

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 815 628**

51 Int. Cl.:

C07K 14/37	(2006.01)
C12P 21/02	(2006.01)
C12N 9/02	(2006.01)
C12N 9/24	(2006.01)
C12N 9/42	(2006.01)
C12N 9/48	(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **19.02.2016 PCT/FI2016/050108**

87 Fecha y número de publicación internacional: **25.08.2016 WO16132021**

96 Fecha de presentación y número de la solicitud europea: **19.02.2016 E 16708433 (4)**

97 Fecha y número de publicación de la concesión europea: **01.07.2020 EP 3259364**

54 Título: **Métodos para controlar la producción de proteasas**

30 Prioridad:

20.02.2015 FI 20155112

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

30.03.2021

73 Titular/es:

**ROAL OY (100.0%)
Tykkimäentie 15b
05200 Rajamäki, FI**

72 Inventor/es:

**PALOHEIMO, MARJA;
MÄKINEN, SUSANNA;
PUNT, PETER;
JUNTUNEN, KARI;
PURANEN, TERHI y
VEHMAANPERÄ, JARI**

74 Agente/Representante:

VALLEJO LÓPEZ, Juan Pedro

ES 2 815 628 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 controlar la producción de proteasas

5 **Campo de la invención**

La presente descripción se refiere al campo de la producción de proteasas. Más específicamente, desvela un regulador novedoso de la expresión de proteasas y su uso en la producción de proteínas de interés en células hospedadoras.

10 **Antecedentes**

Se usan ampliamente microorganismos, tales como hongos y hongos filamentosos, como células hospedadoras para la expresión y la secreción extracelular de proteínas de interés, tales como proteínas recombinantes. Una desventaja que se encuentra frecuentemente con los microorganismos, cuando se usan como células hospedadoras, es su inherente producción y secreción de enzimas proteolíticas que degradan la proteína de interés. Este problema es particularmente difícil cuando se producen proteínas de interés que son sensibles, inestable o ambas cosas. Por tanto, las proteasas endógenas de la célula hospedadora al menos reducen el rendimiento de la proteína de interés e incluso pueden impedir su producción. Adicionalmente, la actividad proteolítica de las proteasas endógenas puede conducir a la formación de proteínas fragmentadas o degradadas, lo que disminuye la calidad de las proteínas producidas en las células hospedadoras. La autenticidad de la proteína puede verse afectada por proteólisis debido al recorte de aminoácidos N y/o C terminales por exopeptidasas. Además, la presencia de proteasas endógenas disminuye la estabilidad y el período de caducidad de composiciones proteínicas cuando las proteasas endógenas están presentes en las composiciones proteínicas. En caso de que se desee un mayor período de caducidad o estabilidad de la composición proteínica, las proteasas endógenas deben retirarse de la composición proteínica o debe inhibirse su actividad proteasa, por ejemplo, mediante inhibidores de proteasas.

Se han previsto diversas soluciones para evitar los problemas anteriores. Por ejemplo, se podrían suprimir o inactivar los genes que codifican las diversas proteasas endógenas, si las proteasas se identifican y caracterizan adecuadamente. El documento WO 90/00192 describe hospedadores hongos filamentosos mutantes que se han vuelto incapaces de secretar una aspártico proteasa enzimáticamente activa. Mediante dicha mutación, se demostró que el rendimiento del polipéptido heterólogo, quimosina bovina, podría aumentarse.

El documento WO2013/102674 describe células fúngicas filamentosas que son deficientes en al menos tres proteasas endógenas, y en donde las proteasas endógenas se inactivan mediante una mutación en los genes que codifican las proteasas endógenas. También se ha hecho intentos de inactivar las proteasas endógenas mediante mutagénesis aleatoria, pero pueden conducir a efectos pleiotrópicos desconocidos y no deseados en el rendimiento de la fermentación, tales como problemas en la expresión génica y una tasa de crecimiento escasa de la célula hospedadora. La mutagénesis aleatoria produce mutaciones no específicamente en todo el genoma de la célula hospedadora. Los genes mutados que producen características deseables o no deseables para la célula hospedadora no pueden identificarse fácilmente. Las cepas mutantes resultantes deben usarse tal cual, aunque algunas de las mutaciones puedan conducir a resultados no deseados en cuanto a las características de la cepa y/o sus productos.

Otro enfoque para prevenir los problemas de las proteasas endógenas ha sido optimizar las materias primas y las condiciones de cultivo de manera que se redujese o impidiese la producción de proteasas endógenas.

Sin embargo, es bien sabido que los hongos producen un gran número de proteasas endógenas. Por tanto, la adaptación de cepas mediante inactivación individual de cada proteasa endógena es poco práctica. Además, se ha demostrado que la disrupción de un gen de proteasa puede conducir a un aumento compensatorio en la expresión y la producción de otro gen o genes de proteínica. En consecuencia, existe un interés en desarrollar para su uso industrial cepas de hongos filamentosos que no presenten, o que presenten niveles muy bajos, de actividad proteolítica originada en proteasas endógenas. Además, sería ventajoso proporcionar métodos que permitan prevenir la producción de proteasas endógenas en células hospedadoras. En particular, *Trichoderma reesei* con baja actividad proteasa endógena sería particularmente deseable porque es una célula hospedadora adecuada para muchas proteínas recombinantes.

Algunas enzimas son excepcionalmente sensibles incluso a cantidades bajas de proteasas y pueden necesitar modificaciones adicionales para permanecer estables en productos tales como en composiciones enzimáticas. Por ejemplo, muchas proteínas que tienen una estructura multidominio en donde los dominios están unidos por regiones enlazadoras flexibles, tales como celulasas con un resto de unión a celulosa, puede ser particularmente susceptible a la escisión por proteasas. En consecuencia, dichas enzimas pueden ser difíciles de convertir en productos con un período de caducidad aceptable y con frecuencia requieren una modificación por ingeniería genética cuidadosa de la secuencia de unión, además del uso de un hospedador de proteasa baja y la optimización de las condiciones de cultivo.

65 Owen P. Ward, *Biotechnology Advances*, vol. 30, N.º 5, 1 de septiembre de 2012, páginas 1119-1139, desvela métodos para producir proteínas recombinantes en hongos filamentosos.

Sharma R. *et al.*, *World J Microbiol Biotechnol*, vol. 25, N.º 12, 2 de agosto de 2009, páginas 2083-2094, desvela métodos para producir proteínas heterólogas en hongos filamentosos, así como métodos para optimizar una producción de este tipo.

5 La entrada de la base de datos Uniprot A0A024S366_HYPJR (9 de julio de 2014) desvela una proteína no caracterizada de *Hypocrea jecorina*.

10 El documento WO2006/040312 A2 desvela activadores transcripcionales fúngicos que tienen actividad transcripcional en un promotor de proteasa.

El documento WO01/68864 A1 desvela un activador transcripcional fúngico de *Aspergillus niger*.

Sumario

15 Es un objetivo resolver, al menos parcialmente, los problemas anteriores de la técnica anterior. Un objetivo relacionado es mejorar la producción de proteínas, especialmente dichas proteínas que son sensibles a las proteasas del hospedador o son inestables cuando se producen en un sistema de expresión fúngica.

20 También es un objetivo proporcionar un método para regular la expresión de proteasas endógenas en microorganismos.

Otro objetivo es proporcionar un regulador de proteasas, un gen que lo codifique, y un vector que comprenda dicho gen.

25 Es otro objetivo proporcionar una variante de regulador de proteasas, un gen que la codifique, y un vector que comprenda dicho gen.

30 Es otro objetivo proporcionar un método de producción de una proteína de interés en una célula hospedadora.

Otro objetivo más es proporcionar un polinucleótido alternativo y un polipéptido que regule la expresión de proteasas endógenas en una célula hospedadora.

35 Los presentes inventores han descubierto sorprendentemente que la expresión endógena de varias proteasas puede suprimirse en una célula hospedadora inactivando un gen que codifica un regulador de proteasas denominado *pea1* por los presentes inventores. La supresión de las proteasas endógenas impidiendo la acción de *pea1* en una célula hospedadora dio como resultado, por ejemplo, a una mejora del rendimiento y la estabilidad de las proteínas recombinantes producidas en la célula hospedadora.

40 De acuerdo con un primer aspecto se proporciona un polinucleótido modificado que comprende una secuencia de nucleótidos que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado; y en donde el polinucleótido modificado contiene al menos una modificación que da como resultado la incapacidad, al menos parcial, de un producto génico que puede obtenerse mediante transcripción y/o traducción de un gen cromosómico que comprende el polinucleótido modificado para inducir la expresión de proteasas endógenas en una célula hospedadora. Los inventores han demostrado que el polinucleótido modificado del primer aspecto es responsable de producir un producto génico que regula la expresión de muchas proteasas endógenas fúngicas. Por tanto, el gen se denomina en el presente documento un regulador de proteasas, que afecta a la expresión de proteasas 1 (del inglés *protease expression affecting 1*), o *pea1*, y se caracteriza al menos por la presencia de la secuencia que codifica los restos de la región altamente conservada 402-533 del SEQ ID NO: 13. El correspondiente producto génico *pea1* (cuando es un polipéptido) se denomina en el presente documento *Pea1*. La inhibición del *pea1* dio como resultado niveles reducidos de expresión de proteasas endógenas, como se muestra en los Ejemplos a continuación. Mediante la represión, la regulación negativa, la inactivación o la inhibición de la expresión de *pea1* se demostró que era posible suprimir, es decir, regular negativamente, la expresión de varias proteasas endógenas de la célula hospedadora fúngica. La disminución drástica de la actividad de proteasas endógenas dio como resultado una degradación menor de las proteínas expresadas por la célula hospedadora y, en consecuencia, un rendimiento mayor de proteínas de interés, tales como proteínas recombinantes heterólogas producidas por la célula hospedadora. Otra ventaja puede ser que pueden producirse menos proteínas inactivas o fragmentadas de interés porque se producen y secretan menos proteasas endógenas. La proteína de interés producida por la célula hospedadora también puede ser menos propensa a la degradación, lo que conduce a una autenticidad, una estabilidad y un período de caducidad mejorados. Las variantes, los fragmentos y los nucleótidos que son hibridables pueden usarse, por ejemplo, para detectar la presencia del regulador de proteasas o una secuencia similar al mismo. El polinucleótido de acuerdo con el primer aspecto y el producto génico codificado por el mismo son útiles en la producción industrial de proteínas.

En el presente documento se desvela un fragmento o una variante del polinucleótido del primer aspecto.

El polinucleótido modificado del primer aspecto codifica una forma inactiva o un fragmento del regulador de proteasas codificado por el polinucleótido. Puede usarse para inactivar la función normal del regulador de proteasas y, en consecuencia, suprimir la expresión de proteasas endógenas en una célula hospedadora.

5 De acuerdo con un cuarto aspecto se proporciona un vector que comprende el polinucleótido modificado del primer aspecto. El polinucleótido puede insertarse en el genoma de una célula hospedadora, por ejemplo, en un vector. En determinadas realizaciones, el polinucleótido puede codificar una forma activa o inactiva de *Pea1* y puede comprender elementos genéticos necesarios para insertar el polinucleótido aislado en la región del genoma (locus) que codifica la proteína activa por cruce doble o recombinación de reemplazo. Por tanto, puede usarse un polinucleótido de este tipo
10 en un método para activar o inactivar el gen que codifica el regulador de proteasas del primer aspecto. En una realización, el vector es un plásmido o un vector de fago. Dichos polinucleótidos y vectores pueden comprender regiones sin traducir 5' y 3', secuencias reguladoras de *pea1* para incorporar la construcción genética en el genoma del hospedador y, opcionalmente, al menos un marcador.

15 De acuerdo con un quinto aspecto se proporciona una célula hospedadora que comprende al menos un gen cromosómico inactivado en donde el gen cromosómico inactivado comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13.

20 La célula hospedadora del quinto aspecto puede producir menos proteasas endógenas con respecto a las que normalmente produciría cuando el gen cromosómico está activo, o no está inactivo. Por tanto, la actividad de degradación de proteínas de las proteasas endógenas de las células hospedadoras puede evitarse, al menos parcialmente, en la célula hospedadora del quinto aspecto.

25 En el presente documento se desvela una preparación proteínica que comprende la proteína producida en la célula hospedadora del quinto aspecto. En determinadas realizaciones, la preparación proteínica comprende células hospedadoras de acuerdo con el quinto aspecto.

30 La preparación proteínica puede tener un contenido mayor de la proteína que una preparación proteínica correspondiente producida usando la misma célula hospedadora con un *pea1* intacto. Por tanto, cuando se usa la preparación proteínica, puede requerirse una cantidad volumétrica total inferior de la preparación proteínica para obtener el mismo efecto con respecto a la que se requeriría cuando se usa una preparación proteínica producida de forma correspondiente pero en la que el efecto biológico del regulador de proteasas es el mismo que el de un regulador
35 de proteasas nativo. Además, la autenticidad, la estabilidad y el período de caducidad de la preparación proteínica pueden mejorar cuando la preparación proteínica contiene menos proteasas endógenas de la célula hospedadora.

40 En el presente documento se desvela un uso de la preparación proteínica del sexto aspecto para el procesamiento de biomasa o en la industria de biocombustibles, almidón, tejidos, detergentes, pulpa y papel, alimentos, panificación, piensos, bebidas o la industria farmacéutica.

45 El uso anterior tiene la ventaja de que como la preparación proteínica comprende más proteína, puede obtenerse más actividad proteínica a partir de una cantidad dada de la preparación proteínica y puede disminuirse la cantidad total de la preparación proteínica utilizada. También pueden evitarse problemas relacionados con la actividad de proteasa endógena en dichos procesos industriales.

De acuerdo con un octavo aspecto se proporciona un método de producción de una proteína que comprende

50 a. cultivar la célula hospedadora del quinto aspecto en condiciones adecuadas para producir la proteína, en donde la célula hospedadora comprende al menos un gen cromosómico inactivado que comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13;
el gen cromosómico inactivado se inactiva por disrupción;
y la célula hospedadora tiene actividad proteasa reducida en comparación con la célula hospedadora sin dicha
55 inactivación; y opcionalmente

b. recuperar la proteína.

60 El método del octavo aspecto proporciona un mejor rendimiento y estabilidad de la proteína. Además, el método permite producir proteínas que son difíciles o en algunos casos incluso imposibles de producir en una célula hospedadora debido a su sensibilidad a proteasas endógenas de la célula hospedadora. En determinadas realizaciones la proteína es una proteína recombinante.

65 En el presente documento se desvela una composición que comprende al menos una de entre: la preparación proteínica del sexto aspecto; y la proteína que puede obtenerse mediante el método del octavo aspecto. En determinados aspectos, la composición puede comprender al menos un componente adicional, tal como un tampón,

sal, disolvente, agua o detergente.

La composición es ventajosa en el sentido de que puede tener un mayor contenido de la proteína en comparación con una composición producida según corresponda, pero en una célula hospedadora con un *pea1* activo capaz de inducir la expresión de proteasas endógenas. Además, la composición puede tener un contenido bajo de proteasas endógenas. En determinadas realizaciones, la proteína puede ser sensible a la degradación por proteasas y la obtención de una composición estable producida en una célula hospedadora con un *pea1* activo requeriría etapas de purificación para retirar las proteasas endógenas inducidas por *pea1*. En un caso de este tipo, la composición puede ser más fácil de obtener con el método del octavo aspecto, porque el nivel inicial de proteasas endógenas es bajo. Asimismo, la composición puede tener un período de caducidad y una estabilidad mejoradas.

De acuerdo con un décimo aspecto se proporciona un método para generar una célula hospedadora para la producción de proteínas que comprende suprimir la expresión del gen de proteasas endógenas en una célula hospedadora inhibiendo al menos parcialmente la transcripción o traducción del polinucleótido de:

un polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado; o un polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado, seleccionándose el polinucleótido entre el grupo que consiste en:

- a) un polinucleótido que comprende una secuencia que tiene al menos un 55 % de identidad de secuencia con los nucleótidos 1141-3889 del SEQ ID NO: 11;
- b) el polinucleótido del SEQ ID NO: 12 o la secuencia codificante del mismo;
- c) el polinucleótido del SEQ ID NO: 11 o la secuencia codificante del mismo; y
- d) una secuencia de nucleótidos hibridable con una secuencia de nucleótidos que es complementaria con una cualquiera de a) a c) en condiciones altamente rigurosas.

El método es ventajoso porque puede usarse para suprimir muchas proteasas endógenas simultáneamente. La célula hospedadora resultante puede usarse para producir rendimientos mayores de cualquier proteína, tal como proteínas endógenas, proteínas recombinantes, proteínas heterólogas o cualquier proteína producida y opcionalmente secretada por la célula hospedadora. En la Tabla 2 se enumeran ejemplos no limitantes de tipos de proteasas cuya expresión puede suprimirse al menos parcialmente. En determinada realización, el método proporciona una célula hospedadora que tiene un nivel de expresión reducido de al menos una proteasa.

De acuerdo con un undécimo aspecto se proporciona una célula hospedadora que puede obtenerse usando el método del décimo aspecto.

De acuerdo con un duodécimo aspecto se proporciona un regulador de proteasas seleccionado entre el grupo que consiste en

- a) un polipéptido o un producto génico codificado por la secuencia codificante del polinucleótido del primer o del segundo aspecto;
- b) un polipéptido o un producto génico codificado por la secuencia codificante del polinucleótido del tercer aspecto;
- c) un polipéptido codificado por el SEQ ID NO: 11 o 12;
- d) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13; y
- e) una variante o un fragmento de un polipéptido o un producto génico de uno cualquiera de a) a d).

El regulador de proteasas del duodécimo aspecto puede proporcionarse en una célula hospedadora para inducir o suprimir la expresión de proteasas endógenas: un regulador de proteasas que tenga el efecto biológico de un regulador de proteasas nativo puede inducir la expresión de proteasas endógenas, mientras que un regulador de proteasas inactivado puede suprimir la expresión de proteasas endógenas. Además, pueden usarse fragmentos y variantes para interactuar con compañeros de unión del producto del gen *pea1* nativo, por ejemplo, para unir en una célula hospedadora un fragmento o una variante inactivos de un producto del gen *pea1* a un compañero de unión natural de un producto del gen *pea1*. En el presente documento se desvela un anticuerpo que tiene especificidad de unión al regulador de proteasas del duodécimo aspecto.

El anticuerpo puede producirse mediante métodos conocidos en la técnica. El anticuerpo puede usarse para unirse específicamente al regulador de proteasas. Por tanto, la presencia del regulador de proteasas puede detectarse, por ejemplo, en un inmunoensayo cuando el anticuerpo está directa o indirectamente unido a un marcador detectable.

Como alternativa, el anticuerpo puede usarse para unirse al regulador de proteasas para impedir la unión de un compañero de unión al regulador de proteasas. En una realización adicional, cuando se usa un anticuerpo que se une a una parte del regulador de proteasas que no participa en la unión con su compañero de unión, el regulador de proteasas con su compañero de unión pueden unirse en un complejo con el anticuerpo, y el compañero de unión puede identificarse con métodos conocidos en la técnica de la química de proteínas. Por tanto, en una realización el anticuerpo puede usarse como una herramienta de investigación para identificar las biomoléculas que participan en la regulación de la expresión de proteasas.

En el presente documento se desvela un método para inducir la expresión de proteasas en una célula hospedadora proporcionando el regulador de proteasas del punto a), c), d) o el punto e) que se refiere al punto a), c) o d) del duodécimo aspecto dentro o en contacto con la célula hospedadora. En determinadas realizaciones, el método puede comprender la expresión del regulador de proteasas en la célula hospedadora bajo el control del promotor.

Las realizaciones de la presente divulgación proporcionan determinados beneficios. Dependiendo de la realización, puede conseguirse uno o varios de los siguientes beneficios: producción mejorada de proteínas, posibilidad de producir proteínas que sean sensibles a proteasas o inestables de otro modo, autenticidad, estabilidad y período de caducidad mejorados de las composiciones, disminución del consumo de productos químicos, disminución de la necesidad de agentes estabilizantes y disminución de las cantidades de consumo de productos químicos, agua y energía cuando se usa en procesos industriales.

LISTADOS DE SECUENCIAS

SEQ ID NO: 1: Secuencia de nucleótidos del genoma QM6a v2.0 ID del gen: 123125

SEQ ID NO: 2: Secuencia de nucleótidos del genoma QM6a v2.0 ID: ADNc 123125

SEQ ID NO: 3: Secuencia de aminoácidos del genoma QM6a v2.0 ID: 123125

SEQ ID NO: 4: Secuencia de nucleótidos del gen del genoma RutC-30 v1.0 ID: 85889

SEQ ID NO: 5: Secuencia de nucleótidos del genoma RutC-30 v1.0 ID: ADNc 85889

SEQ ID NO: 6: Secuencia de aminoácidos del genoma RutC-30 v1.0 ID: 85889

SEQ ID NO: 7: Secuencia de nucleótidos del gen *pea1* en la cepa 33SP n.º 9

SEQ ID NO: 8: Secuencia de nucleótidos del gen *pea1* en la cepa 31SP n.º 4

SEQ ID NO: 9: Secuencia de nucleótidos del gen *pea1* en la cepa 31UV n.º 22

SEQ ID NO: 10: Secuencia de nucleótidos del gen *pea1* en la cepa A21

SEQ ID NO: 11: Secuencia de nucleótidos del gen *pea1* clonado a partir de QM6a (incluyendo secuencias de 1140 pb corriente arriba y 821 pb corriente abajo)

SEQ ID NO: 12: Secuencia de nucleótidos del ADNc de *pea1* determinada a partir de QM6a (incluyendo secuencias 5'UTR de 654 pb y 3'-UTR de 821 pb)

SEQ ID NO: 13: La secuencia de aminoácidos deducida de la proteína Pea1 de longitud completa

SEQ ID NO: 14: La secuencia de aminoácidos deducida de la proteína Pea1 en la cepa 33SP n.º 9

SEQ ID NO: 15: La secuencia de aminoácidos deducida de la proteína Pea1 en la cepa 31SP n.º 4

SEQ ID NO: 16: La secuencia de aminoácidos deducida de la proteína Pea1 en la cepa 31UV n.º 22

SEQ ID NO: 17: La secuencia de aminoácidos deducida de la proteína Pea1 en la cepa A21

SEQ ID NO: 18: La proteína Pea1 truncada codificada por pALK4106

Breve descripción de las figuras

La Fig. 1, panel A, muestra el análisis por SDS-PAGE de sobrenadantes de cultivo de cultivos en matraz de agitación de transformantes que producen la proteína 20K+CBD. Calles 1 - 3, muestras que derivan del cultivo de un hospedador de proteasa no baja del mismo linaje de la cepa que el hospedador de transformación después de 3, 5 y 7 días de cultivo, respectivamente; 4 - 6, muestras de los transformantes de 33SP n.º 9 pALK1769 n.º 2,

n.º 6 y n.º 7, respectivamente. Se cargaron cantidades iguales de los sobrenadantes de cultivo en cada calle.

La Fig. 1, panel A, muestra el análisis por SDS-PAGE de sobrenadantes de cultivo de cultivos por lotes en biorreactor de transformantes que producen la proteína 20K+CBD a partir del casete de expresión pALK1769. Calle 1, muestra que deriva del cultivo de un hospedador de proteasa no baja del mismo linaje de cepa que los hospedadores de transformación; 2 - 6, muestras de los cultivos de cepas transformadas con PALK1769; un transformante 33UV n.º 82, dos transformadores paralelos 33SP n.º 9, un transformante 33UV n.º 48 y uno 33SP n.º 11, respectivamente. Las muestras se tomaron después de cuatro días de cultivo en biorreactores. Se cargaron cantidades iguales de los sobrenadantes de cultivo en cada calle.

La Fig. 2 muestra esquemáticamente las anotaciones del QM_ID123125, Rut_ID85889 y la anotación deducida del ADNc derivado de QM6a (*pea1*). La ubicación de las mutaciones en las cepas 33SP n.º 9, 31SP n.º 4, 31UV n.º 22 y A21 se muestran con triángulos en el esquema de anotación de *pea1*.

La Fig. 3 muestra la secuencia de nucleótidos del gen *pea1* (nucleótidos 1141-3889 del SEQ ID NO: 11) y la secuencia de aminoácidos deducida. La longitud y la ubicación de los intrones se determinaron a partir del análisis de ADNc y se muestran en letras cursivas subrayadas.

La Fig. 4 muestra la alineación de los aminoácidos de la región altamente conservada de Pea1 (aminoácidos 402-533 del SEQ ID NO: 13) con las correspondientes regiones de secuencias similares de múltiples especies. Debajo de la alineación hay un símbolo que representa restos idénticos (*), restos conservadores (:) y restos no conservadores () de acuerdo con una alineación de secuencia realizada con Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>).

La Fig. 5A muestra el casete pALK4104 para la delección del gen *pea1* de longitud completa, el fragmento *EcoRI* - *PstI* de 6748 pb se desprendió del plásmido pALK4104. Se muestra una selección de sitios de enzimas de restricción. *pea1_5'* y *pea1_3'*, regiones flanqueantes 5' y 3' del gen *pea1*, respectivamente, utilizadas para dirigir el casete de delección al locus de *pea1* para el reemplazo del gen *pea1* con el gen marcador; *syn-amdS*, gen *amdS* sintético que codifica acetamidasa para la selección de transformantes; Rut_ID120107 y Rut_ID10852, la ubicación y los números de ID de genes anotados de acuerdo con la secuencia del genoma público RutC-30; QM_ID66437, la ubicación y el número de ID de un gen anotado, de acuerdo con la secuencia del genoma público QM6a.

La Fig. 5B muestra el casete pALK4106 para el truncamiento de *pea1*, el fragmento *EcoRI* - *PstI* de 6595 pb se desprendió del plásmido pALK4106. Se muestra una selección de sitios de enzimas de restricción. *pea1'*, un gen *pea1* truncado; *pea1_3'*, *syn-amdS*, Rut_ID120107, Rut_ID10852 y QM_ID66437; genes/regiones idénticas a las descritas para el casete pALK4104 (Fig. 5A).

La Fig. 6 muestra el casete pALK4107 para la delección del gen *pea1* de longitud completa usando el gen marcador *ble*, el fragmento *EcoRI* - *PstI* de 7615 pb se desprendió del plásmido pALK4107. Se muestra una selección de sitios de enzimas de restricción. *pea_5'*, *pea_3'*, Rut_ID120107, Rut_ID10852 y QM_ID66437, idénticas a las descritas para el casete PALK4104 (Fig. 5A); *ble*, gen originado a partir de *Streptoalloteichus hindustanus* y que codifica ShBle, proporcionando resistencia a antibióticos de la familia de la fleomicina; *pgpdA* y *ttrpC*, originados a partir de *Aspergillus nidulans*, el promotor del gen de la gliceraldehído-3-fosfato deshidrogenasa y el terminador de un gen que codifica el polipéptido que actúa en la biosíntesis de triptófano, respectivamente. El *ble* con promotor y terminador se aislaron del plásmido pAN8-1 (Mattern *et al.*, 1988; NCBI gi: 475899).

La Fig. 7 muestra una alineación de secuencia de las secuencias de aminoácidos deducidas de las proteínas Pea1 truncadas en las cepas 33SP n.º 9 (SEQ ID NO: 14), 31SP n.º 4 (SEQ ID NO: 15), 31UV n.º 22 (SEQ ID NO: 16) y A21 (SEQ ID NO: 17) y la secuencia de aminoácidos deducida de la proteína Pea1 truncada (SEQ ID NO: 18) codificada por el *pea1* truncado en pALK4106 (Fig. 5B). Los aminoácidos que no coinciden con la secuencia de aminoácidos de la Pea1 nativa (SEQ ID NO: 13), es decir, los aminoácidos generados mediante un desplazamiento de marco, están subrayados.

Fig. 8. Análisis por SDS-PAGE de transformantes Δ *pea1* y hospedador que produce una proteína celulosa recombinante. Las muestras se ejecutaron en un gel de poliacrilamida-SDS al 12 % de sobrenadantes de cultivo de fermentaciones a escala de laboratorio ejecutadas durante cuatro días (la misma cantidad de muestra de cada fermentación). El gel se tiñó con Azul Coomassie. 1, marcador de masa molecular; 2, sobrenadante de cultivo de cultivo de RF5969; 3 - 7, sobrenadantes de cultivo de cultivos de cinco transformantes RF5969 separados con delección de *pea1*.

Fig. 9. Diseño del enfoque de marcador dividido para inactivar los homólogos de *pea1* de especies de *Fusarium*. P*pea1*, región promotora del gen *pea1* de *Fusarium oxysporum*; *pea1*, gen *pea1* de *F. oxysporum*; T*pea1*, región terminadora del gen *pea1* de *F. oxysporum*; P*gpdA*, promotor del gen de gliceraldehído-3-fosfato deshidrogenasa de *Aspergillus*; *hph*, gen que codifica higromicina fosfotransferasa (para la resistencia a la higromicina); T*trpC*, terminador del gen *trpC* (triptófano C) de *Aspergillus*. Las regiones para posibles recombinaciones homólogas se muestran mediante cruces. Los tamaños de los fragmentos no están a escala.

Fig. 10. Actividades proteasa de los sobrenadantes de cultivo de *Fusarium oxysporum*, *F. fujikuroi* y sus transformantes con el gen *pea1* inactivado. OXY WT y OXY-03, OXY-09, OXY-24, OXY-38, la actividad proteasa analizada a partir de los sobrenadantes de cultivo de *F. oxysporum* Fo47 y sus cuatro transformantes, respectivamente; FUJI WT y FUJI-01, FUJI-08, FUJI-11 y FUJI-31, la actividad proteasa es resultado de los sobrenadantes de cultivo de *F. fujikuroi* IMI58289 y sus cuatro transformantes, respectivamente.

DEPÓSITOS

Se realizaron los siguientes depósitos de cepas de acuerdo con el Tratado de Budapest sobre el Reconocimiento Internacional del Depósito de Microorganismos con Fines de Procedimiento de Patente:

La cepa de *E. coli* RF11697, que incluye el plásmido pALK3535, se depositó en la Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Inhoffenstrasse 7 b, D-38124 Braunschweig, Alemania el 4 de febrero de 2015 y se le asignó el número de acceso DSM 32007.

La cepa de *E. coli* RF11698, que incluye el plásmido pALK3536, se depositó en la Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Inhoffenstrasse 7 b, D-38124 Braunschweig, Alemania el 4 de febrero de 2015 y se le asignó el número de acceso DSM 32008.

Descripción detallada

Los presentes inventores han identificado y caracterizado un regulador de la expresión de proteasas fúngico y han modificado satisfactoriamente mediante ingeniería genética una célula hospedadora adecuada para su uso industrial que carece del regulador de la expresión de proteasas funcional o en la que el regulador de proteasas está inactivado. Sin quedar ligados a teoría alguna, la presente divulgación muestra que mediante la inactivación del regulador de proteasas, los niveles de expresión de varias proteasas endógenas de la célula hospedadora pueden reducirse significativamente. Por tanto, cuando el regulador de proteasas endógenas se suprime en una célula hospedadora, la producción de una proteína de interés puede potenciarse, dando como resultado un rendimiento mejorado y una degradación proteolítica reducida de las proteínas producidas y/o secretadas. Simultáneamente, las capacidades de rendimiento de fermentación, proliferación y producción de proteínas de la célula hospedadora puede mantenerse a los niveles requeridos en la producción industrial de proteínas.

Como se usa en el presente documento, "*pea1*" significa un polinucleótido que comprende la secuencia del SEQ ID NO: 11 nucleótidos 1141 - 3889, así como la secuencia de la región codificante en el SEQ ID NO: 12 y las secuencias que tienen similitud con dichas SEQ ID NO. El gen *pea1* codifica un producto génico cuya supresión da como resultado una expresión reducida de muchas proteasas endógenas fúngicas. Por tanto, el gen se denomina regulador de proteasas, que afecta a la expresión de proteasas 1 (del inglés *protease expression affecting 1*), o *pea1*. Las regiones sin traducir 5' y 3', regiones promotoras, intrones, exones y secuencias reguladoras pueden tener un efecto sobre la función de *pea1*.

En determinadas realizaciones, el polinucleótido o el polipéptido de cualquier aspecto o realización es un polinucleótido aislado o un polipéptido aislado.

Como se usa en el presente documento, "aislado" significa una sustancia en una forma o entorno que no se produce en la naturaleza. Los ejemplos no limitantes de sustancias aisladas incluyen (a) cualquier sustancia de origen no natural, (2) cualquier sustancia, incluyendo cualquier enzima, variante, ácido nucleico, proteína, péptido o cofactor, que se retira al menos parcialmente de uno o más o de todos los componentes naturales a los que se asocia en la naturaleza; (3) cualquier sustancia modificada por la mano del hombre con respecto a esa sustancia hallada en la naturaleza; o (4) cualquier sustancia modificada aumentando o disminuyendo la cantidad de la sustancia con respecto a otros componentes a los que se asocia de forma natural (por ejemplo, producción recombinante en una célula hospedadora; múltiples copias de un gen que codifica la sustancia; y uso de un promotor alternativo al promotor asociado de forma natural al gen que codifica la sustancia).

Como se usa en el presente documento, la expresión "que comprende" incluye los significados más amplios de "que incluye", "que contiene" y "que comprende", así como las expresiones más limitadas "que consistente en" y "que consistente solamente en".

Como se usa en el presente documento, "fragmento" significa una proteína o un polinucleótido que tiene uno o más aminoácidos o nucleótidos suprimidos. En el contexto del ADN, un fragmento incluye ADN tanto monocatenario como bicatenario de cualquier longitud. Un fragmento puede ser un fragmento activo que tenga la función biológica, tal como la actividad enzimática o la actividad reguladora, de la proteína o el polinucleótido. Un fragmento también puede ser un fragmento inactivo, es decir, no tiene uno o más efectos biológicos de la proteína o del polinucleótido nativos.

Como se usa en el presente documento, "variante" significa un fragmento de secuencia (nucleótido o aminoácido) insertado o suprimido por uno o más nucleótidos/aminoácidos o que está modificado químicamente.

Como se usa en el presente documento, un "péptido" y un "polipéptido" son secuencias de aminoácidos que incluyen una pluralidad de restos de aminoácidos polimerizados consecutivos. Para los fines de la presente invención, los

péptidos son moléculas que incluyen hasta 20 restos de aminoácidos, y los polipéptidos incluyen más de 20 restos de aminoácidos. El péptido o polipéptido puede incluir restos de aminoácidos modificados, restos de aminoácidos de origen natural no codificados por un codón, y restos de aminoácidos de origen no natural. Como se usa en el presente documento, una "proteína" puede referirse a un péptido o un polipéptido de cualquier tamaño. Una proteína puede ser una enzima, una proteína, un anticuerpo, una proteína de membrana, una hormona peptídica, regulador, o cualquier otra proteína.

Como se usa en el presente documento, "modificación", "modificado" y términos similares en el contexto de los polinucleótidos se refieren a la modificación en una región codificante o no codificante del polinucleótido, tal como una secuencia reguladora, región sin traducir 5', región sin traducir 3', elemento genético de regulación positiva, elemento genético de regulación negativa, región potenciadora, supresora, promotora, exónica o intrónica. En algunas realizaciones, la modificación puede ser solamente estructural, que no tiene ningún efecto sobre el efecto biológico, la acción o la función del polinucleótido. En otras realizaciones, la modificación es una modificación estructural que proporciona un cambio en el efecto biológico, la acción o la función del polinucleótido. Una modificación de este tipo puede potenciar, suprimir o cambiar la función biológica del polinucleótido.

Como se usa en el presente documento, "identidad" significa el porcentaje de coincidencias exactas de restos de aminoácidos entre dos secuencias alineadas sobre el número de posiciones donde hay restos presentes en ambas secuencias. Cuando una secuencia tiene un resto sin el resto correspondiente en la otra secuencia, el programa de alineación permite un hueco en la alineación, y esa posición no se cuenta en el denominador del cálculo de identidad. En este caso, la identidad es un valor determinado con la herramienta de Alineación de Secuencias por Pares EMBOSS Needle en el sitio web de EMBL-EBI (www.ebi.ac.uk/Tools/psa/emboss_needle/).

Como se usa en el presente documento, "similitud" significa el porcentaje de coincidencias entre dos secuencias sobre la región alineada publicada. Además de los aminoácidos que coinciden de forma idéntica (identidad), la similitud permite que las sustituciones conservadoras (cambio a un aminoácido con propiedades físicoquímicas similares) se tengan en cuenta en el valor porcentual. En este caso, la similitud es un valor determinado con la herramienta de Alineación de Secuencias por Pares EMBOSS Needle en el sitio web de EMBL-EBI (www.ebi.ac.uk/Tools/psa/emboss_needle/).

Como se usa en el presente documento, "célula hospedadora" significa cualquier tipo de célula que sea susceptible de transformación, transfección, transducción o similar con una construcción de ácido nucleico o vector de expresión que comprende un polinucleótido. La expresión "célula hospedadora" abarca cualquier descendencia que no sea idéntica debido a mutaciones que se producen durante la replicación. Son ejemplos no limitantes de una célula hospedadora las células fúngicas, las células fúngicas filamentosas de la División *Ascomycota*, Subdivisión *Pezizomycotina*; preferentemente del grupo que consiste en miembros de la Clase *Sordariomycetes*, Subclase *Hypocreomycetidae*, Órdenes *Hypocreales* y *Microascales* y *Aspergillus*, *Chrysosporium*, *Myceliophthora* y *Humicola*; más preferentemente del grupo que consiste en las Familias *Hypocreaceae*, *Nectriaceae*, *Clavicipitaceae*, *Microascaceae* y Géneros *Trichoderma* (anamorfo de *Hypocrea*), *Fusarium*, *Gibberella*, *Nectria*, *Stachybotrys*, *Claviceps*, *Metarhizium*, *Villosiclava*, *Ophiocordyceps*, *Cephalosporium* y *Scedosporium*; más preferentemente del grupo que consiste en *Trichoderma reesei* (*Hypocrea jecorina*), *T. citrinoviridae*, *T. longibrachiatum*, *T. virens*, *T. harzianum*, *T. asperellum*, *T. atroviridae*, *T. parareesei*, *Fusarium oxysporum*, *F. graminearum*, *F. pseudograminearum*, *F. venenatum*, *Gibberella fujikuroi*, *G. moniliformis*, *G. zeaeae*, *Nectria* (*Haematonectria*) *haematococca*, *Stachybotrys chartarum*, *S. chlorohalonata*, *Claviceps purpurea*, *Metarhizium acridum*, *M. anisopliae*, *Villosiclava virens*, *Ophiocordyceps sinensis*, *Acremonium* (*Cephalosporium*) *chrysogenum*, y *Scedosporium apiospermum* y *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Myceliophthora thermophila*, *Humicola insolens* y *Humicola grisea*, mucho más preferentemente *Trichoderma reesei*. En una realización, la célula hospedadora se selecciona entre el siguiente grupo de cepas que pueden obtenerse de colecciones públicas: QM6a, ATCC13631; RutC-30, ATCC56765; QM9414, ATCC26921, y derivados de los mismos.

Como se usa en el presente documento, condiciones de rigurosidad baja significa sondas de al menos 100 nucleótidos en condiciones de longitud correspondientes a la hibridación en la prehibridación y la hibridación a 55 °C en 5x SSC, N-lauroilsarcosina al 0,1 %, SDS al 0,02 %, reactivo bloqueante al 1 % (Roche 11 096 176 001), siguiendo los procedimientos convencionales de transferencia Southern durante de 12 a 24 horas. El material de vehículo finalmente se lava dos o tres veces cada uno durante 15 minutos usando 2X SSC, SDS al 0,1 % a 55 °C.

Como se usa en el presente documento, condiciones de rigurosidad alta significa sondas de al menos 100 nucleótidos en condiciones de longitud correspondientes a la hibridación en la prehibridación y la hibridación a 65 °C en 5x SSC, N-lauroilsarcosina al 0,1 %, SDS al 0,02 %, reactivo bloqueante al 1 % (Roche 11 096 176 001), siguiendo los procedimientos convencionales de transferencia Southern durante de 12 a 24 horas. El material de vehículo finalmente se lava dos o tres veces cada uno durante 15 minutos usando 0.1X SSC, SDS al 0,1 % a 65 °C.

Como se usa en el presente documento, "expresión" incluye cualquier etapa implicada en la producción de un polipéptido en una célula hospedadora incluyendo, pero sin limitación, transcripción, traducción, modificación postraduccional y secreción. La expresión puede ir seguida de la recogida, es decir, la recuperación, de las células hospedadoras o el producto expresado.

Como se usan en el presente documento, inhibir, inactivar, suprimir y regular negativamente significan, al menos parcialmente, evitar la acción biológica del gen *pea1* o el producto del gen. Como se entiende en la técnica, esto puede lograrse a nivel transcripcional, traduccional o proteínico, es decir, evitando la lectura o la expresión del gen *pea1*, evitando la traducción correcta de la proteína Pea1 o evitando que el producto del gen *pea1* se una a su compañero o compañeros de unión que en condiciones naturales participan en la acción del producto del gen *pea1*.

Como se usa en el presente documento, una proteasa inducida por el regulador de proteasas del primer aspecto puede ser cualquier proteasa cuya expresión sea inducida por el regulador de proteasas, y cuya expresión y/o actividad proteasa se reduzcan cuando el regulador de proteasas esté inactivo. Son ejemplos no limitantes de dichas proteasas las aspártico proteasas, serina proteasas, glutámico proteasas y metaloproteasas (Tabla 2). Por tanto, un efecto biológico de *pea1* puede ser regular la expresión de proteasas endógenas.

Como se usa en el presente documento, un "producto génico" es el ARN o la proteína resultante de la expresión de un polinucleótido. Los ejemplos de productos genéticos incluyen ARNm, ARNip, ADNc, proteína, polipéptido y péptido.

En una realización de ejemplo del primer aspecto, la célula hospedadora es *Trichoderma*.

En una realización de ejemplo del primer aspecto, la secuencia de nucleótidos codifica una proteína que comprende una secuencia de aminoácidos con un 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13. En otra realización, la secuencia de nucleótidos codifica una proteína que comprende una secuencia de aminoácidos con un 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de similitud de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13.

En una realización de ejemplo del primer aspecto, el polinucleótido se selecciona entre el grupo que consiste en la secuencia codificante de las SEQ ID NO: 11 y 12.

En una realización de ejemplo del primer aspecto, el polinucleótido se selecciona entre el grupo que consiste en:

a) un polinucleótido que comprende una secuencia que tiene al menos un 55 %, 56 %, 57 %, 58 %, 59 %, 60 %, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad de secuencia con los nucleótidos 1141-3889 del SEQ ID NO: 11;

b) el polinucleótido del SEQ ID NO: 12 o la secuencia codificante del mismo;

c) el polinucleótido del SEQ ID NO: 11 o la secuencia codificante del mismo; y

d) una secuencia de nucleótidos hibridable con una secuencia de nucleótidos que es complementaria con una cualquiera de a) a c) en condiciones altamente rigurosas.

En una realización de ejemplo del primer aspecto, el polinucleótido o su región no codificante contiene al menos una modificación. En determinadas realizaciones del primer aspecto, la modificación lo hace estructuralmente diferente en comparación con cualquier regulador de proteasas de origen natural, o la modificación hace que su expresión y/o traducción sea diferente, por ejemplo, en términos de eficiencia o estabilidad en comparación con los de cualquier regulador de proteasas de origen natural. La modificación puede tener un efecto sobre una función biológica u otra propiedad del regulador de proteasas. En otra realización, la modificación no cambia sustancialmente una función biológica u otra propiedad del regulador de proteasas. Por tanto, en determinadas realizaciones la modificación no disminuye sustancialmente la capacidad del polinucleótido del primer aspecto para inducir la expresión de proteasas endógenas en una célula hospedadora.

En una realización de ejemplo, el polinucleótido del primer aspecto comprende elementos genéticos para permitir su transcripción y/o traducción en una célula hospedadora. En otra realización, el polinucleótido comprende adicionalmente elementos genéticos que permiten secretar la proteína fuera de la célula hospedadora.

En una realización de ejemplo del cuarto aspecto, el vector comprende elementos genéticos para incorporar el polinucleótido del segundo aspecto o de la realización anterior en el genoma de una célula hospedadora. En determinadas realizaciones, los elementos genéticos comprenden una región 5' sin traducir y/o una región 3' sin traducir, opcionalmente en forma de un casete.

En una realización de ejemplo del quinto aspecto, la célula hospedadora se selecciona entre el grupo que consiste en células fúngicas filamentosas de la División *Ascomycota*, Subdivisión *Pezizomycotina*; preferentemente del grupo que consiste en miembros de la Clase *Sordariomycetes*, Subclase *Hypocreomycetidae*, Órdenes *Hipocreales* y *Microascales* y *Aspergillus*, *Chrysosporium*, *Myceliophthora* y *Humicola*; más preferentemente del grupo que consiste en las Familias *Hypocreaceae*, *Nectriaceae*, *Clavicipitaceae*, *Microasceae* y Géneros *Trichoderma* (anamorfo de

Hypocrea), *Fusarium*, *Gibberella*, *Nectria*, *Stachybotrys*, *Claviceps*, *Metarhizium*, *Villosiclava*, *Ophiocordyceps*, *Cephalosporium* y *Scedosporium*; más preferentemente del grupo que consiste en *Trichoderma reesei* (*Hypocrea jecorina*), *T. citrinoviridae*, *T. longibrachiatum*, *T. virens*, *T. harzianum*, *T. asperellum*, *T. atroviridae*, *T. parareesei*, *Fusarium oxysporum*, *F. graminearum*, *F. pseudograminearum*, *F. venenatum*, *Gibberella fujikuroi*, *G. moniliformis*, *G. zeaeae*, *Nectria* (*Haematonectria*) *haematococca*, *Stachybotrys chartarum*, *S. chlorohalonata*, *Claviceps purpurea*, *Metarhizium acridum*, *M. anisopliae*, *Villosiclava virens*, *Ophiocordyceps sinensis*, *Acremonium* (*Cephalosporium*) *chrysogenum*, y *Scedosporium apiospermum* y *Aspergillus niger*, *A. awamori*, *A. oryzae*, *Chrysosporium lucknowense*, *Mycelohphthora thermophila*, *Humicola insolens*, *Humicola grisea*, mucho más preferentemente *Trichoderma reesei*. En una realización, la célula hospedadora se selecciona entre el siguiente grupo de cepas que pueden obtenerse de colecciones públicas: QM6a, ATCC13631; RutC-30, ATCC56765; QM9414, ATCC26921, y derivados de los mismos.

En una realización de ejemplo del quinto aspecto, el gen cromosómico inactivado comprende el polinucleótido del primer aspecto.

En una realización de ejemplo del quinto aspecto el gen cromosómico inactivado se inactiva por interrupción, por ejemplo, con un marcador seleccionable, inhibición de la traducción o transcripción del gen cromosómico, delección al menos parcial, truncamiento, delección, inserción, mutación o silenciamiento, mediante iARN o mediante tecnología CRISPR/Cas9. Cuando se usa iARN, puede usarse ARN bicatenario para silenciar postraduccionalmente los niveles de expresión de un gen específico, tal como *pea1*, debido a la degradación específica de la secuencia mediada por ARN bicatenarios pequeños. Por ejemplo, pueden diseñarse y usarse moléculas de ARNbc y ARNip sintetizadas *in vitro* o ARNbc o ARN de horquilla de tallo-bucle sintetizados *in vivo* como desencadenantes para el direccionamiento. Cuando la tecnología de CRISPR/Cas9 se usa en la inactivación, se entregan en la célula la proteína Cas9 y ARN guía apropiados (de acuerdo con la secuencia objetivo, tal como *pea1*), dando como resultado una escisión en el lugar deseado.

En una realización de ejemplo del quinto aspecto, la célula hospedadora comprende elementos genéticos para permitir la expresión, en condiciones adecuadas para promover la expresión, al menos una proteína de interés codificada por un polinucleótido recombinante. Está dentro del nivel de habilidad en la técnica elegir las condiciones adecuadas, incluyendo reactivos y condiciones para la expresión de ARN a partir de la construcción de expresión, seguida de la traducción del polipéptido codificado. En los ejemplos a continuación se describen reactivos y condiciones de ejemplo. Los métodos de la presente realización también pueden realizarse en un sistema de traducción sin células o *in vivo*. En una realización preferida, la expresión de proteína se realiza en una célula hospedadora recombinante.

En una realización de ejemplo del quinto aspecto, la proteína de interés se selecciona entre la lista que consiste en una proteína farmacológicamente activa, anticuerpo, fragmento de anticuerpo, proteína terapéutica, biosimilar, proteína de dominios múltiples, hormona peptídica, péptido antimicrobiano, péptido, módulo de unión a hidratos de carbono, enzima tal como celulasa, proteasa, inhibidor de proteasas, aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, quitinasa, cutinasa, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, invertasa, laccasa, lipasa, mananasa, mutanasa, oxidasa, enzima pectinolítica, peroxidasa, fosfolipasa, fitasa, fosfatasa, polifenoloxidasas, enzima redox, enzima proteolítica, ribonucleasa, transglutaminasa y xilanasas. La misma célula hospedadora puede expresar una o más proteínas de interés.

En una realización de ejemplo del sexto aspecto, la preparación proteínica comprende al menos un componente adicional seleccionado entre estabilizante, conservante, fragancia, tampón, sal y colorante.

En una realización de ejemplo del décimo aspecto, la inhibición se proporciona realizando una modificación inactivadora en el gen que comprende la secuencia del polinucleótido del primer aspecto. La modificación puede ser delección, truncamiento o mutación de al menos una parte del regulador de proteasas, incluyendo su secuencia de control, que da como resultado la supresión o la inhibición al menos parcial de la capacidad del regulador de proteasas para inducir la expresión de proteasas endógenas en la célula hospedadora. En otra realización, la función del gen regulador de proteasas se inactiva postraduccionalmente, por ejemplo, inhibiendo la interacción proteína-proteína o inhibiendo la unión del regulador de proteasas a cualquiera de sus compañeros de unión naturales. En otra realización más, el regulador de proteasas se inactiva mediante la supresión de un promotor u otra región reguladora del presente regulador de proteasas.

En una realización de ejemplo del décimo aspecto, la inhibición se consigue mediante mutación, delección, inserción, interferencia de ARN, anticuerpo o inhibidor de molécula pequeña.

En una realización de ejemplo del undécimo aspecto, la célula hospedadora comprende adicionalmente un ácido nucleico que codifica una proteína heteróloga.

En una realización de ejemplo del undécimo aspecto, la célula hospedadora es una célula fúngica, preferentemente una célula fúngica filamentosa, tal como *Trichoderma* o *Trichoderma reesei*.

Ejemplos

Ejemplo 1. Aislamiento de mutantes de proteasa baja de cepas de *Trichoderma reesei*.

5 La A21 de *Trichoderma reesei* es un mutante UV de proteasa baja que deriva del linaje de la cepa QM9414 de *T. reesei*. La cepa A21 se cribó de los mutantes obtenidos después de la irradiación con luz ultravioleta de lotes de esporas parentales usando un ensayo en placa de leche desnatada. Produjo un halo reducido en la placa de selección en comparación con su progenitor, indicando una producción reducida de proteasas. Se confirmó que A21 producía cantidades claramente reducidas de actividades proteasa en sus sobrenadantes de cultivo en comparación con su progenitor, tanto en cultivos en matraz de agitación como en biorreactores a escala de laboratorio en medio inductor de celulasa. El análisis por FPLC demostró que A21 carece, por ejemplo, de un pico proteínico que en la cepa parental mostraba actividad proteasa que podía ser inhibida por la pepstatina A, lo que indica que no hay producción o que hay una producción menor de al menos un tipo de aspártico de una proteasa o proteasas, en comparación con el progenitor.

15 Para desarrollar mutantes con niveles de producción reducidos de proteasas nativas a partir de un linaje de cepa mutante de *T. reesei* diferente, se eligieron las cepas industriales patentadas A31 y A33 para un programa de desarrollo de cepas. A31 es una cepa mutante de *T. reesei* con una alta capacidad de producción de proteínas (celulasa). A33 es un derivado genéticamente modificado de A31 del cual se han suprimido los cuatro genes principales nativos que codifican celulasas *cbh1 (cel7A)*, *cbh2 (cel6A)*, *egl1 (cel7B)* y *egl2 (cel6A)* usando el método de contra selección de *pyr4* (para obtener información sobre el método, véase Seidl y Seiboth, 2010). Los mutantes A31 y A33 se generaron mediante el uso de mutagénesis UV y mediante selección de mutantes espontáneos de proteasa baja mediante el método del suicidio (SUI) (Braaksma y Punt, 2008) desarrollado en TNO (Países Bajos). Este método se basa en un producto químico SUI patentado frente al que las cepas que producen cantidades reducidas de proteasas son más resistentes que las cepas parentales. Mediante el uso del enfoque SUI, la detección de mutantes de proteasa baja (cepas) es rápida y eficiente. Sin embargo, el cribado de dichos mutantes también puede realizarse mediante el cultivo en placa directo de las esporas mutadas (o esporas) en leche desnatada u otras placas de detección de proteasa adecuadas.

30 Las cepas de *T. reesei* se inocularon y cultivaron en placas PD (agar de dextrosa de patata) para generar esporas para la mutagénesis. La mutagénesis por UV se realizó usando la cámara BioRad UV y un tiempo de irradiación de 40 - 80 s (con una tasa de supervivencia del 5 - 50 %). Los lotes de esporas no mutadas y tratadas con UV se cultivaron en placas de agar a base de medio mínimo de *Trichoderma* (TMM; Penttilä *et al.*, 1987) que contenían diferentes concentraciones del reactivo SUI (50-500 µg/ml) y placas AMM_{NH4} (Bennet y Lasure, 1991) con 25 - 500 µg/ml de SUI para seleccionar los mutantes de proteasa baja.

35 De las dos cepas de *T. reesei* se cribaron en las placas de SUI aproximadamente 5×10^7 esporas no mutadas y $1 - 2 \times 10^7$ mutadas. Después de la primera ronda de selección por SUI, se volvieron a cribar 200 - 300 colonias resistentes a SUI de cada cepa en placas de SUI. Aproximadamente el 75 % de las cepas aún mostraban un fenotipo resistente a SUI. El tipo anterior de cepas después se analizó en placas de TMM-NO₃ + leche desnatada (100 ml de leche desnatada al 10 % añadida a TMM después del tratamiento en autoclave, (NH₄)₂SO₄ reemplazado por 6 g/l de NaNO₃). En las placas de leche desnatada, aproximadamente 20 - 40 cepas (aproximadamente el 15 % tanto de A31 como de A33) no mostraron ningún halo o un halo reducido en comparación con las cepas parentales, lo que indica una producción de proteasa muy baja o baja. Una selección de cepas se purificó a través de esporas individuales. Estas cepas se caracterizaron adicionalmente en placas de celulosa (Walseth al 0,5 %) y xilano (xilano de escanda de avena al 0,5 %) (cepas derivadas de A31) o en placas de xilano (cepas derivadas de A33) para confirmar que seguían siendo capaces de producir actividades celulasa y/o xilanasa. Una selección de cepas con producción de proteasa reducida, pero producción similar de celulasa y/o xilanasa en las placas en comparación con los progenitores, se eligió para un análisis y caracterización adicionales. Se analizaron sus niveles de crecimiento y de producción de proteínas y proteasas en cultivos en matraces de agitación y biorreactores (Ejemplo 2). También se sometió a ensayo la idoneidad de las cepas elegidas como hospedadores para la producción de proteínas sensibles a proteasas (Ejemplo 3).

Ejemplo 2. Caracterización de las cepas mutantes de proteasa baja.

55 Se cultivó una selección de mutantes de proteasa baja, basándose en los resultados del ensayo en placa, en matraces de agitación usando medio mínimo a base de lactosa inductor de celulasa (Bailey *et al.*, 2002). Las actividades proteasa se midieron a partir de los sobrenadantes de cultivo usando como sustrato caseína dimetilada o BSA (albúmina sérica bovina), basándose en el procedimiento descrito por Holm (1980) y usando glicina para la calibración. Para el ensayo de caseína, los pH utilizados en las mediciones de actividad fueron 5,5, 7,0 y 8,5 y para el ensayo de BSA se usaron los pH 4,0 y 6,0. Se observaron diversos niveles de actividad proteasa en los sobrenadantes de cultivo entre las cepas mutantes. Sin embargo, varios mutantes (pero no todos) que habían mostrado una actividad proteasa reducida en el ensayo del halo de leche también mostraron una actividad proteasa reducida en los cultivos líquidos. Algunos de los mutantes seleccionados mostraron actividades celulasa y/o xilanasa similares o mejores en comparación con el hospedador. Sin embargo, algunos de los mutantes seleccionados mostraron actividades celulasa y/o la xilanasa reducidas, lo que indica una deficiencia general en la secreción de proteínas en estas cepas.

65 Basándose en los resultados del ensayo de placa de leche desnatada y el cultivo en medio mínimo, en total se eligieron

22 cepas de proteasa baja derivadas de A31 y 23 derivadas de A33 para su cultivo en matraces de agitación usando un medio complejo inductor de celulasa a base de lactosa (Joutsjoki *et al.*, 1993) tamponado con KH_2PO_4 al 5 %. La selección de cepas incluyó tanto mutantes espontáneos como de UV. Se cuantificaron las actividades proteasa, así como las cantidades de proteínas secretadas y las actividades enzimáticas pertinentes (por ejemplo, celulasa, xilanasas) a partir de los sobrenadantes de cultivo para confirmar que las actividades proteasa se habían reducido en comparación con la cepa parental, pero que las cantidades de otras proteínas secretadas no lo estaban. Las cepas se inocularon desde medios inclinados de PD en matraces de agitación (50 ml de volumen de medio en un matraz de 250 ml). Cada una de las cepas se cultivó en dos matraces con el pH del medio ajustado (antes del tratamiento en autoclave del medio de cultivo) a 5,5 y 6,0. Los cultivos se realizaron a 30 °C, 250 rpm durante 7 días. Las muestras se tomaron y analizaron después de 3, 5 y 7 días de cultivo. El pH (que representa el crecimiento de la cepa), la cantidad de proteínas secretadas (método BioRad DC), las actividades celulasa (hidroxietilcelulosa y 4-metilumbeliferil- β -D-lactósido como sustratos), la actividad xilanasas (xilano de abedul como sustrato; Bailey *et al.*, 1992) y las actividades proteasa se midieron a partir de los sobrenadantes de cultivo. Las actividades proteasa se midieron usando hemoglobina (4,0 g en 100 ml de agua; a un pH de 4,7, 40 °C, reacción de 30 min; dando como resultado unidades de actividad HUT) y caseína (1,2 g en 100 ml de tampón de fosfato de amonio 30 mM; pH 7,0, 30 °C, 60 min de reacción) como sustratos.

Algunas de las cepas produjeron una actividad proteasa claramente reducida en comparación con sus progenitores (Tabla 1). Asimismo, una selección de las cepas produjo cantidades al menos similares de proteínas secretadas, actividades celulasa y/o xilanasas como su cepa parental. Algunas de las cepas incluso produjeron cantidades mayores de proteínas y actividades celulasa/xilanasas en comparación con su progenitor. No se detectaron diferencias obvias entre el progenitor y las cepas de proteasa baja en los patrones de proteína de los sobrenadantes de cultivo en los geles de SDS-PAGE al 12 %. (Criterio XT, Biorad).

Tabla 1. Actividades proteasa relativas (HUT) medidas a partir de los sobrenadantes de cultivo de los mutantes de proteasa baja cultivados en matraces de agitación durante 7 días. A. Mutantes de proteasa baja derivados de A31. B. Mutantes de proteasa baja derivados de A33. Se incluyen resultados de una selección de cepas que produjeron menos actividades proteasa, pero cantidades al menos similares de proteínas secretadas y actividades enzimáticas con respecto a la cepa parental en el cultivo. Las cepas con el código NSP o SP son mutantes espontáneos, las que tienen el código UV derivan de esporas tratadas con irradiación UV. TMM y AMM, placa de selección utilizada (véase el Ejemplo 1 para más detalles); SUI50 - SUI500, concentración del reactivo SUI en la placa utilizada en el cribado primario. pH 5,5 y pH 6,0, el pH del medio de cultivo, ajustado antes del tratamiento en autoclave.

A.			
N.º de cepa	Placa de cribado primario	Actividad proteasa (HUT relativa)	
		pH 5,5	pH 6,0
A31		100	100
31NSP n.º 1	TMM-SUI50	105	35
31SP n.º 4	TMM-SUI100	21	41
31SP n.º 7	TMM-SUI500	37	49
31UV n.º 22	TMM-SUI50	14	95
31NSP n.º 6	TMM-SUI50	36	44
31NSP n.º 7	TMM-SUI50	35	107
31NSP n.º 8	TMM-SUI50	62	42
B.			
N.º de cepa	Cribado primario	Actividad proteasa (HUT relativa) pH	
		5,5	pH 6,0
A33		100	100
33SP n.º 9	AMM-SUI25	29	43
33SP n.º 11	AMM-SUI100	27	44
33SP n.º 12	AMM-SUI150	26	34
33UV n.º 48	TMM-SUI50	26	50
33UV n.º 64	AMM-SUI50	27	39
33UV n.º 68	AMM-SUI50	30	37
33UV n.º 82	AMM-SUI50	37	45

Una selección de cepas derivadas de A31 y A33 se cultivó en biorreactores a escala de laboratorio en medio complejo inductor de celulasa. Las cantidades de proteínas secretadas, las actividades enzimáticas pertinentes (por ejemplo, actividades celulasa y xilanasas) y las actividades proteasa se analizaron a partir de los medios de cultivo gastados. Los resultados obtenidos confirmaron las características de proteasa baja de la mayoría de las cepas. Las muestras elegidas de las fermentaciones se usaron además para el análisis y la identificación de las proteasas secretadas en

los medios de cultivo mediante el uso de separación de proteínas, IEF y análisis de zimograma y cartografiado de masa peptídica (Ejemplo 4). Se recogieron muestras de micelios fúngicos de las fermentaciones para el análisis de expresión por transferencia Northern (Ejemplo 4).

5 **Ejemplo 3. Cepas de proteasa baja como hospedadores para la producción de proteínas homólogas y heterólogas.**

Se sometieron a ensayo cepas elegidas de proteasa baja derivadas de A31 (31SP n.º 4, 31UV n.º 22 y 31NSP n.º 6) y A33 (33SP n.º 11, 33UV n.º 82, 33SP n.º 9 y 33UV n.º 48) como hospedadores para expresar dos genes que codifican proteínas heterólogas, que se sabe, por estudios de expresión realizados anteriormente, que son sensibles a proteasas cuando se produce en cepas de *T. reesei*. Los genes expresados en las cepas de proteasa baja elegidas fueron los siguientes: endoglucanasa modificada derivada de *Melanocarpus albomyces* denominada 20K+CBD (con un enlazador sensible a proteasas "WGEI"; expresada a partir del casete PALK1769; documento EP1874927) y transglutaminasa de *Streptomyces mobaraensis* (TGasa; Washizu *et al.*, 1994). Los genes se expresaron a partir del promotor nativo *cbh1* (*cel7A*) de *T. reesei*. El gen que codifica 20K+CBD se fusionó directamente con el promotor *cbh1*, pero el gen de TGasa (región codificante de proteína pro/madura) se fusionó en 3' con la secuencia codificante de polipéptido transportador de *T. reesei* Man5A (fusionado con el promotor *cbh1*) de manera similar a como se describió para la expresión del gen de xilanasa en Paloheimo *et al.* (2003). El gen *amdS* (acetamidasa) se usó como marcador en los dos casetes de expresión.

Los casetes de expresión lineales se aislaron de las estructuras principales vectoriales y se transformaron en protoplastos preparados a partir de las cepas de proteasa baja. Las transformaciones se realizaron como en Penttilä *et al.* (1987) con las modificaciones descritas en Karhunen *et al.* (1993). Los transformantes se purificaron en placas de selección de acetamida a través de conidios simples antes de esporularlos en PD. Los transformantes se inocularon desde los medios inclinados de PD en matraces de agitación que contenían 50 ml de medio complejo inductor de celulasa a base de lactosa (Joutsjoki *et al.*, 1993) tamponado con un KH_2PO_4 al 5 % y pH ajustado a 5,5 o 6,0. Se analizó la producción de enzimas de los transformantes a partir de los sobrenadantes de cultivo después de haberlos cultivado durante 7 días a 30 °C, 250 rpm. Los transformantes elegidos se cultivaron también en biorreactores a escala de laboratorio usando un medio complejo inductor de celulasa y se realizó un análisis de la producción enzimática. La producción de proteínas recombinantes y su estabilidad en los caldos de cultivo se analizaron a partir de los sobrenadantes de cultivo mediante ensayos de actividad enzimática y la ejecución de muestras en geles de SDS-PAGE. Para la detección de TGasa también se realizó un análisis por transferencia Western usando en la detección un anticuerpo comercial para la transglutaminasa bacteriana. La estabilidad de la proteína recombinante se analizó incubando muestras de los sobrenadantes de cultivo a diferentes temperaturas durante diferentes períodos de tiempo y analizándolas usando el método de SDS-PAGE (y/o transferencia Western).

Varios de los transformantes obtenidos produjeron una mayor cantidad de proteína de 20K+CBD de longitud completa, en comparación con la cepa parental (Fig. 1A). En el hospedador de proteasa baja la proteína 20K+CBD no se degradó después de 7 días de cultivo, como se demostró cuando se usó un hospedador del mismo linaje de cepa (pero no un mutante de proteasa baja) para la producción de la misma proteína. La estabilidad claramente mejor de la 20K+CBD producida en los hospedadores de proteasa baja también se demostró en el análisis de los cultivos de fermentación (Fig. 1B). En éstos, la 20K+CBD permaneció en forma de longitud completa mientras que la CBD se escindió en el hospedador no de proteasa baja, dando como resultado una forma de proteína 20K. De acuerdo con el análisis por SDS-PAGE y transferencia Western, las cantidades de TGasa producidas por los transformantes de las cepas de proteasa baja fueron algo mayores que las cantidades producidas por los progenitores A31 y A33. Asimismo, la TGasa producida por las cepas de proteasa baja era más estable, ya que en las muestras de fermentación de estas cepas eran visibles menos productos de degradación de la TGasa en comparación con las muestras correspondientes producidas por las cepas parentales.

Además de su uso como hospedadores para proteínas heterólogas, las cepas de proteasa baja se han utilizado satisfactoriamente como hospedadores para proteínas homólogas de *T. reesei*.

Ejemplo 4. Proteasas producidas y expresadas por las cepas mutantes de proteasa baja.

55 **Identificación de proteasas que no se producen o que son menos abundantes en las cepas de proteasa baja**

Varias cepas mutantes de proteasa baja mostraron actividades proteasa muy reducidas en comparación con su progenitor (Ejemplos 1 - 3). Se realizó un estudio sobre inhibidores de proteasas para analizar con más detalle qué tipo o tipos de proteasas no se producían o eran menos abundantes en los sobrenadantes de cultivo de las cepas de proteasa baja en comparación con sus progenitores. Se realizó un análisis de las actividades proteasa de los sobrenadantes de cultivo en ausencia y en presencia de inhibidores de proteasas, E64 0,01 mM, EDTA 10 mM, leupeptina 0,04 mg/ml, Pefabloc 1 mM, pepstatina 0,01 mM y 0,02 comprimidos/ml de Complete™ para inhibir cisteína proteasa, proteasa dependiente de catión divalente, serina/cisteína proteasa, serina proteasa, aspartil proteasa y diversas clases de proteasas, respectivamente. Un efecto reducido de un inhibidor específico de la actividad proteasa indicó que la cepa mutante era deficiente para el tipo de proteasa que se sabe que es inhibida por este inhibidor. Los resultados obtenidos indicaron que las principales actividades proteasa en los sobrenadantes de cultivo de *T. reesei*

se debían a las proteasas de tipo aspartil y serina proteasa. Estas actividades se redujeron claramente en los sobrenadantes de cultivo de varias cepas mutantes. No se observó ninguna inhibición de la actividad proteasa en los sobrenadantes de cultivo de varios mutantes por pepstatina (a un pH de 5,6), Pefabloc (a un pH de 4,0) o leupeptina (a un pH de 5,6), lo que indica que en estas cepas las aspartil y/o serina proteasas estaban en gran parte ausentes.

5 Los resultados obtenidos mostraron que varias de las cepas de proteasa baja se vieron afectadas en múltiples proteasas. Además de los estudios sobre inhibidores de proteasas descritos anteriormente, se realizaron diversos enfoques de separación de proteínas para identificar de los progenitores las proteasas que no se producían o que eran menos abundantes en los sobrenadantes de cultivo de los mutantes de proteasa baja. Estos métodos incluían ejecuciones en gel de SDS-poliacrilamida, PAGE nativa, análisis en gel de IEF (enfoque isoeléctrico) y análisis de zimograma usando geles proteínicos a base de caseína. Para reducir el fondo de celulasas y hemicelulasas y permitir una mejor identificación de las bandas de proteínas restantes, las muestras para ejecuciones en gel/IEF y análisis de zimograma se prepurificaron (pre-absorción) en primer lugar usando matrices de celulosa. De acuerdo con el análisis por SDS-PAGE de la fracción de proteína no unida, se descubrió que varias bandas de proteína estaban ausentes en las muestras derivadas de los mutantes de proteasa en comparación con los progenitores. Sin embargo, también aparecieron nuevas bandas en las muestras de los cultivos de mutantes de proteasa. Las diferencias en los patrones de proteínas secretadas entre las muestras de los progenitores y las cepas de proteasa baja también se detectaron en el análisis por IEF. Para analizar si las bandas diferenciales identificadas en los geles de SDS-PAGE y IEF eran proteasas o correspondían, por ejemplo, a proteínas procesadas incorrectamente, o a proteínas que en las muestras de tipo silvestre habían experimentado procesamiento proteolítico de dominios proteínicos específicos (por ejemplo, módulos de CBM), se realizó un análisis de zimograma basado en la actividad proteasa. Usando este tipo de análisis pudieron identificarse al menos seis bandas de proteínas diferentes con actividad proteolítica. El patrón del zimograma del tipo silvestre y de las muestras mutantes reveló varias diferencias entre estas muestras. Para algunas de las bandas de proteasa no estaba claro si estaban ausentes de las cepas de proteasa baja o si solo eran menos abundantes.

25 Las bandas de proteínas que diferían en las cepas se extrajeron de los geles y se realizó un análisis por EM/EM. En la identificación de las proteasas se usaron datos de secuencia de proteínas que pueden obtenerse del genoma QM6a versión 2.0 (Trire2) de *Trichoderma reesei* en <http://genome.jgi-psf.org/Trire2/Trire2.home.html> (en lo sucesivo en el presente documento, los números de ID derivados de este genoma se denominarán con un prefijo QM_). En total se identificaron ocho proteasas diferentes, cuatro de las cuales estaban claramente ausentes en una o más de las cepas de proteasa baja.

30 Para encontrar proteasas adicionales que faltan o que se están produciendo a niveles más bajos por las cepas de proteasa baja, también se realizó un análisis por nano-CL-EM (Proxeon nLC2 y Orbitrab Elite, Thermo Fischer) para el conjunto completo de proteínas de los sobrenadantes de cultivo de varias cepas de *T. reesei*, incluyendo, por ejemplo, 31UV n.º 22. Los datos de EM obtenidos se analizaron usando el programa Proteome Discoverer frente a la secuencia del genoma público de *T. reesei* (Trire2). En este análisis se identificaron en total 13 proteasas de *T. reesei* secretadas. De éstas, al menos cinco proteasas estaban claramente ausentes o se producían en niveles muy bajos en la cepa mutante de proteasa baja en comparación con el progenitor.

40 **Expresión de proteasas endógenas en cepas de proteasa baja, análisis por transferencia Northern y micromatriz**

45 Para analizar los niveles de expresión de los genes que codifican proteínas elegidos, se aisló ARN de muestras recogidas de siete fermentaciones a escala de laboratorio (progenitores y cinco cepas de proteasa baja), de cuatro puntos temporales (incluyendo las fases logarítmica y estacionaria) de los cultivos. Las cepas elegidas para el análisis fueron las siguientes: A31, A31SP n.º 4, A31UV n.º 22, A33, A33SP n.º 9, A33UV n.º 48 y A33UV n.º 82. Se estudió la expresión de las ocho proteasas, anteriormente identificadas a partir de los sobrenadantes de cultivo de *T. reesei*. Las sondas se prepararon mediante PCR, basándose en secuencias de la base de datos pública de *T. reesei*. Las sondas tenían aproximadamente 600 pb de longitud, en cada caso, y consistían en fragmentos internos de la secuencia codificante de los 8 genes de proteasa respectivos. Como sonda de referencia, se usó un fragmento de PCR *gpdI* de aproximadamente 600 pb (QM_ID119735).

55 Los resultados del análisis por transferencia Northern mostraron que la expresión de siete de los ocho genes que codifican proteasa se había visto afectada (sin niveles de expresión o con niveles muy bajos) en todos los mutantes sometidos a ensayo. Se realizó un perfil transcripcional adicional de una de las cepas de proteasa baja usando una micromatriz de oligonucleótidos (Roche NimbleGen Inc., EE.UU.). Se recogieron micelios en tres puntos temporales de tres fermentaciones replicadas a escala de laboratorio de las cepas 33SP n.º 9 y de la cepa de tipo silvestre A33 y se extrajo ARN total de las muestras. La síntesis de ADNc, el marcaje, la hibridación, el cribado de micromatrices y la detección de señales de las muestras se realizaron de acuerdo con las instrucciones de Roche NimbleGen. Se usaron portaobjetos de micromatrices personalizados que contenían sondas de 60 monómeros diseñadas basándose en la secuencia del genoma público de *T. reesei* de <http://genome.jgi-psf.org/Trire2/Trire2.home.html>. Los datos de micromatrices se analizaron para determinar genes de proteasa expresados diferencialmente con un corte de significación estadística en $P < 0,01$ usando los paquetes R Oligo, Limma y Mfuzz (<http://www.bioconductor.org/>).

65 Basándose en los resultados de las micromatrices, la expresión de varios genes de proteasa estaba regulada

negativamente en el mutante de proteasa baja 33SP n.º 9. Además de las proteasas identificadas anteriormente, en total se descubrieron al menos 18 proteasas adicionales con una expresión claramente reducida.

Los resultados obtenidos del análisis de proteína y ARN se resumen en la Tabla 2.

5

Tabla 2. Proteasas que están ausentes o son menos abundantes en los sobrenadantes de cultivo y/o que tienen un nivel de expresión más bajo en los mutantes de proteasa baja en comparación con sus progenitores. Las proteasas se agruparon de acuerdo con la base de datos de péptidos MEROPS (<http://merops.sanger.ac.uk/>). N.º, número de proteasas individuales que pertenecen al grupo.

Grupo de Proteasas (MEROPS)	N.º	Familias representadas (MEROPS)
Metalo Peptidasa (M)	11	M1, M3, M6, M14, M18, M28
Serina Peptidasa (S)	7	S1, S28, S8/S53
Aspártico Peptidasa (A)	5	A1
Glutámico Peptidasa (G)	2	G1
Peptidasa mixta (P)	1	P1

10

Ejemplo 5. Secuenciación y comparación del genoma.

Se aisló ADN genómico de micelio liofilizado y molido de cepas seleccionadas de proteasa baja con el Mini Kit de ADN Fúngico E.Z.N.A.® SP (Omega Bio-Tek Inc., EE.UU.) de acuerdo con las instrucciones del fabricante. Los genomas se secuenciaron usando el método Illumina (Solexa) y los borradores de genomas se ensamblaron frente al genoma público de *Trichoderma reesei* RutC-30 versión 1.0 (TrireRUTC30_1) disponible en http://genome.jgi.doe.gov/TrireRUTC30_1/TrireRUTC30_1.home.html. Los números de ID derivados de este genoma se denominarán en lo sucesivo en el presente documento con un prefijo Rut_. Se analizaron todas las diferencias de los genomas frente al genoma público y se compararon los perfiles de mutación entre las cepas de proteasa baja. De acuerdo con la secuenciación del genoma, tres cepas individuales de proteasa baja tenían mutaciones en la región codificante de un gen previsto Rut_ID85889 (SEQ ID NO: 4-6). El gen correspondiente en la cepa 33SP n.º 9 (SEQ ID NO: 7) contenía una inserción de dos nucleótidos dentro de la región codificante del gen previsto 840 pb corriente abajo del codón de inicio. En la cepa 31UV n.º 22 (SEQ ID NO: 9), el gen tenía una delección de un nucleótido de la región codificante 968 pb corriente abajo del codón de inicio. De acuerdo con la anotación del gen Rut_ID85889, tanto la inserción como la delección descritas anteriormente dan como resultado un desplazamiento de marco y la formación de un codón de parada precoz corriente abajo de las mutaciones. La mutación en 31SP n.º 4 (SEQ ID NO: 8) es una mutación puntual simple 1224 pb corriente abajo del codón de inicio que da como resultado la formación de un codón de parada precoz.

15

20

25

30

Para la cepa A21, el correspondiente gen de longitud completa Rut_ID85889 se amplificó mediante PCR a partir del ADN genómico de A21 y se secuenció directamente a partir del fragmento de PCR usando el analizador genético ABI PRISM® 310 de Applied Biosystems (Thermo Fisher Scientific Inc., EE.UU.). Se descubrió que la secuencia de nucleótidos del gen correspondiente en la cepa A21 (SEQ ID NO: 10) contenía una mutación puntual simple 952 pb corriente abajo del codón de inicio que dio como resultado la formación de un codón de parada precoz.

35

Todas las mutaciones descritas anteriormente inactivan el marco de lectura abierto de longitud completa del gen Rut_ID85889 y los genes mutados, cuando se traducen, codifican productos proteínicos truncados. El supuesto Rut_ID85889 se denominó gen que afecta a la expresión de proteasas, *pea1*. El gen *pea1* en el genoma de TrireRUTC30_1 tiene 2749 pb de largo incluyendo el codón de parada y contiene dos intrones, un intrón de 191 pb de largo 1029 pb corriente abajo del codón de inicio y un intrón de 80 pb de largo 1402 pb corriente abajo del codón de inicio. La anotación del gen *pea1* de RutC-30 difiere de la anotación del gen en la región del genoma correspondiente en el genoma de Trire2, QM_ID123125 (SEQ ID NO: 1-3). La secuencia del hipotético gen de QM6a QM_ID123125 corresponde a la secuencia de nucleótidos C-terminal del Rut_ID85889. QM_ID123125 tiene 961 pb de largo y tiene un intrón de 42 pb de 383 pb corriente abajo del sitio de inicio (Fig. 2). Debido a las discrepancias en la anotación de los genes de Rut_ID85889 y QM_ID123125, la síntesis y la secuenciación del ADNc de *pea1* ADNc se realizó a partir de una muestra de ARN de QM6a (Ejemplo 6).

45

Ejemplo 6. La anotación del gen *pea1* y comparación de secuencias

Con el fin de confirmar la secuencia de nucleótidos del gen *pea1* y su locus, se clonó un fragmento de 4,7 kb mediante PCR usando ADN genómico de QM6a como molde. El fragmento se amplificó usando los cebadores S-*ppea1* (cebador sentido CGTTGCTCGAGGCAACTGC) y AS-3UTRout16 (cebador antisentido TGTCATCATGTCTTATTCA). Las mezclas de reacción de PCR contenían 1 x tampón Phusion HF (Thermo Fisher Scientific Inc., EE.UU.), dNTPs 0,23 mM, cada cebador 1,3 µM y 1,3 unidades de ADN polimerasa Phusion High-Fidelity (Thermo Fisher Scientific Inc., EE.UU.) por cada 50 µl de volumen de reacción. Las condiciones para las reacciones de PCR fueron las siguientes: 1 min de desnaturalización inicial a 98 °C, seguido de 29 ciclos de 10 s a 98 °C, 30 s de hibridación a 63 °C, 1 min de extensión a 72 °C y una extensión final a 72 °C durante 5 min. El fragmento de PCR de 4,7 kb resultante se cortó del gel de agarosa y se aisló usando el kit de extracción de gel QIAquick (Qiagen GmbH, Alemania). El fragmento

50

55

purificado se clonó en el vector PCR®4 BluntTOPO® usando el kit de clonación de PCR Zero Blunt® TOPO® (Thermo Fisher Scientific Inc., EE.UU.). El plásmido resultante se denominó pALK3535 y la cepa de *Escherichia coli* (TOP10) que incluía el plásmido, RF11697, se depositó en la colección DSM con el número de acceso DSM32007. El fragmento de PCR en pALK3535 contiene el gen de RutC-30 de longitud completa ID: 85889 y secuencias 1140 pb corriente arriba y 821 pb corriente abajo (SEQ ID NO: 11). Este fragmento se secuenció usando la tecnología ABI PRISM® 310 como se describe en el Ejemplo 5. La secuencia fue idéntica a la secuencia de nucleótidos de los genomas públicos Trire2 y TrireRUTC30_1.

Para el análisis del ADNc, se aisló ARN total del micelio congelado de QM6a cultivado en medio inductor de celulosa (Joutsjoki *et al.*, 1993) con el Mini Kit RNeasy® Plant (Qiagen GmbH, Alemania) y la traducción de ARNm a ADNc del ARN aislado se realizó con el Kit de Síntesis de ADNc Transcriptor First Strand ADNc Synthesis Kit (Roche Diagnostics GmbH, Alemania) de acuerdo con las instrucciones del fabricante. El ADNc se amplificó mediante PCR usando los cebadores específicos S-5'UTR26 (cebador sentido CCAGAACAGCTCCGTCCTGG) y AS-3'UTRout16. Las mezclas de reacción de PCR contenían 1 x Q5 tampón de reacción (New England Biolabs Inc., EE.UU.), dNTPs 0,2 mM, cada cebador 0,5 µM y 2 unidades de ADN polimerasa de alta fidelidad Q5® (New England Biolabs Inc., EE.UU.) y aproximadamente 2 µl de ADNc por cada 50 µl de volumen de reacción. Las condiciones para las reacciones de PCR fueron las siguientes: 1 min de desnaturalización inicial a 98 °C, seguido de 31 ciclos de 10 s a 98 °C, 30 s de hibridación a 63 °C, 1 min 20 s de extensión a 72 °C y una extensión final a 72 °C durante 4 min. El fragmento resultante de PCR de 4,1 kb se cortó y se aisló del gel de agarosa. El fragmento purificado se clonó en el vector PCR®4 BluntTOPO® usando el kit de clonación de PCR Zero Blunt® TOPO® (Thermo Fisher Scientific Inc., EE.UU.). El plásmido resultante se denominó pALK3536 y la cepa de *Escherichia coli* (TOP10) que incluía el plásmido, RF11698, se depositó en la colección DSM con el número de acceso DSM32008. El ADNc en el pALK3536 incluye 654 pb de la 5'UTR (región no traducida) y 821 pb de la 3'UTR (SEQ ID NO: 12). El fragmento se secuenció y la secuencia se comparó con el correspondiente gen *pea1* clonado a partir de QM6a (SEQ ID NO: 11). Los resultados mostraron que los sitios de inicio y parada del gen *pea1* y el segundo intrón eran como se predijo para el Rut_ID85889, pero contrarios a la anotación de Rut_ID85889, el primer intrón de *pea1* tiene 62 pb de largo y se ubica 1158 pb corriente abajo del codón de inicio (Fig 3).

La secuencia de nucleótidos del gen *pea1* de longitud completa (SEQ ID NO: 11, nucleótidos 1141-3889) y la secuencia de aminoácidos deducida (SEQ ID NO: 13) se usaron para buscar secuencias similares en fuentes públicas. Las búsquedas se realizaron usando las herramientas de búsqueda FASTA en el sitio web de EMBL-EBI, usando la base de datos de secuencias ENA para la búsqueda de nucleótidos (www.ebi.ac.uk/Tools/sss/fasta/nucleotide.html) y la base de conocimientos UniProt para la búsqueda de proteínas (www.ebi.ac.uk/Tools/sss/fasta/). Las búsquedas se hicieron usando los valores por defecto. Además, se hicieron búsquedas a partir de las secuencias genómicas disponibles de las cepas pertenecientes al género *Trichoderma*. Las secuencias genómicas de *Trichoderma* utilizadas en las búsquedas fueron las siguientes: *Trichoderma citrinoviride* (<http://genome.jgi.doe.gov/Trici1/Trici1.home.html>), *Trichoderma longibrachiatum* (<http://genome.jgi.doe.gov/Trilo1/Trilo1.home.html>), *Trichoderma virens* (http://genome.jgi-psf.org/TriviGv29_8_2/TriviGv29_8_2.home.html), *Trichoderma harzianum* (<http://genome.jgi.doe.gov/Triha1/Triha1.home.html>), *Trichoderma asperellum* (<http://genome.jgi.doe.gov/Trias1/Trias1.home.html>), *Trichoderma atroviride* (<http://genome.jgi.doe.gov/Triat2/Triat2.home.html>). Los valores de identidad (%) con las secuencias más similares identificadas en las búsquedas se determinaron usando la Herramienta de Alineación de Secuencias por Pares del sitio web de EMBL-EBI (para secuencias de nucleótidos: www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html; para las secuencias de proteínas: www.ebi.ac.uk/Tools/psa/emboss_needle/ usando los valores predeterminados (Abertura de hueco: 10 y extensión de hueco: 0,5). Los resultados se muestran en las Tablas 3A y 3B. Las identidades más altas eran con las secuencias homólogas de otras especies de *Trichoderma*. El porcentaje de identidad más alto con una secuencia no de *Trichoderma* se obtuvo con una secuencia hipotética de *Ophiocordyceps sinensis* OCS_06053, con un 59,6 % de identidad en el nivel de los nucleótidos y un 58,3 % de identidad en el nivel de las proteínas.

La secuencia de aminoácidos de Pea1 se alineó con las secuencias homólogas obtenidas de otras especies de *Trichoderma* y con secuencias que tenían más del 50 % de identidad con la proteína Pea1, de acuerdo con los resultados de la búsqueda de proteínas FASTA. Se detectó una región altamente conservada a partir de la alineación. Se seleccionó una secuencia por género de los resultados de la búsqueda para su análisis adicional. La identidad entre las especies de *Trichoderma* en la región altamente conservada de Pea1, de Arg₄₀₂ a Pro₅₃₃ (132 restos), es de al menos un 97 % y una similitud del 99 %, mientras que esta región tenía al menos un 90 % de identidad y un 96 % de similitud con las secuencias derivadas de otras especies fúngicas filamentosas, seleccionadas entre los resultados de búsqueda FASTA (Tabla 3C). Para determinar el grado de identidad se usaron las regiones de secuencia correspondientes, como se muestra en la Fig. 4. Taxonómicamente (<http://www.mycobank.org>), todas las secuencias seleccionadas se originan en especies pertenecientes a los *Sordariomycetes*, subclase *Hypocreomycetidae* y orden *Hypocreales*, lo que indica que esta región está altamente conservada, especialmente en *Hypocreales*. También se encontraron valores altos, un 90,2 % de identidad y un 96,2 % de similitud, por ejemplo, en la secuencia de *Scedosporium apiospermum* (SAPIO_CDS0483). La especie *S. apiospermum* también pertenece a la subclase *Hypocreomycetidae*, Orden *Microascales*.

La región altamente conservada de Pea1 contiene un dominio pfam previsto Clr5 (PF14420). El dominio Clr5 se ubica

en la posición Ala₄₁₀ a Lys₄₆₂ (53 restos) en la secuencia de Pea1. Se ha demostrado que el dominio Clr5 está implicado en el silenciamiento en levadura de fisión (Hansen *et al.*, 2011).

- 5 **Tabla 3A. Los valores de identidad (%) obtenidos a partir de la Alineación de Secuencia por Pares de la secuencia de nucleótidos del gen *pea1* de longitud completa (SEQ ID NO: 11, nucleótidos 1141-3889).** Se usó EMBOSS Needle (EMBL-EBI, EMBOSS-Needle - Alineación de Secuencias por Pares, Matriz DNFull, abertura de hueco 10, extensión de hueco 0,5) en www.ebi.ac.uk/Tools/psa/emboss_needle/nucleotide.html para determinar el grado de identidad.

Nombre	Identidad
Rut_ID85889	100
<i>Trichoderma citrinoviride</i> ID: 7704 (v1.0)	90,0
<i>Trichoderma longibrachiatum</i> ID: 60713 (v1.0)	89,7
<i>Trichoderma virens</i> ID: 58331 (v2.0)	81,2
<i>Trichoderma harzianum</i> ID: 235354 (v1.0)	80,6
<i>Trichoderma asperellum</i> ID: 84188 (v1.0)	76,9
<i>Trichoderma atroviride</i> ID: 280821 (v2.0)	76,7
<i>Ophiocordyceps sinensis</i> OCS_06053	59,6

- 10 **Tabla 3B. Los valores de identidad y similitud (%) obtenidos a partir de la Alineación de Secuencia por Pares de la secuencia de aminoácidos de Pea1 de longitud completa (SEQ ID NO: 13, aminoácidos 1-868).** Se usó EMBOSS Needle (EMBL-EBI, EMBOSS-Needle - Alineación de Secuencias por Pares, Matriz BLOSUM62, abertura de hueco 10, extensión de hueco 0,5) en www.ebi.ac.uk/Tools/psa/emboss_needle/ para determinar el grado de identidad y similitud.

Nombre	Identidad	Similitud
Rut_ID85889	95,0	95,0
<i>Trichoderma citrinoviride</i> ID: 7704 (v1.0)	96,0	96,0
<i>Trichoderma longibrachiatum</i> ID: 60713 (v1.0)	91,5	93,9
<i>Trichoderma harzianum</i> ID: 235354 (v1.0)	88,3	93,3
<i>Trichoderma virens</i> ID: 58331 (v2.0)	85,9	90,7
<i>Trichoderma atroviride</i> ID: 280821 (v2.0)	83,0	89,6
<i>Trichoderma asperellum</i> ID: 84188 (v1.0)	82,9	90,1
<i>Ophiocordyceps sinensis</i> OCS_06053	58,3	69,9

- 15 **Tabla 3C. Los valores de identidad y similitud (%) obtenidos a partir de la Alineación de Secuencia por Pares de los aminoácidos de la región altamente conservada de Pea1 (SEQ ID NO: 13, aminoácidos 402-533) con la región correspondiente en otras secuencias.** Se usó EMBOSS Needle (EMBL-EBI, EMBOSS-Needle - Alineación de Secuencias por Pares, Matriz BLOSUM62, abertura de hueco 10, extensión de hueco 0,5) en www.ebi.ac.uk/Tools/psa/emboss_needle/ para determinar el grado de identidad y similitud.

Nombre	Identidad	Similitud
<i>Trichoderma citrinoviride</i> ID: 7704 (v1.0)	100	100
<i>Trichoderma longibrachiatum</i> ID: 60713 (v1.0)	100	100
<i>Trichoderma atroviride</i> ID: 280821 (v2.0)	98,5	100
<i>Trichoderma asperellum</i> ID: 84188 (v1.0)	98,5	100
<i>Trichoderma harzianum</i> ID: 235354 (v1.0)	97,0	100
<i>Trichoderma virens</i> ID: 58331 (v2.0)	97,0	99,2
<i>Fusarium oxysporum</i> FOVG_08585	95,5	97,7
<i>Gibberela fujikuroi</i> FFUJ_12153	95,5	97,7
<i>Stachybotrys chartarum</i> S40293_07230	94,7	100
<i>Claviceps purpurea</i> CPUR_05697	92,4	97,0
<i>Ophiocordyceps sinensis</i> OCS_06053	91,7	98,5
<i>Nectria haematococca</i> NECHADRAFT 85885	91,7	98,5
<i>Metarhizium acridum</i> MAC_08836	91,7	97,7
<i>Villosiclava virens</i> UV8b_6262	91,7	96,2
<i>Acremonium chrysogenum</i> ACRE_079620	90,2	97,0

Ejemplo 7. Construcción de casetes para suprimir el gen *pea1* de longitud completa y parcial de *T. reesei*.

En total se planificaron y construyeron tres casetes de delección, pALK4104 (Fig. 5A), pALK4106 (Fig. 5B) y pALK4107 (Fig. 6). pALK4104 y pALK4107 se construyeron para suprimir el gen *pea1* de longitud completa y pALK4106 para la delección parcial de *pea1* (truncamiento) de los genomas de las cepas hospedadoras de *T. reesei*. La longitud de la secuencia de aminoácidos deducida de la *Pea1* truncada codificada por el pALK4106 (297 aminoácidos) está en el intervalo de la proteína mutante *Pea1* deducida en las cepas A21, 31SP n.º 4, 33SP n.º 9 y 31UV n.º 22 (Fig 7). Todos los casetes contienen un marcador de selección rodeado de regiones flanqueantes para dirigir el casete a una ubicación prevista en el genoma de *T. reesei*. Para obtener detalles, véase a continuación.

El vector pUC19 se usó como estructura principal en las construcciones de plásmidos. Los métodos comunes de biología molecular se usaron en los tratamientos enzimáticos del ADN, la PCR (reacción en cadena de la polimerasa), las transformaciones de *E. coli* y el aislamiento de ADN plasmídico y fragmentos de ADN para ligaduras y transformaciones. Una preparación de ADN genómico aislado de QM6a se usó como molde en todas las reacciones de PCR.

El casete de delección pALK4104 contiene:

- Una región flanqueante 5' de *pea1* para dirigir el casete dentro del locus de *pea1* para el reemplazo del gen, junto con la región flanqueante 3' (véase a continuación). La región flanqueante 5' es el fragmento genómico *Sall* - *XbaI* de 1578 pb, ubicándose el sitio *XbaI* 531 pb aguas arriba del inicio del gen *pea1* (primer ATG que codifica Met). El fragmento se sintetizó mediante PCR.
- *amdS* sintética (acetamidasa) que codifica el marcador de selección de acetamidasa. En el casete de delección se usó un ADNc del gen nativo *amdS* de *Aspergillus nidulans* con modificaciones adicionales (delección de los sitios de restricción elegidos). El gen codifica la secuencia original de aminoácidos de *AmdS*.
- Una región flanqueante 3' de *pea1* para dirigir el casete dentro del locus de *pea1* para el reemplazo del gen, junto con la región flanqueante 5' (véase anteriormente). La región flanqueante 3' es el fragmento genómico *KpnI* - *XbaI* de 2676 pb, ubicándose el sitio *KpnI* 60 pb corriente abajo del codón de parada del gen *pea1* (TAG). Este fragmento se sintetizó mediante PCR. Incluye todos los genes anotados en esta región, de acuerdo con las secuencias genómicas públicas Trire2 o TrireRUTC30_1. El sitio *XbaI* en el extremo 3' del fragmento no está disponible en la construcción final debido al relleno en la reacción (por Klenow) realizado al construir el plásmido.

El casete de delección pALK4106 contiene:

- Una región flanqueante 5' de *pea1* para dirigir el casete dentro del locus/gen *pea1* para el reemplazo del gen, junto con la región flanqueante 3' (véase a continuación). La región flanqueante 5' contiene un promotor parcial de *pea1*, comenzando inmediatamente después del sitio *XbaI* en la región promotora (526 pb antes del codón de inicio del gen, el sitio *XbaI* no está incluido) y terminando inmediatamente antes del sitio *EcoRI* interno en el gen *pea1* (892 pb desde el inicio del gen, el sitio *EcoRI* no está incluido). Este fragmento se sintetizó mediante PCR. Codifica un producto truncado de 297 aminoácidos *Pea1* (SEQ ID NO: 18).
- *amdS* sintética (acetamidasa) que codifica el marcador de selección de acetamidasa. En el casete de delección se usó un ADNc del gen nativo *amdS* de *Aspergillus nidulans* con modificaciones adicionales (delección de los sitios de restricción elegidos). El gen codifica la secuencia original de aminoácidos de *AmdS*.
- Una región flanqueante 3' de *pea1* para dirigir el casete dentro del locus de *pea1* para el reemplazo del gen, junto con la región flanqueante 5' (véase anteriormente) era la misma que el fragmento flanqueante 3' utilizado en el pALK4104 (véase anteriormente).

El casete de delección pALK4107 contiene las regiones flanqueantes 5' y 3' idénticas a las incluidas en pALK4104. El gen *syn-amdS* en el pALK4104 (digestión por *XbaI* de pALK4104, rellena por Klenow) se reemplazó por el gen marcador de selección *ble* (con un promotor y terminador originarios de *Aspergillus nidulans*) que deriva del fragmento pAN8-1 (fragmento *BglII* - *XbaI* de 3313 pb, los extremos rellenos usando Klenow) y que codifica la resistencia a la fleomicina (para más detalles, véase la descripción de la Fig. 6). El casete de delección pALK4107 se usó para suprimir el gen de longitud completa *pea1* de dichas cepas de *T. reesei* que ya incluían el gen marcador *amdS*, debido, por ejemplo, a la transformación previa de un casete de expresión génica en la cepa.

Los casetes de delección pALK4104 de 6756 pb y pALK4106 de 6595 pb para las transformaciones de *T. reesei* se escindieron de las estructuras principales de los vectores mediante digestiones por *PstI* - *EcoRI*, se aislaron de geles de agarosa y se transformaron (como se describe en el Ejemplo 3) en protoplastos de una selección de cepas hospedadoras de *T. reesei*, en concreto, QM6a, RutC-30 y A33. Los transformantes se seleccionaron en placas de acetamida y se purificaron a través de esporas individuales antes de cultivarlos en estrías en las placas de PD.

Las transformaciones realizadas usando el casete de delección pALK4107 se describen en el Ejemplo 9.

Ejemplo 8. Caracterización del transformante pALK4104 y pALK4106

La producción de proteasa de una selección de transformantes QM6a, RutC-30 y A33 se analizaron cultivando las cepas en placas de leche desnatada. Las cepas hospedadoras se usaron como controles. De cada conjunto de transformantes se descubrieron transformantes que producían cantidades menores de proteasa en el ensayo en placa en comparación con su hospedador (Tabla 4). El locus *pea1* de los genomas de una selección de estos transformantes se analizó mediante el método de transferencia Southern. Se descubrió que el gen *pea1* se suprimió de los genomas de todos los transformantes de proteasa baja pALK4104 y estaba truncado en los genomas de todos los transformantes de proteasa baja pALK4106 analizados mediante transferencia Southern. De cada conjunto de transformantes se descubrieron cepas con un reemplazo satisfactorio del gen *pea1* con una copia (reemplazo de una sola copia) del marcador de selección *syn-amdS* (en transformantes pALK4104) y reemplazo del gen *pea1* parcial con *syn-amdS* (en transformantes pALK4106, lo que provocó el truncamiento del gen *pea1* en estas cepas) (Tabla 4).

Tabla 4. Sumario sobre los transformantes pALK4104 (delección del *pea1* de longitud completa) y pALK4106 (supresión/truncamiento parcial del *pea1*) analizados en placas de leche desnatada y mediante transferencia Southern. Se muestran cantidades de cepas de proteasa baja (halo reducido en comparación con el hospedador) y cepas de una sola copia (reemplazo correcto) de todos los transformantes analizados.

Cepa hospedadora	Casete de delección transformado	Transformantes de proteasa baja (ensayo de placa de leche desnatada)	Cepas de reemplazo de una sola copia (análisis por transferencia Southern)
QM6a	pALK4104	4/18	3/4
RutC-30	"	11/33	6/6
A33	"	14/31	6/6
QM6a	pALK4106	4/18	4/4
RutC-30	"	10/29	4/6
A33	"	12/30	6/6

Se eligieron tres cepas de reemplazo de una sola copia de cada transformación y se almacenaron en la colección de cultivos Roal. El fenotipo de proteasa baja de estas cepas se analiza adicionalmente cultivando los transformantes y sus hospedadores (para su comparación) en biorreactores a escala de laboratorio. En los cultivos se usa un medio complejo inductor de celulasa. Se espera que los resultados se correspondan con los obtenidos anteriormente de los cultivos de los mutantes de proteasa baja (Ejemplo 2): los transformantes con supresión y truncamiento del gen *pea1* producen actividades proteasa menores en comparación con sus hospedadores. Se espera que las cepas modificadas genéticamente con supresión o truncamiento de *pea1* produzcan cantidades similares o mejores de proteínas secretadas y/o actividades celulasa en comparación con sus hospedadores, ya que en estas cepas solo se ha modificado el locus de *pea1*.

Ejemplo 9. Delección de *pea1* de cepas que producen en exceso una enzima celulasa y una laccasa.

El casete de delección pALK4107 para las transformaciones de *T. reesei* se escindió de la estructura principal del vector mediante digestión por *Pst*I - *Eco*RI, se aisló de un gel de agarosa y se transformó en protoplastos de dos cepas construidas anteriormente que producen enzimas recombinantes. Las cepas transformadas fueron las siguientes: RF5969 que produce la 20K+CBD (expresión del casete pALK1769, Ejemplo 2) y RF5597 que produce una laccasa Talcc1, originada a partir de *Thielavia arenaria* (expresión del casete pALK1667, documento US7927849). En ambos casos, el gen que codifica la enzima recombinante se expresó usando el promotor nativo fuerte *cbh1* de *T. reesei* (*cel7A*). La transformación del casete de delección de pALK4107 en los protoplastos RF5969 y RF5597 se hizo como se describe en el Ejemplo 3, pero usando la selección de fleomicina para el cribado de los transformantes (Harkki *et al.*, 1991). Después de la purificación a través de esporas individuales, los transformantes se cultivaron en estrías en medios inclinados de PD.

La producción de proteasa de los transformantes se analizó usando placas de leche desnatada (como se explica en el Ejemplo 1) usando los hospedadores de transformación como controles. De ambas transformaciones se obtuvieron transformantes que producían cantidades menores de proteasas en comparación con su hospedador.

Los transformantes RF5969 pueden someterse a ensayo adicionalmente en placas indicadoras de celulasa que contienen, por ejemplo, Azo-CM-celulosa (Megazyme) y los transformantes RF5597 en placas indicadoras de laccasa que contienen ABTS (Roche) para confirmar la producción de 20K+CBD y laccasa, respectivamente, de estas cepas. Los hospedadores de transformación se usan como controles en los ensayos en placa.

Los transformantes elegidos con producción de proteasa baja y producción confirmada de la enzima recombinante pueden cultivarse en matraces de agitación y/o biorreactores usando condiciones de inducción de celulasa. La producción de proteasa reducida en comparación con los hospedadores puede demostrarse a partir de los sobrenadantes de cultivo mediante ensayo o ensayos de actividad. La producción aumentada y la mejor estabilidad de las enzimas recombinantes en las muestras de sobrenadante de cultivo de las cepas de proteasa baja en

comparación con los hospedadores pueden confirmarse mediante métodos conocidos.

Ejemplo 10. Caracterización de las cepas de producción con delección de *pea1*

5 Se caracterizó adicionalmente un conjunto de transformantes RF5597 y RF5969 que produjeron cantidades menores de proteasas en comparación con sus hospedadores en el ensayo en placa (Ejemplo 9). Un análisis por transferencia Southern confirmó que en todas estas cepas el gen *pea1* se reemplazó por el marcador de selección. Los hospedadores y los transformantes elegidos con delección confirmada del gen *pea1* se cultivaron en biorreactores de 0,5 l usando condiciones de inducción de celulasa. La actividad proteasa y otras actividades enzimáticas pertinentes se midieron a partir de los sobrenadantes de cultivo. La actividad proteasa (HUT) se midió usando sustrato de hemoglobina (como en el Ejemplo 2). Se analizó la actividad celulasa (NCU, "unidad de celulasa neutra") a partir de RF5969 y sus transformantes. Se usó carboximetilcelulosa (Sigma, CMC de viscosidad baja) como sustrato en este análisis. La reacción enzimática se realizó a un pH de 7,0, 50 °C durante 10 minutos y se usó el método de DNS para medir los extremos reductores liberados. Como se usó sustrato soluble en el análisis, no hay grandes diferencias en las actividades específicas entre las formas de celulasa 20K con y sin el dominio de unión (CBD/CBM). La actividad laccasa se midió a partir del RF5597 y sus transformantes a pH 4,5 usando ABTS como sustrato (Niku-Paavola *et al.*, 1988).

20 Los transformantes RF5597 y RF5969 con delección de *pea1* produjeron una actividad proteasa (HUT) claramente menor en comparación con sus hospedadores que tienen el gen *pea1* de tipo silvestre. Las actividades proteasa de los sobrenadantes de cultivo de los transformantes RF5597 fueron, en promedio, solo aproximadamente el 50 % y de los transformantes RF5969, en promedio, solo aproximadamente el 25 % de la actividad medida a partir de los sobrenadantes de cultivo de los hospedadores. La actividad celulasa (NCU) en los sobrenadantes de cultivo de los transformantes RF5969 aumentó hasta el 37 % en comparación con la actividad medida a partir del cultivo de RF5969. Sin embargo, no se detectaron aumentos de las actividades laccasa producidas por los transformantes RF5597, en comparación con RF5597. Para analizar la integridad y estabilidad de los productos de enzimas recombinantes, las muestras de los sobrenadantes de cultivo se ejecutaron en gel de SDS-PAGE. La banda de proteína laccasa TaLcc1 era similar (en masa y cantidad) a la del RF5597 y sus transformantes. Sin embargo, había diferencias claras en la proteína celulasa recombinante producida por RF5969 y sus transformantes con la supresión de *pea1* (Fig. 8). La proteína principal del sobrenadante de cultivo de RF5969 no era la 20K+CBD de longitud completa sino la forma de núcleo 20K de la que se había escindido CBD. Solo una cantidad muy pequeña de la 20K+CBD de longitud completa era detectable en el gel. Los transformantes RF5969 con supresión de *pea1* produjeron principalmente la forma 20K+CBD de longitud completa y solo una cantidad relativa muy baja de la forma 20K. Este resultado confirma que las cepas de supresión de *pea1* fueron capaces de producir cantidades mayores del producto recombinante y que la enzima recombinante en los sobrenadantes de cultivo de las cepas de supresión de *pea1* fue más estable que en el sobrenadante de cultivo del hospedador.

40 Las estabildades de los productos de TaLcc1 se estudiaron adicionalmente incubando muestras de sobrenadantes de cultivo a 30 y 50 °C (a pH 4) durante hasta tres días. Después de las incubaciones, las muestras se ejecutaron en gel de SDS-PAGE. La TaLcc1 recombinante era muy estable en todas las muestras. Sin embargo, después de 3 días de incubación a 50 °C la banda de proteína TaLcc1 se degradó claramente más en el sobrenadante de cultivo del hospedador (RF5597) en comparación con los sobrenadantes de los transformantes Δ *pea1* de RF5597. Este resultado confirma adicionalmente la mayor estabilidad de los productos obtenidos a partir de las cepas con un gen *pea1* no funcional.

45 Se obtuvieron resultados similares a partir de los transformantes RF5597 y RF5969 de los que se suprimió el gen *pea1* usando el casete de supresión pALK4116. En este casete, el gen marcador *ble* en pALK4107 se reemplazó por el gen marcador *hph* que codificaba la resistencia a la higromicina B (Mach *et al.*, 1994).

50 Ejemplo 11. Disrupción del homólogo de *pea1* de la especie *Fusarium*.

Muchas especies fúngicas contienen un homólogo del gen *pea1* de *T. reesei*, como se describe en el Ejemplo 6. Los homólogos de Pea1 de *T. reesei* de longitud completa codificados de *Fusarium oxysporum*, por ejemplo, FOVG_08585 y FOZG_02804 (aminoácidos 1 - 887) tienen valores de identidad del 57,2 y el 57,1 % y valores de similitud del 68,8 y el 68,7 %, respectivamente, con respecto a la Pea1 de *T. reesei* de longitud completa (SEQ ID NO: 13, aminoácidos 1 - 868; alineación realizada usando EMBL-EBI, EMBOSS-Needle - Alineación de Secuencias por Pares, Matriz BLOSUM62, abertura de hueco 10, extensión de hueco 0,5 en www.ebi.ac.uk/Tools/psa/emboss_needle/). Los correspondientes valores de identidad y similitud entre el Pea1 de *T. reesei* y el homólogo de Pea1 de *Fusarium (Gibberela) fujikuroi* de longitud completa (por ejemplo, FFUJ_12153, aminoácidos 1 - 882) son del 57,5 y el 68,9 %, respectivamente. Las secuencias de aminoácidos deducidas de los homólogos de Pea1 de longitud completa de *F. oxysporum* y *F. fujikuroi* son muy similares entre sí, siendo los valores de identidad y similitud entre las secuencias de aminoácidos de longitud completa anteriores del 96,6 y el 97,4 %, respectivamente.

65 Para confirmar que el papel de los homólogos de Pea1 en otros hongos es similar al papel en *T. reesei*, se diseñó un enfoque de marcador dividido (Fig. 9) para inactivar los homólogos de *pea1* de dos especies de *Fusarium*, *F. oxysporum* y *F. fujikuroi*. El fragmento 1 del marcador dividido de ~ 3 kb contenía una región promotora del gen *pea1*

de Fo47 de *F. oxysporum* (1468 pb, nucleótidos de -1483 a -16 desde el codón de inicio, para dirigir el fragmento al locus de *pea1*) y la mitad 5' del gen marcador *hph* (del nucleótido 1 a 615 y el promotor *gpdA* de *Aspergillus*). El fragmento 2 del marcador dividido de ~ 3 kb contenía la mitad 3' del marcador de selección *hph* (del nucleótido 166 al 1026 y la región terminadora de *trpC* de *Aspergillus*) y el gen *pea1* parcial de *F. oxysporum* y su región terminadora (1358 pb; comenzando por el nucleótido 1667 del gen y terminando 380 nucleótidos después del codón de parada de *pea1*, para dirigir el fragmento al locus de *pea1*). Por tanto, ambos fragmentos del marcador dividido incluían la misma parte media de 450 pb del gen *hph*. Cuando los dos fragmentos de marcador dividido se transforman en el mismo hospedador, se recombinan con las correspondientes regiones de *pea1* en el genoma. Cuando también se recombinan entre sí en la región de la parte media común de *hph*, el marcador de selección se vuelve funcional. Usando el enfoque diseñado, se esperaba que un marcador de selección funcional en los transformantes se uniese a un gen *pea1* inactivo a una frecuencia alta.

Como las secuencias de los genes *pea1* de *F. oxysporum* y *F. fujikuroi* y sus regiones 5' y 3' son muy similares (pero no idénticas) entre sí, se usaron los mismos fragmentos de marcador dividido para la interrupción de los genes *pea1* de ambas especies.

Ejemplo 12. Transformación de *Fusarium oxysporum* y *F. fujikuroi* y análisis de los transformantes

Se transformaron cepas Fo47 de *Fusarium oxysporum* y IM158289 de *F. fujikuroi* usando los fragmentos de marcador dividido diseñados y sintetizados (Ejemplo 11). El método descrito en Wiemann *et al.* (*PLos Pathog.* 2013; 9(6): e1003475 y las referencias del mismo) se usó en las transformaciones fúngicas. En total se obtuvieron 96 transformantes de *F. oxysporum* y 46 de *F. fujikuroi*. La modificación del ADN objetivo (disrupción del homólogo de *pea1*) se analizó a partir de 20 transformantes de *F. oxysporum* y 10 transformantes de *F. fujikuroi* usando PCR de diagnóstico. Los cebadores en la reacción de PCR se diseñaron a partir del extremo del flanco 5' de *pea1* en el fragmento de marcador dividido 1 (del promotor de *pea1*, nucleótidos de -38 a -21 del ATG) y el comienzo del flanco 3' en el fragmento marcador dividido 2 (nucleótidos 1716 - 1695 del gen *pea1* de *F. oxysporum*). La reacción PCR de diagnóstico diseñada da como resultado un fragmento de 1,6 kb del gen nativo (completo) de *pea1* de *Fusarium*, mientras que la longitud del producto de un gen inactivado es de 2,5 kb.

De la mayoría de los transformantes se obtuvo un único producto de PCR de 2,5 kb, lo que indica una integración satisfactoria del marcador de longitud completa en el locus de *pea1* y la inactivación del gen *pea1*. Los fragmentos flanqueantes de *pea1* de *F. oxysporum* podrían usarse para la interrupción del *pea1* de ambas especies de *Fusarium*.

Se purificó una selección de transformantes que, según se demostró mediante PCR de diagnóstico, contenían un gen *pea1* inactivado. Se cultivaron cuatro transformantes de cada especie y sus cepas parentales en matraces de agitación en un medio de inducción a base de caseína (FusP) con y sin complemento de 0,5 g/l de CasAminoácidos. El medio FusP contenía (por 1000 ml): 20 ml de sales 50XFusP (26 g/l de KCl, K₂HPO₄ 82 g/l, NaH₂PO₄xH₂O 43 g/l, pH ajustado a 7,5 usando NaOH), glucosa 10 g/l, caseína 5 g/l (Sigma C8654), 2 ml de MgSO₄ 1 M, 1 ml de solución de 1000x oligoelementos (contiene, por 100 ml: 2,2 g de ZnSO₄x7H₂O, 1,1 g de H₃BO₃, 0,5 g de MnCl₂x4H₂O, 0,5 g de FeSO₄x7H₂O, 0,17 g de CoCl₂x6H₂O, 0,16 g de CuSO₄x5H₂O, 0,15 g de Na₂MoO₄x2H₂O, 5,0 g de Na₂EDTAx2H₂O, pH ajustado a 6,5 usando KOH). De forma interesante, los transformantes con el gen *pea1* inactivado apenas mostraron crecimiento en el medio que no se complementó con los CasAminoácidos, lo que indica que estas cepas fueron incapaces de usar caseína como fuente de nitrógeno. Todas las cepas, sin embargo, crecieron bien en el medio complementado con los CasAminoácidos. Se tomaron muestras de estos cultivos después de 6 días de cultivo a 25 °C. Las actividades proteolíticas extracelulares se midieron a partir de los sobrenadantes de los cultivos basándose en un procedimiento descrito por Holm (1980).

Las actividades proteasa determinadas a partir de los sobrenadantes de cultivo de los ocho transformantes con el gen *pea1* inactivado fueron muy bajas en comparación con las actividades de los sobrenadantes de cultivo de las cepas parentales (Fig. 10). La actividad proteasa en los sobrenadantes de cultivo de los transformantes de *F. fujikuroi* fue aproximadamente 10 veces inferior a la del sobrenadante de cultivo de la cepa parental. La actividad proteasa medida a partir de los sobrenadantes de cultivo de los transformantes de *F. oxysporum* fue aproximadamente 40 veces inferior a la de la cepa parental.

La interrupción del homólogo de *pea1* de la especie de *Fusarium* fue satisfactoria con el método utilizado. Los transformantes de *Fusarium* con *pea1* inactivado muestran un fenotipo deficiente en proteasa distintivo, como el de las cepas de *Trichoderma* con el gen *pea1* no funcional.

Los resultados muestran que las cepas de *Trichoderma reesei* de proteasa baja que carecen de *pea1* funcional proporcionan beneficios cuando se usan como hospedadores para la producción de proteínas, y especialmente de proteínas sensibles a proteasa. Con estas cepas pueden alcanzarse rendimientos de producción de proteínas al menos similares, o en varios casos incluso superiores, en comparación con los rendimientos obtenidos cuando los progenitores de estas cepas se usan como hospedadores de los mismos productos enzimáticos. Además, los productos enzimáticos obtenidos de las cepas que carecen de un *pea1* funcional son más estables en comparación con los productos correspondientes de los progenitores de estas cepas.

Los homólogos de *pea1* pueden encontrarse en el genoma de varias especies de hongos. Los resultados de los presentes inventores muestran que la disrupción de los homólogos de *pea1* de especies distintas de *T. reesei* conduce a fenotipos deficientes en proteasa similares como los que se muestran para las cepas de *T. reesei* que carecen de un *pea1* funcional. Los resultados confirman el papel de *pea1* y sus homólogos como factores importantes que afectan a la expresión de proteasas. Pueden conseguirse mejoras significativas en el rendimiento de proteínas y la estabilidad de los productos mediante la disrupción del *pea1* de las cepas de producción de diferentes especies.

LISTADO DE SECUENCIAS

10 <110> Roal Oy

<120> MÉTODOS PARA CONTROLAR LA PRODUCCIÓN DE PROTEASAS

15 <130> 31256FI-1f / 1135 PREGU

<160> 18

<170> PatentIn versión 3.5

20 <210> 1
<211> 885
<212> ADN
<213> *Trichoderma reesei*

25 <400> 1

atgggttggg gtgctggatc gagtgaactc ctggtcgagg ccaactttta cttcgaagct 60

cgcgacgcag atcaaggcgg tgactacctg atgcgagctt tcaagcagct cgagctggac 120

ctcagaaagc tctcgccaca aggcatcatg gaactaatcc tgggtatgat caatcgggat 180

cccggcatga tgacggccct ttgcaagtac cttgcagcat attcgacaac caatctggag 240

cgaaccatc ctctccgga aatcttctact tgtctgtatg aagtgcaaca aaagcatggg 300

gcgcagacgt tgtctgagct cctgtggact agcatctcga caattgcgga ggaactcgag 360

gccatctatg gacgcaagca tccgtatgtg gctcgcacat gggccgatct tgcgctgttt 420

tacagccagg tgaaccgga aaggctggag aagttggttg ttgagcttcg tgtgctccag 480

aggcagctcg agcaacgaca tgggcattcc agtgtcgaag tggtttccat ccgatatgcc 540

attctgctgt tggctctatgc gtcgtctccc cagtcggatg cctcgaagca agccgcaaat 600

gattattgga acctgctgcg gaatatgaac accatgtttc ccatgcgcga ctcccgtccg 660

aatagttact gctatcacag cccgctcaag gtcgatccgt ggacaaagag gtgccgcagg 720

cggtacgaca cactcgtcac catattcgag gagcatgtag gcgttagaat caatccctat 780

ttcgaagagg acttccacac gaccgagcac gctcaagaaa cgcaggatgc ctggggcgga 840

gctctgcaaa tgggttcgac gaatagatct tggggcttca tctag 885

30 <210> 2
<211> 1344
<212> ADN
<213> *Trichoderma reesei*

35 <400> 2

ES 2 815 628 T3

atggtttggg gtgctggatc gagtgaactc ctggctcgagg ccaactttta cttcgaagct 60
 cgcgacgcag atcaaggcgg tgactacctg atgcgagctt tcaagcagct cgagctggac 120
 ctcagaaagc tctcgccaca aggcacatg gaactaatcc tgggtatgat caatcgggat 180
 cccggcatga tgacggccct ttgcaagtac cttgcagcat attcgacaac caatctggag 240
 cgaacccatc ctctccggca aatcttcact tgtctgtatg aagtgcaaca aaagcatggg 300
 gcgcagacgt tgtctgagct cctgtggact agcatctcga caattgcgga ggaactcgag 360
 gccatctatg gacgcaagca tccccaggtg aaccggaaa ggctggagaa gttggttgtt 420
 gagcttcgtg tgctccagag gcagctcgag caacgacatg ggcatccag tgtcgaagtg 480
 gtttccatcc gatatgccat tctgctgttg gtctatgcgt cgtctcccca gtcggatgcc 540
 tcgaagcaag ccgcaaatga ttattggaac ctgctgcgga atatgaacac catgtttccc 600
 atgcgcgact cccgtccgaa tagttactgc taccacagcc cgctcaaggt cgatccgtgg 660
 acaaagaggt gccgcaggcg gtacgacaca ctcgtcacca tattcgagga gcatgtaggc 720
 gttagaatca atccctatct cgaagaggac ttccacacga ccgagcacgc tcaagaaacg 780
 caggatgcct gggcggcagc tctgcaaatg ggttcgacga atagatcttg gggcttcac 840
 tagcggtttc gctttcagaa tggtgggtgt gttccacgcc cagctggcgc aaatgtcagg 900
 tacciaagat accttgtctc tcgtctccta gcgcgcgggtg cagcgagctt ctttttcgag 960
 atttcttttt ccggaatcta gcgggcggtt aatacccagc atcagaagga ttgcggcgca 1020
 tctacctagc gacagttttt ttcggttaag tctgttttgc cacctcacat cgagcctgga 1080
 gttttggaca gttgacaatt ctgccatcg cggaacattt tcctttctca tgtcgccgac 1140
 gctgacaagt cgatttcctt ggtctttcca cgcaagttgt tgcacctctc gagccgggag 1200
 ttagggaggg ctttttcccc cttggatgag tcgtcatccg tgtataatag tatccgttgt 1260
 tttctgtttt cttttctgc gtttgatatg cgatgttcag ctttcttgat actgaagacc 1320
 acgagggaga ggcgggaaac gagt 1344

5 <210> 3
 <211> 280
 <212> PRT
 <213> *Trichoderma reesei*

10 <400> 3

ES 2 815 628 T3

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

Tyr Phe Glu Ala Arg Asp Ala Asp Gln Gly Gly Asp Tyr Leu Met Arg
20 25 30

Ala Phe Lys Gln Leu Glu Leu Asp Leu Arg Lys Leu Ser Pro Gln Gly
35 40 45

Ile Met Glu Leu Ile Leu Gly Met Ile Asn Arg Asp Pro Gly Met Met
50 55 60

Thr Ala Leu Cys Lys Tyr Leu Ala Ala Tyr Ser Thr Thr Asn Leu Glu

ES 2 815 628 T3

tttgtcaccg aggcctggga gagatatggc ctgggcagtg tgcctcggc tccgcctgcg 120
 atgcctccgc agcccatcca gaggaccgga cgggcctatt ccaccgcttc ggtaatggac 180
 tggacgcccc agatcgatac cacctttcac caccacggcg acggagtggc catgtcggcg 240
 gtggaacacg tcgaactgga cagccttggg tgcctcagg gagggggagg ggttggccgc 300
 ttggtcggcg actttgagaa taaaggctat gtcccgccat tgcctcggcg gcccatcaac 360
 aatcagccgc atcaaccgca tcacatgggc cacagtcctt caatgtcgtc gcagtttggc 420
 agtctaaact ctgtagccca tacgcatagg atatcgagtc ccggtgctag cccaggtgaa 480
 tcccaatatg gctcgcctgg ggtgcattcg cccccaacac ccacctcaat cgtcaccaac 540
 atgaaccgct ccggcagtat aagctatggc agcttccatg aactcagca ggtgcatagt 600
 cccggtattg gtacgccttt tgggagcatg gatggcttta tgagttctgc ccgctgaac 660
 agcccgatgg tggcgacacc aatggcatcg acgccatga tggccagtcc catggccgct 720
 cggggagtgc ctggtacccc tggattcgag atttggcgac ctccctccctc gatgactccc 780
 aaacctgaac cctctcaaat caccaaccga gccaatctcg gcggctatct cagacctcca 840
 gtgccgacga ctcccagcc ggtagtcaac acaggtaatc agttcatctt ggaattcaac 900
 cccagcggca aggcagccaa agggaaagct ccggcaaagc ctcccaggcc tcgattgcct 960
 cctcggaaac ccaacctttc acaatcacia ctctcggcac aaaccaccagc agcacctgtg 1020
 caagcgtcgg tttcaacacc ggcaccggca tcggcaccga caccgacacc agccgttgcc 1080
 tcaactccag gtccaatgtc tcctcctcct aaaccgcctc gcccttcaga accgtcttct 1140
 tcatccacac ccattgttgt gagcaatgcc gctaaaccaa gagaatcctt gcagctaggt 1200
 gtactaacgc gcatgtctag gcacagcggc gggattcgat ggctcttcgt gccaggcag 1260
 gtactaggcc gtcacgggaa caggttcctg cagaggcctg ggagcaatac aatccacta 1320
 tccgtaccct ctatctcgag gagagaaaac ccttgaaaga agtcatgagt gtcattgctg 1380
 aacaatatgg gtttcaagca acgtgagtcg aagattgatt cctcttctgc ttcattcttc 1440
 gggttcttgg ttcgttgtaa actaattcga cggcctttct aggccaaaga tgtataaaac 1500
 aaggttctct caatggggtt ttgtgaagaa taacacggag gaagaagtga aacggctgtt 1560
 gtcgatgaag ttccagcgag atgccgaggg caaagtttcc gagtttgctt gaaacggcag 1620
 ggtggtgaac ctaggtaact atttgaaacg gaaaggagtg acggagtatg acctcgttga 1680
 tttcgaacta ccggccgacc ttccagcaca tatccgatgc agaaccacca caccacctcc 1740
 ggctctgca tcaccggatc tgctccgtgc acaggaggta gtcggtggaa atatgcgcaa 1800
 ggcattccta cactgtcggc aattcgagat ggagactgag actcagattg gttggccatc 1860
 aaccatggtt tggggtgctg gatcgagtga actcctggtc gagccaact tttacttcca 1920
 agctcggcgc gcagatcaag gcggtgacta cctgatgcga gctttcaagc agctcagct 1980

ES 2 815 628 T3

ggacctcaga aagctctcgc cacaaggcat catggaacta atcctgggta tgatcaatcg 2040
 ggatcccggc atgatgacgg ccctttgcaa gtaccttgca gcatattcga caaccaatct 2100
 ggagcgaacc catcctctcc ggcaaatctt cacttgtctg tatgaagtgc aacaaaagca 2160
 tggggcgcag acgttgtctg agctcctgtg gactagcatc tcgacaattg cggaggaact 2220
 cgaggccatc tatggacgca agcatccgta tgtggctcgc acatgggccc atcttgcgct 2280
 gttttacagc caggtgaacc cggaaaggct ggagaagttg gttgttgagc ttcgtgtgct 2340
 ccagaggcag ctcgagcaac gacatgggca ttccagtgtc gaagtggttt ccatccgata 2400
 tgccattctg ctgttggctc atgcgtcgtc tcccagtcg gatgcctcga agcaagccgc 2460
 aatgattat tggaacctgc tgcggaatat gaacaccatg tttcccatgc gcgactcccg 2520
 tccgaatagt tactgctatc acagcccgtc caaggtcgat ccgtggacaa agaggtgccg 2580
 caggcggtag gacacactcg tcaccatatt cgaggagcat gtaggcgta gaatcaatcc 2640
 ctatttcgaa gaggacttcc acacgaccga gcacgctcaa gaaacgcagg atgcctgggc 2700
 ggcagctctg caaatgggtt cgacgaatag atcttggggc ttcattctag 2749

<210> 5
 <211> 2478
 <212> ADN
 <213> *Trichoderma reesei*

5

<400> 5

atgttccatc aaccgccgga aaagcctcgc gcgcctggct tcgacgatga tccgttcggg 60
 tttgtcacc caggcctggga gagatatggc ctgggcagtg tgtcgtcgcc tccgcctcgc 120
 atgcctccgc agcccatcca gaggaccgga cgggcctatt ccaccgcttc ggtaatggac 180
 tggacgcccg agatcgatac cacctttcac caccacggcg acggagtggc catgtcggcg 240
 gtggaacacg tcgaactgga cagccttggc tcgcctcagg gagggggagg ggttggccgc 300
 ttggtcgcgc actttgagaa taaaggctat gtcccgccat tgcctccgcg gcccatcaac 360
 aatcagccgc atcaaccgca tcacatgggc cacagtcctt caatgtcgtc gcagtttggc 420
 agtctaaact ctgtagccca tacgcatagg atatcgagtc ccgttgctag cccaggtgaa 480
 tcccaatatg gctcgtcgcg ggtgcattcg cccccaacac ccacctcaat cgtcaccaac 540
 atgaaccgct ccggcagtat aagctatggc agcttccatg aactcagca ggtgcatagt 600
 cccggtattg gtacgccttt tgggagcatg gatggcttta tgagttctgc ccgcgtgaac 660
 agcccgatgg tggcgacacc aatggcatcg acgcccata tgcccagtcc catggccgct 720
 ccgggagtg cctgtacccc tggattcgag atttggcgac ctctccctc gatgactccc 780
 aaacctgaac cctctcaaat caccaacca gccaatctcg gcggctatct cagacctcca 840
 gtgccgacga ctcccaagcc ggtagtcaac acaggtaatc agttcatctt ggaattcaac 900

10

ES 2 815 628 T3

cccagcgcca aggcagccaa agggaaagct ccggcaaagc ctcccaggcc tcgattgcct 960
 cctcggaaac ccaacctttc acaatcacia ctctcggcac aaacaccagc agcacctgtg 1020
 caagcgtcgg cacagcggcg ggattcgatg gctcttcgtg cccaggcagg tactaggccg 1080
 tcacgggaac aggttcctgc agaggcctgg gagcaatata aatccactat ccgtaccctc 1140
 tatctcgagg agagaaaacc cttgaaagaa gtcattgagtg tcatggctga acaatatggg 1200
 tttcaagcaa cgccaaagat gtataaaaca aggttctctc aatggggttt tgtgaagaat 1260
 aacacggagg aagaagtgaa acggctgttg tcgatgaagt tccagcgaga tgccgagggc 1320
 aaagtttccg agtttgctcg aaacggcagg gtggtgaacc taggtacctt tttgaaacgg 1380
 aaaggagtga cggagtatga cctcgttgat ttcgaactac cggccgacct tccagcacat 1440
 atccgatgca gaacaccgac accacctccg gctctgcgat caccgatctt gctccgtgca 1500
 caggaggtag tcggtgaaa tatgcgcaag gcattcctac actgtcggca attcgagatg 1560
 gagactgaga ctgagattgg ttggccatca accatggttt ggggtgctgg atcgagtgaa 1620
 ctctggtcgg aggccaaactt ttacttcgaa gctcgcgacg cagatcaagg cggtgactac 1680
 ctgatgcgag ctttcaagca gctcgagctg gacctcagaa agctctcgcc acaaggcatc 1740
 atggaactaa tcctgggtat gatcaatcgg gatcccggca tgatgacggc ctttgcaag 1800
 taccttgagc catattcgac aaccaatctg gagcgaacct atcctctccg gcaaactctc 1860
 acttgctctg atgaagtgca acaaaagcat ggggcgcaga cgttgctctga gctcctgtgg 1920
 actagcatct cgacaattgc ggaggaactc gaggccatct atggacgcaa gcatccgtat 1980
 gtggctcgca catgggcccga tcttgcgctg ttttacagcc aggtgaacct ggaaaggctg 2040
 gagaagttgg ttggtgagct tcgtgtgctc cagaggcagc tcgagcaacg acatgggcat 2100
 tccagtgtcg aagtggtttc catccgatat gccattctgc tgttggtcta tgcgtcgtct 2160
 ccccagtcgg atgcctcgaa gcaagccgca aatgattatt ggaacctgct gcggaatatg 2220
 aacaccatgt ttcccatgcg cgactcccgt ccgaatagtt actgctatca cagcccgtct 2280
 aaggtcgatc cgtggacaaa gaggtgccgc aggcggtacg acacactcgt caccatattc 2340
 gaggagcatg taggcgttag aatcaatccc tatttcgaag aggacttcca cacgaccgag 2400
 cacgctcaag aaacgcagga tgcctgggcg gcagctctgc aatgggttc gacgaataga 2460
 tcttggggct tcattctag 2478

<210> 6
 <211> 825
 <212> PRT
 <213> *Trichoderma reesei*

5

<400> 6

ES 2 815 628 T3

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
 1 5 10 15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
 20 25 30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
 35 40 45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
 50 55 60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
 65 70 75 80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
 85 90 95

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
 115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser
 130 135 140

Val Ala His Thr His Arg Ile Ser Ser Pro Val Ala Ser Pro Gly Glu
 145 150 155 160

Ser Gln Tyr Gly Ser Leu Gly Val His Ser Pro Pro Thr Pro Thr Ser
 165 170 175

Ile Val Thr Asn Met Asn Arg Ser Gly Ser Ile Ser Tyr Gly Ser Phe
 180 185 190

His Asp Thr Gln Gln Val His Ser Pro Gly Ile Gly Thr Pro Phe Gly
 195 200 205

Ser Met Asp Gly Phe Met Ser Ser Ala Arg Val Asn Ser Pro Met Val
 210 215 220

Ala Thr Pro Met Ala Ser Thr Pro Met Met Pro Ser Pro Met Ala Ala
 225 230 235 240

Pro Gly Val Pro Gly Thr Pro Gly Phe Glu Ile Trp Arg Pro Pro Pro
 245 250 255

ES 2 815 628 T3

Ser Met Thr Pro Lys Pro Glu Pro Ser Gln Ile Thr Asn Pro Ala Asn
260 265 270

Phe Gly Gly Tyr Phe Arg Pro Pro Val Pro Thr Thr Pro Lys Pro Val
275 280 285

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

Ala Ala Lys Gly Lys Ala Pro Ala Lys Pro Pro Arg Pro Arg Leu Pro
305 310 315 320

Pro Arg Lys Pro Asn Leu Ser Gln Ser Gln Leu Ser Ala Gln Thr Pro
325 330 335

Ala Ala Pro Val Gln Ala Ser Ala Gln Arg Arg Asp Ser Met Ala Leu
340 345 350

Arg Ala Gln Ala Gly Thr Arg Pro Ser Arg Glu Gln Val Pro Ala Glu
355 360 365

Ala Trp Glu Gln Tyr Lys Ser Thr Ile Arg Thr Leu Tyr Leu Glu Glu
370 375 380

Arg Lys Pro Leu Lys Glu Val Met Ser Val Met Ala Glu Gln Tyr Gly
385 390 395 400

Phe Gln Ala Thr Pro Lys Met Tyr Lys Thr Arg Phe Ser Gln Trp Gly
405 410 415

Phe Val Lys Asn Asn Thr Glu Glu Glu Val Lys Arg Leu Leu Ser Met
420 425 430

Lys Phe Gln Arg Asp Ala Glu Gly Lys Val Ser Glu Phe Val Arg Asn
435 440 445

Gly Arg Val Val Asn Leu Gly Thr Tyr Leu Lys Arg Lys Gly Val Thr
450 455 460

Glu Tyr Asp Leu Val Asp Phe Glu Leu Pro Ala Asp Leu Pro Ala His
465 470 475 480

Ile Arg Cys Arg Thr Pro Thr Pro Pro Pro Ala Leu Arg Ser Pro Asp
485 490 495

Leu Leu Arg Ala Gln Glu Val Val Val Gly Asn Met Arg Lys Ala Phe
500 505 510

ES 2 815 628 T3

Leu His Cys Arg Gln Phe Glu Met Glu Thr Glu Thr Gln Ile Gly Trp
 515 520 525

Pro Ser Thr Met Val Trp Gly Ala Gly Ser Ser Glu Leu Leu Val Glu
 530 535 540

Ala Asn Phe Tyr Phe Glu Ala Arg Asp Ala Asp Gln Gly Gly Asp Tyr
 545 550 555 560

Leu Met Arg Ala Phe Lys Gln Leu Glu Leu Asp Leu Arg Lys Leu Ser
 565 570 575

Pro Gln Gly Ile Met Glu Leu Ile Leu Gly Met Ile Asn Arg Asp Pro
 580 585 590

Gly Met Met Thr Ala Leu Cys Lys Tyr Leu Ala Ala Tyr Ser Thr Thr
 595 600 605

Asn Leu Glu Arg Thr His Pro Leu Arg Gln Ile Phe Thr Cys Leu Tyr
 610 615 620

Glu Val Gln Gln Lys His Gly Ala Gln Thr Leu Ser Glu Leu Leu Trp
 625 630 635 640

Thr Ser Ile Ser Thr Ile Ala Glu Glu Leu Glu Ala Ile Tyr Gly Arg
 645 650 655

Lys His Pro Tyr Val Ala Arg Thr Trp Ala Asp Leu Ala Leu Phe Tyr
 660 665 670

Ser Gln Val Asn Pro Glu Arg Leu Glu Lys Leu Val Val Glu Leu Arg
 675 680 685

Val Leu Gln Arg Gln Leu Glu Gln Arg His Gly His Ser Ser Val Glu
 690 695 700

Val Val Ser Ile Arg Tyr Ala Ile Leu Leu Leu Val Tyr Ala Ser Ser
 705 710 715 720

Pro Gln Ser Asp Ala Ser Lys Gln Ala Ala Asn Asp Tyr Trp Asn Leu
 725 730 735

Leu Arg Asn Met Asn Thr Met Phe Pro Met Arg Asp Ser Arg Pro Asn
 740 745 750

Ser Tyr Cys Tyr His Ser Pro Leu Lys Val Asp Pro Trp Thr Lys Arg

ES 2 815 628 T3

cttcatccac acccattggt gtgagcaatg ccgctaaacc aagagaatcc ttgcagctag 1200
 gtgtactaac gcgcatgtct aggcacagcg gcgggattcg atggctcttc gtgcccaggg 1260
 aggtactagg ccgtcacggg aacaggttcc tgcagaggcc tgggagcaat acaaatccac 1320
 tatccgtacc ctctatctcg aggagagaaa accottgaaa gaagtcatga gtgtcatggc 1380
 tgaacaatat gggtttcaag caacgtgagt cgaagattga ttcctcttct gcttcatcgt 1440
 tcgggttctt ggttcgttgt aaactaattc gacgcccttt ctaggccaaa gatgtataaa 1500
 acaaggttct ctcaatgggg ttttgtgaag aataacacgg aggaagaagt gaaacggctg 1560
 ttgtcgatga agttccagcg agatgccgag ggcaaagttt ccgagtttgt tcgaaacggc 1620
 aggggtggtga acctaggtac ctatttgaaa cggaaaggag tgacggagta tgacctcgtt 1680
 gatttcgaac taccggccga ccttccagca catatccgat gcagaacacc gacaccacct 1740
 ccggctctgc gatcaccgga tctgctccgt gcacaggagg tagtcgttgg aatatgcgc 1800
 aaggcattcc tacactgtcg gcaattcgag atggagactg agactcagat tggttggcca 1860
 tcaaccatgg tttgggtgct tggatcgagt gaactcctgg tcgaggccaa cttttacttc 1920
 gaagctcgcg acgcagatca aggcggtgac tacctgatgc gagctttcaa gcagctcagag 1980
 ctggacctca gaaagctctc gccacaaggc atcatggaac taatcctggg tatgatcaat 2040
 cgggatcccg gcatgatgac ggccctttgc aagtacctg cagcatattc gacaaccaat 2100
 ctggagcgaa cccatcctct ccggcaaatc ttcacttgct tgtatgaagt gcaacaaaag 2160
 catggggcgc agacgttgct tgagctcctg tggactagca tctcgacaat tgcggaggaa 2220
 ctcgaggcca tctatggacg caagcatccg tatgtggctc gcacatgggc cgatcttgcg 2280
 ctgttttaca gccaggtgaa cccggaaagg ctggagaagt tggttggtga gcttcgtgtg 2340
 ctccagaggc agctcgagca acgacatggg cattccagtg tcgaagtggg ttccatccga 2400
 tatgccattc tgctggttgt ctatgcgtcg tctcccagc cggatgcctc gaagcaagcc 2460
 gcaaatgatt attggaacct gctgcggaat atgaacacca tgtttcccat gcgcgactcc 2520
 cgtccgaata gttactgcta tcacagcccg ctcaaggctg atccgtggac aaagaggtgc 2580
 cgcaggcggc acgacacact cgtcaccata ttcgaggagc atgtaggcgt tagaatcaat 2640
 ccctatttcg aagaggactt ccacacgacc gagcacgctc aagaaacgca ggatgcctgg 2700
 gcggcagctc tgcaaatggg ttcgacgaat agatcttggg gcttcatcta g 2751

<210> 8
 <211> 2749
 <212> ADN
 <213> *Trichoderma reesei*

<400> 8

10 atgttccatc aaccgccgga aaagcctcgc gcgcctggct tcgacgatga tccgttcggg 60

ES 2 815 628 T3

tttgtcacc aggcctggga gagatatggc ctgggcagtg tgtcgtcgcc tccgcctgcg 120
 atgcctccgc agcccatcca gaggaccgga cgggcctatt ccaccgcttc ggtaatggac 180
 tggacgcccg agatcgatac cacctttcac caccacggcg acggagtggc catgtcgccg 240
 gtggaacacg tcgaactgga cagccttggg tgcctcagg gagggggagg ggttggccgc 300
 ttggtcgcgc actttgagaa taaaggctat gtcccgccat tgcctccgcg gcccatcaac 360
 aatcagccgc atcaaccgca tcacatgggc cacagtcctt caatgtcgtc gcagtttggc 420
 agtctaaact ctgtagccca tacgcatagg atatcgagtc ccgttgctag cccaggtgaa 480
 tcccaatatg gctcgtcgg ggtgcattcg ccccaaacac ccacctcaat cgtcaccaac 540
 atgaaccgct ccggcagtat aagctatggc agcttccatg aactcagca ggtgcatagt 600
 cccggtattg gtacgccttt tgggagcatg gatggcttta tgagtctgc ccggtgaa 660
 agcccgatgg tggcgacacc aatggcatcg acgcccata tgcccagtc catggccgct 720
 ccgggagtg ctggtacccc tggattcag atttggcgac ctctccctc gatgactccc 780
 aaacctgaac cctctcaaat caccaacca gccaatctcg gcggctatct cagacctca 840
 gtgccgacga ctccaagcc ggtagtcaac acaggtaatc agttcatctt ggaattcaac 900
 cccagcgcca aggcagccaa agggaaagct ccggcaaagc ctcccaggcc tcgattgcct 960
 cctcggaaac ccaacctttc acaatcacia ctctcggcac aacaccagc agcacctgtg 1020
 caagcgtcgg tttcaacacc ggcaccggca tcggcaccga caccgacacc agccgttgcc 1080
 tcaactccag gtccaatgtc tcctcctcct aaaccgcctc gcccttcaga accgtcttct 1140
 tcatccacac ccattgttgt gagcaatgcc gctaaaccaa gagaatcctt gcagctaggt 1200
 gtactaacgc gcatgtctag gcatagcggc gggattcgat ggctcttctg gcccaggcag 1260
 gtactaggcc gtcacgggaa caggttcctg cagaggcctg ggagcaatac aaatccacta 1320
 tccgtaccct ctatctcgag gagagaaaac ccttgaaaga agtcatgagt gtcattgctg 1380
 aacaatatgg gtttcaagca acgtgagtcg aagattgatt cctcttctgc ttcattgctc 1440
 ggttcttgg ttcggtgtaa actaattcga cgcctttct agccaaaga tgtataaac 1500
 aaggttctct caatggggtt ttgtgaagaa taacacggag gaagaagtga aacggctgtt 1560
 gtcgatgaag ttccagcgag atgccgaggg caaagtttc gagtttgtc gaaacggcag 1620
 ggtggtgaac ctaggtacct atttgaacg gaaaggagt acggagtatg acctcgttga 1680
 tttcgaacta ccggccgacc ttccagcaca tatccgatgc agaaccga caccacctcc 1740
 ggctctgcca tcaccgatc tgcctcgtc acaggaggta gtcgttgaa atatgccc 1800
 ggcatccta cactgtcggc aattcgagat ggagactgag actcagattg gttggccatc 1860
 aaccatggtt tggggtgctg gatcgagtga actcctggtc gaggccaact tttacttca 1920
 agctcgcgac gcagatcaag gcggtgacta cctgatgcga gctttcaagc agctcagct 1980

ES 2 815 628 T3

ggacctcaga aagctctcgc cacaaggcat catggaacta atcctgggta tgatcaatcg 2040
 ggatcccggc atgatgacgg ccctttgcaa gtaccttgca gcatattcga caaccaatct 2100
 ggagcgaacc catcctctcc ggcaaactct cacttgtctg tatgaagtgc aacaaaagca 2160
 tggggcgag acgttgtctg agctcctgtg gactagcatc tcgacaattg cggaggaact 2220
 cgaggccatc tatggacgca agcatccgta tgtggctcgc acatgggccg atcttgcgct 2280
 gttttacagc caggtgaacc cggaaaggct ggagaagttg gttgttgagc ttcgtgtgct 2340
 ccagaggcag ctcgagcaac gacatgggca ttccagtgtc gaagtggttt ccatccgata 2400
 tgccattctg ctgttggtct atgcgtcgtc tcccagtcg gatgcctcga agcaagccgc 2460
 aatgattat tggaaacctg tgcggaatat gaacaccatg tttcccatgc gcgactcccg 2520
 tccgaatagt tactgctatc acagcccgtc caaggtcgat ccgtggacia agaggtgccg 2580
 caggcggtag gacacactcg tcaccatatt cgaggagcat gtaggcgta gaatcaatcc 2640
 ctatttcgaa gaggacttcc acacgaccga gcacgctcaa gaaacgcagg atgcctgggc 2700
 ggcagctctg caaatggggt cgacgaatag atcttggggc ttcacttag 2749

<210> 9
 <211> 2748
 <212> ADN
 <213> *Trichoderma reesei*

5

<400> 9

atgttccatc aaccgccgga aaagcctcgc gcgcctggct tcgacgatga tccgttcggg 60
 tttgtcacc aggcctggga gagatatggc ctgggcagtg tgtcgtcgcc tccgcctgag 120
 atgcctccgc agcccatcca gaggaccgga cgggcctatt ccaccgcttc ggtaatggac 180
 tggacgccc agatcgatac cacctttcac caccacggcg acggagtggc catgtcggc 240
 gtggaacacg tcgaactgga cagccttggc tcgcctcagg gagggggagg ggttggccgc 300
 ttggtcgcgc actttgagaa taaaggctat gtcccgccat tgcctccgcg gcccatcaac 360
 aatcagccgc atcaaccgca tcacatgggc cacagtcctt caatgtcgtc gcagtttggc 420
 agtctaaact ctgtagccca tacgcatagg atatcgagtc ccgttgctag cccaggtgaa 480
 tcccaatatg gctcgtcggc ggtgcattcg cccccaacac ccacctcaat cgtcaccaac 540
 atgaaccgct cgggcagtat aagctatggc agcttccatg aactcagca ggtgcatagt 600
 cccggtattg gtacgccttt tgggagcatg gatggcttta tgagttctgc ccgctgaac 660
 agcccgatgg tggcgacacc aatggcatcg acgcccata tggccagtc catggccgct 720
 cccggagtgc ctggtacccc tggattcgag atttggcgac ctccctccctc gatgactccc 780
 aaacctgaac cctctcaaat caccaacca gccaatctcg gccgctatct cagacctcca 840
 gtgccgacga ctccaagcc ggtagtcaac acaggtaatc agttcatctt ggaattcaac 900

10

ES 2 815 628 T3

cccagcgcca aggcagccaa agggaaagct ccggcaaagc ctcccaggcc tcgattgcct 960
 cctcggaacc caacctttca caatcacaac tctcggcaca aacaccagca gcacctgtgc 1020
 aagcgtcggg ttcaacaccg gcaccggcat cggcaccgac accgacacca gccgttgctt 1080
 caactccagg tccaatgtct cctcctccta aaccgcctcg cccttcagaa ccgtcttctt 1140
 catccacacc cattgtttgt agcaatgccg ctaaaccaag agaatccttg cagctagggtg 1200
 tactaacgcg catgtctagg cacagcggcg ggattcgatg gctcttcgtg cccaggcagg 1260
 tactaggccg tcacgggaac aggttcctgc agaggcctgg gagcaataca aatccactat 1320
 ccgtaccctc tatctcgagg agagaaaacc cttgaaagaa gtcgatgagtg tcatggctga 1380
 acaatatggg tttcaagcaa cgtgagtcga agattgattc ctcttctgct tcatcgttcg 1440
 ggttcttggg tcggttgtaa ctaattcgac gccctttcta ggccaaagat gtataaaaca 1500
 aggttctctc aatgggggtt tgtgaagaat aacacggagg aagaagtga acggctgttg 1560
 tcgatgaagt tccagcgaga tgccgagggc aaagtttccg agtttgttcg aaacggcagg 1620
 gtggtgaacc taggtacctt tttgaaacgg aaaggagtga cggagtatga cctcgttgat 1680
 ttogaactac cggccgacct tccagcacat atccgatgca gaacaccgac accacctccg 1740
 gctctgcgat caccgatctt gctccgtgca caggaggtag tcgttggaaa tatgcgcaag 1800
 gcattcctac actgtcggca attcgagatg gagactgaga ctgagattgg ttggccatca 1860
 accatggttt ggggtgctgg atcgagtga ctccctggcg aggccaaactt ttacttcgaa 1920
 gctcgcgacg cagatcaagg cgtgactac ctgatgagc ctttcaagca gctcgagctg 1980
 gacctcagaa agctctcgcc acaaggcatc atggaactaa tcctgggtat gatcaatcgg 2040
 gatcccgcca tgatgacggc cctttgcaag taccttgagc catattcgac aaccaatctg 2100
 gagcgaaccc atcctctccg gcaaatcttc acttgtctgt atgaagtgca acaaaagcat 2160
 ggggocgaga cgttgtctga gctcctgtgg actagcatct cgacaattgc ggaggaactc 2220
 gaggccatct atggacgcaa gcatccgtat gtggctcgca catgggccga tcttgcgctg 2280
 ttttacagcc aggtgaaccc ggaaaggctg gagaagttgg ttggtgagct tcgtgtgctc 2340
 cagaggcagc tcgagcaacg acatgggcat tocagtgtcg aagtggtttc catccgatat 2400
 gccattctgc tgttggtcta tgcgtcgtct cccagtcgg atgcctcgaa gcaagccgca 2460
 aatgattatt ggaacctgct gcggaatatg aacaccatgt tcccatgcg cgactcccgt 2520
 ccgaatagtt actgctatca cagcccgtc aaggctgatc cgtggacaaa gaggtgccgc 2580
 aggcggtacg acacactcgt caccatattc gaggagcatg taggcgtag aatcaatccc 2640
 tatttcgaag aggacttcca cacgaccgag caogctcaag aaacgcagga tgcctgggcg 2700
 gcagctctgc aatgggttc gacgaataga tcttggggct tcatctag 2748

<210> 10
 <211> 2749

ES 2 815 628 T3

<212> ADN

<213> *Trichoderma reesei*

<400> 10

5

atgttccatc	aaccgccgga	aaagcctcgc	gcgccctggct	tcgacgatga	tccgttcggg	60
tttgtcacc	aggcctggga	gagatatggc	ctgggcagtg	tgtcgtcgc	tccgcctcgc	120
atgcctccgc	agcccatcca	gaggaccgga	cgggcctatt	ccaccgcttc	ggtaatggac	180
tggacgccc	agatcgatac	cacctttcac	caccacggcg	acggagtggc	catgtcgcgc	240
gtggaacac	tcgaactgga	cagccttggt	tcgcctcagg	gagggggagg	ggttggccgc	300
ttggtcgcgc	actttgagaa	taaaggctat	gtcccgccat	tgcctccgcg	gcccataaac	360
aatcagccgc	atcaaccgca	tcacatgggc	cacagtcctt	caatgtcgtc	gcagtttggc	420
agtctaaact	ctgtagccca	tacgcatagg	atatcgagtc	ccgttgctag	cccaggtgaa	480
tcccaatatg	gctcgcctcg	ggtgcattcg	cccccaacac	ccacctcaat	cgtcaccaac	540
atgaaccgct	ccggcagtat	aagctatggc	agcttccatg	aaactcagca	ggtgcatagt	600
cccggatttg	gtacgccttt	tgggagcatg	gatggcttta	tgagttctgc	ccgcgtgaa	660
agcccgatgg	tggcgacacc	aatggcatcg	acgcccatag	tgccaggtcc	catggcccgt	720
ccgggagtgc	ctggtacccc	tggattcgag	atttggcgac	ctcctccctc	gatgactccc	780
aaacctgaac	cctctcaaat	caccaacca	gccaatctcg	gcggctatct	cagacctcca	840
gtgccgacga	ctcccaagcc	ggtagtcaac	acaggtaatc	agttcatctt	ggaattcaac	900
cccagcgcca	aggcagccaa	agggaaagct	ccggcaaagc	ctcccaggcc	ttgattgcct	960
cctcggaaac	ccaacctttc	acaatcacia	ctctcggcac	aaacaccagc	agcacctgtg	1020
caagcgtcgg	tttcaacacc	ggcaccggca	tcggcaccca	caccgacacc	agccggtgcc	1080
tcaactccag	gtccaatgtc	tcctcctcct	aaaccgcctc	gcccttcaga	accgtcttct	1140
tcattccacac	ccattgttgt	gagcaatgcc	gctaaaccaa	gagaatcctt	gcagctaggt	1200
gtactaacgc	gcatgtctag	gcacagcggc	gggattcgat	ggctcttcgt	gccagggcag	1260
gtactaggcc	gtcacgggaa	caggttcctg	cagaggcctg	ggagcaatac	aaatccacta	1320
tccgtaccct	ctatctcgag	gagagaaaac	ccttgaaaga	agtcatgagt	gtcatggctg	1380
aacaatatgg	gtttcaagca	acgtgagtcg	aagattgatt	cctcttctgc	ttcatcgctc	1440
gggttcttgg	ttcgttgtaa	actaattcga	cgccctttct	aggccaaaga	tgtataaaac	1500
aaggttctct	caatgggggt	ttgtgaagaa	taacacggag	gaagaagtga	aacggctggt	1560
gtcgatgaag	ttccagcgag	atgccgaggg	caaagtttcc	gagtttggtc	gaaacggcag	1620
ggtggtgaac	ctaggtacct	atttgaaacg	gaaaggagtg	acggagtatg	acctcgttga	1680

ES 2 815 628 T3

tttcgaacta ccggccgacc ttccagcaca tatccgatgc agaacaccga caccacctcc 1740
 ggctctgcga tcacccgatc tgctccgtgc acaggaggta gtcgttggaa atatgcgcaa 1800
 ggcattccta cactgtcggc aattcgagat ggagactgag actcagattg gttggccatc 1860
 aaccatggtt tggggtgctg gatcgagtga actcctggtc gaggccaact tttacttcga 1920
 agctcgcgac gcagatcaag gcggtgacta cctgatgoga gctttcaagc agctcgagct 1980
 ggacctcaga aagctctcgc cacaaggcat catggaacta atcctgggta tgatcaatcg 2040
 ggatcccggc atgatgacgg ccctttgcaa gtaccttgca gcatattcga caaccaatct 2100
 ggagcgaacc catcctctcc ggcaaactct cacttgtctg tatgaagtgc aacaaaagca 2160
 tggggcgcag acgttgtctg agctcctgtg gactagcatc tcgacaattg cggaggaact 2220
 cgaggccatc tatggacgca agcatccgta tgtggctcgc acatgggccc atcttgcgct 2280
 gttttacagc caggtgaacc cggaaaggct ggagaagttg gttgttgagc ttcgtgtgct 2340
 ccagaggcag ctcgagcaac gacatgggca ttccagtgtc gaagtggttt ccatccgata 2400
 tgccattctg ctgttggctc atgcgtcgtc tcccagtcg gatgcctcga agcaagccgc 2460
 aatgattat tggaaacctg tgcggaatat gaacaccatg tttcccatgc gcgactcccg 2520
 tccgaatagt tactgctatc acagcccgct caaggtcgat ccgtggacia agaggtgccg 2580
 caggcggtag gacacactcg tcacatatt cgaggagcat gtaggcgta gaatcaatcc 2640
 ctatttcgaa gaggacttcc acacgaccga gcacgctcaa gaaacgcagg atgcctgggc 2700
 ggcagctctg caaatgggtt cgacgaatag atcttggggc ttcacttag 2749

<210> 11
 <211> 4710
 <212> ADN
 <213> *Trichoderma reesei*

5

<400> 11

cgttggctcg aggcaactgc gtccgtggag gcagaatgcc atcatcgatc ggtcccagca 60
 cagtgtcctt gagcgtacgg gcaccagggg gttggtatca gagcgtgcc attgggcat 120
 gtgtcctggc gagtggtagc ttgcgttcat atcaccttct ttttctggcc gatggtggca 180
 ttcacacgcc gacccccctg gctactccag ctcaatggcc ggcagcccta ctgggtttcc 240
 ccggcctgat tatttgcttc gccatcgctg ttgcaatggt cgttagccac cagcctgcat 300
 taggttacat cccctccaat ccaacctctg gctcaccatg ctgctgtctc ttgtctgcta 360
 gaggcccctc tctccacctg ctctcgtgcc taggctgtcc ccgtcgtttt gatcccatcg 420
 tactagcagc ggcatcagga ctcttagagg caccactcgt actagcccgt agcacaccag 480
 ggccagccag aacagctccg tcctggccca gcctgtatct tgagcatctc tggctctctt 540
 atctcagcgg gggccttcta aattgaccct tctcgaaaga cggcacatgg gatttcctt 600

10

ES 2 815 628 T3

cttggcactc tagacgaggg aacttttgct cgctctcttc gctgctggc ttggcttgg 660
 ttctcccac gccgctgctg atgcttttg agctggctgc tcttggtgtt tgtttgtcca 720
 aaccacacag cggctgttgc cctatcaagg cgttttccca catcggccga tatctggggg 780
 aacgatacgg atatctgcc ctagccagcc agccagccag ccagccagcc agccagccct 840
 tgctgtact ttcgtgctgc gcccgctcca gtaggcctct ccgccgctg gattccctct 900
 ctctttacta tccctctaca ccaggcgatt atgggcacct ttgccgcct ccatacacg 960
 accttcgttt gactgagcaa gcacgctgc cactctcaga gagtgatgaa tcaccgctg 1020
 ggagactgat tattccagcc gaggcctacg acttgcccag gtccgtcagt tcctgacaga 1080
 caagacagcc ggcggactga gcaccactgc tctgtctgct cgggtgtgtg gcacgttgac 1140
 atgttccatc aaccgccgga aaagcctcgc ggcctggct tcgacgatga tccgttcggg 1200
 tttgtcacc aggctggga gagatatggc ctgggcagtg tctcgtcgc tccgctcgc 1260
 atgcctcgc agccatcca gaggaccgga cggcctatt ccaccgctc ggtaatggac 1320
 tggacgccc agatcgata cacctttcac caccacggcg acggagtggc catgtcgcg 1380
 gtggaacacg tcgaactgga cagccttggc tcgcctcagg gagggggagg ggttgccgc 1440
 ttggtcgcgc actttgagaa taaaggctat gtcccgcct tgcctccgc gccatcaac 1500
 aatcagccgc atcaaccgca tcacatggc cacagtcctt caatgtcgc gcagtttggc 1560
 agtctaaact ctgtagcca tacgcatagg atatcgagtc ccgttgctag ccaggtgaa 1620
 tcccaatatg gctcgtcgg ggtgcattcg ccccaacac ccacctcaat cgtcaccaac 1680
 atgaaccgct ccggcagtat aagctatggc agcttccatg aactcagca ggtgcatagt 1740
 cccggtattg gtacgcctt tgggagcatg gatggcttta tgagttctgc ccgctgaa 1800
 agcccgatgg tggcgacacc aatggcatcg acgccatga tgcccagtcc catggccgct 1860
 ccgggagtgc ctggtacccc tggattcgag atttggcgac ctctccctc gatgactccc 1920
 aaacctgaac cctctcaaat caccaaccca gccaatctc gggctattt cagacctcca 1980
 gtgccgacga ctccaagcc gtagtcaac acaggtaatc agttcatctt ggaattcaac 2040
 cccagcgcga aggcagccaa agggaaagct ccggcaaagc ctcccaggcc tcgattgcct 2100
 cctcggaac ccaaccttc acaatcaca ctctcggcac aaacaccagc agcacctgtg 2160
 caagcgtcgg tttcaacacc ggcaccgga tcggcaccga caccgacacc agccgttgcc 2220
 tcaactccag gtccaatgtc tcctcctcct aaaccgcct gcccttcaga accgtcttct 2280
 tcatccacac ccattgttgt gagcaatgcc gctaaaccaa gagaatcctt gcagctaggt 2340
 gtactaacgc gcatgtctag gcacagcggc gggattcgat ggctcttcgt gccaggcag 2400
 gtactaggcc gtcacgggaa caggttcctg cagaggcctg ggagcaatac aaatccacta 2460
 tccgtacct ctatctcgag gagagaaaac ccttgaaaga agtcatgagt gtcatggctg 2520

ES 2 815 628 T3

aacaatatgg gtttcaagca acgtgagtcg aagattgatt cctcttctgc ttcacgttc 2580
gggttcttgg ttcgttgtaa actaattcga cgccctttct aggcceaaga tgtataaaac 2640
aaggttctct caatggggtt ttgtgaagaa taacacggag gaagaagtga aacggctgtt 2700
gtcgatgaag ttccagcgag atgccgaggg caaagtttcc gagtttgttc gaaacggcag 2760
ggtggtgaac ctaggtacct atttgaaacg gaaaggagtg acggagtatg acctcgttga 2820
tttcgaacta ccggccgacc ttccagcaca tatccgatgc agaacaccga caccacctcc 2880
ggctctgcga tcacccgatc tgctccgtgc acaggaggta gtcgttggaa atatgcgcaa 2940
ggcattccta cactgtcggc aattcgagat ggagactgag actcagattg gttggccatc 3000
aaccatggtt tggggtgctg gatcgagtga actcctggtc gaggccaact tttacttcca 3060
agctcgcgac gcagatcaag gcggtgacta cctgatgcga gctttcaagc agctcgagct 3120
ggacctcaga aagctctcgc cacaaggcat catggaacta atcctgggta tgatcaatcg 3180
ggatcccggc atgatgacgg ccctttgcaa gtaccttgca gcatattcga caaccaatct 3240
ggagcgaacc catcctctcc ggcaaattct cacttgtctg tatgaagtgc aacaaaagca 3300
tggggcgag acgttgtctg agctcctgtg gactagcatc tcgacaattg cggaggaact 3360
cgaggccatc tatggacgca agcatccgta tgtggctcgc acatgggccg atcttgcgct 3420
gttttacagc caggtgaacc cggaaaggct ggagaagttg gttgttgagc ttcgtgtgct 3480
ccagaggcag ctcgagcaac gacatgggca ttccagtgtc gaagtggttt ccatccgata 3540
tgccattctg ctgttggtct atgcgtcgtc tcccagtcg gatgcctcga agcaagccgc 3600
aaatgattat tggaacctgc tgcggaatat gaacaccatg tttcccatgc gcgactcccg 3660
tccgaatagt tactgctatc acagcccgtc caaggtcgat ccgtggacaa agaggtgccg 3720
caggcggtag gacacactcg tcaccatatt cgaggagcat gtaggcgta gaatcaatcc 3780
ctatttcgaa gaggacttcc acacgaccga gcacgctcaa gaaacgcagg atgcctgggc 3840
ggcagctctg caaatgggtt cgacgaatag atcttggggc ttcacttagc ggtttcgctt 3900
tcagaatggt ggggtgtgtc cacgcccagc tggcgcaaat gtcaggtacc caagatacct 3960
tgtctctcgt ctccatagcgc gcggtgcagc gagcttcttt ttcgagattt ctttttccgg 4020
aatctagcgg gcggttaata cccagcatca gaaggattgc ggcgcatcta cctagcgaca 4080
gtttttttcg gttacgtcct gtttgccacc tcacatcgag cctggagttt tggacagttg 4140
acaattctcg ccatcgcgga acattttcct ttctcatgtc gccgacgctg acaagtcgat 4200
ttccttggtc tttccacgca agttgttgca cctctcgagc cgggcgttag ggagggcctt 4260
ttcccccttg gatgcgtcgt catccgtgta taatagtatc cgttgttttc tgttttcctt 4320
tcctgcggtt gatatgcgat gttcagcttt cttgatactg aagaccacga gggagaggcg 4380

ES 2 815 628 T3

ggaaacgagt gttttgaaag aagttggcag tcagactttt tttgttacag ggcgaacgag 4440
gagttgacgg gaattgagca gaagaaaaaa aataccttta tcgagcattt cctgcttcat 4500
cgggaaggag aaggcaggcc ggagtttga tggcaggaac ggacgaggca cttgtcgatt 4560
gtcgtcattg cagatggtta gaccagaccg accgtccttc tgggaaggag cgggagcaca 4620
ctgattttct catcgcgcct tcttttcgac ttgacttgcc gttagtataa aaaacttcaa 4680
ctggtattaa tgaataaaga catgatgaca 4710

<210> 12
<211> 4082
<212> ADN
<213> *Trichoderma reesei*

5

<400> 12

ccagaacagc tccgtcctgg cccagcctgt attttgagca tctctggctc tcttatctca 60
gcgggggctt tctaaattga cccttctcga aagacggcac atgggatttc ccttcttggc 120
actctagacg agggaaacttt tgctcgcctc cttcgcctgc tggcttggct tggtttcctc 180
ccacgccgct gctgatgctt ttggagctgg ctgctcttgt tgtttgtttg tccaaaccac 240
acagcggctg ttgccctatc aaggcgtttt cccacatcgg ccgatatctg ggggaacgat 300
acggatatct gcccttagcc agccagccag ccagccagcc agccagccag ccctgcctg 360
tactttcgtg ctgcccctgt cccagtaggc ctctccgccc cgtgattcc ctctctcttt 420
actatccctc tacaccaggc gattatgggc acctttgccc gcctcccata cacgacctc 480
gtttgactga gcaagcacgc tcgccactct cagagagtga tgaatcaccg cgtcggagac 540
tgattattcc agccgagtcc tacgacttgc ccaggtccgt cagttcctga cagacaagac 600
agccggcgga ctgagcacca ctgctctgtc tgctcgggtg gtgggcacgt tgacatgttc 660
catcaaccgc cggaaaagcc tcgcgcgcct ggcttcgacg atgatccgtt cgggtttgtc 720
accagggcct gggagagata tggcctgggc agtgtgtcgt cgcctccgcc tcgatgcct 780
ccgcagccca tccagaggac cggacgggcc tattccaccg cttcggtaat ggactggacg 840
cccagatcgc ataccacctt tcaccaccac ggcgacggag tggccatgtc gccggtggaa 900
cacgtcgaac tggacagcct tggttcgcct cagggagggg gagggggttg ccgcttggtc 960
gcgcactttg agaataaagg ctatgtcccg ccattgcctc cgcggcccat caacaatcag 1020
ccgcatcaac cgcatacat gggccacagt ccttcaatgt cgtcgcagtt tggcagtcta 1080
aactctgtag cccatacgc taggatatcg agtcccgttg ctagcccagg tgaatcccaa 1140
tatggctcgc tcggggtgca ttcgccccca acaccacct caatcgtcac caacatgaac 1200
cgctccggca gtataagcta tggcagcttc catgacactc agcaggtgca tagtcccgg 1260
attggtacgc cttttgggag catggatggc tttatgagtt ctgcccgcgt gaacagcccg 1320

10

ES 2 815 628 T3

atggtggcga caccaatggc atcgacgcc atgatgccca gtcccatggc cgctccggga 1380
 gtgcctggta cccctggatt cgagatttgg cgacctcctc cctcgatgac tcccaaacct 1440
 gaaccctctc aatcaccaa cccagccaat ttggcgggct atttcagacc tccagtgccg 1500
 acgactccca agccggtagt caacacaggt aatcagttca tcttgaatt caaccccagc 1560
 gccaaaggcag ccaaagggaa agctccggca aagcctccca ggctcggatt gcctcctcgg 1620
 aaaccaacc tttcacaatc acaactctcg gcacaaacac cagcagcacc tgtgcaagcg 1680
 tcggtttcaa caccggcacc ggcatcggca cgcacaccga caccagccgt tgcctcaact 1740
 ccaggtccaa tgtctcctcc tcctaaaccg cctcgccctt cagaaccgtc ttcttcatcc 1800
 acaccattg ttgcacagcg gcgggattcg atggctcttc gtgccaggc aggtactagg 1860
 ccgtcacggg aacaggttcc tgcagaggcc tgggagcaat acaaatccac tatccgtacc 1920
 ctctatctcg aggagagaaa acccttgaaa gaagtcatga gtgtcatggc tgaacaatat 1980
 gggtttcaag caacgcaaaa gatgtataaa acaaggttct ctcaatgggg ttttgtgaag 2040
 aataacacgg aggaagaagt gaaacggctg ttgtcgatga agttccagcg agatgccgag 2100
 ggcaaagttt ccgagtttgt tcgaaacggc aggggtggta acctaggtac ctatgtgaaa 2160
 cggaaaggag tgacggagta tgacctggt gatttcgaac taccggccga ccttccagca 2220
 catatccgat gcagaacacc gacaccacct ccggctctgc gatcaccgga tctgctccgt 2280
 gcacaggagg tagtctgttg aatatgcgc aaggcattcc taccctgtcg gcaattcgag 2340
 atggagactg agactcagat tggttggcca tcaaccatgg tttgggtgct tggatcgagt 2400
 gaactcctgg tcgaggccaa cttttacttc gaagctcgcg acgcagatca aggcggtgac 2460
 tacctgatgc gagctttcaa gcagctcgag ctggacctca gaaagctctc gccacaaggc 2520
 atcatggaac taatcctggg tatgatcaat cgggatcccg gcatgatgac ggcccttctg 2580
 aagtacctg cagcatattc gacaaccaat ctggagcgaa cccatcctct ccggcaaatc 2640
 ttcacttgtc tgtatgaagt gcaacaaaag catggggcgc agacgttgtc tgagctcctg 2700
 tggactagca tctcgacaat tgcggaggaa ctcgaggcca tctatggacg caagcatccg 2760
 tatgtggctc gcacatggc cgatcttgcg ctgttttaca gccagtgaa cccggaaagg 2820
 ctggagaagt tggttgttga gcttcgtgtg ctccagaggc agctcgagca acgacatggg 2880
 cattccagtg tcgaagtggg ttccatccga tatgccattc tgctgttggg ctatgcgtcg 2940
 tctccccagt cggatgcctc gaagcaagcc gcaaatgatt attggaacct gctgcggaat 3000
 atgaacacca tgtttcccat gcgcgactcc cgtccgaata gttactgcta tcacagcccg 3060
 ctcaaggtcg atccgtggac aaagaggtgc cgcaggcggg acgacacact cgtcaccata 3120
 ttcgaggagc atgtaggcgt tagaatcaat ccctatttctg aagaggactt ccacacgacc 3180
 gagcacgctc aagaaacgca ggatgcctgg gcggcagctc tgcaaatggg ttcgacgaat 3240

ES 2 815 628 T3

```

agatcttggg gcttcatcta gcggtttcgc tttcagaatg ttgggtgtgt tccacgcca 3300
gctggcgcaa atgtcaggta cccaagatac cttgtctctc gtctcctagc gcgcggtgca 3360
gcgagcttct ttttcgagat ttctttttcc ggaatctagc gggcggtaa taccagcat 3420
cagaaggatt gggcgcatc tacctagcga cagttttttt cggttacgtc ctgtttgcca 3480
cctcacatcg agcctggagt tttggacagt tgacaattct cgccatcgcg gaacattttc 3540
ctttctcatg tcgccgacgc tgacaagtcg atttccttgg tctttccacg caagttgttg 3600
cacctctcga gccgggcggt agggagggcc ttttccccct tggatgcgtc gtcacccgtg 3660
tataatagta tccgttgttt tctgttttcc tttcctgcgt ttgatatgcg atgttcagct 3720
ttcttgatac tgaagaccac gagggagagg cgggaaacga gtgttttgaa agaagttggc 3780
agtcagactt tttttgttac agggcgaacg aggagttgac ggggaattgag cagaagaaaa 3840
aaaatacctt tatcgagcat ttctgcttc atcgggaagg agaaggcagg ccggagtttg 3900
gatggcagga acggacgagg cacttgcga ttgtcgtcat tgcagatggt tagaccagac 3960
cgaccgtcct tctgggaagg agcgggagca cactgatttt ctcatcgcgc cttcttttcg 4020
acttgacttg ccgtagtat aaaaaacttc aactggtatt aatgaataaa gacatgatga 4080
ca 4082

```

<210> 13
 <211> 868
 <212> PRT
 <213> *Trichoderma reesei*

5

<400> 13

```

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
1           5           10           15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
20           25           30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
35           40           45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
50           55           60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
65           70           75           80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
85           90           95

```

10

ES 2 815 628 T3

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser
130 135 140

Val Ala His Thr His Arg Ile Ser Ser Pro Val Ala Ser Pro Gly Glu
145 150 155 160

Ser Gln Tyr Gly Ser Leu Gly Val His Ser Pro Pro Thr Pro Thr Ser
165 170 175

Ile Val Thr Asn Met Asn Arg Ser Gly Ser Ile Ser Tyr Gly Ser Phe
180 185 190

His Asp Thr Gln Gln Val His Ser Pro Gly Ile Gly Thr Pro Phe Gly
195 200 205

Ser Met Asp Gly Phe Met Ser Ser Ala Arg Val Asn Ser Pro Met Val
210 215 220

Ala Thr Pro Met Ala Ser Thr Pro Met Met Pro Ser Pro Met Ala Ala
225 230 235 240

Pro Gly Val Pro Gly Thr Pro Gly Phe Glu Ile Trp Arg Pro Pro Pro
245 250 255

Ser Met Thr Pro Lys Pro Glu Pro Ser Gln Ile Thr Asn Pro Ala Asn
260 265 270

Phe Gly Gly Tyr Phe Arg Pro Pro Val Pro Thr Thr Pro Lys Pro Val
275 280 285

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

Ala Ala Lys Gly Lys Ala Pro Ala Lys Pro Pro Arg Pro Arg Leu Pro
305 310 315 320

Pro Arg Lys Pro Asn Leu Ser Gln Ser Gln Leu Ser Ala Gln Thr Pro
325 330 335

Ala Ala Pro Val Gln Ala Ser Val Ser Thr Pro Ala Pro Ala Ser Ala
340 345 350

ES 2 815 628 T3

Pro Thr Pro Thr Pro Ala Val Ala Ser Thr Pro Gly Pro Met Ser Pro
 355 360 365

Pro Pro Lys Pro Pro Arg Pro Ser Glu Pro Ser Ser Ser Ser Thr Pro
 370 375 380

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

Thr Arg Pro Ser Arg Glu Gln Val Pro Ala Glu Ala Trp Glu Gln Tyr
 405 410 415

Lys Ser Thr Ile Arg Thr Leu Tyr Leu Glu Glu Arg Lys Pro Leu Lys
 420 425 430

Glu Val Met Ser Val Met Ala Glu Gln Tyr Gly Phe Gln Ala Thr Pro
 435 440 445

Lys Met Tyr Lys Thr Arg Phe Ser Gln Trp Gly Phe Val Lys Asn Asn
 450 455 460

Thr Glu Glu Glu Val Lys Arg Leu Leu Ser Met Lys Phe Gln Arg Asp
 465 470 475 480

Ala Glu Gly Lys Val Ser Glu Phe Val Arg Asn Gly Arg Val Val Asn
 485 490 495

Leu Gly Thr Tyr Leu Lys Arg Lys Gly Val Thr Glu Tyr Asp Leu Val
 500 505 510

Asp Phe Glu Leu Pro Ala Asp Leu Pro Ala His Ile Arg Cys Arg Thr
 515 520 525

Pro Thr Pro Pro Pro Ala Leu Arg Ser Pro Asp Leu Leu Arg Ala Gln
 530 535 540

Glu Val Val Val Gly Asn Met Arg Lys Ala Phe Leu His Cys Arg Gln
 545 550 555 560

Phe Glu Met Glu Thr Glu Thr Gln Ile Gly Trp Pro Ser Thr Met Val
 565 570 575

Trp Gly Ala Gly Ser Ser Glu Leu Leu Val Glu Ala Asn Phe Tyr Phe
 580 585 590

Glu Ala Arg Asp Ala Asp Gln Gly Gly Asp Tyr Leu Met Arg Ala Phe
 595 600 605

ES 2 815 628 T3

Lys Gln Leu Glu Leu Asp Leu Arg Lys Leu Ser Pro Gln Gly Ile Met
 610 615 620

Glu Leu Ile Leu Gly Met Ile Asn Arg Asp Pro Gly Met Met Thr Ala
 625 630 635 640

Leu Cys Lys Tyr Leu Ala Ala Tyr Ser Thr Thr Asn Leu Glu Arg Thr
 645 650 655

His Pro Leu Arg Gln Ile Phe Thr Cys Leu Tyr Glu Val Gln Gln Lys
 660 665 670

His Gly Ala Gln Thr Leu Ser Glu Leu Leu Trp Thr Ser Ile Ser Thr
 675 680 685

Ile Ala Glu Glu Leu Glu Ala Ile Tyr Gly Arg Lys His Pro Tyr Val
 690 695 700

Ala Arg Thr Trp Ala Asp Leu Ala Leu Phe Tyr Ser Gln Val Asn Pro
 705 710 715 720

Glu Arg Leu Glu Lys Leu Val Val Glu Leu Arg Val Leu Gln Arg Gln
 725 730 735

Leu Glu Gln Arg His Gly His Ser Ser Val Glu Val Val Ser Ile Arg
 740 745 750

Tyr Ala Ile Leu Leu Leu Val Tyr Ala Ser Ser Pro Gln Ser Asp Ala
 755 760 765

Ser Lys Gln Ala Ala Asn Asp Tyr Trp Asn Leu Leu Arg Asn Met Asn
 770 775 780

Thr Met Phe Pro Met Arg Asp Ser Arg Pro Asn Ser Tyr Cys Tyr His
 785 790 795 800

Ser Pro Leu Lys Val Asp Pro Trp Thr Lys Arg Cys Arg Arg Arg Tyr
 805 810 815

Asp Thr Leu Val Thr Ile Phe Glu Glu His Val Gly Val Arg Ile Asn
 820 825 830

Pro Tyr Phe Glu Glu Asp Phe His Thr Thr Glu His Ala Gln Glu Thr
 835 840 845

Gln Asp Ala Trp Ala Ala Ala Leu Gln Met Gly Ser Thr Asn Arg Ser

ES 2 815 628 T3

850

855

860

Trp Gly Phe Ile
865

<210> 14

<211> 288

<212> PRT

<213> *Trichoderma reesei*

5

<400> 14

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
1 5 10 15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
20 25 30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
35 40 45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
50 55 60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
65 70 75 80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
85 90 95

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser
130 135 140

Val Ala His Thr His Arg Ile Ser Ser Pro Val Ala Ser Pro Gly Glu
145 150 155 160

Ser Gln Tyr Gly Ser Leu Gly Val His Ser Pro Pro Thr Pro Thr Ser
165 170 175

Ile Val Thr Asn Met Asn Arg Ser Gly Ser Ile Ser Tyr Gly Ser Phe
180 185 190

His Asp Thr Gln Gln Val His Ser Pro Gly Ile Gly Thr Pro Phe Gly

10

ES 2 815 628 T3

195

200

205

Ser Met Asp Gly Phe Met Ser Ser Ala Arg Val Asn Ser Pro Met Val
 210 215 220

Ala Thr Pro Met Ala Ser Thr Pro Met Met Pro Ser Pro Met Ala Ala
 225 230 235 240

Pro Gly Val Pro Gly Thr Pro Gly Phe Glu Ile Trp Arg Pro Pro Pro
 245 250 255

Ser Met Thr Pro Lys Pro Glu Pro Ser Gln Ile Thr Asn Pro Ala Asn
 260 265 270

Phe Gly Gly Tyr Phe Arg Pro Pro Gln Cys Arg Arg Leu Pro Ser Arg
 275 280 285

<210> 15
 <211> 387
 <212> PRT
 <213> *Trichoderma reesei*

5

<400> 15

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
 1 5 10 15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
 20 25 30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
 35 40 45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
 50 55 60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
 65 70 75 80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
 85 90 95

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
 115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser

10

ES 2 815 628 T3

130																			
Val	Ala	His	Thr	His	Arg	Ile	Ser	Ser	Pro	Val	Ala	Ser	Pro	Gly	Glu				
145					150					155					160				
Ser	Gln	Tyr	Gly	Ser	Leu	Gly	Val	His	Ser	Pro	Pro	Thr	Pro	Thr	Ser				
				165					170					175					
Ile	Val	Thr	Asn	Met	Asn	Arg	Ser	Gly	Ser	Ile	Ser	Tyr	Gly	Ser	Phe				
			180					185					190						
His	Asp	Thr	Gln	Gln	Val	His	Ser	Pro	Gly	Ile	Gly	Thr	Pro	Phe	Gly				
		195					200					205							
Ser	Met	Asp	Gly	Phe	Met	Ser	Ser	Ala	Arg	Val	Asn	Ser	Pro	Met	Val				
	210					215					220								
Ala	Thr	Pro	Met	Ala	Ser	Thr	Pro	Met	Met	Pro	Ser	Pro	Met	Ala	Ala				
225					230					235					240				
Pro	Gly	Val	Pro	Gly	Thr	Pro	Gly	Phe	Glu	Ile	Trp	Arg	Pro	Pro	Pro				
				245					250					255					
Ser	Met	Thr	Pro	Lys	Pro	Glu	Pro	Ser	Gln	Ile	Thr	Asn	Pro	Ala	Asn				
			260					265					270						
Phe	Gly	Gly	Tyr	Phe	Arg	Pro	Pro	Val	Pro	Thr	Thr	Pro	Lys	Pro	Val				
		275					280					285							
Val	Asn	Thr	Gly	Asn	Gln	Phe	Ile	Leu	Glu	Phe	Asn	Pro	Ser	Ala	Lys				
	290					295					300								
Ala	Ala	Lys	Gly	Lys	Ala	Pro	Ala	Lys	Pro	Pro	Arg	Pro	Arg	Leu	Pro				
305					310					315				320					
Pro	Arg	Lys	Pro	Asn	Leu	Ser	Gln	Ser	Gln	Leu	Ser	Ala	Gln	Thr	Pro				
				325					330					335					
Ala	Ala	Pro	Val	Gln	Ala	Ser	Val	Ser	Thr	Pro	Ala	Pro	Ala	Ser	Ala				
			340					345					350						
Pro	Thr	Pro	Thr	Pro	Ala	Val	Ala	Ser	Thr	Pro	Gly	Pro	Met	Ser	Pro				
		355					360					365							
Pro	Pro	Lys	Pro	Pro	Arg	Pro	Ser	Glu	Pro	Ser	Ser	Ser	Ser	Thr	Pro				
	370					375						380							

ES 2 815 628 T3

Ile Val Ala
385

5
<210> 16
<211> 430
<212> PRT
<213> *Trichoderma reesei*

<400> 16

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
1 5 10 15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
20 25 30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
35 40 45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
50 55 60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
65 70 75 80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
85 90 95

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser
130 135 140

Val Ala His Thr His Arg Ile Ser Ser Pro Val Ala Ser Pro Gly Glu
145 150 155 160

Ser Gln Tyr Gly Ser Leu Gly Val His Ser Pro Pro Thr Pro Thr Ser
165 170 175

Ile Val Thr Asn Met Asn Arg Ser Gly Ser Ile Ser Tyr Gly Ser Phe
180 185 190

His Asp Thr Gln Gln Val His Ser Pro Gly Ile Gly Thr Pro Phe Gly
195 200 205

10

ES 2 815 628 T3

Ser Met Asp Gly Phe Met Ser Ser Ala Arg Val Asn Ser Pro Met Val
 210 215 220

Ala Thr Pro Met Ala Ser Thr Pro Met Met Pro Ser Pro Met Ala Ala
 225 230 235 240

Pro Gly Val Pro Gly Thr Pro Gly Phe Glu Ile Trp Arg Pro Pro Pro
 245 250 255

Ser Met Thr Pro Lys Pro Glu Pro Ser Gln Ile Thr Asn Pro Ala Asn
 260 265 270

Phe Gly Gly Tyr Phe Arg Pro Pro Val Pro Thr Thr Pro Lys Pro Val
 275 280 285

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

Ala Ala Lys Gly Lys Ala Pro Ala Lys Pro Pro Arg Pro Arg Leu Pro
 305 310 315 320

Pro Arg Asn Pro Thr Phe His Asn His Asn Ser Arg His Lys His Gln
 325 330 335

Gln His Leu Cys Lys Arg Arg Phe Gln His Arg His Arg His Arg His
 340 345 350

Arg His Arg His Gln Pro Leu Pro Gln Leu Gln Val Gln Cys Leu Leu
 355 360 365

Leu Leu Asn Arg Leu Ala Leu Gln Asn Arg Leu Leu His Pro His Pro
 370 375 380

Leu Leu His Ser Gly Gly Ile Arg Trp Leu Phe Val Pro Arg Gln Val
 385 390 395 400

Leu Gly Arg His Gly Asn Arg Phe Leu Gln Arg Pro Gly Ser Asn Thr
 405 410 415

Asn Pro Leu Ser Val Pro Ser Ile Ser Arg Arg Glu Asn Pro
 420 425 430

<210> 17
 <211> 317
 <212> PRT
 <213> *Trichoderma reesei*

5

<400> 17

ES 2 815 628 T3

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
1 5 10 15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
20 25 30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
35 40 45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
50 55 60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
65 70 75 80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
85 90 95

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser
130 135 140

Val Ala His Thr His Arg Ile Ser Ser Pro Val Ala Ser Pro Gly Glu
145 150 155 160

Ser Gln Tyr Gly Ser Leu Gly Val His Ser Pro Pro Thr Pro Thr Ser
165 170 175

Ile Val Thr Asn Met Asn Arg Ser Gly Ser Ile Ser Tyr Gly Ser Phe
180 185 190

His Asp Thr Gln Gln Val His Ser Pro Gly Ile Gly Thr Pro Phe Gly
195 200 205

Ser Met Asp Gly Phe Met Ser Ser Ala Arg Val Asn Ser Pro Met Val
210 215 220

Ala Thr Pro Met Ala Ser Thr Pro Met Met Pro Ser Pro Met Ala Ala
225 230 235 240

Pro Gly Val Pro Gly Thr Pro Gly Phe Glu Ile Trp Arg Pro Pro Pro
245 250 255

ES 2 815 628 T3

Ser Met Thr Pro Lys Pro Glu Pro Ser Gln Ile Thr Asn Pro Ala Asn
 260 265 270

Phe Gly Gly Tyr Phe Arg Pro Pro Val Pro Thr Thr Pro Lys Pro Val
 275 280 285

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

Ala Ala Lys Gly Lys Ala Pro Ala Lys Pro Pro Arg Pro
 305 310 315

<210> 18
 <211> 297
 <212> PRT
 <213> *Trichoderma reesei*

5

<400> 18

Met Phe His Gln Pro Pro Glu Lys Pro Arg Ala Pro Gly Phe Asp Asp
 1 5 10 15

Asp Pro Phe Gly Phe Val Thr Gln Ala Trp Glu Arg Tyr Gly Leu Gly
 20 25 30

Ser Val Ser Ser Pro Pro Pro Ala Met Pro Pro Gln Pro Ile Gln Arg
 35 40 45

Thr Gly Arg Ala Tyr Ser Thr Ala Ser Val Met Asp Trp Thr Pro Glu
 50 55 60

Ile Asp Thr Thr Phe His His His Gly Asp Gly Val Ala Met Ser Pro
 65 70 75 80

Val Glu His Val Glu Leu Asp Ser Leu Gly Ser Pro Gln Gly Gly Gly
 85 90 95

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

Pro Leu Pro Pro Arg Pro Ile Asn Asn Gln Pro His Gln Pro His His
 115 120 125

Met Gly His Ser Pro Ser Met Ser Ser Gln Phe Gly Ser Leu Asn Ser
 130 135 140

Val Ala His Thr His Arg Ile Ser Ser Pro Val Ala Ser Pro Gly Glu
 145 150 155 160

10

ES 2 815 628 T3

Ser Gln Tyr Gly Ser Leu Gly Val His Ser Pro Pro Thr Pro Thr Ser
 165 170 175

Ile Val Thr Asn Met Asn Arg Ser Gly Ser Ile Ser Tyr Gly Ser Phe
 180 185 190

His Asp Thr Gln Gln Val His Ser Pro Gly Ile Gly Thr Pro Phe Gly
 195 200 205

Ser Met Asp Gly Phe Met Ser Ser Ala Arg Val Asn Ser Pro Met Val
 210 215 220

Ala Thr Pro Met Ala Ser Thr Pro Met Met Pro Ser Pro Met Ala Ala
 225 230 235 240

Pro Gly Val Pro Gly Thr Pro Gly Phe Glu Ile Trp Arg Pro Pro Pro
 245 250 255

Ser Met Thr Pro Lys Pro Glu Pro Ser Gln Ile Thr Asn Pro Ala Asn
 260 265 270

Phe Gly Gly Tyr Phe Arg Pro Pro Val Pro Thr Thr Pro Lys Pro Val
 275 280 285

Val Asn Thr Gly Asn Gln Phe Ile Leu
 290 295

REIVINDICACIONES

1. Una célula hospedadora que comprende al menos un gen cromosómico inactivado en donde el gen cromosómico inactivado comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13; el gen cromosómico inactivado se inactiva por disrupción; y la célula hospedadora tiene actividad proteasa reducida en comparación con la célula hospedadora sin dicha inactivación.
2. La célula hospedadora de la reivindicación 1, en donde la célula hospedadora se selecciona entre el grupo que consiste en células de hongos filamentosos de la División *Ascomycota*, Subdivisión *Pezizomycotina*; preferentemente del grupo que consiste en miembros de la Clase *Sordariomycetes*, Subclase *Hypocreomycetidae*, Órdenes *Hipocreales* y *Microascales* y *Aspergillus*, *Chrysosporium*, *Myceliophthora* y *Humicola*; más preferentemente del grupo que consiste en las Familias *Hypocreaceae*, *Nectriaceae*, *Clavicipitaceae*, *Microascaceae* y Géneros *Trichoderma*, *Hypocrea*, *Fusarium*, *Gibberella*, *Nectria*, *Stachybotrys*, *Claviceps*, *Metarhizium*, *Villosiclava*, *Ophiocordyceps*, *Cephalosporium* y *Scedosporium*; más preferentemente entre el grupo que consiste en *Trichoderma reesei*, *Hypocrea jecorina*, *T. citrinoviridae*, *T. longibrachiatum*, *T. virens*, *T. harzianum*, *T. asperellum*, *T. atroviridae*, *T. parareesei*, *Fusarium oxysporum*, *F. graminearum*, *F. pseudograminearum*, *F. venenatum*, *Gibberella fujikuroi*, *G. moniliformis*, *G. zeaeae*, *Nectria haematococca*, *Stachybotrys chartarum*, *S. chlorohalonata*, *Claviceps purpurea*, *Metarhizium acridum*, *M. anisopliae*, *Villosiclava virens*, *Ophiocordyceps sinensis*, *Acremonium chrysogenum*, *Scedosporium apiospermum*, *Aspergillus niger*, *A. awamori*, *A. oryzae*, *Chrysosporium lucknowense*, *Myceliophthora thermophila*, *Humicola insolens* y *Humicola grisea*, mucho más preferentemente *Trichoderma reesei*.
3. La célula hospedadora de las reivindicaciones 1 o 2, en donde el gen cromosómico inactivado comprende:
- un polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado; o
- un polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado, seleccionándose el polinucleótido entre el grupo que consiste en:
- a) un polinucleótido que comprende una secuencia que tiene al menos un 55 % de identidad de secuencia con los nucleótidos 1141-3889 del SEQ ID NO: 11;
- b) el polinucleótido del SEQ ID NO: 12 o la secuencia codificante del mismo;
- c) el polinucleótido del SEQ ID NO: 11 o la secuencia codificante del mismo; y
- d) una secuencia de nucleótidos hibridable con una secuencia de nucleótidos que es complementaria con una cualquiera de a) a c) en condiciones altamente rigurosas.
4. La célula hospedadora de una cualquiera de las reivindicaciones 1-3 que comprende elementos genéticos para permitir expresar, en condiciones adecuadas para promover la expresión, al menos una proteína de interés codificada por un polinucleótido recombinante.
5. La célula hospedadora de la reivindicación 4, en donde la proteína de interés se selecciona entre el grupo que consiste en una proteína, anticuerpo, fragmento de anticuerpo, proteína terapéutica, biosimilar, proteína de dominios múltiples, hormona peptídica, péptido antimicrobiano, péptido, módulo de unión a hidratos de carbono, enzima, celulasa, proteasa, inhibidor de proteasas, aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, quitinasa, cutinasa, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, invertasa, laccasa, lipasa, mananasa, mutanasa, oxidasa, enzima pectinolítica, peroxidasa, fosfolipasa, fitasa, fosfatasa, polifenoloxidasas, enzima redox, enzima proteolítica, ribonucleasa, transglutaminasa y xilanasas farmacológicamente activos.
6. Un método de producción de proteínas que comprende
- a. cultivar una célula hospedadora en condiciones adecuadas para producir proteína, en donde la célula hospedadora comprende al menos un gen cromosómico inactivado que comprende una secuencia de ácido nucleico que codifica un polipéptido que comprende una secuencia que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13; el gen cromosómico inactivado se inactiva por disrupción; y la célula hospedadora tiene actividad proteasa reducida en comparación con la célula hospedadora sin dicha inactivación; y opcionalmente

b. recuperar la proteína producida.

7. El método de producción de proteínas de la reivindicación 6, en donde la célula hospedadora comprende elementos genéticos para permitir expresar, en condiciones adecuadas para promover la expresión, al menos una proteína de interés codificada por un polinucleótido recombinante, y en donde la proteína de interés se selecciona entre el grupo que consiste en una proteína, anticuerpo, fragmento de anticuerpo, proteína terapéutica, biosimilar, proteína de dominios múltiples, hormona peptídica, péptido antimicrobiano, péptido, módulo de unión a hidratos de carbono, enzima, celulasa, proteasa, inhibidor de proteasas, aminopeptidasa, amilasa, carbohidrasa, carboxipeptidasa, catalasa, quitinasa, cutinasa, desoxirribonucleasa, esterasa, alfa-galactosidasa, beta-galactosidasa, glucoamilasa, alfa-glucosidasa, beta-glucosidasa, invertasa, laccasa, lipasa, mananasa, mutanasa, oxidasa, enzima pectinolítica, peroxidasa, fosfolipasa, fitasa, fosfatasa, polifenoloxidasa, enzima redox, enzima proteolítica, ribonucleasa, transglutaminasa y xilanasa farmacológicamente activos.
8. Un método para generar una célula hospedadora para la producción de proteínas, que comprende suprimir la expresión de genes de proteasas endógenas en una célula hospedadora inhibiendo al menos parcialmente la transcripción o la traducción de un polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado; o un polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado, seleccionándose el polinucleótido entre el grupo que consiste en:
- a) un polinucleótido que comprende una secuencia que tiene al menos un 55 % de identidad de secuencia con los nucleótidos 1141-3889 del SEQ ID NO: 11;
 - b) el polinucleótido del SEQ ID NO: 12 o la secuencia codificante del mismo;
 - c) el polinucleótido del SEQ ID NO: 11 o la secuencia codificante del mismo; y
 - d) una secuencia de nucleótidos hibridable con una secuencia de nucleótidos que es complementaria con una cualquiera de a) a c) en condiciones altamente rigurosas.
9. El método de la reivindicación 8, en donde la inhibición es mediante mutación, delección, inserción, interferencia de ARN, anticuerpo, inhibidor de moléculas pequeñas o CRISPR/Cas9.
10. Una célula hospedadora que puede obtenerse usando el método de las reivindicaciones 8 o 9.
11. La célula hospedadora de la reivindicación 10, en donde la célula hospedadora comprende adicionalmente un ácido nucleico que codifica una proteína heteróloga.
12. Un polinucleótido modificado que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado; y en donde el polinucleótido modificado contiene al menos una modificación que da como resultado la incapacidad, al menos parcial, de un producto génico que puede obtenerse mediante transcripción y/o traducción de un gen cromosómico que comprende el polinucleótido modificado para inducir la expresión de proteasas endógenas en una célula hospedadora.
13. El polinucleótido modificado de la reivindicación 12 que comprende elementos genéticos para permitir su transcripción y/o traducción en una célula hospedadora.
14. Un vector que comprende el polinucleótido modificado de las reivindicaciones 12 o 13.
15. El vector de la reivindicación 14 que comprende elementos genéticos para incorporar el polinucleótido modificado de la reivindicación 12 en un genoma de una célula hospedadora.
16. Un método para inducir la expresión de proteasas en una célula hospedadora proporcionando un regulador de proteasas dentro de o en contacto con la célula hospedadora; en donde el regulador de proteasas se selecciona entre el grupo que consiste en:
- a) un polipéptido o un producto génico codificado por la secuencia codificante del polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende

el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado;

5 b) un polipéptido o un producto génico codificado por la secuencia codificante del polinucleótido que codifica una proteína que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13, en donde la inactivación de un gen cromosómico que comprende el polinucleótido da como resultado la supresión de la producción de proteasas endógenas de la célula hospedadora en comparación con una célula hospedadora en donde el gen cromosómico que comprende el polinucleótido no está inactivado, seleccionándose el polinucleótido entre el grupo que consiste en:

10 i) un polinucleótido que comprende una secuencia que tiene al menos un 55 % de identidad de secuencia con los nucleótidos 1141-3889 del SEQ ID NO: 11;

ii) el polinucleótido del SEQ ID NO: 12 o la secuencia codificante del mismo;

15 iii) el polinucleótido del SEQ ID NO: 11 o la secuencia codificante del mismo; y

iv) una secuencia de nucleótidos hibridable con una secuencia de nucleótidos que es complementaria con una cualquiera de i) a iii) en condiciones altamente rigurosas;

c) un polipéptido o un producto génico codificado por el polinucleótido de los SEQ ID NO: 11 o 12; y

20 d) un polipéptido que comprende una secuencia de aminoácidos que tiene al menos un 90 % de identidad de secuencia con los aminoácidos 402-533 del SEQ ID NO: 13.

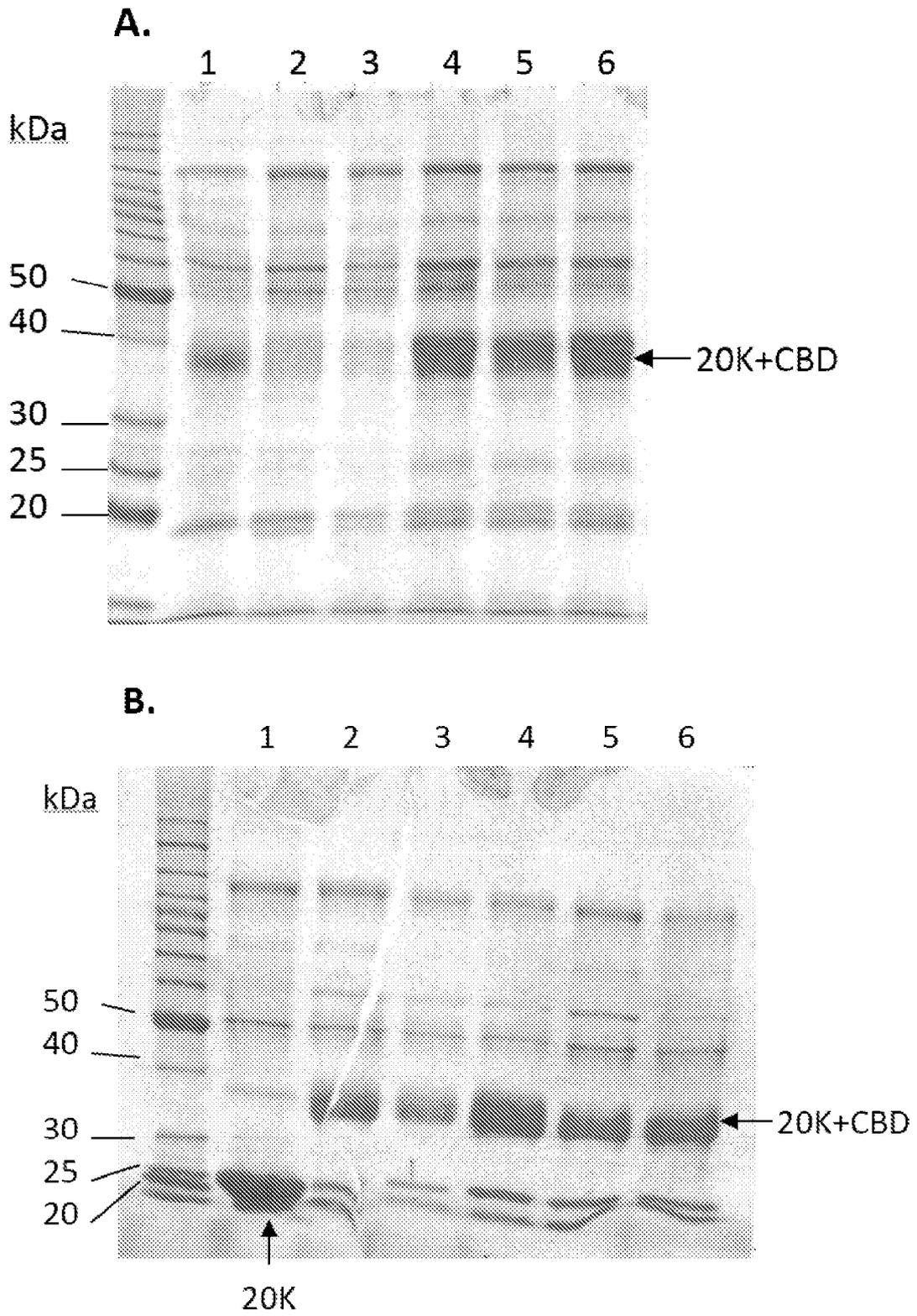


Fig. 1

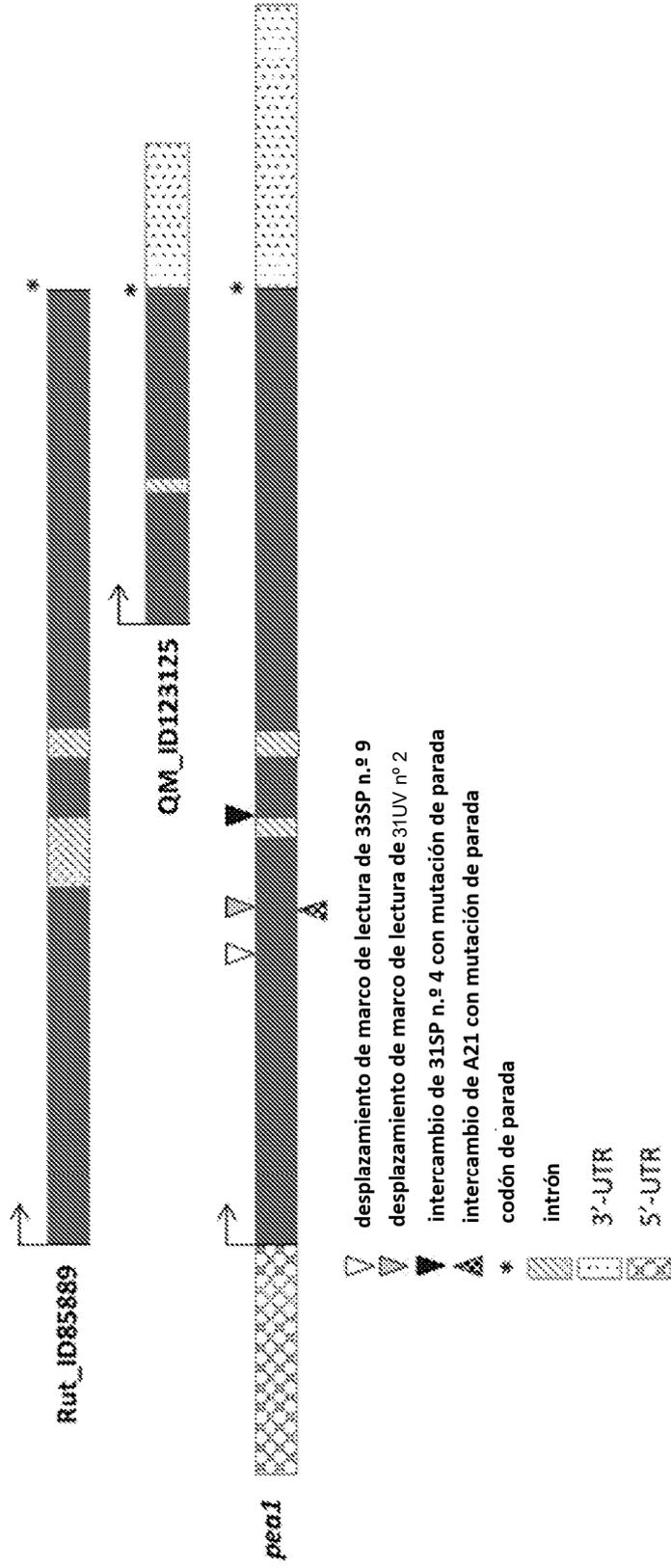


Fig. 2

ES 2 815 628 T3

1141 atgttccatc aaccgccgga aaagcctcgc gcgcctggct tcgacgatga tccggttcggg
 M F H Q P P E K P R A P G F D D D P F G
 1201 ttgtgcaccc aggcctggga gagatatggc ctgggcagtg tgctcgtcgc tccgcctgcg
 F V T Q A W E R Y G L G S V S S P P P A
 1261 atgcctccgc agcccatcca gaggaccgga cgggcctatt ccaccgcttc ggtaatggac
 M P P Q P I Q R T G R A Y S T A S V M D
 1321 tggacgcccg agatcgatac cacctttcac caccacggcg acggagtggc catgtcgcgc
 W T P E I D T T F H H H G D G V A M S P
 1381 gtggaacacg tcgaaactgga cagccttggc tcgcctcagg gagggggagg ggttggccgc
 V E H V E L D S L G S P Q G G G G V G R
 1441 ttggtcgcgc actttgagaa taaaggctat gtcccgccat tgcctccgcg gcccatcaac
 L V A H F E N K G Y V P P L P P R P I N
 1501 aatcagccgc atcaaccgca tcacatgggc cacagtcctt caatgtcgtc gcagtttggc
 N Q P H Q P H H M G H S P S M S S Q F G
 1561 agtctaaact ctgtagccca tacgcatagg atatcgagtc ccggttctag ccaggtgaa
 S L N S V A H T H R I S S P V A S P G E
 1621 tcccaatatg gctcgtcctg ggtgcattcg cccccaacac ccacctcaat cgtcaccaac
 S Q Y G S L G V H S P P T P T S I V T N
 1681 atgaaccgct ccggcagtat aagctatggc agcttccatg acactcagca ggtgcatagt
 M N R S G S I S Y G S F H D T Q Q V H S
 1741 cccggtattg gtacgccttt tgggagcatg gatggcttta tgagttctgc ccgcgtgaac
 P G I G T P F G S M D G F M S S A R V N
 1801 agcccgatgg tggcgacacc aatggcatcg acgcccata tgcccagtc catggccgct
 S P M V A T P M A S T P M M P S P M A A
 1861 cggggagtgc ctggtacccc tggattcgag atttggcgac ctccctccctc gatgactccc
 P G V P G T P G F E I W R P P P S M T P
 1921 aaacctgaac cctctcaaat caccaacca gcccaatttcg gcggctatatt cagacctcca
 K P E P S Q I T N P A N F G G Y F R P P
 1981 gtgccgacga ctcccaagcc ggtagtcaac acaggtaatc agttcatctt ggaattcaac
 V P T T P K P V V N T G N Q F I L E F N
 2041 cccagcgcca aggcagccaa agggaaagct ccggcaaacg ctcccaggcc tcgattgcct
 P S A K A A K G K A P A K P P R P R L P
 2101 cctcggaaac ccaacctttc acaatcacia ctctcggcac aaacaccage agcacctgtg
 P R K P N L S Q S Q L S A Q T P A A P V
 2161 caagcgtcgg tttcaacacc ggcaccggca tcggcaccga caccgacacc agccgttgcc
 Q A S V S T P A P A S A P T P T P A V A
 2221 tcaactccag gtccaatgtc tctcctcct aaaccgcctc gcccttcaga accgttctct
 S T P G P M S P P P K P P R P S E P S S
 2281 tcatccacac ccattgttgt gagcaatgcc gctaaaccaa gagaatcctt qcagctaggt
 S S T P I V
 2341 gtactaacgc qcatgtctag gcacagcggc gggattcgat ggctcttcgt gcccaggcag
 A Q R R D S M A L R A Q A
 2401 gtactaggcc gtcacgggaa caggttcctg cagaggcctg ggagcaatac aaatccacta
 G T R P S R E Q V P A E A W E Q Y K S T
 2461 tccgtaccct ctatctcgag gagagaaaac ccttgaaga agtcatgagt gtcattggctg
 I R T L Y L E E R K P L K E V M S V M A
 2521 aacaatatgg gtttcaagca acgtgagtcg aagattgatt cctcttctgc ttcattcgttc
 E Q Y G F Q A
 2581 gggttcttgg ttcggtgtaa actaattcga cgccctttct aggccaaaga tgtataaaac
 T P K M Y K
 2641 aaggttctct caatgggggtt ttgtgaagaa taacacggag gaagaagtga aacggctggt
 T R F S Q W G F V K N N T E E E V K R L

Fig. 3

ES 2 815 628 T3

2701 gtcgatgaag ttccagcgag atgccgaggg caaagtttcc gagtttgttc gaaacggcag
 L S M K F Q R D A E G K V S E F V R N G
 2761 ggtgggtgaac ctaggtacct atttgaacg gaaaggagtg acggagtatg acctcgttga
 R V V N L G T Y L K R K G V T E Y D L V
 2821 tttcgaacta ccggccgacc ttccagcaca tatccgatgc agaacaccga caccacctcc
 D F E L P A D L P A H I R C R T P T P P
 2881 ggctctgcga tcacccgatc tgctccgtgc acaggaggta gtcgttggaa atatgcgcaa
 P A L R S P D L L R A Q E V V V G N M R
 2941 ggcatccta cactgtcggc aattcgagat ggagactgag actcagattg gttggccatc
 K A F L H C R Q F E M E T E T Q I G W P
 3001 aaccatgggt tggggtgctg gatcgagtga actcctggtc gaggccaact tttacttcga
 S T M V W G A G S S E L L V E A N F Y F
 3061 agctcgcgac gcagatcaag gcggtgacta cctgatgcga gctttcaagc agctcgagct
 E A R D A D Q G G D Y L M R A F K Q L E
 3121 ggacctcaga aagctctcgc cacaaggcat catggaacta atcctgggta tgatcaatcg
 L D L R K L S P Q G I M E L I L G M I N
 3181 ggatcccggc atgatgacgg ccctttgcaa gtaccttgca gcatattcga caaccaatct
 R D P G M M T A L C K Y L A A Y S T T N
 3241 ggagcgaacc catcctctcc ggcaaatctt cacttgctctg tatgaagtgc aacaaaagca
 L E R T H P L R Q I F T C L Y E V Q Q K
 3301 tggggcgcag acgttgctctg agctcctgtg gactagcatc tcgacaattg cggaggaact
 H G A Q T L S E L L W T S I S T I A E E
 3361 cgaggccatc tatggacgca agcatccgta tgtggctcgc acatgggccg atcttgcgct
 L E A I Y G R K H P Y V A R T W A D L A
 3421 gttttacagc caggtgaacc cggaaaggct ggagaagttg gttgttgagc ttcgtgtgct
 L F Y S Q V N P E R L E K L V V E L R V
 3481 ccagaggcag ctcgagcaac gacatgggca ttccagtgtc gaagtggttt ccatccgata
 L Q R Q L E Q R H G H S S V E V V S I R
 3541 tgccattctg ctgttggtct atgcgtcgtc tccccagtcg gatgcctcga agcaagccgc
 Y A I L L L V Y A S S P Q S D A S K Q A
 3601 aatgattat tggaacctgc tgcggaatat gaacaccatg tttcccatgc gcgactcccg
 A N D Y W N L L R N M N T M F P M R D S
 3661 tccgaatagt tactgctatc acagcccgtc caaggctgat ccgtggacaa agaggtgccg
 R P N S Y C Y H S P L K V D P W T K R C
 3721 caggcgggtac gacacactcg tcacatatt cgaggagcat gtaggcgtta gaatcaatcc
 R R R Y D T L V T I F E E H V G V R I N
 3781 ctatttcgaa gaggacttcc acacgaccga gcacgtcaa gaaacgcagg atgcctgggc
 P Y F E E D F H T T E H A Q E T Q D A W
 3841 ggcagctctg caaatgggtt cgacgaatag atcttggggc ttcacttag 3889
 A A A L Q M G S T N R S W G F I -

Fig. 3 (continuación)

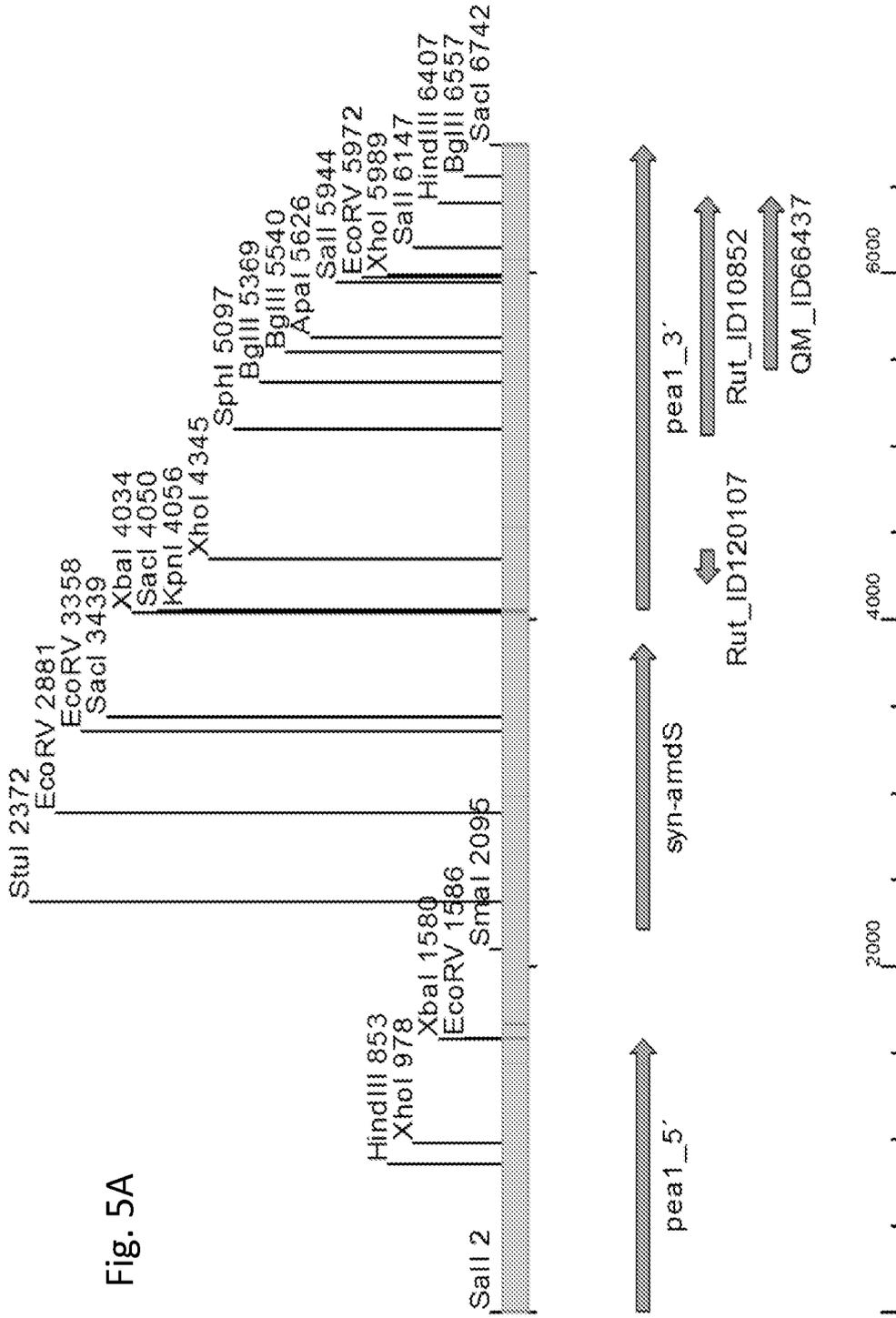


Fig. 5A

Casete de delección de pALK4104 (6748 pb)

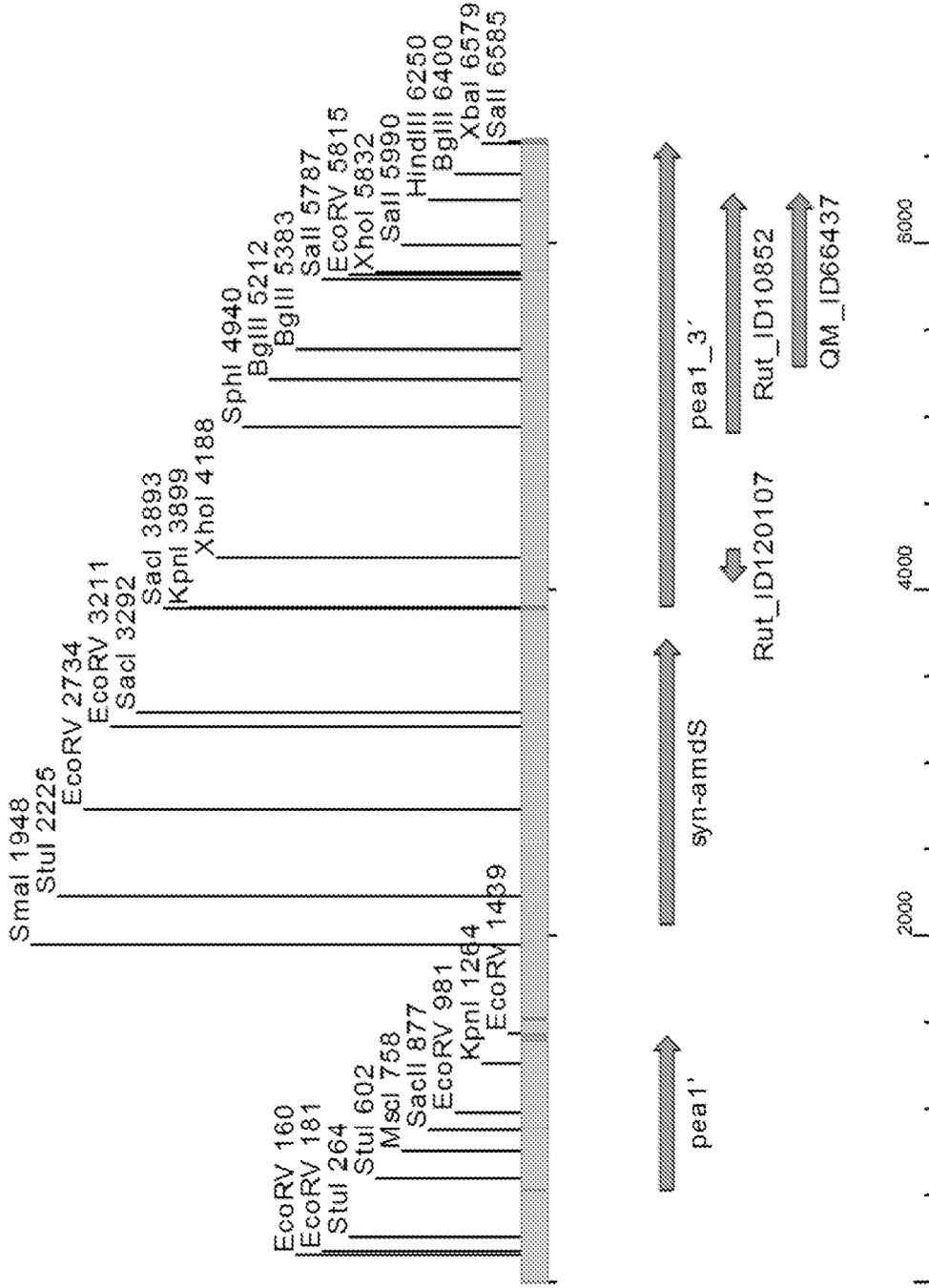


Fig. 5B

Casete de delección de pALK4106 (6595 pb)

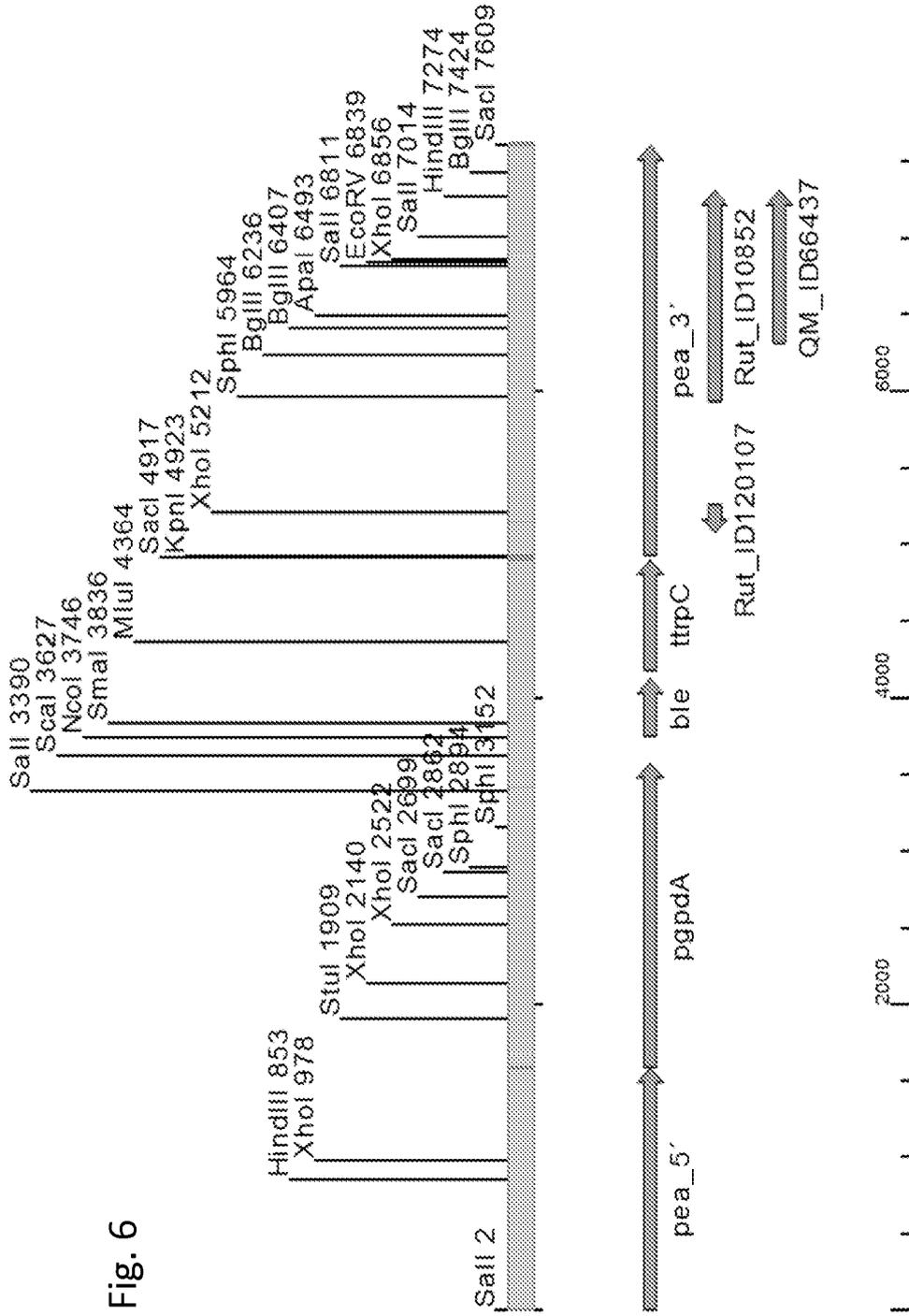


Fig. 6

Casete de delección de pALK4107 (7615 pb)

ES 2 815 628 T3

31UV n.° 22	MFHQPPPEKPRAPGFDDDDPFGFVTQAWERYGLGSVSSPPPAMPPQPIQRTGRAYSTASVMD
31SP n.° 4	MFHQPPPEKPRAPGFDDDDPFGFVTQAWERYGLGSVSSPPPAMPPQPIQRTGRAYSTASVMD
A21	MFHQPPPEKPRAPGFDDDDPFGFVTQAWERYGLGSVSSPPPAMPPQPIQRTGRAYSTASVMD
33SP n.° 9	MFHQPPPEKPRAPGFDDDDPFGFVTQAWERYGLGSVSSPPPAMPPQPIQRTGRAYSTASVMD
pALK4106	MFHQPPPEKPRAPGFDDDDPFGFVTQAWERYGLGSVSSPPPAMPPQPIQRTGRAYSTASVMD
31UV n.° 22	WTPEIDTTFHGGDGVAMSPVEHVELDSLGSPPQGGGGVGRVLAHFENKGYVPPPLPPRPIN
31SP n.° 4	WTPEIDTTFHGGDGVAMSPVEHVELDSLGSPPQGGGGVGRVLAHFENKGYVPPPLPPRPIN
A21	WTPEIDTTFHGGDGVAMSPVEHVELDSLGSPPQGGGGVGRVLAHFENKGYVPPPLPPRPIN
33SP n.° 9	WTPEIDTTFHGGDGVAMSPVEHVELDSLGSPPQGGGGVGRVLAHFENKGYVPPPLPPRPIN
pALK4106	WTPEIDTTFHGGDGVAMSPVEHVELDSLGSPPQGGGGVGRVLAHFENKGYVPPPLPPRPIN
31UV n.° 22	NQPHQPHHMGHSPSMSSQFGSLNSVAHTHRISPPVSPGESQYGLSLGVHSPPTPTSIVTN
31SP n.° 4	NQPHQPHHMGHSPSMSSQFGSLNSVAHTHRISPPVSPGESQYGLSLGVHSPPTPTSIVTN
A21	NQPHQPHHMGHSPSMSSQFGSLNSVAHTHRISPPVSPGESQYGLSLGVHSPPTPTSIVTN
33SP n.° 9	NQPHQPHHMGHSPSMSSQFGSLNSVAHTHRISPPVSPGESQYGLSLGVHSPPTPTSIVTN
pALK4106	NQPHQPHHMGHSPSMSSQFGSLNSVAHTHRISPPVSPGESQYGLSLGVHSPPTPTSIVTN
31UV n.° 22	MNRSGSISYGSFHDTQQVHSPGIGTFPGSMDGMSSARVNSPMVATPMASPPMMPSPMAA
31SP n.° 4	MNRSGSISYGSFHDTQQVHSPGIGTFPGSMDGMSSARVNSPMVATPMASPPMMPSPMAA
A21	MNRSGSISYGSFHDTQQVHSPGIGTFPGSMDGMSSARVNSPMVATPMASPPMMPSPMAA
33SP n.° 9	MNRSGSISYGSFHDTQQVHSPGIGTFPGSMDGMSSARVNSPMVATPMASPPMMPSPMAA
pALK4106	MNRSGSISYGSFHDTQQVHSPGIGTFPGSMDGMSSARVNSPMVATPMASPPMMPSPMAA
31UV n.° 22	PGVPGTPGFEIWRPPSMTPKPEPSQITNPANFGGYFRPPVPTTPKPVVNTGNQFIFLNFN
31SP n.° 4	PGVPGTPGFEIWRPPSMTPKPEPSQITNPANFGGYFRPPVPTTPKPVVNTGNQFIFLNFN
A21	PGVPGTPGFEIWRPPSMTPKPEPSQITNPANFGGYFRPPVPTTPKPVVNTGNQFIFLNFN
33SP n.° 9	PGVPGTPGFEIWRPPSMTPKPEPSQITNPANFGGYFRPP <u>QCRRLPSR</u> -----
pALK4106	PGVPGTPGFEIWRPPSMTPKPEPSQITNPANFGGYFRPPVPTTPKPVVNTGNQFIL---
31UV n.° 22	PSAKAAKGKAPAKPPRRLPPRNP <u>TFHNNHNSRHKHQHLCRRFQHRHRHRHRHRHQPLP</u>
31SP n.° 4	PSAKAAKGKAPAKPPRRLPPRKPNSQSLSAQTAAAPVQASVSTPAPASAPTPTPAVA
A21	PSAKAAKGKAPAKPPR-----
33SP n.° 9	-----
pALK4106	-----
31UV n.° 22	<u>QLQVQCLLLNRLALQNRLHHPHLLHSGGIRWLFVPRQVLGRHGNRFLORPGSNTNPLS</u>
31SP n.° 4	STPGPMSPPPKPPRPSEPPSSSSTPIVA-----
A21	-----
33SP n.° 9	-----
pALK4106	-----
31UV n.° 22	<u>VPSISRREN</u>
31SP n.° 4	-----
A21	-----
33SP n.° 9	-----
pALK4106	-----

Fig. 7

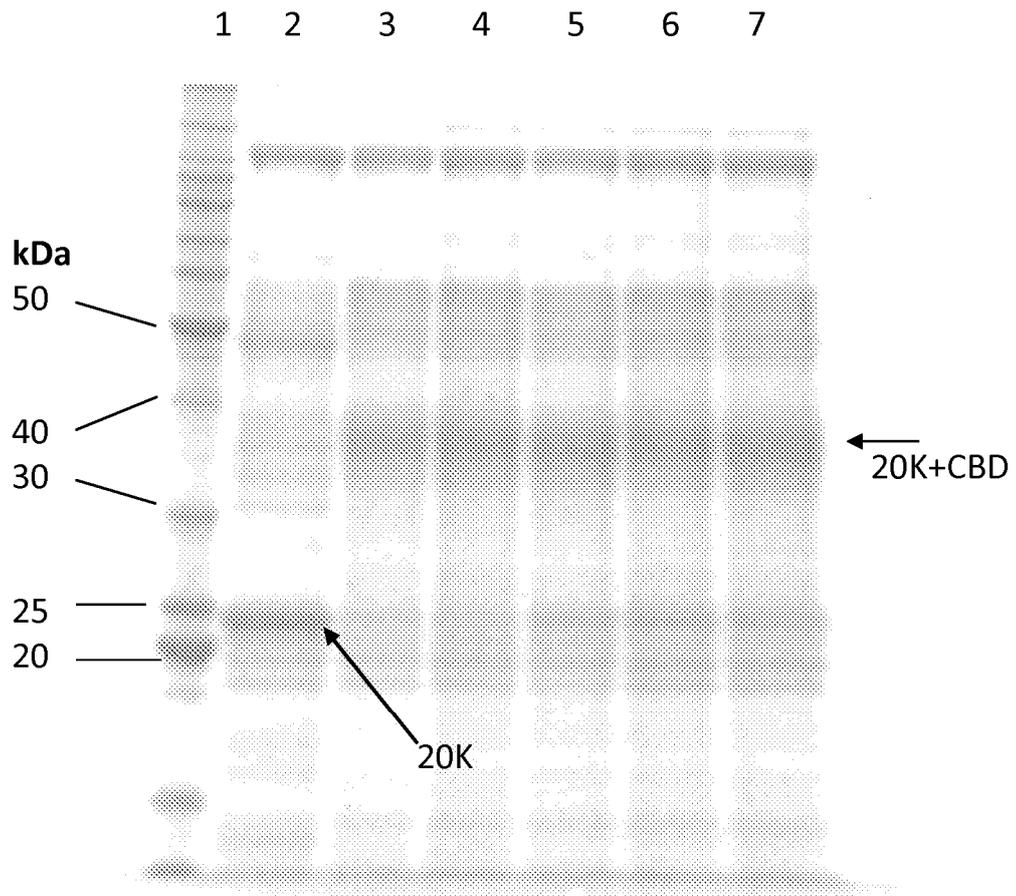


Fig. 8

Fig. 9

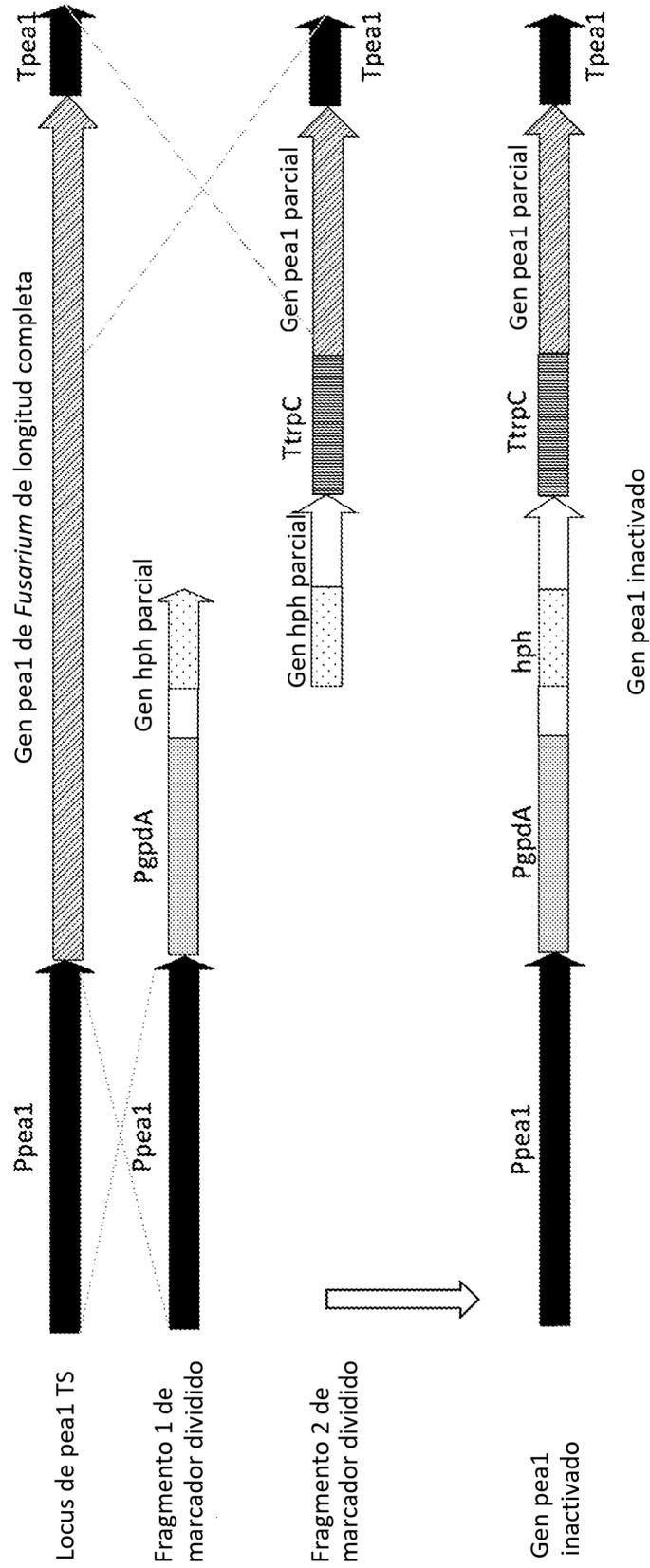


Fig. 10 **Actividad proteasa**

