thr Val Ser
 305 310 315 320

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

Arg Ile Met Thr Ala Tyr Tyr Lys Val Gly Arg Asp Arg Leu Arg Ile
 340 345 350

Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly Trp Glu His
 355 360 365

Ser Ala Val Ser Glu Gly Ala Trp Thr Lys Val Asn Asp Phe Val Asn
 370 375 380

Val Gln Arg Ser His Ser Gln Ile Ile Arg Glu Ile Gly Ala Ala Ser
 385 390 395 400

Thr Val Leu Leu Lys Asn Thr Gly Ala Leu Pro Leu Thr Gly Lys Glu
 405 410 415

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

Ala Asn Gly Cys Pro Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met
 435 440 445

Ala Trp Gly Ser Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu
 450 455 460

Gln Ala Ile Gln Arg Glu Val Ile Ser Asn Gly Gly Asn Val Phe Ala
 465 470 475 480

ES 2 538 360 T3

Val Thr Asp Asn Gly Ala Leu Ser Gln Met Ala Asp Val Ala Ser Gln
485 490 495

Ser Ser Val Ser Leu Val Phe Val Asn Ala Asp Ser Gly Glu Gly Phe
500 505 510

Ile Ser Val Asp Gly Asn Glu Gly Asp Arg Lys Asn Leu Thr Leu Trp
515 520 525

Lys Asn Gly Glu Ala Val Ile Asp Thr Val Val Ser His Cys Asn Asn
530 535 540

Thr Ile Val Val Ile His Ser Val Gly Pro Val Leu Ile Asp Arg Trp
545 555 560

Tyr Asp Asn Pro Asn Val Thr Ala Ile Ile Trp Ala Gly Leu Pro Gly
565 570 575

Gln Glu Ser Gly Asn Ser Leu Val Asp Val Leu Tyr Gly Arg Val Asn
580 585 590

Pro Ser Ala Lys Thr Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr
595 600 605

Gly Ala Pro Leu Leu Thr Glu Pro Asn Asn Gly Asn Gly Ala Pro Gln
610 615 620

Asp Asp Phe Asn Glu Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys
625 630 635 640

Arg Asn Glu Thr Pro Ile Tyr Glu Phe Gly His Gly Leu Ser Tyr Thr
645 650 655

Thr Phe Gly Tyr Ser His Leu Arg Val Gln Ala Leu Asn Ser Ser Ser
660 665 670

Ser Ala Tyr Val Pro Thr Ser Gly Glu Thr Lys Pro Ala Pro Thr Tyr
675 680 685

Gly Glu Ile Gly Ser Ala Ala Asp Tyr Leu Tyr Pro Glu Gly Leu Lys
690 695 700

Arg Ile Thr Lys Phe Ile Tyr Pro Trp Leu Asn Ser Thr Asp Leu Glu
705 710 715 720

Asp Ser Ser Asp Asp Pro Asn Tyr Gly Trp Glu Asp Ser Glu Tyr Ile

ES 2 538 360 T3

725

730

735

Pro Glu Gly Ala Arg Asp Gly Ser Pro Gln Pro Leu Leu Lys Ala Gly
740 745 750

Gly Ala Pro Gly Gly Asn Pro Thr Leu Tyr Gln Asp Leu Val Arg Val
755 760 765

Ser Ala Thr Ile Thr Asn Thr Gly Asn Val Ala Gly Tyr Glu Val Pro
770 775 780

Gln Leu Tyr Val Ser Leu Gly Gly Pro Asn Glu Pro Arg Val Val Leu
785 790 795 800

Arg Lys Phe Asp Arg Ile Phe Leu Ala Pro Gly Glu Gln Lys Val Trp
805 810 815

Thr Thr Thr Leu Asn Arg Arg Asp Leu Ala Asn Trp Asp Val Glu Ala
820 825 830

Gln Asp Trp Val Ile Thr Lys Tyr Pro Lys Lys Val His Val Gly Ser
835 840 845

Ser Ser Arg Lys Leu Pro Leu Arg Ala Pro Leu Pro Arg Val Tyr
850 855 860

<210> 29

<211> 2800

<212> ADN

<213> Penicillium brasilianum

<400> 29

5

ES 2 538 360 T3

tgaaaatgca gggttctaca atctttctgg ctttcgcctc atgggcgagc caggttgctg	60
ccattgcgca gcccatacag aagcacgagg tttgttttat cttgctcatg gacgtgcttt	120
gacttgacta attgttttac atacagcccg gatttctgca cgggcccacaa gccatagaat	180
cgttctcaga accgttctac ccgtcgccct ggatgaatcc tcacgccgag ggctgggagg	240
ccgcatatca gaaagctcaa gattttgtct cgcaactcac tatcttggag aaaataaatc	300
tgaccaccgg tgttgggtaa gtctctccga ctgcttctgg gtcacgggtc gacgagccac	360
tgactttttg aagctgggaa aatgggccgt gtgtaggaaa cactggatca attcctcgtc	420
tcggattcaa aggattttgt acccaggatt caccacaggg tgttcgggtc gcagattatt	480
cctccgcttt cacatctagc caaatggccg ccgcaacatt tgaccgctca attctttatc	540
aacgaggcca agccatggca caggaacaca aggctaaggg tatcacaatt caattgggcc	600
ctgttgccgg ccctctcggc cgcacccccg agggcggccg caactgggaa ggattctccc	660

ES 2 538 360 T3

ctgatcctgt cttgactggt atagccatgg ctgagacaat taagggcatg caggatactg 720
 gagtgattgc ttgcgctaaa cattatattg gaaacgagca ggagcacttc cgtcaagtgg 780
 gtgaagctgc gggtcacgga tacactatth ccgatactat ttcactaat attgacgacc 840
 gtgctatgca tgagctatac ttgtggccat ttgctgatgc cgttcgcgct ggtgtgggtt 900
 ctttcatgtg ctcatactct cagatcaaca actcctacgg atgccaaaac agtcagaccc 960
 tcaacaagct cctcaagagc gaattgggct tccaaggctt tgtcatgagc gattgggggtg 1020
 cccatcactc tggagtgtca tcggcgctag ctggacttga tatgagcatg ccgggtgata 1080
 ccgaatttga ttctggcttg agcttctggg gctctaact caccattgca attctgaacg 1140
 gcacggttcc cgaatggcgc ctggatgaca tggcgatgcg aattatggct gcatacttca 1200
 aagttggcct tactattgag gatcaaccag atgtcaact caatgcctgg acccatgaca 1260
 cctacggata taaatacgct tatagcaagg aagattacga gcaggcaac tggcatgtcg 1320
 atgttcgag cgaccacaat aagctcattc gcgagactgc cgcgaagggt acagttctgc 1380
 tgaagaacaa ctttcatgct ctccctctga agcagcccag gttcgtggcc gtcggtggtc 1440
 aggatgccgg gccaaaacccc aagggcccta acggctgcmc agaccgagga tgcgaccaag 1500
 gcactctcgc aatgggatgg ggctcagggt ctaccgaatt cccttacctg gtcactcctg 1560
 aactgctat tcagtcaaag gtccctgaat acgggggtcg atacgagagt atttttgata 1620
 actatgacga caatgctatc ttgtcgcttg tctcacagcc tgatgcaacc tgtatcgth 1680
 ttgcaaatgc cgattccggt gaaggctaca tcaactgtcga caacaactgg ggtgaccgca 1740
 acaatctgac cctctggcaa aatgccgatc aagtgattag cactgtcagc tcgcatgca 1800
 acaacacaat cgttgttctc cactctgtcg gaccagtgtt gctaaatgg atatatgagc 1860
 acccgaacat cacagctatt gtctgggcag ggatgccagg cgaagaatct ggcaatgctc 1920
 tcgtggatat tctttggggc aatgttaacc ctgccggctg cactccgttc acctgggcca 1980
 aaagtcgaga ggactatggc actgatataa tgtacgagcc caacaacggc cagcgtgcmc 2040
 ctcagcagga tttcaccgag agcatctacc tcgactaccg ccatttcgac aaagctggta 2100
 tcgagccaat ttacgagtht ggattcggcc tctcctatac caccttcgaa tactctgacc 2160
 tccgtgttgt gaagaagtat gttcaaccat acagtcccac gaccggcacc ggtgctcaag 2220
 caccttccat cggacagcca cctagccaga acctggatac ctacaagthc cctgctacat 2280
 acaagtacat caaaaccttc atttatccct acctgaacag cactgtctcc ctccgcgctg 2340
 cttccaagga tcccgaatac ggtcgtacag actttatccc accccacgcm cgtgatggct 2400
 cccctcaacc tctcaacccc gctggagacc cagtggccag tgggtgaaac aacatgctct 2460
 acgacgaact ttacgagthc actgcacaga tcaaaaacac tggcgacgth gccggcgacg 2520
 aagtcgtcca gctttacgta gatctcgggg gtgacaaccc gcctcgtcag ttgagaaact 2580

ES 2 538 360 T3

ttgacagggtt ttatctgctg cccggtcaga gctcaacatt ccgggctaca ttgacgcgcc 2640
gtgatttgag caactgggat attgaggcgc agaactggcg agttacggaa tcgcctaaga 2700
gagtgtatgt tggacggtcg agtcgggatt tgccgctgag ctcacaattg gagtaatgat 2760
catgtctacc aatagatggt gaatgtctgg tgtggatatt 2800

<210> 30

<211> 878

<212> PRT

5 <213> Penicillium brasilianum

<400> 30

Met Gln Gly Ser Thr Ile Phe Leu Ala Phe Ala Ser Trp Ala Ser Gln
 1 5 10 15

Val Ala Ala Ile Ala Gln Pro Ile Gln Lys His Glu Pro Gly Phe Leu
 20 25 30

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

Pro Trp Met Asn Pro His Ala Glu Gly Trp Glu Ala Ala Tyr Gln Lys
 50 55 60

Ala Gln Asp Phe Val Ser Gln Leu Thr Ile Leu Glu Lys Ile Asn Leu
 65 70 75 80

Thr Thr Gly Val Gly Trp Glu Asn Gly Pro Cys Val Gly Asn Thr Gly
 85 90 95

Ser Ile Pro Arg Leu Gly Phe Lys Gly Phe Cys Thr Gln Asp Ser Pro
 100 105 110

Gln Gly Val Arg Phe Ala Asp Tyr Ser Ser Ala Phe Thr Ser Ser Gln
 115 120 125

Met Ala Ala Ala Thr Phe Asp Arg Ser Ile Leu Tyr Gln Arg Gly Gln
 130 135 140

Ala Met Ala Gln Glu His Lys Ala Lys Gly Ile Thr Ile Gln Leu Gly
 145 150 155 160

Pro Val Ala Gly Pro Leu Gly Arg Ile Pro Glu Gly Gly Arg Asn Trp
 165 170 175

Glu Gly Phe Ser Pro Asp Pro Val Leu Thr Gly Ile Ala Met Ala Glu
 180 185 190

ES 2 538 360 T3

Thr Ile Lys Gly Met Gln Asp Thr Gly Val Ile Ala Cys Ala Lys His
 195 200 205

Tyr Ile Gly Asn Glu Gln Glu His Phe Arg Gln Val Gly Glu Ala Ala
 210 215 220

Gly His Gly Tyr Thr Ile Ser Asp Thr Ile Ser Ser Asn Ile Asp Asp
 225 230 235 240

Arg Ala Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg
 245 250 255

Ala Gly Val Gly Ser Phe Met Cys Ser Tyr Ser Gln Ile Asn Asn Ser
 260 265 270

Tyr Gly Cys Gln Asn Ser Gln Thr Leu Asn Lys Leu Leu Lys Ser Glu
 275 280 285

Leu Gly Phe Gln Gly Phe Val Met Ser Asp Trp Gly Ala His His Ser
 290 295 300

Gly Val Ser Ser Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp
 305 310 315 320

Thr Glu Phe Asp Ser Gly Leu Ser Phe Trp Gly Ser Asn Leu Thr Ile
 325 330 335

Ala Ile Leu Asn Gly Thr Val Pro Glu Trp Arg Leu Asp Asp Met Ala
 340 345 350

Met Arg Ile Met Ala Ala Tyr Phe Lys Val Gly Leu Thr Ile Glu Asp
 355 360 365

Gln Pro Asp Val Asn Phe Asn Ala Trp Thr His Asp Thr Tyr Gly Tyr
 370 375 380

Lys Tyr Ala Tyr Ser Lys Glu Asp Tyr Glu Gln Val Asn Trp His Val
 385 390 395 400

Asp Val Arg Ser Asp His Asn Lys Leu Ile Arg Glu Thr Ala Ala Lys
 405 410 415

Gly Thr Val Leu Leu Lys Asn Asn Phe His Ala Leu Pro Leu Lys Gln
 420 425 430

Pro Arg Phe Val Ala Val Val Gly Gln Asp Ala Gly Pro Asn Pro Lys

435					440					445					
Gly	Pro	Asn	Gly	Cys	Ala	Asp	Arg	Gly	Cys	Asp	Gln	Gly	Thr	Leu	Ala
	450					455					460				
Met	Gly	Trp	Gly	Ser	Gly	Ser	Thr	Glu	Phe	Pro	Tyr	Leu	Val	Thr	Pro
465					470					475					480
Asp	Thr	Ala	Ile	Gln	Ser	Lys	Val	Leu	Glu	Tyr	Gly	Gly	Arg	Tyr	Glu
				485					490					495	
Ser	Ile	Phe	Asp	Asn	Tyr	Asp	Asp	Asn	Ala	Ile	Leu	Ser	Leu	Val	Ser
			500					505					510		
Gln	Pro	Asp	Ala	Thr	Cys	Ile	Val	Phe	Ala	Asn	Ala	Asp	Ser	Gly	Glu
		515					520					525			
Gly	Tyr	Ile	Thr	Val	Asp	Asn	Asn	Trp	Gly	Asp	Arg	Asn	Asn	Leu	Thr
	530					535					540				
Leu	Trp	Gln	Asn	Ala	Asp	Gln	Val	Ile	Ser	Thr	Val	Ser	Ser	Arg	Cys
545					550					555					560
Asn	Asn	Thr	Ile	Val	Val	Leu	His	Ser	Val	Gly	Pro	Val	Leu	Leu	Asn
				565					570					575	
Gly	Ile	Tyr	Glu	His	Pro	Asn	Ile	Thr	Ala	Ile	Val	Trp	Ala	Gly	Met
			580					585					590		
Pro	Gly	Glu	Glu	Ser	Gly	Asn	Ala	Leu	Val	Asp	Ile	Leu	Trp	Gly	Asn
		595					600					605			
Val	Asn	Pro	Ala	Gly	Arg	Thr	Pro	Phe	Thr	Trp	Ala	Lys	Ser	Arg	Glu
	610					615					620				
Asp	Tyr	Gly	Thr	Asp	Ile	Met	Tyr	Glu	Pro	Asn	Asn	Gly	Gln	Arg	Ala
625					630					635					640
Pro	Gln	Gln	Asp	Phe	Thr	Glu	Ser	Ile	Tyr	Leu	Asp	Tyr	Arg	His	Phe
				645					650					655	
Asp	Lys	Ala	Gly	Ile	Glu	Pro	Ile	Tyr	Glu	Phe	Gly	Phe	Gly	Leu	Ser
			660					665					670		
Tyr	Thr	Thr	Phe	Glu	Tyr	Ser	Asp	Leu	Arg	Val	Val	Lys	Lys	Tyr	Val
		675					680					685			

ES 2 538 360 T3

Gln Pro Tyr Ser Pro Thr Thr Gly Thr Gly Ala Gln Ala Pro Ser Ile
690 695 700

Gly Gln Pro Pro Ser Gln Asn Leu Asp Thr Tyr Lys Phe Pro Ala Thr
705 710 715 720

Tyr Lys Tyr Ile Lys Thr Phe Ile Tyr Pro Tyr Leu Asn Ser Thr Val
725 730 735

Ser Leu Arg Ala Ala Ser Lys Asp Pro Glu Tyr Gly Arg Thr Asp Phe
740 745 750

Ile Pro Pro His Ala Arg Asp Gly Ser Pro Gln Pro Leu Asn Pro Ala
755 760 765

Gly Asp Pro Val Ala Ser Gly Gly Asn Asn Met Leu Tyr Asp Glu Leu
770 775 780

Tyr Glu Val Thr Ala Gln Ile Lys Asn Thr Gly Asp Val Ala Gly Asp
785 790 795 800

Glu Val Val Gln Leu Tyr Val Asp Leu Gly Gly Asp Asn Pro Pro Arg
805 810 815

Gln Leu Arg Asn Phe Asp Arg Phe Tyr Leu Leu Pro Gly Gln Ser Ser
820 825 830

Thr Phe Arg Ala Thr Leu Thr Arg Arg Asp Leu Ser Asn Trp Asp Ile
835 840 845

Glu Ala Gln Asn Trp Arg Val Thr Glu Ser Pro Lys Arg Val Tyr Val
850 855 860

Gly Arg Ser Ser Arg Asp Leu Pro Leu Ser Ser Gln Leu Glu
865 870 875

<210> 31
<211> 2235
<212> ADN
<213> Trichoderma reesei

5

<400> 31

ES 2 538 360 T3

atgcgttacc gaacagcagc tgcgctggca cttgccactg ggccctttgc tagggcagac 60
agtcactcaa catcgggggc ctcggctgag gcagttgtac ctctgcagg gactccatgg 120
ggaaccgctg acgacaaggc gaaggccgca ttggcaaagc tcaatctcca agataaggtc 180
ggcatcgtga gcggtgtcgg ctggaacggc ggtccttgcg ttggaaacac atctccggcc 240
tccaagatca gctatccatc gctatgcctt caagacggac ccctcgggtg togataactcg 300

ES 2 538 360 T3

acaggcagca cagcctttac gccgggcggt caagcggcct cgacgtggga tgtcaatttg 360
 atccgcgaac gtggacagtt catcgggtgag gaggtgaagg cctcgggggat tcatgtcata 420
 cttggtcctg tggctgggcc gctgggaaag actccgcagg gcggtcgcaa ctgggagggc 480
 ttcggtgtcg atccatatct cacgggcatt gccatgggtc aaaccatcaa cggcatccag 540
 tcggtaggcg tgcagggcag agcgaagcac tatatcctca acgagcagga gctcaatcga 600
 gaaaccattt cgagcaaccc agatgaccga actctccatg agctgtatac ttggccattt 660
 gccgacgcgg ttcaggccaa tgtcgttct gtcatgtgct cgtacaacaa ggtcaatacc 720
 acctgggcct gcgaggatca gtacacgctg cagactgtgc tgaaagacca gctggggttc 780
 ccaggctatg tcatgacgga ctggaacgca cagcacacga ctgtccaaag cgcaattct 840
 gggcttgaca tgtcaatgcc tggcacagac ttcaacggta acaatcggct ctgggggtcca 900
 gctctcacca atgcggtaaa tagcaatcag gtccccacga gcagagtcga cgatatgggtg 960
 actcgtatcc tcgccgcatg gtacttgaca ggccaggacc aggcaggcta tccgtcgttc 1020
 aacatcagca gaaatgttca aggaaaccac aagaccaatg tcagggcaat tgccagggac 1080
 ggcacgttc tgctcaagaa tgacgccaac atcctgccgc tcaagaagcc cgctagcatt 1140
 gccgtcgttg gatctgccgc aatcattggt aaccacgcca gaaactcgcc ctctgtcaac 1200
 gacaaaggct gcgacgacgg ggccttgggc atgggttggg gttccggcgc cgtcaactat 1260
 ccgtacttgc tcgcgcccta cgatgccatc aataccagag cgtcttcgca gggcaccag 1320
 gttaccttga gcaacaccga caacacgtcc tcaggcgcac ctgcagcaag aggaaaggac 1380
 gtcgccatcg tcttcatcac cgccgactcg ggtgaaggct acatcacctg ggagggcaac 1440
 gcgggagatc gcaacaacct ggatccgtgg cacaacggca atgccctggt ccaggcgggtg 1500
 gccggtgcca acagcaacgt cattgttgtt gtccactccg ttggcgccat cattctggag 1560
 cagattcttg ctcttcgca ggtcaaggcc gttgtctggg cgggtcttcc ttctcaggag 1620
 agcggcaatg cgctcgtcga cgtgctgtgg ggagatgtca gcccttctgg caagctggtg 1680
 tacaccattg cgaagagccc caatgactat aacactcgca tcgtttccgg cggcagtgac 1740
 agcttcagcg agggactggt catcgactat aagcaactcg acgacgcaa tatcacgccg 1800
 cggtacgagt tcggctatgg actgtcttac accaagttca actactcacg cctctccgtc 1860
 ttgtcgaccg ccaagtctgg tcttgcgact ggggcccgtt tgccgggagg cccgagtgat 1920
 ctgttccaga atgtcgcgac agtcaccgtt gacatcgcaa actctggcca agtgactggt 1980
 gccgaggtag cccagctgta catcacctac ccatcttcag caccaggac ccctccgaag 2040
 cagctgcgag gctttgcaa gctgaacctc acgcctggtc agagcggaac agcaacgttc 2100
 aacatccgac gacgagatct cagctactgg gacacggctt cgagaaatg ggtggtgccg 2160

ES 2 538 360 T3

tcggggtcgt ttggcatcag cgtgggagcg agcagccggg atatcaggct gacgagcact 2220

ctgtcggtag cgtag 2235

5 <210> 32
<211> 744
<212> PRT
<213> Trichoderma reesei

<400> 32

ES 2 538 360 T3

Met Arg Tyr Arg Thr Ala Ala Ala Leu Ala Leu Ala Thr Gly Pro Phe
1 5 10 15

Ala Arg Ala Asp Ser His Ser Thr Ser Gly Ala Ser Ala Glu Ala Val
20 25 30

Val Pro Pro Ala Gly Thr Pro Trp Gly Thr Ala Tyr Asp Lys Ala Lys
35 40 45

Ala Ala Leu Ala Lys Leu Asn Leu Gln Asp Lys Val Gly Ile Val Ser
50 55 60

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

Ser Lys Ile Ser Tyr Pro Ser Leu Cys Leu Gln Asp Gly Pro Leu Gly
85 90 95

Val Arg Tyr Ser Thr Gly Ser Thr Ala Phe Thr Pro Gly Val Gln Ala
100 105 110

Ala Ser Thr Trp Asp Val Asn Leu Ile Arg Glu Arg Gly Gln Phe Ile
115 120 125

Gly Glu Glu Val Lys Ala Ser Gly Ile His Val Ile Leu Gly Pro Val
130 135 140

Ala Gly Pro Leu Gly Lys Thr Pro Gln Gly Gly Arg Asn Trp Glu Gly
145 150 155 160

Phe Gly Val Asp Pro Tyr Leu Thr Gly Ile Ala Met Gly Gln Thr Ile
165 170 175

Asn Gly Ile Gln Ser Val Gly Val Gln Ala Thr Ala Lys His Tyr Ile
180 185 190

Leu Asn Glu Gln Glu Leu Asn Arg Glu Thr Ile Ser Ser Asn Pro Asp
195 200 205

Asp Arg Thr Leu His Glu Leu Tyr Thr Trp Pro Phe Ala Asp Ala Val
 210 215 220

Gln Ala Asn Val Ala Ser Val Met Cys Ser Tyr Asn Lys Val Asn Thr
 225 230 235 240

Thr Trp Ala Cys Glu Asp Gln Tyr Thr Leu Gln Thr Val Leu Lys Asp
 245 250 255

Gln Leu Gly Phe Pro Gly Tyr Val Met Thr Asp Trp Asn Ala Gln His
 260 265 270

Thr Thr Val Gln Ser Ala Asn Ser Gly Leu Asp Met Ser Met Pro Gly
 275 280 285

Thr Asp Phe Asn Gly Asn Asn Arg Leu Trp Gly Pro Ala Leu Thr Asn
 290 295 300

Ala Val Asn Ser Asn Gln Val Pro Thr Ser Arg Val Asp Asp Met Val
 305 310 315 320

Thr Arg Ile Leu Ala Ala Trp Tyr Leu Thr Gly Gln Asp Gln Ala Gly
 325 330 335

Tyr Pro Ser Phe Asn Ile Ser Arg Asn Val Gln Gly Asn His Lys Thr
 340 345 350

Asn Val Arg Ala Ile Ala Arg Asp Gly Ile Val Leu Leu Lys Asn Asp
 355 360 365

Ala Asn Ile Leu Pro Leu Lys Lys Pro Ala Ser Ile Ala Val Val Gly
 370 375 380

Ser Ala Ala Ile Ile Gly Asn His Ala Arg Asn Ser Pro Ser Cys Asn
 385 390 395 400

Asp Lys Gly Cys Asp Asp Gly Ala Leu Gly Met Gly Trp Gly Ser Gly
 405 410 415

Ala Val Asn Tyr Pro Tyr Phe Val Ala Pro Tyr Asp Ala Ile Asn Thr
 420 425 430

Arg Ala Ser Ser Gln Gly Thr Gln Val Thr Leu Ser Asn Thr Asp Asn
 435 440 445

Thr Ser Ser Gly Ala Ser Ala Ala Arg Gly Lys Asp Val Ala Ile Val
 450 455 460

Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Glu Gly Asn
 465 470 475 480

Ala Gly Asp Arg Asn Asn Leu Asp Pro Trp His Asn Gly Asn Ala Leu
 485 490 495

Val Gln Ala Val Ala Gly Ala Asn Ser Asn Val Ile Val Val Val His
 500 505 510

Ser Val Gly Ala Ile Ile Leu Glu Gln Ile Leu Ala Leu Pro Gln Val
 515 520 525

Lys Ala Val Val Trp Ala Gly Leu Pro Ser Gln Glu Ser Gly Asn Ala
 530 535 540

Leu Val Asp Val Leu Trp Gly Asp Val Ser Pro Ser Gly Lys Leu Val
 545 550 555 560

Tyr Thr Ile Ala Lys Ser Pro Asn Asp Tyr Asn Thr Arg Ile Val Ser
 565 570 575

Gly Gly Ser Asp Ser Phe Ser Glu Gly Leu Phe Ile Asp Tyr Lys His
 580 585 590

Phe Asp Asp Ala Asn Ile Thr Pro Arg Tyr Glu Phe Gly Tyr Gly Leu
 595 600 605

Ser Tyr Thr Lys Phe Asn Tyr Ser Arg Leu Ser Val Leu Ser Thr Ala
 610 615 620

Lys Ser Gly Pro Ala Thr Gly Ala Val Val Pro Gly Gly Pro Ser Asp
 625 630 635 640

Leu Phe Gln Asn Val Ala Thr Val Thr Val Asp Ile Ala Asn Ser Gly
 645 650 655

Gln Val Thr Gly Ala Glu Val Ala Gln Leu Tyr Ile Thr Tyr Pro Ser
 660 665 670

Ser Ala Pro Arg Thr Pro Pro Lys Gln Leu Arg Gly Phe Ala Lys Leu
 675 680 685

Asn Leu Thr Pro Gly Gln Ser Gly Thr Ala Thr Phe Asn Ile Arg Arg
 690 695 700

Arg Asp Leu Ser Tyr Trp Asp Thr Ala Ser Gln Lys Trp Val Val Pro

705

710

715

720

Ser Gly Ser Phe Gly Ile Ser Val Gly Ala Ser Ser Arg Asp Ile Arg
 725 730 735

Leu Thr Ser Thr Leu Ser Val Ala
 740

5 <210> 33
 <211> 29
 <212> ADN
 <213> Trichoderma reesei

10 <400> 33
 aacgtaatt aaggaatcgt tttgtgtt 29

15 <210> 34
 <211> 29
 <212> ADN
 <213> Trichoderma reesei

20 <400> 34
 agtactagta gctccgtggc gaaagcctg 29

25 <210> 35
 <211> 31
 <212> ADN
 <213> Saccharomyces cerevisiae

30 <400> 35
 ttgaattgaa aatagattga tttaaaactt c 31

35 <210> 36
 <211> 25
 <212> ADN
 <213> Saccharomyces cerevisiae

40 <400> 36
 ttgcatgctg aatcatggtc atagc 25

45 <210> 37
 <211> 26
 <212> ADN
 <213> Saccharomyces cerevisiae

50 <400> 37
 ttgaattcat gggaataaac tgatat 26

55 <210> 38
 <211> 32
 <212> ADN
 <213> Saccharomyces cerevisiae

<400> 38
 aaatcaatct atttcaatt caattcatca tt 32

<210> 39
 <211> 11
 <212> ADN
 <213> Aspergillus niger

<400> 39
 gtactaaaac c 11

5 <210> 40
 <211> 11
 <212> ADN
 <213> Aspergillus niger

10 <400> 40
 ccgtaaatt t 11

15 <210> 41
 <211> 45
 <212> ADN
 <213> Aspergillus niger

<400> 41
 ggatgctgtt gactccggaa attaacggt ttggtctgc atccc 45

20 <210> 42
 <211> 14
 <212> ADN
 <213> Aspergillus niger

25 <400> 42
 atgcaattta aact 14

30 <210> 43
 <211> 14
 <212> ADN
 <213> Aspergillus niger

35 <400> 43
 cggcaattta acgg 14

40 <210> 44
 <211> 44
 <212> ADN
 <213> Aspergillus niger

<400> 44
 ggtattgtcc tgacagcggc aattaacgg cttctgcgaa tcgc 44

45 <210> 45
 <211> 29
 <212> ADN
 <213> Humicola insolens

50 <400> 45
 aagcttaagc atgcttctct cccccctcc 29

55 <210> 46
 <211> 32
 <212> ADN
 <213> Humicola insolens

<400> 46
 ctgcagaatt ctacaggcac tgatggtacc ag 32

60 <210> 47
 <211> 32
 <212> ADN
 <213> Trichoderma reesei

65 <400> 47
 ctgcagaatt ctacaggcac tgatggtacc ag 32

<210> 48
 <211> 36
 <212> ADN
 5 <213> Trichoderma reesei

 <400> 48
 accgaggact gcgcatcatg cggtcctccc ccctcc 36

 10 <210> 49
 <211> 29
 <212> ADN
 <213> Trichoderma reesei

 15 <400> 49
 aaacgtcgac cgaatgtagg attgttatc???29

 <210> 50
 <211> 17
 20 <212> ADN
 <213> Trichoderma reesei

 <400> 50
 gatgcgcagt ccgcggt 17

 25 <210> 51
 <211> 29
 <212> ADN
 <213> Trichoderma reesei

 30 <400> 51
 aaacgtcgac cgaatgtagg attgttatc 29

 <210> 52
 <211> 36
 <212> ADN
 <213> Trichoderma reesei

 35 <400> 52
 ggagggggga ggaacgcatg atgcgcagtc cgcggt 36

 <210> 53
 <211> 29
 <212> ADN
 45 <213> Trichoderma reesei

 <400> 53
 aaacgtcgac cgaatgtagg attgttatc 29

 50 <210> 54
 <211> 32
 <212> ADN
 <213> Trichoderma reesei

 55 <400> 54
 ctgcagaatt ctacaggcac tgatgtacc ag 32

 <210> 55
 <211> 46
 60 <212> ADN
 <213> Aspergillus oryzae

 <400> 55
 atagtcaacc gcggactgcg catcatgaag ctggttga tcgagg 46

 65 <210> 56

<211> 26
 <212> ADN
 <213> Aspergillus oryzae

5 <400> 56
 actagtttac tgggccttag gcagcg 26

<210> 57
 <211> 26
 <212> ADN
 <213> Trichoderma reesei

10

<400> 57
 gtcgactoga agcccgaatg taggat 26

15

<210> 58
 <211> 45
 <212> ADN
 <213> Trichoderma reesei

20

<400> 58
 cctgatcca accaagcttc atgatgcgca gtccgcggtt gacta 45

<210> 59
 <211> 57
 <212> ADN
 <213> Aspergillus oryzae

25

<400> 59
 atgaagcttg gttgatcga ggtggccgca ttggcggctg cctcagtagt cagtgcc 57

30

<210> 60
 <211> 19
 <212> PRT
 <213> Aspergillus oryzae

35

<400> 60

Met Lys Leu Gly Trp Ile Glu Val Ala Ala Leu Ala Ala Ala Ser Val
1 5 10 15

Val Ser Ala

40

<210> 61
 <211> 42
 <212> ADN
 <213> Aspergillus oryzae

45

<400> 61
 tgccggtggt ggccttgcc aaggatgatc tcgctactc cc 42

<210> 62
 <211> 28
 <212> ADN
 <213> Aspergillus oryzae

50

<400> 62
 gactagtctt actggcctt aggcagcg 28

55

<210> 63
 <211> 63
 <212> ADN
 <213> Humicola insolens

60

<400> 63
atgctgtct cccccctct ccgctccgcc gttgtggccg ccctgccggt gttggcctt 60

5 Gcc 63

<210> 64
<211> 21
<212> PRT

10 <213> Humicola insolens

<400> 64

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro
1 5 10 15

Val Leu Ala Leu Ala
20

15 <210> 65
<211> 30
<212> ADN
<213> Trichoderma reesei

20 <400> 65
acgcgtcgac cgaatgtagg atgttatcc 30

<210> 66
<211> 42
25 <212> ADN
<213> Trichoderma reesei

<400> 66
gggagtacgc gagatcatcc ttggcaaggg ccaacaccgg ca 42

30 <210> 67
<211> 20
<212> ADN
<213> Trichoderma reesei

35 <400> 67
cccaagctta gccaagaaca 20

<210> 68
<211> 29
40 <212> ADN
<213> Trichoderma reesei

<400> 68
45 gggggaggaa cgcatggat ctggacggc 29

<210> 69
<211> 30
<212> ADN
50 <213> Aspergillus oryzae

<400> 69
gccgtccaga tccccatgct ttctcccc 30

55 <210> 70
<211> 20
<212> ADN
<213> Aspergillus oryzae

ES 2 538 360 T3

<400> 70
ccaagcttgt tcagagtttc 20

5 <210> 71
<211> 20
<212> ADN
<213> Aspergillus oryzae

10 <400> 71
ggactgcgca gcatgcgttc 20

15 <210> 72
<211> 30
<212> ADN
<213> Aspergillus oryzae

20 <400> 72
agttaattaa ttactgggcc ttaggcagcg 30

25 <210> 73
<211> 3294
<212> ADN
<213> Aspergillus oryzae

25 <400> 73

ES 2 538 360 T3

atgcgttcct cccccctcct ccgctccgcc gttgtggccg ccctgccggt gttggccctt 60
 gccgctgatg gcaggtccac ccgctactgg gactgctgca agccttcgtg cggctgggccc 120
 aagaaggctc ccgtgaacca gcctgtcttt tcctgcaacg ccaacttcca gcgtatcacg 180
 gacttcgacg ccaagtccgg ctgcgagccg ggcggtgtcg cctactcgtg cgccgaccag 240
 accccatggg ctgtgaacga cgacttcgcg ctcggttttg ctgccacctc tattgccggc 300
 agcaatgagg cgggctggtg ctgcgccctgc tacgagctca ccttcacatc cggctcctggt 360
 gctggcaaga agatggctgt ccagtcacc agcactggcg gtgatcttgg cagcaaccac 420
 ttgatctca acatccccgg cggcggcgtc ggcatcttcg acggatgcac tccccagttc 480
 ggtggtctgc ccggccagcg ctacggcggc atctcgtccc gcaacgagtg cgatcggttc 540
 cccgacgccc tcaagcccgg ctgctactgg cgcttcgact ggttcaagaa cgccgacaat 600
 ccgagcttca gttcccgta ggtccagtgc ccagccgagc tcgtcgctcg caccggatgc 660
 cgccgcaacg acgacggcaa ctccctgcc gtccagatcc ccatgcgttc ctccccctc 720
 ctccgctccg ccgttgtggc cgccctgccg gtgttggccc ttgccaagga tgatctcgcg 780
 tactccccct ctttctaccc ttccccatgg gcagatggtc aggggtgaatg ggcggaagta 840
 tacaacgcg ctgtagacat agtttcccag atgacggtga cagagaaagt caacttaacg 900
 actggaacag gatggcaact agagaggtgt gttggacaaa ctggcagtgt tcccagactc 960
 aacatccccca gcttgtggtt gcaggatagt cctcttggtg ttcgtttctc ggactacaat 1020
 tcagctttcc ctgcggtgtg taatgtcgct gccacctggg acaagacgct cgcctacctt 1080
 cgtggtcagg caatgggtga ggagttcagt gataagggtg ttgacgttca gctgggtcct 1140
 gctgctggcc ctctcgggtc tcatccggat ggcggtagaa actgggaagg tttctcacca 1200
 gatccagccc tcaccggtgt actttttgcg gagacgatta agggattca agatgctggt 1260
 gtcattgcga cagctaagca ttatatcatg aacgaacaag agcatttccg ccaacaaccc 1320
 gaggctgcgg gttacggatt caacgtaagc gacagtttga gttccaacgt tgatgacaag 1380
 actatgcatg aattgtacct ctggcccttc gcggatgcag tacgcgctgg agtcggtgct 1440
 gtcattgtgct cttacaacca aatcaacaac agctacggtt gcgagaatag cgaaactctg 1500

ES 2 538 360 T3

aacaagcttt tgaaggcggg gcttggtttc caaggcttcg tcatgagtga ttggaccgct 1560
catcacagcg gcgtaggcgc tgctttagca ggtctggata tgtcgatgcc cggtgatggt 1620
accttcgata gtggtacgtc tttctggggt gcaaacttga cggtcggtgt ccttaacggt 1680
acaatcccc aatggcgtgt tgatgacatg gctgtccgta tcatggccgc ttattacaag 1740
gttggccgcg acaccaaata caccctccc aacttcagct cgtggaccag ggacgaatat 1800
ggtttcgcg ataaccatgt ttcggaagg gcttacgaga gggcaacga attcgtggac 1860
gtgcaacgcg atcatgccga cctaatecgt cgcacggcg cgcagagcac tgttctgctg 1920
aagaacaagg gtgccttgcc cttgagccgc aaggaaaagc tggtcgccct tctgggagag 1980
gatgcggtt ccaactcgtg gggcgctaac ggctgtgatg accgtggtg cgataacggt 2040
acccttgcca tggcctgggg tagcgggtact gcgaatttcc catacctcgt gacaccagag 2100
caggcgattc agaacgaagt tcttcagggc cgtggtaatg tcttcgccgt gaccgacagt 2160
tgggcgctcg acaagatcgc tgcggctgcc cgccaggcca gcgtatctct cgtgttcgtc 2220
aactccgact caggagaagg ctatcttagt gtggatggaa atgagggcga tcgtaacaac 2280
atcactctgt ggaagaacgg cgacaatgtg gtcaagaccg cagcgaataa ctgtaacaac 2340
accgttgtca tcatccactc cgtcggacca gttttgatcg atgaatggta tgaccacccc 2400
aatgtcactg gtattctctg ggctggctctg ccaggccagg agtctggtaa ctccattgcc 2460
gatgtgctgt acggtcgtgt caaccctggc gccaaagtctc ctttcacttg gggcaagacc 2520
cgggagtcgt atggtttccc cttggtcaag gatgccaaca atggcaacgg agcgccccag 2580
tctgatttca cccaggtgt tttcatcgat taccgccatt tcgataagtt caatgagacc 2640
cctatctacg agtttggcta cggcttgagc tacaccacct tcgagctctc cgacctccat 2700
gttcagcccc tgaacgcgtc ccgatacact cccaccagtg gcatgactga agctgcaaag 2760
aactttggtg aaattggcga tgcgtcggag tacgtgtatc cggaggggct ggaaaggatc 2820
catgagttta tctatccctg gatcaactct accgacctga aggcacgctc tgacgattct 2880
aactacggct gggaagactc caagtatatt cccgaaggcg ccacggatgg gtctgcccag 2940
ccccgtttgc ccgctagtgg tggtgccgga ggaaaccccc gtctgtacga ggatcttttc 3000
cgcgtctctg tgaaggtcaa gaacacgggc aatgtcgccg gtgatgaagt tcctcagctg 3060
tacgtttccc taggcggccc gaatgagccc aaggtggtac tgcgcaagtt tgagcgtatt 3120
cacttgccc cttcgcagga ggccgtgtgg acaacgacc ttaccgctcg tgacctgca 3180
aactgggacg tttcggctca ggactggacc gtcactcctt accccaagac gatctacgtt 3240
ggaaactcct cacggaaact gccgctccag gcctcgtgct ctaaggccca gtaa 3294

<210> 74
<211> 1097
<212> PRT
<213> *Aspergillus oryzae*

5

<400> 74

ES 2 538 360 T3

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro
1 5 10 15

Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
20 25 30

Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
35 40 45

Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
50 55 60

Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
65 70 75 80

Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
85 90 95

Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
100 105 110

Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
115 120 125

Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
130 135 140

Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
145 150 155 160

Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
165 170 175

Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
180 185 190

Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
195 200 205

Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
210 215 220

Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Met Arg Ser Ser Pro Leu
225 230 235 240

ES 2 538 360 T3

Leu Arg Ser Ala Val Val Ala Ala Leu Pro Val Leu Ala Leu Ala Lys
 245 250 255
 Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser Pro Trp Ala Asp
 260 265 270
 Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala Val Asp Ile Val
 275 280 285
 Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr Thr Gly Thr Gly
 290 295 300
 Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser Val Pro Arg Leu
 305 310 315 320
 Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu Gly Ile Arg Phe
 325 330 335
 Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn Val Ala Ala Thr
 340 345 350
 Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Gln Ala Met Gly Glu Glu
 355 360 365
 Phe Ser Asp Lys Gly Ile Asp Val Gln Leu Gly Pro Ala Ala Gly Pro
 370 375 380
 Leu Gly Ala His Pro Asp Gly Gly Arg Asn Trp Glu Gly Phe Ser Pro
 385 390 395 400
 Asp Pro Ala Leu Thr Gly Val Leu Phe Ala Glu Thr Ile Lys Gly Ile
 405 410 415
 Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr Ile Met Asn Glu
 420 425 430
 Gln Glu His Phe Arg Gln Gln Pro Glu Ala Ala Gly Tyr Gly Phe Asn
 435 440 445
 Val Ser Asp Ser Leu Ser Ser Asn Val Asp Asp Lys Thr Met His Glu
 450 455 460
 Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ala
 465 470 475 480
 Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Glu Asn

ES 2 538 360 T3

				485					490							495
Ser	Glu	Thr	Leu	Asn	Lys	Leu	Leu	Lys	Ala	Glu	Leu	Gly	Phe	Gln	Gly	
			500					505					510			
Phe	Val	Met	Ser	Asp	Trp	Thr	Ala	His	His	Ser	Gly	Val	Gly	Ala	Ala	
		515					520					525				
Leu	Ala	Gly	Leu	Asp	Met	Ser	Met	Pro	Gly	Asp	Val	Thr	Phe	Asp	Ser	
	530					535					540					
Gly	Thr	Ser	Phe	Trp	Gly	Ala	Asn	Leu	Thr	Val	Gly	Val	Leu	Asn	Gly	
545					550					555					560	
Thr	Ile	Pro	Gln	Trp	Arg	Val	Asp	Asp	Met	Ala	Val	Arg	Ile	Met	Ala	
				565					570					575		
Ala	Tyr	Tyr	Lys	Val	Gly	Arg	Asp	Thr	Lys	Tyr	Thr	Pro	Pro	Asn	Phe	
			580					585					590			
Ser	Ser	Trp	Thr	Arg	Asp	Glu	Tyr	Gly	Phe	Ala	His	Asn	His	Val	Ser	
		595					600					605				
Glu	Gly	Ala	Tyr	Glu	Arg	Val	Asn	Glu	Phe	Val	Asp	Val	Gln	Arg	Asp	
	610					615					620					
His	Ala	Asp	Leu	Ile	Arg	Arg	Ile	Gly	Ala	Gln	Ser	Thr	Val	Leu	Leu	
625					630					635					640	
Lys	Asn	Lys	Gly	Ala	Leu	Pro	Leu	Ser	Arg	Lys	Glu	Lys	Leu	Val	Ala	
				645					650					655		
Leu	Leu	Gly	Glu	Asp	Ala	Gly	Ser	Asn	Ser	Trp	Gly	Ala	Asn	Gly	Cys	
			660					665					670			
Asp	Asp	Arg	Gly	Cys	Asp	Asn	Gly	Thr	Leu	Ala	Met	Ala	Trp	Gly	Ser	
		675					680					685				
Gly	Thr	Ala	Asn	Phe	Pro	Tyr	Leu	Val	Thr	Pro	Glu	Gln	Ala	Ile	Gln	
	690					695					700					
Asn	Glu	Val	Leu	Gln	Gly	Arg	Gly	Asn	Val	Phe	Ala	Val	Thr	Asp	Ser	
705					710					715					720	
Trp	Ala	Leu	Asp	Lys	Ile	Ala	Ala	Ala	Ala	Arg	Gln	Ala	Ser	Val	Ser	
				725						730					735	

ES 2 538 360 T3

Leu Val Phe Val Asn Ser Asp Ser Gly Glu Gly Tyr Leu Ser Val Asp
 740 745 750

Gly Asn Glu Gly Asp Arg Asn Asn Ile Thr Leu Trp Lys Asn Gly Asp
 755 760 765

Asn Val Val Lys Thr Ala Ala Asn Asn Cys Asn Asn Thr Val Val Ile
 770 775 780

Ile His Ser Val Gly Pro Val Leu Ile Asp Glu Trp Tyr Asp His Pro
 785 790 795 800

Asn Val Thr Gly Ile Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly
 805 810 815

Asn Ser Ile Ala Asp Val Leu Tyr Gly Arg Val Asn Pro Gly Ala Lys
 820 825 830

Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr Gly Ser Pro Leu
 835 840 845

Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln Ser Asp Phe Thr
 850 855 860

Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys Phe Asn Glu Thr
 865 870 875 880

Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu Leu
 885 890 895

Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg Tyr Thr Pro Thr
 900 905 910

Ser Gly Met Thr Glu Ala Ala Lys Asn Phe Gly Glu Ile Gly Asp Ala
 915 920 925

Ser Glu Tyr Val Tyr Pro Glu Gly Leu Glu Arg Ile His Glu Phe Ile
 930 935 940

Tyr Pro Trp Ile Asn Ser Thr Asp Leu Lys Ala Ser Ser Asp Asp Ser
 945 950 955 960

Asn Tyr Gly Trp Glu Asp Ser Lys Tyr Ile Pro Glu Gly Ala Thr Asp
 965 970 975

Gly Ser Ala Gln Pro Arg Leu Pro Ala Ser Gly Gly Ala Gly Gly Asn
 980 985 990

Pro Gly Leu Tyr Glu Asp Leu Phe Arg Val Ser Val Lys Val Lys Asn
 995 1000 1005

Thr Gly Asn Val Ala Gly Asp Glu Val Pro Gln Leu Tyr Val Ser
 1010 1015 1020

Leu Gly Gly Pro Asn Glu Pro Lys Val Val Leu Arg Lys Phe Glu
 1025 1030 1035

Arg Ile His Leu Ala Pro Ser Gln Glu Ala Val Trp Thr Thr Thr
 1040 1045 1050

Leu Thr Arg Arg Asp Leu Ala Asn Trp Asp Val Ser Ala Gln Asp
 1055 1060 1065

Trp Thr Val Thr Pro Tyr Pro Lys Thr Ile Tyr Val Gly Asn Ser
 1070 1075 1080

Ser Arg Lys Leu Pro Leu Gln Ala Ser Leu Pro Lys Ala Gln
 1085 1090 1095

<210> 75
 <211> 3294
 <212> ADN
 <213> Aspergillus oryzae

5

<400> 75

ES 2 538 360 T3

atgcgttctt	ccccctcct	cegetccgcc	gttgtggccg	ccctgccggt	gttggccctt	60
gccgctgatg	gcaggtccac	ccgctactgg	gactgctgca	agccttcgtg	cggctgggcc	120
aagaaggctc	ccgtgaacca	gcctgtcttt	tcctgcaacg	ccaacttcca	gcgtatcacg	180
gacttcgacg	ccaagtccgg	ctgcgagccg	ggcgggtgctg	cctactcgtg	cgccgaccag	240
accccatggg	ctgtgaacga	cgacttcgcg	ctcggttttg	ctgccacctc	tattgcccgc	300
agcaatgagg	cgggctgggtg	ctgcccctgc	tacgagctca	ccttcacatc	cggctcctggt	360
gctggcaaga	agatggtcgt	ccagtccacc	agcactggcg	gtgatcttgg	cagcaaccac	420
ttcgatctca	acatccccgg	cggcggcgtc	ggcatcttcg	acggatgcac	tccccagttc	480
ggtggtctgc	ccggccagcg	ctacggcggc	atctcgtccc	gcaacgagtg	cgatcggttc	540
cccgaacgcc	tcaagcccgg	ctgctactgg	cgcttcgact	ggttcaagaa	cgccgacaat	600
ccgagcttca	gcttccgtca	ggtccagtgc	ccagccgagc	tcgtcgtctg	caccggatgc	660
cgccgcaacg	acgacggcaa	cttcctgcc	gtccagatcc	ccatgcggtc	ctccccctc	720
ctccgctccg	ccgttgtggc	cgccctgccg	gtggtggccc	ttgccaagga	tgatctcgcg	780
tactcccctc	ctttctacct	ttccccatgg	gcagatggtc	agggatgaatg	ggcgggaagta	840

ES 2 538 360 T3

tacaaaocgcg ctgtagacat agtttcccag atgacgttga cagagaaagt caacttaacg 900
actggaacag gatggcaact agagaggtgt gttggacaaa ctggcagtggt tcccagactc 960
aacatcccca gcttgtgttt gcaggatagt cctcttggtta ttcgtttctc ggactacaat 1020
tcagctttcc ctgcggtgt taatgtcgtt gccacctggg acaagacgct cgcctacctt 1080
cgtggtcagg caatgggtga ggagttcagt gataagggta ttgacgttca gctgggtcct 1140
gctgctggcc ctctcgggtgc tcatccggat ggcggtagaa actgggaaag tttctcacca 1200
gatccagccc tcaccggtgt actttttgcg gagacgatta agggatttca agatgctggt 1260
gtcattgoga cagctaagca ttatatcatg aacgaacaag agcatttccg ccaacaaccc 1320
gaggctgcgg gttacggatt caacgtaagc gacagtttga gttccaacgt tgatgacaag 1380
actatgcatg aattgtacct ctggcccttc gcggtatgcag tacgctgctg agtcggtgct 1440
gttatgtgct cttacaacca aatcaacaac agctacggtt gcgagaatag cgaaactctg 1500
aacaagcttt tgaaggcggg gcttggtttc caaggcttcg tcatgagtga ttggaccgct 1560
caacacagcg gcgtaggcgc tgctttagca ggtctggata tgtcgtatgcc cggatgatgt 1620
accttcgata gtggtacgtc tttctggggg gcaaacttga cggtcggtgt ccttaacggt 1680
acaatcccc aatggcgtgt tgatgacatg gctgtccgta tcatggccgc ttattacaag 1740
gttggccgcg acaccaaata caccctccc aacttcagct cgtggaccag ggacgaatat 1800
ggtttcgocg ataacctatg ttcggaaggt gcttacgaga gggtaaacga attcgtggac 1860
gtgcaacgcg atcatgccga cctaaccgt cgcacgagcg cgcagagcac tgttctgctg 1920
aagaacaagg gtgccttgcc cttgagccgc aaggaaaagc tggtcgccct tctgggagag 1980
gatgcggggt ccaactcgtg gggcgctaac ggctgtgatg accgtggttg cgataacggt 2040
acccttgcca tggcctgggg tagcgggtact gcgaatttcc catacctcgt gacaccagag 2100
caggcgattc agaacgaagt tcttcagggc cgtggtaatg tcttcgccgt gaccgacagt 2160
tgggcgctcg acaagatcgc tgcggctgcc cgccaggcca gcgtatctct cgtgttcgtc 2220
aactccgact caggagaagg ctatcttagt gtggatggaa atgagggcga tcgtaacaac 2280
atcactctgt ggaagaacgg cgacaatgtg gtcaagaccg cagcgaataa ctgtaacaac 2340
accgttgtca tcatccactc cgtcggacca gttttgatcg atgaatggtg tgaccacccc 2400
aatgtcactg gtattctctg ggctggtctg ccaggccagg agtctggtta ctccattgcc 2460
gatgtgctgt acggtcgtgt caaccctggc gccaagtctc ctttcaactg gggcaagacc 2520
cgggagtcgt atggttctcc cttggtcaag gatgccaca atggcaacgg agcgcgccag 2580
tctgatttca cccaggggtgt tttcatcgat taccgccatt tcgataagtt caatgagacc 2640
cctatctacg agtttggtta cggcttgagc tacaccacct tcgagctctc cgacctccat 2700
gttcagcccc tgaacgcgtc ccgatacact cccaccagtg gcatgactga agctgcaaag 2760

ES 2 538 360 T3

aactttggtg aaattggcga tgcgtcggag tacgtgtatc cggaggggct ggaaaggatc 2820
catgagttta tctatccctg gatcaactct accgacctga aggcatcgtc tgacgattct 2880
aactacggct ggggaagactc caagtatatt cccgaaggcg ccacggatgg gtctgcccag 2940
ccccgtttgc cgcctagtgg tggtgccgga ggaaaccccc gtctgtacga ggatcttttc 3000
cgcgtctctg tgaaggtcaa gaacacgggc aatgtcgccg gtgatgaagt tcctcagctg 3060
tacgtttccc taggcggccc gaatgagccc aaggtggtac tgcgcaagtt tgagcgtatt 3120
cacttgcccc cttcgcagga ggccgtgtgg acaacgacc ttaccctcgc tgaccttgca 3180
aactgggacg tttcggctca ggactggacc gtcactcctt accccaagac gatctacgtt 3240
ggaaactcct cacggaaact gccgctccag gcctcgtcgc ctaaggcca gtaa 3294

5 <210> 76
<211> 1097
<212> PRT
<213> Aspergillus oryzae

10 <400> 76

Met Arg Ser Ser Pro Leu Leu Arg Ser Ala Val Val Ala Ala Leu Pro
 1 5 10 15

Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
 20 25 30

Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
 35 40 45

Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
 50 55 60

Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
 65 70 75 80

Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
 85 90 95

Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
 100 105 110

Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
 115 120 125

Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
 130 135 140

Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
145 150 155 160

Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
165 170 175

Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
180 185 190

Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
195 200 205

Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
210 215 220

Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Met Arg Ser Ser Pro Leu
225 230 235 240

Leu Arg Ser Ala Val Val Ala Ala Leu Pro Val Leu Ala Leu Ala Lys
245 250 255

Asp Asp Leu Ala Tyr Ser Pro Pro Phe Tyr Pro Ser Pro Trp Ala Asp
260 265 270

Gly Gln Gly Glu Trp Ala Glu Val Tyr Lys Arg Ala Val Asp Ile Val
275 280 285

Ser Gln Met Thr Leu Thr Glu Lys Val Asn Leu Thr Thr Gly Thr Gly
290 295 300

Trp Gln Leu Glu Arg Cys Val Gly Gln Thr Gly Ser Val Pro Arg Leu
305 310 315 320

Asn Ile Pro Ser Leu Cys Leu Gln Asp Ser Pro Leu Gly Ile Arg Phe
325 330 335

Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly Val Asn Val Ala Ala Thr
340 345 350

Trp Asp Lys Thr Leu Ala Tyr Leu Arg Gly Gln Ala Met Gly Glu Glu
355 360 365

Phe Ser Asp Lys Gly Ile Asp Val Gln Leu Gly Pro Ala Ala Gly Pro
370 375 380

Leu Gly Ala His Pro Asp Gly Gly Arg Asn Trp Glu Ser Phe Ser Pro
385 390 395 400

ES 2 538 360 T3

Asp Pro Ala Leu Thr Gly Val Leu Phe Ala Glu Thr Ile Lys Gly Ile
 405 410 415

Gln Asp Ala Gly Val Ile Ala Thr Ala Lys His Tyr Ile Met Asn Glu
 420 425 430

Gln Glu His Phe Arg Gln Gln Pro Glu Ala Ala Gly Tyr Gly Phe Asn
 435 440 445

Val Ser Asp Ser Leu Ser Ser Asn Val Asp Asp Lys Thr Met His Glu
 450 455 460

Leu Tyr Leu Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ala
 465 470 475 480

Val Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Glu Asn
 485 490 495

Ser Glu Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu Gly Phe Gln Gly
 500 505 510

Phe Val Met Ser Asp Trp Thr Ala Gln His Ser Gly Val Gly Ala Ala
 515 520 525

Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Val Thr Phe Asp Ser
 530 535 540

Gly Thr Ser Phe Trp Gly Ala Asn Leu Thr Val Gly Val Leu Asn Gly
 545 550 555 560

Thr Ile Pro Gln Trp Arg Val Asp Asp Met Ala Val Arg Ile Met Ala
 565 570 575

Ala Tyr Tyr Lys Val Gly Arg Asp Thr Lys Tyr Thr Pro Pro Asn Phe
 580 585 590

Ser Ser Trp Thr Arg Asp Glu Tyr Gly Phe Ala His Asn His Val Ser
 595 600 605

Glu Gly Ala Tyr Glu Arg Val Asn Glu Phe Val Asp Val Gln Arg Asp
 610 615 620

His Ala Asp Leu Ile Arg Arg Ile Gly Ala Gln Ser Thr Val Leu Leu
 625 630 635 640

Lys Asn Lys Gly Ala Leu Pro Leu Ser Arg Lys Glu Lys Leu Val Ala
 645 650 655

Leu Leu Gly Glu Asp Ala Gly Ser Asn Ser Trp Gly Ala Asn Gly Cys
 660 665 670
 Asp Asp Arg Gly Cys Asp Asn Gly Thr Leu Ala Met Ala Trp Gly Ser
 675 680 685
 Gly Thr Ala Asn Phe Pro Tyr Leu Val Thr Pro Glu Gln Ala Ile Gln
 690 695 700
 Asn Glu Val Leu Gln Gly Arg Gly Asn Val Phe Ala Val Thr Asp Ser
 705 710 715 720
 Trp Ala Leu Asp Lys Ile Ala Ala Ala Ala Arg Gln Ala Ser Val Ser
 725 730 735
 Leu Val Phe Val Asn Ser Asp Ser Gly Glu Gly Tyr Leu Ser Val Asp
 740 745 750
 Gly Asn Glu Gly Asp Arg Asn Asn Ile Thr Leu Trp Lys Asn Gly Asp
 755 760 765
 Asn Val Val Lys Thr Ala Ala Asn Asn Cys Asn Asn Thr Val Val Ile
 770 775 780
 Ile His Ser Val Gly Pro Val Leu Ile Asp Glu Trp Tyr Asp His Pro
 785 790 795 800
 Asn Val Thr Gly Ile Leu Trp Ala Gly Leu Pro Gly Gln Glu Ser Gly
 805 810 815
 Asn Ser Ile Ala Asp Val Leu Tyr Gly Arg Val Asn Pro Gly Ala Lys
 820 825 830
 Ser Pro Phe Thr Trp Gly Lys Thr Arg Glu Ser Tyr Gly Ser Pro Leu
 835 840 845
 Val Lys Asp Ala Asn Asn Gly Asn Gly Ala Pro Gln Ser Asp Phe Thr
 850 855 860
 Gln Gly Val Phe Ile Asp Tyr Arg His Phe Asp Lys Phe Asn Glu Thr
 865 870 875 880
 Pro Ile Tyr Glu Phe Gly Tyr Gly Leu Ser Tyr Thr Thr Phe Glu Leu
 885 890 895
 Ser Asp Leu His Val Gln Pro Leu Asn Ala Ser Arg Tyr Thr Pro Thr

REIVINDICACIONES

1. Método para producir una enzima de beta-glucosidasa, que comprende:

5 (a) transformación de una célula huésped fúngica con un constructo de proteína de fusión que codifica una proteína de fusión, donde el constructo de proteína de fusión comprende:

(i) un primer polinucleótido que comprende una secuencia de nucleótidos que codifica un péptido señal obtenido a partir de un gen de endoglucanasa (cel45) de *Humicola insolens* CEL45;

10 (ii) un segundo polinucleótido que comprende una secuencia de nucleótidos que codifica al menos un dominio catalítico de una endoglucanasa obtenida a partir de un gen de endoglucanasa (cel45) de *Humicola insolens* CEL45; y

(iii) un tercer polinucleótido que comprende una secuencia de nucleótidos que codifica una enzima de beta-glucosidasa obtenida a partir de *Aspergillus oryzae*,

15 donde la proteína de fusión es SEC ID n°: 74,

(b) cultivo de la célula huésped fúngica transformada bajo condiciones adecuadas para la producción de la proteína de fusión; y

20 (c) recuperación de la proteína de fusión del medio de cultivo.

2. Célula huésped fúngica transformada con una proteína de fusión que comprende el polinucleótido según la reivindicación 1.

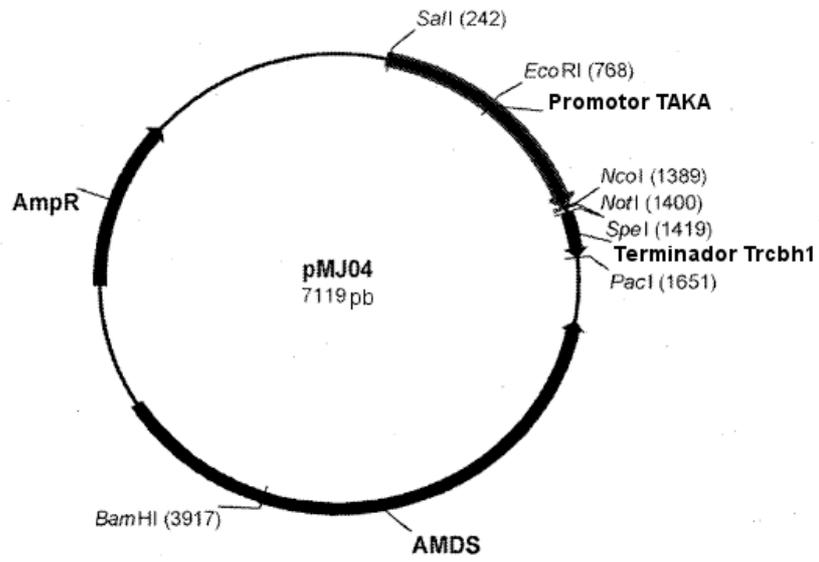


Fig. 1

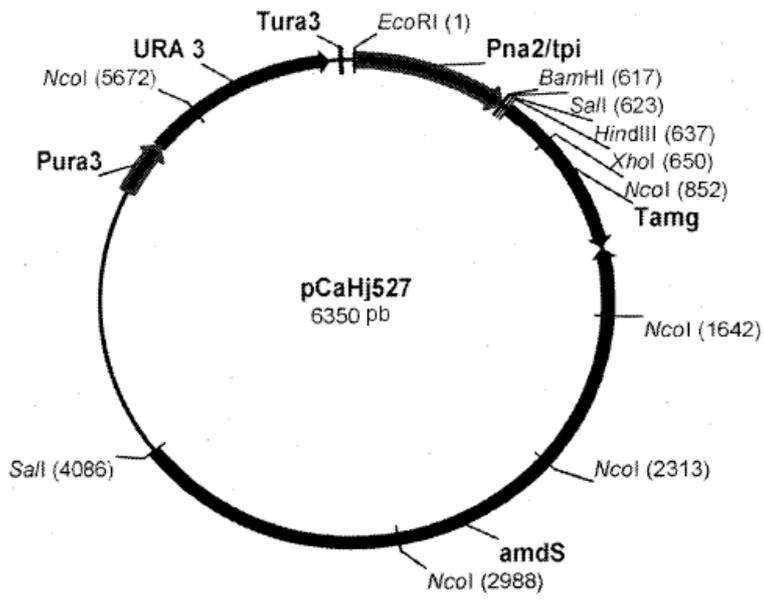


Fig. 2

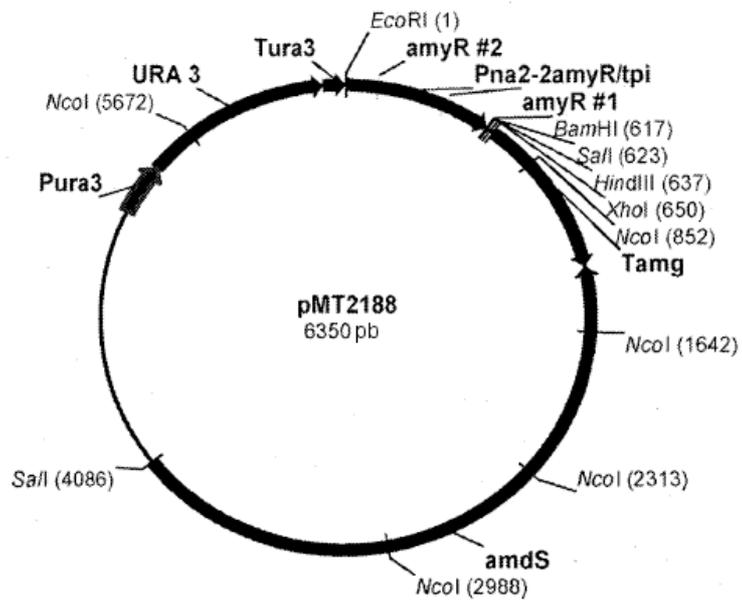


Fig. 3

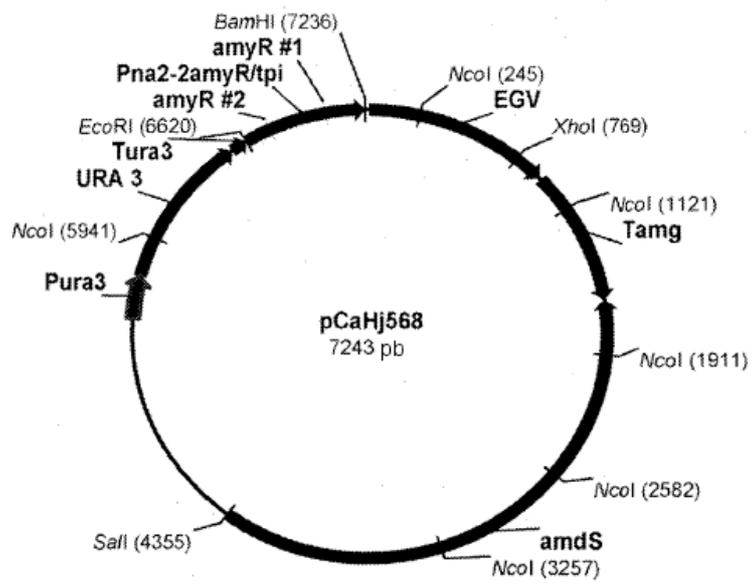


Fig. 4

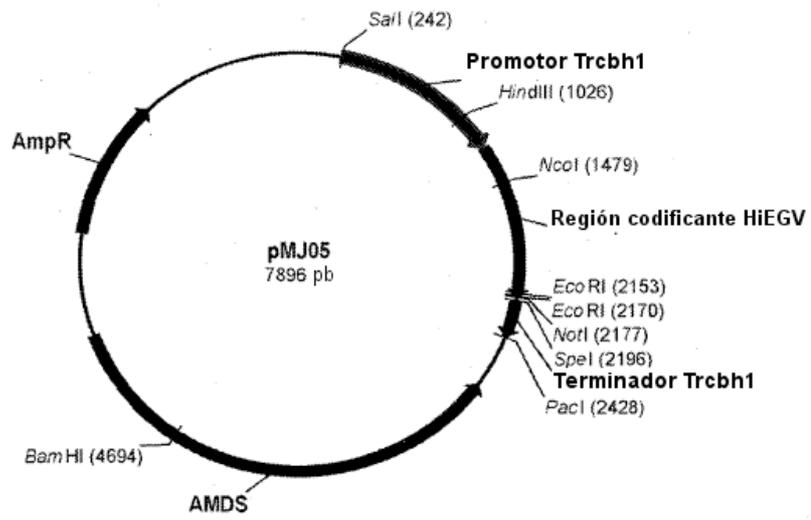


Fig. 5

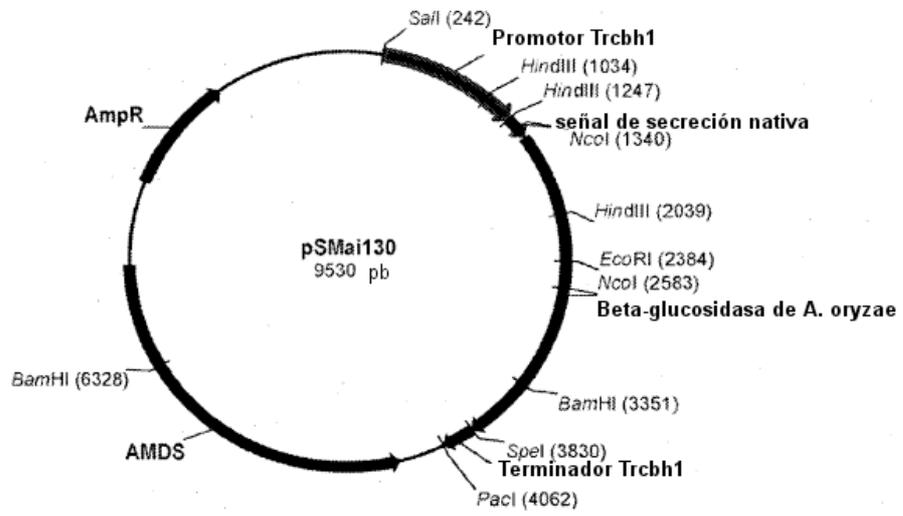


Fig. 6

ATG AAG CTT GGT TGG ATC GAG GAG GAG GAG GGT GCG TCA GAA GTC AAT GCC
M K L C W I K Y A A E A A S Y V S A

Fig. 7

ATG CTT TCC TCC CCC CTC CTC CAC TCC GCC GTT GAG GGC GGC CTT GCG CCG GCG TTT GCG CTT GCC
M R E S P L L L R S A V V A A L P Y L A L A

Fig. 8

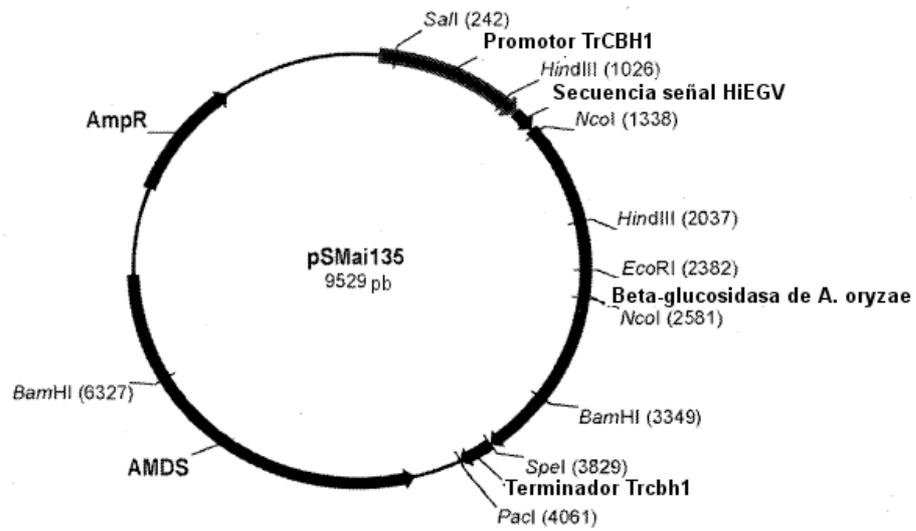


Fig. 9

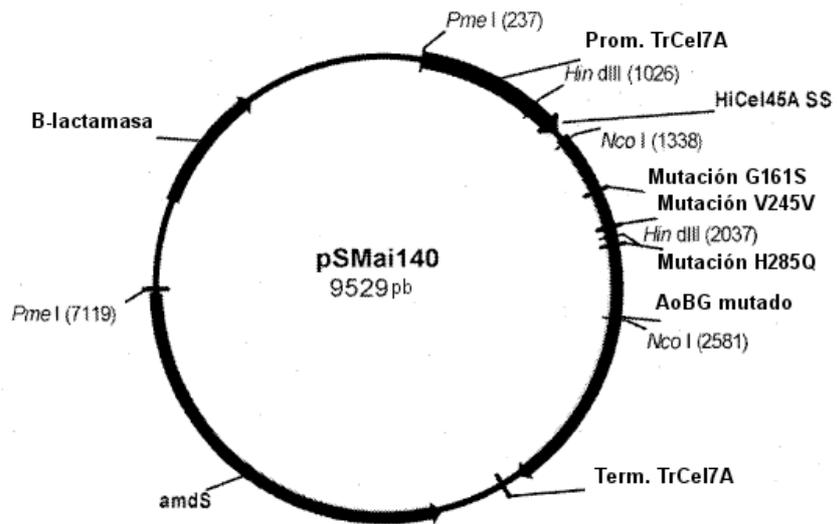


Fig. 10

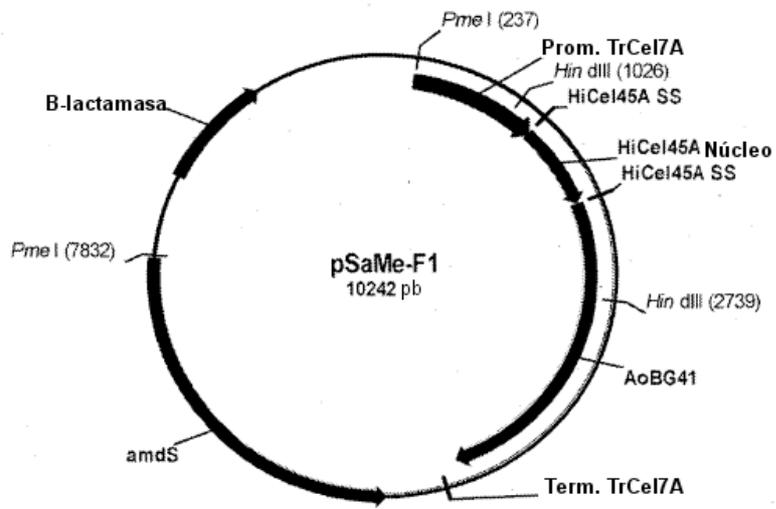


Fig. 11

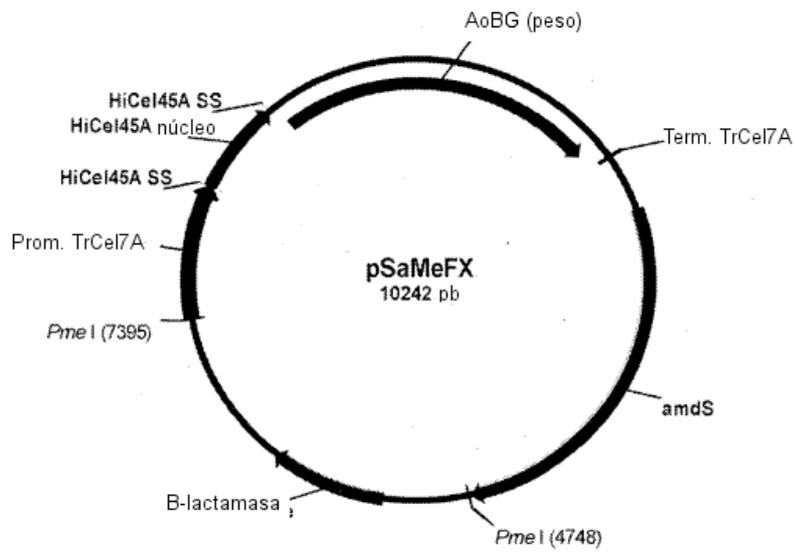


Fig. 12

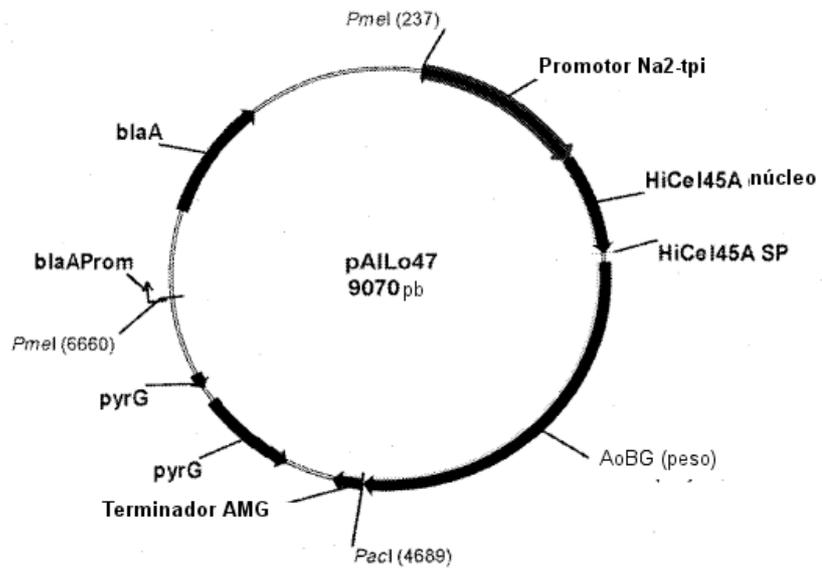


Fig. 13

M R S H P L L S

1201 ATGCGTTCCTCCGCGCTTCCTCCGCTC
 TAAGCGAGGGAGCGCGGACGACGCGGAG

· A V V A A L P V L A L A B D S R S T R V

1261 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · W D C C K P S C G M A K K A P V N Q P V

1321 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · F S C N A N E Q R I T D F D A K S G C E

1381 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · F G G V A Y E C A D Q T P W A V N D D F

1441 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · A L Q F A A T S I A G E N S A G W C C A

1501 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · C Y K L T P T S G P V A G Y K M V Y Q S

1561 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · T S I E G E L G S N R F D L N Y P G G Q

1621 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · V S I F D G C T P Q P G G L P G Q P Y Q

1681 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · G I S S R N E C D R F P D A L K S G C Y

1741 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · W R F D W P K N A D N P S F S P R Q V Q

1801 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · C P R E L V A R T G C R R W D G M F P

1861 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · A V Q I F M R S S P L L N S A V V A L

1921 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · F V L R L A K D D L A Y S P P F Y P S P

1981 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · W A D Q Q G E N A E V Y K R A V D I V S

2041 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · Q K T L T E K V N L T F G T G W Q L R R

2101 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 GCGCGTTCCTCCGCGCTTCCTCCGCTC
 · C V G Q T G S V P R L N Y P S L C L Q D

Fig. 14A


```

4141 TATTCATGAGACGGCCACGGATGCTGTTGCCCAGCCCGCTTTCGGCCCTAGTGGTGGTTC
AATAGGGCTTNDGCGGTGCCATACCCAGAGCGGCTCGGGGGGAAAGCCCGCATCACCAACCGS
. G S N P G L Y E D L F R V S V E V E N T
4201 GCGAGGAAACXXXGCTCTGTACGAAGATCTTTCCGCGTTCCTGTGAAGGTCAGAAACAC
. G N V A G D E V P Q L V V S L G G P N E
4261 GGGCAATGTCCCGGNAATGAGATTCTCTCAGTGTATAGTTCCTAGCGCGCCCGCATGA
CCGTTACAGCGCGCACTACTTCAAGAGATCGGCAATGCAAGAGCATCCGCGCGGCTTACT
. P K V V L R H F E R I H L A P S Q D A V
4321 GCGCAGGTTGCTACTCGCGAGTTTGGAGCGTATTCACITGACCCCTTCCAGGAGGCGCT
CGGCTTCACCGCTGACCGCTTCGCACTCGCATAGTGAAGCGCGGAGCGCTCCCTCCCGCA
. W F T T L E R R E L A N W D V S A Q D W
4381 GTGACAAAGAACCTTACCGCTCGTACGCTTGCAAAAGCTGAAAGCTTTCCGCTCAGGACTG
CACTGTGCTGGGAATGGCAAGCACTGGAAGCTTTGACCCCTGCAAGCCGAGTCCCTGAC
. F V T P Y P K T Y V G N S S R E L P L
4441 GACCGTCACTGCTTACCCAGACAGTCTACTTTGAAAGTCTCCGCGGAGACCTGGCGCT
CTGCGAGGAGGAAATGGGTTCTGCTGAGNAGACCTTTGAGGAGTGCCTTTGACCGGAA
. Q R S L P K A Q *
4501 CCAGCGCTCGCTGCTTAAAGCCCGATAA
GGTCCCGGAGCGAGCGATTCXGGGTCATT

```

Fig. 14D

M R S R P L L R S

1291 ATGCGTTCCCTCCGCTCTCTCCCTC
 TACCGAAGGAGGGGGAGGAGGCGAG
 · A V V A A L P V L A L A D G R S T R Y
 1291 CCGCTGTGTGCGCCGCTGCGGCTGTGCGCCCTGCGCGCTGATVACRGGTCCRCCGCTA
 GCGCAACACCGCGCGGACCGCCACRACCGGAAAGGAGGATACCGTCCAGGTGCGCGAT
 · M D C C K P S C G W A K K A P V N Q P V
 1321 CTTGGGCTGCTGTGAGGCTTTCGTGGGCTGCGCCRAGAGGCTTCCCGTAA CCGAGCTGT
 TAGCCTGACGACCTTTCGAAACACCGCCGACCGCGCTTCTTCCGAGGCGACTTCCGTGGGAC
 · F S C N A N P Q R I T E F D A E B O C E
 1381 CTTTCCCTGCAAGCGCACTTCCAGGCTATCACGCACTTCCGACCGCAAGTCCCGGCTGCA
 GAAAAGCACTTTCGCTTTCAGGCTGCGATACGCGCTGAGACTCCGCTTCCGCTTCCGCGACGCT
 · P G G V A Y S C A D Q T P W A V M D D F
 1441 GCGCGCGCTGTGCGCTACTGCTGGTGGCGGACCGGACCGCAATGCGCTGCTGAGAGGAGCTT
 CCGCGCGCAACAGCGGATGCGCACCGCGCTGCTGCTGGGCTACCGACACTTCCGTGCGAA
 · A L G F A T S I A G S H E A G N C R
 1501 CCGCGCTGCGCTTTCGCTGCGCTCTATTCGCGCGAGGCAATGAGGCGCGCTGCTGCTGCGC
 GCGCGGCGCAAAAGGAGGCTGAGGATACCGCGCTGCTTACTTCCGCGCGGACCGCAAGCGC
 · C Y E L E F T S P V A G K K M V V Q S
 1561 CTTCCAGAGCTGCGCTTCAAGTCCGCTTCGCTGCTGCGGAGAGGATGCGCTGCGCTGCGCT
 GAGCAAGCTCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT
 · T S T G G D L G S N H F D L N I P G G G
 1621 CACCGCGCTGCGCGGCTTTCGCGAGCAACGACTTCCGATTCGACGCTGCGCGCGCGCGCG
 CTGCTGCTGCGCGCGCTGAGGAGCGCTTTCGCTGAGGCTGAGGCTTTCGCGCGCGCGCGCG
 · V Q I P D Q C Y P Q F G G L P Q Q R Y Q
 1681 CCTGCGCTTTCGCGAGGATGCGCTGCGCGCTTCCGCTGCTGCGCGCGCGCGCGCGCGCGCG
 GCGCGCTGAGGAGGCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT
 · G I S R R N E D E F P D A L K P R C Y
 1741 GCGCATCTGCTGCGCAAGCGCTGCGCTGCGCTTCCGCGCGCGCTGCGCGCGCGCGCGCGCGCG
 GCGCTGAGGAGGCTGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
 · W R F D W F K N A D N P S P S P R Q V Q
 1801 CTGCGCTTTCGAGCTGCTTTCGAGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
 GCGCGCGGAGGCTGAGGAGGCTTCTTCCGCGCTGCTGAGGCTGCGCGCGCGCGCGCGCGCGCGCG
 · C P A H L V A R E T G C R R N D D S H F P
 1861 GTGCGCGCGGAGCTGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
 CCGCGCTGCGCTGCG
 · K V Q I P M R S S P L L R S A V V A K L
 1921 TGCGCTGCGAGTCCCGGATGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
 ACGCGCGCTTTCG
 · P V L A L A K N D L A Y S P P P Y P S P
 1981 GCGCGCTTTCG
 CCG
 · H A D E Q G S E W A E V Y K R A V D I U S
 2041 ATGCGCGAGTGTGAGGCTGAAATGCGCGGAGGATACGAAACGCGCTGAGCAAGGCTTTC
 TAGCGCTTACCGCTCCGACTTACCGCGCTTCCGATGCTTTCGCGCGCGCGCGCGCGCGCGCGCG
 · Q M I L T E K V N L T T G T G W Q L E R

Fig. 15A

4021 GGACTACGCTGTATCCGGAGGCGCTTGGAAAGGATCCATACAGCTTATCTATCCCTGGATCAW
 CTTCAATGACATAGGCTTCCXCCGACCTTTCCTAGGTACCTCAATAGATAGCGGACCTAGCTT
 · S T D L K A S S D D S N Y G W E D S K Y ·
 4081 CTCTACCCGACCTGAGGSCATGCTCTGACGCTTCTAAGCTACGCTCCGAGAGCTCCAGTA
 GAGATGGCTGGACTTCCGTAGGAGGCTGCTAAGATTGATTCGCGCCCTTCTGAAGTCTCT
 · I F E G A T D G S A Q P R L P A S G A ·
 4141 TATTCCGAGAGCCGACCGATGGCTTGGCCGAGCCCTTTCGCGCTTGGTGGCTGCTC
 ATAAGGCTTTCGCGCTTCCCTACCCGACCGCTTCCGGGCGAACAAGGCGATCCGCGCCGCG
 · G G N P G L Y D D L F R V S V K V K N T ·
 4201 CCGAGGAACCGGCTCTGTACGAGATCCTTTCGCGCTTCTGTAAGGCTTAGAAGAC
 GCTCTCTTTCGCGCCAGGCACTGCTCCTAGAAAGCCGACAGGACTTCCGCTTCTTCTG
 · G N V A G R E V P Q L Y V S L G G P N E ·
 4261 GGCGAATGTCGCGCTGATGGAATTCCTCGGCTTTCCTTAGCCGCGCCGAGTA
 CCGCTTACAGGCGCCACTTCTCAGGAGCTCCGATGCGAAGGATTCGCGCGCTTACT
 · P K V V L R K K E R R I H L A P S Q R A V ·
 4321 GCGGAGCTTGGTCCGCGGAGCTTTCGAGGTAATTCGCTTCCGCGCTTCTGAGGAGGCTT
 CCGCTTCCGACATGACGCTTCAACTCCGATAGGAGGCGCGGAGGCTTCTGCGCA
 · W T T F L T R R D L A N D U S A Q D W ·
 4381 GTGGACAGGACCGCTTACCGCTGCTGACCTTCCAAACTGGGACCTTTCGCGCTCAGGACTG
 CAGCTGCTTCCGCGGATGGGAGGACTTGGAGGCTTTCGCGCTGCGAGGCGGCTTCTGAC
 · T V T P Y P K T I Y V G N S S P K L P L ·
 4441 GACCTGCTTCTTACGCGAAGATCTACTGGAAGCTTCTGCGGAACTGCTGCGGAACTGCGCT
 CTGCGCACTAGGAGTGGGCTTCTGCTAGATGCGACTTTCGCGGAGTCTTTCGCGGAA
 · Q A S L P K A Q ·
 4501 CCGGCTTCTGCGCTTACGCGGAGTAA
 GCTGCGGAGCGGAGGATTCGCGGCTT

Fig. 15D