

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 539 955**

51 Int. Cl.:

**C12N 15/56** (2006.01)

**C12N 9/42** (2006.01)

**C12N 1/15** (2006.01)

**C12N 1/19** (2006.01)

**C12N 15/63** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **07.12.2004** **E 10180094 (4)**

97 Fecha y número de publicación de la concesión europea: **08.04.2015** **EP 2330199**

54 Título: **Celulasa tolerante a los tensioactivos y procedimiento de modificación de la misma**

30 Prioridad:

**08.12.2003 JP 2003409692**

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

**07.07.2015**

73 Titular/es:

**MEIJI SEIKA PHARMA CO., LTD. (100.0%)**  
**4-16, Kyobashi 2-chome, chuo-ku**  
**Tokyo 104-8002, JP**

72 Inventor/es:

**WATANABE, MANABU;**  
**YANAI, KOJI y**  
**TSUYUKI, YUMIKO**

74 Agente/Representante:

**CARPINTERO LÓPEZ, Mario**

ES 2 539 955 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

Celulasa tolerante a los tensioactivos y procedimiento de modificación de la misma

## Campo técnico

- 5 La presente invención se refiere a un procedimiento para modificar el extremo N-terminal de una proteína que tiene actividad endoglucanasa (en especial, una celulasa perteneciente a la familia 45 y que tiene una actividad endoglucanasa) a ácido piroglutámico, para convertir la proteína en celulasa que tiene actividad endoglucanasa, cuya reducción en presencia de un tensioactivo es pequeña y está relacionada con la celulasa.

## Antecedentes de la técnica

- 10 Se dice que la biomasa de celulosa es el recurso más abundante entre los recursos naturales y por lo tanto en diversos campos se desea una aplicación eficiente de sistemas de celulasas que descompongan la biomasa de celulosa. En este proceso de desarrollo, se purificaron y caracterizaron diversas celulasas y además, se han clonado diversos genes de celulasa y se han clasificado en familias analizando la homología de secuencias (véase referencia no de patente 1).

- 15 En otro aspecto, las celulasas se utilizan, basándose en sus propiedades, en diversos campos industriales, en particular en el campo de procesamiento de tejidos. Por ejemplo, se lleva a cabo un tratamiento con celulasa para mejorar el tacto y/o el aspecto de tejidos que contienen celulosa, o para un "biolavado", lo cual confiere un aspecto de "lavado a la piedra" a un tejido que contiene celulosa coloreado, proporcionando de este modo al tejido variaciones de color localizadas. Además, en el procedimiento para fabricar liocell, la celulasa se usa para retirar la pelusa generada en el procedimiento de la superficie del tejido. Al respecto, el liocell es un tejido de celulosa regenerado derivado de la pulpa de madera, que recientemente ha llamado la atención por sus propiedades (tales como la alta resistencia o la absorción de agua) y como un procedimiento de producción que provoca menos
- 20 polución ambiental.

- Hasta ahora, se ha considerado que la celulasa descompone la celulosa por el efecto colaborativo de varias enzimas, es decir, el efecto sinérgico. El grupo de celulasas que consiste en varias enzimas contiene enzimas que
- 25 tienen propiedades inadecuadas para el campo del procesamiento de tejidos (tales como una enzima que disminuye la resistencia de la fibra). Por consiguiente, se ha llevado a cabo un intento para separar componentes enzimáticos adecuados para el procesamiento de tejidos del grupo de las celulasas y para producir los componentes enzimáticos, utilizando técnicas de separación de proteínas y/o técnicas de ingeniería genética. En particular, se han sometido a un estudio serio las celulasas derivadas de microorganismos que pertenecen a los hongos filamentosos
- 30 tales como el género *Trichoderma* o el género *Humicola*. Por ejemplo, como componentes de celulasa, se aislaron CBH I, EG V (véase referencia de patente 1), NCE2, NCE4 y NCE5 en el género *Humicola* y CBH I, CBH II, EG II y EG III en el género *Trichoderma* y de este modo, se pueden producir preparaciones de celulasa que contienen como los componentes principales uno o más componentes de celulasa específicos apropiados para cada propósito por la preparación de enzimas sobreexpresadas o enzimas monocomponentes usando técnicas de ingeniería genética.
- 35 Además, se ha aclarado que las celulasas que pertenecen a la familia 45, tales como NCE4 (véase referencia de patente 2), NCE5 (véase referencia de patente 3), RCE1 (véase referencia de patente 4) o STCE1 (véase Solicitud Internacional N° PCT/JP2004/15733), son muy útiles en los campos anteriores.

- En otro aspecto más, cuando se usan las celulasas como un detergente para la ropa, no solamente se desea una mejora cuantitativa de los componentes de celulasa usados sino también una cualitativa. Más particularmente, un detergente para la ropa contiene diversos tensioactivos y una solución obtenida por la solubilización del detergente para la ropa en agua es alcalina (pH 10 a pH 11). Por consiguiente, es necesario que las celulasas contenidas en un detergente para la ropa sean resistentes a diversos tensioactivos, en condiciones alcalinas. Como un informe en el que una reducción de la actividad en presencia de un tensioactivo se suprime, Otzen, D. E. y col. informaron que cuando se introdujo una mutación en la secuencia interna de aminoácidos de Cel45 derivada de *Humicola insolens*,
- 40 la actividad de la misma a pH 7 en presencia de benceno sulfonato de alquilo lineal (LAS) era aproximadamente 3,3 veces superior a la del tipo natural (véase referencia no relacionada con patentes 2). Sin embargo, se encontró que la supresión de la reducción de la actividad en presencia del tensioactivo, proporcionada por la mutación, está limitada a Cel45 o a sus proteínas homólogas y que no es aplicable a las endoglucanasas pertenecientes a la familia 45 que tienen baja homología con Cel45.

- 50 (referencia de patente 1) Publicación Internacional WO91/17243

(referencia de patente 2) Publicación Internacional WO98/0366

(referencia de patente 3) Publicación Internacional WO01/90375

(referencia de patente 4) Publicación Internacional WO00/24879 (referencia no de patente 1) Henrissat B., Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* 316: 695-696 (1996)

- 55 (referencia no de patente 2) Daniel E. Otzen, Lars Christiansen, Martin Schulein. A comparative study of the

unfolding of the endoglucanase Cel45 from *Humicola insolens* in denaturant and surfactant. *Protein Sci.* 8: 1878-1887 (1999)

El documento WO 94/07998 se refiere a una variante de celulasa derivada de una celulasa parental, por ejemplo perteneciente a la familia 45 tal como una endoglucanasa de 43 kD de *Humicola insolens*. Varias propiedades de la variante de celulasa, tales como por ejemplo actividad alcalina, compatibilidad con ingredientes de composición detergente, retirada de suelo particulado, aclarado de color, eliminación de pelusa, depilación, reducción de aspereza y sensibilidad a tensioactivos aniónicos y sistemas blanqueadores de peroxidasa se mejoran eliminando o sustituyendo uno o más aminoácidos del dominio catalíticamente activo (CAD), el dominio de unión a celulasa (CBD), o la región de unión entre ambos dominios y/o por adición de uno o más aminoácidos a la región de unión y/o por adición de otro CBD en el extremo opuesto del CAD.

## **Divulgación de la invención**

### **Problemas a resolverse por la invención**

Un objeto de la presente invención es para proporcionar un procedimiento para convertir una proteína que tiene una actividad endoglucanasa (en particular, una proteína perteneciente a la familia 45 y que tiene una actividad endoglucanasa) en una proteína que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña; un vector que se utiliza en el procedimiento; una proteína que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña; y un polinucleótido que codifique la misma. Otro objeto de la presente invención es proporcionar, utilizando lo anterior, un microorganismo que produce de manera eficiente una proteína útil como una enzima para el lavado de ropa.

### **Medios para solucionar el problema**

Los presentes inventores realizaron estudios intensivos y como resultado, encontraron que las proteínas en las que se añadió ácido piroglutámico (en adelante denominado algunas veces pQ en el presente documento) o un péptido que contenía pQ al extremo N-terminal de cada proteína perteneciente a la familia 45 y que tenían una actividad endoglucanasa, es decir, celulasas con adición en el extremo N-terminal, tenían una actividad endoglucanasa que no presentaba ninguna reducción significativa en la actividad en presencia de un tensioactivo, en comparación con las celulasas de tipo natural.

La característica de que se mantenga una actividad endoglucanasa alta en presencia de un agente tensioactivo aniónico es particularmente útil en el caso de enzimas para el lavado de ropa, pero no se conoce una celulasa de tal tipo. Además, no se ha informado que las funciones inherentes a una enzima puedan mantenerse en presencia de un tensioactivo por la adición a la misma de ácido piroglutámico o de un péptido que lo contenga.

En otro aspecto de la presente invención, con respecto a todas las celulasas en las que el aminoácido N-terminal no está protegido y se desea el mantenimiento de la actividad en presencia de un tensioactivo, es posible, añadiendo ácido piroglutámico o un péptido que lo contiene a las celulasas, para suprimir la reducción de la actividad en presencia de un tensioactivo. La celulasa a la que se añade ácido piroglutámico o un péptido que lo contiene no está particularmente limitada, pero son preferibles las celulasas pertenecientes a la familia 45.

De acuerdo con ello la presente invención incluye lo siguiente:

(1) Una combinación de un polinucleótido y una célula huésped, en la que el polinucleótido está seleccionado del grupo que consiste en:

(a) un polinucleótido que comprende la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39;

(b) un polinucleótido que comprende una secuencia de nucleótidos en la que 1 a 90 uno o varios nucleótidos se han eliminado, sustituido, insertado o añadido en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y que codifica una proteína en la que el extremo N-terminal es ácido glutámico o glutamina y que tiene una actividad endoglucanasa; y

(c) un polinucleótido que hibrida en condiciones rigurosas con un polinucleótido que consiste en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y que codifica una proteína en la que el extremo N-terminal es ácido glutámico o glutamina y que tiene una actividad endoglucanasa;

y en la que la célula huésped es un huésped en el que la formación de piroglutamato del aminoácido N-terminal se lleva a cabo

(2) una combinación de la célula huésped de (1) y un vector de expresión que comprende el polinucleótido de (1);

(3) una célula huésped transformada con el vector de expresión de (2);

(4) la célula huésped de (3), en la que la célula huésped es una levadura o un hongo filamentoso;

(5) la célula huésped de (4), el hongo filamentoso es un microorganismo que pertenece al género *Humicola* o *Trichoderma*;

(6) la célula huésped de (5), el hongo filamentoso es *Humicola insolens* o *Trichoderma viride*;

(7) un procedimiento para producir una proteína seleccionada del grupo que consiste en:

5 (a) una proteína que comprende la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40 en la que el extremo N-terminal es ácido piroglutámico;

(b) una proteína modificada que comprende una secuencia de aminoácidos y en la que 1 a 30 aminoácidos se han eliminado, sustituido, insertado o añadido en la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa, cuya reducción en presencia de un tensioactivo es pequeña, y

10 (c) una proteína homóloga que comprende una secuencia de aminoácidos que tiene al menos el 8,5 % de identidad con una proteína que comprende la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40, en la que el extremo N-terminal es ácido piroglutámico y que tiene una actividad endoglucanasa,

que comprende:

15 cultivar la célula huésped de uno cualquiera de los puntos (3) a (6) y recuperar la proteína de la célula huésped o del cultivo huésped obtenida por la cultivación.

#### Efectos de la invención

De acuerdo con la presente invención, es posible producir de manera eficaz una celulasa nueva que es útil como una enzima para el lavado de ropa y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña.

#### 20 **Mejor modo de llevar a cabo la invención**

Proteína perteneciente a la familia 45 y que tiene actividad endoglucanasa (denominada en adelante algunas veces como "celulasa perteneciente a la familia 45").

#### Familia 45

25 El término "familia 45" como se utiliza en el presente documento significa una proteína clasificada en la familia 45 de acuerdo con el análisis de grupos hidrófobos de enzimas activadoras de hidratos de carbono por B. Henrissat y A. Bairoch [Henrissat B., Bairoch A. Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316: 695-696 (1996)].

#### Proteínas que tienen actividad endoglucanasa

30 La expresión "proteína que tiene una actividad endoglucanasa" como se utiliza en el presente documento significa una enzima que presenta una actividad endoglucanasa, es decir, endo-1,4-β-glucanasa (EC3.2.1.4), que hidroliza el enlace β-1,4-glucopiranosilo en β-1,4-glucano.

#### Tensioactivo

35 El término "tensioactivo", como se utiliza en el presente documento es un componente detergente contenido en un detergente para la ropa y que se clasifica ampliamente en tensioactivos aniónicos, catiónicos y no iónicos. Los tensioactivos aniónicos se utilizan comúnmente. Como tensioactivos aniónicos utilizados de preferencia en la presente invención, se pueden citar, por ejemplo, benceno sulfonato de alquilo lineal (denominado en adelante en el presente documento LAS).

#### Actividad endoglucanasa

40 La expresión "actividad endoglucanasa (denominada en adelante en el presente documento 'EGU') como se utiliza en el presente documento se define como una actividad enzimática obtenida por la medición de la disminución de la viscosidad de una solución de carboximetilcelulosa según el siguiente procedimiento.

45 Como una solución de sustrato, se disolvió carboximetilcelulosa (Hercules) en un tampón Tris-HCl 0,1 mol/l (pH 10,0) (concentración final = 3,5 %). A la solución de sustrato (5 ml) previamente calentada a 40 °C durante 10 minutos se le añadieron 0,15 ml de una solución de la enzima y a continuación se mezcló todo bien para llevar a cabo la reacción a 40 °C durante 30 minutos. La viscosidad de la mezcla de reacción se midió por un viscosímetro de tipo R (RE100; TOKI Sangyo Co., LTD) a 40 °C. Se define "1 unidad" de la actividad enzimática como una cantidad de enzima que reduce la viscosidad inicial al 1/2, en cada condición de reacción. Como un tensioactivo aniónico, se utilizó benceno sulfonato de alquilo lineal (Wako Pure Chemical Industries, Co., Ltd.) y se añadió a la solución de carboximetilcelulosa hasta una concentración final de 800 ppm.

Supresión de la reducción de la actividad endoglucanasa en presencia de tensioactivo

“Que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña”, como se utiliza en el presente documento significa que, cuando la proteína en la que está modificado el extremo N-terminal (proteína del tipo modificación N-terminal) de acuerdo con la presente invención se compara con la proteína original antes de realizar la modificación (denominada a continuación en el presente documento simplemente proteína original) con respecto a la actividad endoglucanasa en presencia del tensioactivo, la actividad endoglucanasa de la proteína del tipo modificación N-terminal es mayor que la de la proteína original.

Proteína original

La proteína original que se puede aplicar al procedimiento de la presente invención no está particularmente limitada, con la condición de que se trate de una proteína que tenga una actividad endoglucanasa en la que el extremo N-terminal es un aminoácido diferente de ácido piroglutámico. Como la proteína original, se pueden citar, por ejemplo, una celulasa (por ejemplo, endoglucanasa, celobiohidrolasa o  $\beta$ -glucosidasa) y resulta de preferencia una celulasa perteneciente a la familia 45. En este sentido, siempre que la proteína original tenga al menos una actividad endoglucanasa, puede ser una proteína que tenga solo la actividad endoglucanasa, o una proteína que tenga otra u otras actividades enzimáticas además de la actividad endoglucanasa. Además, la proteína original puede ser una proteína de origen natural o una proteína modificada genéticamente.

Fuente original de la celulasa perteneciente a la familia 45

La celulasa perteneciente a la familia 45 puede generarse por técnicas de ingeniería genética de uso común, tales como técnicas de ADN recombinante o técnicas de síntesis de polipéptidos, o se pueden obtener a partir de una cepa natural aislada. Además las celulasas incluyen variantes de celulasas de tipo natural pertenecientes a la familia 45.

La celulasa perteneciente a la familia 45 se puede obtener a partir de un microorganismo tal como los hongos filamentosos o zigomicetos. Como hongos filamentosos, se pueden citar, por ejemplo, microorganismos pertenecientes al género *Humicola* (tal como *Humicola insolens*), al género *Trichoderma* (tal como *Trichoderma viride*), al género *Staphylotrichum* (tal como *Staphylotrichum coccosporum*) o del género *Myriococcum* (tal como *Myriococcum thermophilum*). Más concretamente, las celulasas derivadas de género *Humicola* incluyen, por ejemplo, CBH I, EG V, NCE2, NCE4 y NCE5; las celulasas derivadas del género *Trichoderma* incluyen, por ejemplo, CBH I, CBH II, EG II y EG III; las celulasas derivadas del género *Staphylotrichum* incluyen, por ejemplo, STCE1 y STCE3; y las celulasas derivadas del género *Myriococcum* incluyen, por ejemplo, MTE1. Como zigomicetos, se pueden citar, por ejemplo, microorganismos pertenecientes al género *Rhizopus* (tal como *Rhizopus oryzae*), al género *Mucor* (tal como *Mucor circinelloides*) o al género *Phycomyces* (tal como *Phycomyces nitens*). Más concretamente, se pueden citar, por ejemplo, RCE I, RCE II o RCE III derivadas de *Rhizopus oryzae*, MCE I o MCE II derivadas de *Mucor circinelloides*; o PCE I derivada de *Phycomyces nitens* (documento WO 00/24879).

Ácido piroglutámico o péptido que contiene ácido piroglutámico

El término “ácido piroglutámico” como se utiliza en el presente documento significa ácido piroglutámico generado por ciclación de la glutamina o el ácido glutámico del extremo N-terminal de una proteína madura. El ácido piroglutámico tiene la característica de que el grupo amino N-terminal no está expuesto. La formación de piroglutamato se puede realizar in vivo o in vitro. In vivo, un polinucleótido que codifica una proteína modificada, que está genéticamente diseñada para que el extremo N-terminal de una proteína madura sea glutamina o ácido glutámico, puede expresarse en una célula huésped para obtener una proteína de ciclación de piroglutamato. In vitro, una proteína que tiene el extremo N-terminal de glutamina o ácido glutámico pueden ser tratada con una solución ácida tal como una solución de ácido fórmico para obtener una proteína que tiene el extremo N-terminal de ácido piroglutámico.

El término “péptido” como se utiliza en el presente documento, significa un compuesto que consiste en uno o en varios aminoácidos en el que los aminoácidos están polimerizados por enlaces peptídicos. Por lo tanto, la expresión “péptido que contiene ácido piroglutámico” como se utiliza en el presente documento significa un péptido en el que el aminoácido N-terminal es ácido piroglutámico. El péptido que contiene ácido piroglutámico consiste en dos o más (varios) aminoácidos reticulados, por ejemplo, de 2 a 40 aminoácidos, de preferencia de 2 a 30 aminoácidos, de más preferencia de 2 a 20 aminoácidos, de más preferencia aún de 2 a 10, de más preferencia aún de 2 a 5, de la mayor preferencia de 2 a 4 aminoácidos. Los aminoácidos no están particularmente limitados, con la condición de que puedan ser utilizados por los expertos en la técnica para el propósito indicado.

En el procedimiento de la presente invención, un procedimiento para modificar una proteína original a una proteína que tiene el extremo N-terminal de ácido piroglutámico no está particularmente limitado, con la condición de que la modificación de la proteína pueda llevarse a cabo. Como el procedimiento, se pueden citar, por ejemplo, técnicas de ingeniería genética o técnicas químicas.

De acuerdo con las técnicas de ingeniería genética, una proteína original puede ser modificada, por ejemplo, llevando a cabo una adición y/o sustitución de un aminoácido adecuado o secuencia de aminoácidos adecuada por ingeniería genética.

En una forma de realización que utiliza la adición por ingeniería genética, la modificación proteica puede llevarse a cabo añadiendo genéticamente ácido piroglutámico o un péptido que tiene el extremo N-terminal de ácido piroglutámico, al extremo N-terminal de una proteína original (es decir, una proteína que tiene una actividad endoglucanasa en la que el extremo N-terminal no es ácido piroglutámico).

- 5 Más particularmente, la forma de realización que utiliza la adición por ingeniería genética puede comprender, por ejemplo, las etapas de:

(1) añadir un polinucleótido que codifica un aminoácido que puede convertirse en ácido piroglutámico (tal como ácido glutámico o glutamina) o un polinucleótido que codifica un péptido que tiene el extremo N-terminal de un aminoácido que puede convertirse en ácido piroglutámico, al extremo 5' terminal de un polinucleótido que codifica una proteína original y

(2) expresar el polinucleótido resultante en un huésped en el que puede llevarse a cabo la formación de piroglutamato del aminoácido N-terminal.

En una forma de realización que utiliza la sustitución por ingeniería genética, la modificación de proteínas puede llevarse a cabo sustituyendo genéticamente el aminoácido N-terminal o la región N-terminal de una proteína original por ácido piroglutámico o un péptido que tiene el extremo N-terminal de ácido piroglutámico.

Más particularmente, la forma de realización utilizando la sustitución por ingeniería genética puede comprender, por ejemplo, las etapas de:

(1) sustituir el extremo 5' terminal o una región que contiene el mismo de un polinucleótido que codifica una proteína original por un polinucleótido que codifica un aminoácido que puede convertirse en ácido piroglutámico (tal como ácido glutámico o glutamina) o un polinucleótido que codifica un péptido que tiene el extremo N-terminal de un aminoácido que puede convertirse en ácido piroglutámico y

(2) expresar el polinucleótido resultante en un huésped en el que puede llevarse a cabo la formación del piroglutamato del aminoácido N-terminal.

Como formas de realización utilizando las técnicas químicas, se pueden citar, por ejemplo,

(a) una forma de realización en la que el ácido piroglutámico (o un péptido que tiene el extremo N-terminal de ácido piroglutámico) se añade directamente al extremo N-terminal de una proteína original;

(b) una forma de realización en la que un aminoácido que puede convertirse en ácido piroglutámico (o un péptido que tiene el extremo N-terminal de un aminoácido que puede convertirse en ácido piroglutámico) se añade químicamente al extremo N-terminal de una proteína original y a continuación se lleva a cabo químicamente la formación del piroglutamato del aminoácido N-terminal; o

(c) una forma de realización en la que el aminoácido N-terminal de una proteína original que tiene el extremo N-terminal de un aminoácido que puede convertirse en ácido piroglutámico se convierte químicamente en ácido piroglutámico.

Procedimiento para añadir ácido piroglutámico o un péptido que contiene ácido piroglutámico al lado N-terminal de la celulosa perteneciente a la familia 45

El procedimiento de modificación se ilustrará además por una forma de realización que utiliza celulosa perteneciente a la familia 45. Un procedimiento para añadir ácido piroglutámico o un péptido que contiene ácido piroglutámico al lado N-terminal de la celulosa perteneciente a la familia 45 puede llevarse a cabo por técnicas de ingeniería genética. En una producción de celulasas de uso común, se puede unir una región codificadora de un polinucleótido que codifica la celulasa deseada de manera operativa entre un promotor y un terminador funcionable en un huésped tal como un hongo filamentoso y a continuación puede introducirse la cassette de expresión resultante en el huésped. Además, se puede añadir un polinucleótido que codifica una secuencia señal de secreción funcionable en la célula huésped en la cassette. Cuando se introduce la cassette en la célula huésped, la celulasa deseada se secreta en un medio y se puede recoger fácilmente. En este caso, puede añadirse un aminoácido deseado al extremo N-terminal de la celulasa, por la adición de un polinucleótido que codifica el aminoácido inmediatamente secuencia abajo de la secuencia señal de secreción. Además, la modificación del grupo amino del aminoácido N-terminal se puede llevar a cabo utilizando una secuencia señal de secreción en un huésped. Por ejemplo, en cbh1 o cbh2 derivadas de *Trichoderma viride*, o NCE2 o NCE5 derivadas de *Hemicola insolens*, el extremo N-terminal es piroglutamato formado, la modificación puede llevarse a cabo por el uso de estas secuencias señal para la secreción y la expresión en *Trichoderma viride* o *Hemicola insolens*. De acuerdo con la forma de realización de preferencia, la celulasa modificada perteneciente a la familia 45 preparada como se describió anteriormente presenta una característica ventajosa, es decir, tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña.

Según otra forma de realización, todo puede ser sintetizado químicamente, dentro del ámbito de los conocimientos

técnicos comunes de los expertos en la técnica. En este caso, la síntesis puede llevarse a cabo utilizando parte de una proteína de origen natural.

#### Proteína de la presente invención

La proteína de la presente invención se prepara por la modificación del extremo N-terminal de una proteína original que tiene una actividad endoglucanasa a ácido piroglutámico [por ejemplo, preparada por la obtención de celulasa perteneciente a la familia 45 y la adición de ácido piroglutámico (pQ) o un péptido que contiene ácido piroglutámico al lado del aminoácido N-terminal de la proteína madura de la misma] y tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña. Además, la presente invención incluye una proteína que se puede preparar por el procedimiento descrito anteriormente y tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña.

Más concretamente, la proteína de la presente invención incluye una proteína seleccionada del grupo que consiste en las siguientes proteínas:

(a) una proteína que comprende la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40;

(b) una proteína modificada que comprende una secuencia de aminoácidos en la que uno o varios aminoácidos se han eliminado, sustituido, insertado o añadido en la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña; y

(c) una proteína homóloga que comprende una secuencia de aminoácidos que tiene al menos el 85 % de homología con una proteína que comprende la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña.

La secuencia de aminoácidos de la SEQ ID NO: 2 es la secuencia de aminoácidos de NCE4 modificada en el extremo N-terminal (véase Ejemplo A2) en la que se añade un péptido que consiste en cinco aminoácidos (N-terminal: ácido piroglutámico) al extremo N-terminal de la endoglucanasa NCE4 derivada de *Humicola insolens* MN200-1.

La secuencia de aminoácidos de la SEQ ID NO: 4 es la secuencia de aminoácidos de STCE1 modificada en el extremo N-terminal (véase Ejemplo B2) en la que se sustituye el aminoácido N-terminal (Ala) de la endoglucanasa STCE1 derivada de *Staphylotrichum coccosporum* IFO 31817 por un péptido que consiste en cuatro aminoácidos (N-terminal: ácido piroglutámico).

La secuencia de aminoácidos de la SEQ ID NO: 38 es la secuencia de aminoácidos de la STCE1 modificada en el extremo N-terminal (véase Ejemplo B3) en la que se añade ácido piroglutámico al extremo N-terminal de la endoglucanasa STCE1 derivada de *Staphylotrichum coccosporum* IFO 31817.

La secuencia de aminoácidos de la SEQ ID NO: 40 es la secuencia de aminoácidos de la STCE1 modificada en el extremo N-terminal (véase Ejemplo B4) en la que se añade un péptido que consiste en cuatro aminoácidos (N-terminal: ácido piroglutámico) al extremo N-terminal de la endoglucanasa STCE1 derivada de *Staphylotrichum coccosporum* IFO 31817.

El término "proteína modificada" como se utiliza en el presente documento significa una proteína que comprende una secuencia de aminoácidos en la que uno o una pluralidad de aminoácidos (por ejemplo, uno o varios aminoácidos) se han eliminado, sustituido, insertado o añadido en la secuencia de aminoácidos de SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña. El número de aminoácidos a modificar, tal como "eliminar, sustituir, insertar o añadir", es de preferencia de 1 a 30, de más preferencia de 1 a 10, de la mayor preferencia de 1 a 6.

Además, la proteína modificada incluye una proteína que comprende una secuencia de aminoácidos en la que uno o varios aminoácidos están sustituidos de manera conservadora en la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña. La expresión "sustitución conservadora" como se utiliza en el presente documento significa que uno o varios residuos de aminoácidos contenidos en una proteína están sustituidos con diferentes aminoácidos que tienen propiedades químicas similares de modo que las actividades de la proteína no se modifican sustancialmente. Como sustitución conservadora, se pueden citar, por ejemplo, una sustitución de un residuo de aminoácido hidrófobo por otro residuo de aminoácido hidrófobo, o la sustitución de un residuo de aminoácido polar por otro residuo de aminoácido polar que tiene la misma carga. Los aminoácidos que tienen propiedades químicas similares y que pueden estar sustituidos de manera conservadora entre sí son conocidos por los expertos en la técnica. Más en particular, como aminoácidos no polares (hidrófobos), se pueden citar, por ejemplo, alanina, valina, isoleucina, leucina, prolina, triptófano, fenilalanina o metionina. Como aminoácidos polares (neutros), se pueden citar, por ejemplo, glicina, serina, treonina, tirosina, glutamina, asparagina o cisteína. Como aminoácidos básicos que tienen una carga positiva, se pueden citar, por ejemplo, arginina, histidina o lisina. Como aminoácidos ácidos que tienen una carga negativa, se pueden citar, por ejemplo, ácido aspártico o ácido glutámico.

La expresión "proteína homóloga", como se utiliza en el presente documento significa una proteína que comprende una secuencia de aminoácidos que tiene al menos el 85 % (de preferencia el 90 % o más, de la mayor preferencia el 95 % o más) de homología (identidad de secuencia) con una proteína que comprende la secuencia de aminoácidos de SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña. La homología como se utiliza en el presente documento se muestra como el valor (identidad) calculado por FASTA3 [Science, 227, 1435-1441 (1985); Proc. Natl. Acad. Sci. EE.UU., 85, 2444-2498 (1988); <http://www.ddbj.nig.ac.jp/E-mail/homology-j.html>], un programa de búsqueda de homología conocido, de conformidad con los parámetros por defecto.

Como se describió anteriormente, en la "proteína que consiste en la secuencia de aminoácidos de la SEQ ID NO: 2" de la presente invención, la porción salvo el péptido N-terminal que consiste en cinco aminoácidos deriva de la endoglucanasa NCE4. Además, en la "proteína que consiste en la secuencia de aminoácidos de la SEQ ID NO: 4, 38 o 40" de la presente invención, la porción salvo el aminoácido o péptido N-terminal deriva de la endoglucanasa STCE1. Las endoglucanasas NCE4 y STCE1 pertenecen a la familia 45. Como endoglucanasa conocida perteneciente a la familia 45, se pueden citar, por ejemplo, NCE5 derivada del género *Humicola* (documento WO01/90375).

En las figuras 1 y 2, se muestra una alineación de las secuencias de aminoácidos de la endoglucanasa STCE1 [péptido señal (SEQ ID NO: 43) y proteína madura (SEQ ID NO: 44)], la endoglucanasa NCE4 [péptido señal (SEQ ID NO: 45) y proteína madura (SEQ ID NO: 46)] y la endoglucanasa NCE5 [péptido señal (SEQ ID NO: 47) y proteína madura (SEQ ID NO: 48)].

La figura 1 muestra la alineación de la mitad N-terminal y la Figura 2 muestra la de la mitad C terminal. El símbolo "\*" en las figuras 1 y 2 indica un aminoácido común al de STCE1.

Como se muestra en las figuras 1 y 2, las endoglucanasas pertenecientes a la familia 45 contienen el dominio catalítico (1° a 207°) como un dominio común y a veces contienen la región conectora (208° a 258°) y/o el dominio de unión a la celulosa (CDB) (259° a 295°). En este sentido, los números entre paréntesis después de los dominios anteriores representan el número de aminoácidos en la secuencia de aminoácidos (SEQ ID NO: 44) de la endoglucanasa STCE1.

En cada región, hay muchos aminoácidos conservadores entre dos o más endoglucanasas en el dominio catalítico y el dominio de unión a la celulosa, pero no se observa ninguna región conservadora notable en la región del conector. Las regiones que contienen muchos aminoácidos conservadores (por ejemplo, el dominio catalítico o el dominio de unión a la celulosa, en particular, el dominio catalítico), o aminoácidos comunes contenidos en las regiones son consideradas regiones o aminoácidos importantes para la actividad enzimática de las endoglucanasas (tales como STCE1 o NCE4). Por lo tanto, cuando se lleva a cabo una modificación de aminoácidos (por ejemplo, eliminación, sustitución, inserción y/o adición, en particular, sustitución conservadora) en una región o un aminoácido distinto de tales regiones o aminoácidos importantes, puede obtenerse con una alta posibilidad y sin excesiva experimentación una proteína modificada u homóloga que mantiene la actividad de la enzima.

Además, incluso cuando tiene lugar en las regiones que contienen muchos aminoácidos conservadores, una modificación de un aminoácido o aminoácidos no comunes entre dos o más endoglucanasas a aminoácido(s) diferente(s) [de preferencia aminoácido(s) que son similares y que pueden estar sustituidos de manera conservadora] probablemente podrá mantener la actividad de la enzima. Por lo tanto, por una de tales modificaciones, puede obtenerse una proteína modificada u homóloga que mantiene la actividad de la enzima con una alta probabilidad, sin experimentación excesiva.

En relación con lo anterior, incluso si aminoácido(s) común/comunes en la región que contiene muchos aminoácidos conservadores se modifica(n) a aminoácido(s) diferente(s), la actividad de la enzima a veces se mantiene. En particular, en una modificación a aminoácido(s) que es/son similar(es) y que puede(n) estar sustituido(s) de manera conservadora, la posibilidad se incrementa. La proteína modificada u homóloga de la presente invención incluye una proteína en la que uno o más aminoácidos contenidos en cualquier región, tal como el dominio catalítico, la región conectora o el dominio de unión a la celulosa, se modifican, con la condición de que presente una actividad endoglucanasa.

#### Polinucleótido que codifica la proteína de la presente invención

De acuerdo con la presente invención, se puede proporcionar un polinucleótido que codifica una proteína que comprende la secuencia de aminoácidos de las SEQ ID NO 2, 4, 38 o 40, o una de sus proteínas modificadas u homólogas (en adelante denominadas colectivamente como la proteína de la presente invención). Cuando se da la secuencia de aminoácidos de una proteína, se puede seleccionar fácilmente una secuencia de nucleótidos que codifica la secuencia de aminoácidos y por lo tanto pueden seleccionarse diversas secuencias de nucleótidos que codifican la proteína de la presente invención. El término "polinucleótido" como se utiliza en el presente documento, incluye ADN y ARN y resulta de preferencia el ADN.

El polinucleótido de la presente invención incluye un polinucleótido seleccionado del grupo que consiste en los siguientes polinucleótidos:

(a) un polinucleótido que comprende la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39;

(b) un polinucleótido que comprende una secuencia de nucleótidos en la que uno o varios nucleótidos se han eliminado, sustituido, insertado o añadido en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y que codifica una proteína en la que el extremo N-terminal es ácido glutámico o glutamina y que tiene una actividad endoglucanasa cuya reducción en presencia de un tensioactivo es pequeña; y

(c) un polinucleótido que hibrida en condiciones rigurosas con un polinucleótido que consiste en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y que codifica una proteína en la que el extremo N-terminal es ácido glutámico o glutamina y que tiene una actividad endoglucanasa y en la que la célula huésped es un huésped en el que se lleva a cabo formación de piroglutamato del aminoácido N-terminal.

En la secuencia de nucleótidos que se describe en el punto anterior (b), el número de nucleótidos a eliminar, sustituir, insertar o añadir, es por ejemplo, de 1 a 90, de preferencia de 1 a 30, de más preferencia de 1 a 18, de mayor preferencia de 1 a 9.

La expresión "bajo condiciones rigurosas", como se utiliza en el presente documento significa las siguientes condiciones. De acuerdo con un protocolo adjunto de un sistema de detección y marcación directa de ADN/ARN ECL (Amersham), después de que un polinucleótido a probar se prehibrida a 42 °C durante una hora, se añade un polinucleótido que consiste en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y la hibridación se lleva a cabo a 42 °C durante 14 a 16 horas. Después de la hibridación, se repite dos veces un tratamiento de lavado con 0,5 X SSC (1 X SSC; citrato de sodio 15 mmol/l, cloruro de sodio 150 mmol/l) que contiene SDS al 0,4 % y 6 mol/l de urea a 42 °C durante 20 minutos y se lleva a cabo dos veces un tratamiento de lavado con 5 X SSC a temperatura ambiente durante 5 minutos.

El polinucleótido de la presente invención incluye un polinucleótido de origen natural. Además, se puede sintetizar en su totalidad. Además, la síntesis puede llevarse a cabo utilizando parte del polinucleótido de origen natural. Por lo general el polinucleótido de la presente invención puede obtenerse por selección de una biblioteca genómica derivada de un microorganismo deseado según un procedimiento habitual de uso común en la ingeniería genética, por ejemplo, utilizando una sonda de ADN adecuada diseñada en base a la información de una secuencia de aminoácidos parcial.

#### Producción de la proteína de la presente invención

La proteína de la presente invención se pueden producir en células huésped por la transformación de la célula huésped con una molécula de polinucleótido (en particular como una forma de vector de expresión) que comprende un fragmento de polinucleótido que codifica la proteína de manera que la molécula de polinucleótido pueda replicarse y el gen pueda expresarse en la célula huésped.

De acuerdo con la presente invención, se proporciona un vector de expresión que comprende un fragmento de polinucleótido que codifica la proteína de la presente invención de tal forma que el fragmento de polinucleótido pueda ser replicado y la proteína se pueda expresar en un microorganismo huésped.

El vector de expresión se puede construir en base a un vector auto-replicante (tal como un plásmido), que existe como un elemento extracromosómico y que puede replicarse de manera independiente de la replicación de los cromosomas. Como alternativa, el vector de expresión de la presente invención puede ser un vector que está integrado en el cromosoma del microorganismo huésped y que se replica junto con los cromosomas, cuando el huésped se transforma con el vector. La construcción del vector de la presente invención puede llevarse a cabo por procedimientos habituales o procedimientos comúnmente utilizados en la ingeniería genética.

Para expresar una proteína que tiene una actividad deseada por la transformación de un microorganismo huésped con el vector de expresión de la presente invención, es preferible que el vector de expresión contenga, por ejemplo, un polinucleótido capaz de controlar la expresión, o un marcador genético para seleccionar los transformantes, además del fragmento de polinucleótido de la presente invención.

Como el polinucleótido capaz de regular la expresión genética, se pueden utilizar en la presente invención diversas señales para la regulación de la transcripción o de la traducción, tales como un promotor, un terminador o un polinucleótido que codifica un péptido señal para la secreción. La unión de estos polinucleótidos y la inserción de los mismos a un vector pueden llevarse a cabo por un procedimiento habitual.

El promotor que se puede usar en la presente invención no está particularmente limitado, con la condición de que muestre actividad transcripcional en un microorganismo huésped. El promotor se puede obtener como un polinucleótido que regula la expresión de un gen que codifica una proteína igual o diferente de la que deriva del microorganismo huésped. Por ejemplo, se puede usar un promotor tal como un operón de lactosa o un operón de triptófano en *Escherichia coli*; un promotor de un gen de la alcohol deshidrogenasa, un gen de la fosfatasa ácida, un gen de utilización de la galactosa, o un gen de la gliceraldehído 3-fosfato deshidrogenasa en una levadura; y un promotor de un gen de la  $\alpha$ -amilasa, un gen de la glucoamilasa, un gen de la celobiohidrolasa o un gen de la gliceraldehído 3-fosfato deshidrogenasa en un hongo filamentoso.

El péptido señal no está particularmente limitado, con la condición de que contribuya a la secreción de proteínas en un microorganismo huésped. El péptido señal se puede obtener como un polinucleótido derivado de un gen que codifica una proteína igual o diferente de la que deriva del microorganismo huésped.

El marcador seleccionable puede seleccionarse adecuadamente de acuerdo con el procedimiento para seleccionar un transformante. Como el marcador seleccionable, en la presente invención se puede utilizar, por ejemplo, un gen de resistencia a fármacos o un gen que complementa una mutación auxótrofa. Cuando el huésped es una bacteria, se puede utilizar, por ejemplo, un gen de resistencia a la ampicilina, un gen de resistencia a la kanamicina o un gen de resistencia a la tetraciclina. Cuando el huésped es una levadura, se puede utilizar, por ejemplo, un gen de biosíntesis de triptófano (*trpI*, *trpC*), un gen de biosíntesis de uracilo (*ura3*), un gen de biosíntesis de leucina (*leu2*) o un gen de biosíntesis de histidina (*his3*). Cuando el huésped es un hongo filamentoso, se puede utilizar, por ejemplo, un gen de resistencia a la destomicina, un gen de utilización de nitratos (*niaD*), un gen de biosíntesis de arginina (*argB*), un gen de biosíntesis de uracilo (*pyr4*), un gen de resistencia a la higromicina, un gen de resistencia a bialafós, un gen de resistencia a la bleomicina o un gen de resistencia a la aureobasidina.

De acuerdo con la presente invención, se proporciona un microorganismo transformado con el vector de expresión. Un sistema huésped-vector que se puede utilizar en la presente invención no está particularmente limitado. Por ejemplo, se puede usar un sistema que utiliza *E. coli*, *Actinomycetes*, levaduras, u hongos filamentosos, o un sistema para la expresión de una proteína de fusión utilizando un microorganismo tal. Cuando se utiliza una levadura como huésped, se puede citar, por ejemplo, un microorganismo que pertenece al género *Saccharomyces*, *Hansenula* o *Pichia*, de preferencia *Saccharoztyces cerevisiae*. Cuando se utiliza un hongo filamentoso como huésped, se puede utilizar cualquier hongo filamentoso, de preferencia un microorganismo perteneciente al género *Humicola*, *Trichoderma*, *Aspergillus*, *Acremonium* o *Fusarium*, de más preferencia un microorganismo perteneciente al género *Humicola* o *Trichoderma*. Más concretamente, se pueden utilizar *Humicola insolens*, *Humicola thermoidea*, *Trichoderma viride*, *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Staphylotrichum coccosporum*, *Aspergillus niger*, *Aspergillizs oryzae*, *Fusarium oxysporum* o *Acremonium cellulolyticus*, de preferencia *Humicola insolens* o *Trichoderma viride*.

La transformación de un microorganismo con el vector de expresión puede llevarse a cabo según un procedimiento habitual, tal como un procedimiento que utiliza iones de calcio, un procedimiento que utiliza iones de litio, un procedimiento de electroporación, un procedimiento con PEG, un procedimiento de *Agrobacterium* o un procedimiento con pistola de partículas y se puede seleccionar un procedimiento adecuado de acuerdo con el huésped a transformarse.

En la presente invención, se cultiva el transformante de la presente invención y el cultivo resultante se utiliza para obtener la proteína de la presente invención de interés. El transformante de la presente invención se puede cultivar según un procedimiento habitual seleccionado adecuadamente, por ejemplo, un medio o condiciones del cultivo.

El medio se puede complementar con, por ejemplo, fuentes de carbono o fuentes de nitrógeno que pueden anabolizarse o utilizarse, respectivamente, por el transformante de la presente invención, sales inorgánicas, diversas vitaminas, diversos aminoácidos tales como ácido glutámico o asparagina, oligonutrientes, tales como nucleótidos, o agentes de selección tales como antibióticos. Además, se pueden añadir de manera adecuada sustancias orgánicas y/o inorgánicas que ayudan al crecimiento del transformante de la presente invención o promueven la producción de la proteína de la presente invención. Además, si es necesario, se puede utilizar un medio natural o un medio sintético que contenga adecuadamente otros nutrientes.

Se puede utilizar cualquier tipo de fuente de carbono y fuente de nitrógeno en el medio con la condición de que se pueden utilizar por el transformante de la presente invención. Como fuente de carbono anabolizable, se pueden utilizar diversos hidratos de carbono, por ejemplo, sacarosa, almidón, glicerina, glucosa, fructosa, maltosa, lactosa, celulosa, celobiosa, dextrina, aceites animales o aceites vegetales, o hidrolizados de los mismos. En general, la concentración de los mismos en el medio es de preferencia del 0,1 % al 5 %. Como fuente de nitrógeno utilizable, se pueden usar por ejemplo, componentes, animales o vegetales o extractos de los mismos, tales como peptona, extracto de carne, licor de maceración de maíz o polvo de soja desgrasada, sales de amonio de ácidos orgánicos tales como sales de amonio del ácido succínico o sales de amonio del ácido tartárico, urea u otros diversos compuestos que contienen nitrógeno tales como ácidos inorgánicos u orgánicos. Como sales inorgánicas, se pueden utilizar adecuadamente, por ejemplo, las que pueden producir sodio, potasio, calcio, magnesio, cobalto, cloro, ácido fosfórico, ácido sulfúrico, u otros iones.

Se puede utilizar adecuadamente cualquier medio que contenga otros componentes, por ejemplo, células, exudados o extractos de microorganismos tales como levaduras o polvos finos de plantas, con la condición de que no interfieran con el crecimiento de los transformantes de la presente invención ni con la producción y acumulación de la proteína de la presente invención. Cuando se cultiva una cepa mutante que tiene un requerimiento nutricional, por lo general se añade al medio una sustancia para satisfacer su requerimiento nutricional. En relación con lo anterior, es posible que tal nutriente no deba añadirse necesariamente cuando se utiliza un medio que contiene sustancias naturales.

Las condiciones de cultivo, tales como el componente del medio, la fluidez del medio, la temperatura de incubación,

la velocidad de agitación o la tasa de aireación, se pueden seleccionar y controlar adecuadamente de acuerdo con el transformante a utilizarse y las condiciones externas a fin de obtener los resultados preferibles. Si se formara espuma en un medio líquido, puede utilizarse adecuadamente un agente antiespumante, tal como aceites de silicona, aceites vegetales, aceites minerales o tensioactivos.

- 5 La proteína de la presente invención acumulada en el cultivo resultante está contenida en el transformante de la presente invención y en el filtrado del cultivo. De acuerdo con ello, es posible recuperar la proteína de la presente invención a partir tanto del filtrado del cultivo como de las células transformantes por separación del cultivo en cada fracción por centrifugación.

- 10 La proteína de la presente invención se puede recuperar a partir del filtrado del cultivo de acuerdo con un procedimiento habitual. Los procedimientos para recuperar de la proteína de la presente invención a partir del cultivo se pueden realizar de manera individual, en combinación en un determinado orden, o repetidamente. Por ejemplo, se pueden utilizar la extracción por filtración, la centrifugación, la diálisis, la concentración, el secado, la congelación, la adsorción, la desorción, medios para la separación basados en la diferencia de solubilidad en diversos disolventes (tales como la precipitación, la precipitación por saturación de la solución con sales neutras, la cristalización, la
- 15 recristalización, la solución inversa o la cromatografía).

- Además, la proteína de la presente invención se puede obtener a partir del cultivo dentro del transformante de la presente invención. Por ejemplo, la extracción a partir del cultivo (por ejemplo, tratamiento de molienda o alteración por presión), la recuperación (por ejemplo, filtración o centrifugación) o la purificación (por ejemplo, precipitación por saturación de la solución con sales neutras o precipitación del disolvente) pueden llevarse a cabo utilizando un
- 20 procedimiento habitual.

- La sustancia en bruto resultante puede ser purificada de acuerdo con un procedimiento habitual, por ejemplo, por una cromatografía en columna utilizando un vehículo tal como dextrano, celulosa, agarosa, polímeros sintéticos, o gel de sílice. Se puede obtener una proteína pura de la presente invención de interés a partir del cultivo del transformante de la presente invención utilizando los procedimientos anteriormente mencionados, bien por separado
- 25 o bien en una combinación adecuada.

- Como se describió anteriormente, de acuerdo con otra forma de realización de la presente invención, se puede proporcionar el procedimiento para producir la proteína novedosa de la presente invención. El cultivo del transformante (incluyendo las condiciones de cultivo) se puede llevar a cabo de una manera sustancialmente similar al cultivo del huésped original utilizado para preparar el transformante. Como el procedimiento para recuperar la
- 30 proteína de interés después del cultivo del transformante, se pueden llevar a cabo procedimientos de uso común.

#### Cepas depositadas

- Se depositó la cepa de *Escherichia coli* JM109 transformada con el vector de expresión pEGD01 de la presente invención (FERM BP-5973) en el país en el Instituto Nacional Depositario de Organismos de Patentes Internacionales de Ciencia y Tecnología Industrial Avanzadas [(Nombre anterior) Instituto Nacional de Biocencia y
- 35 Agencia de Tecnología Humana de Ciencia y Tecnología Industrial (Dirección: AIST Tsukuba Central 6, 1-1, Higashi 1-chome Tukuba-shi, Ibaraki-ken 305-8566, Japón)] el 12 de julio de 1996 y se trasladó a un depósito internacional el 13 de junio de 1997. El número de depósito internacional (un número entre paréntesis [] después del número de depósito internacional es un número de depósito nacional) es FERM BP-5973 [FERM P-15729].

- Se depositó la cepa de *Escherichia coli* JM109 transformada con el vector de expresión pMKD01 de la presente invención (FERM BP-5974) en el país en el Instituto Nacional Depositario de Organismos de Patentes Internacionales de Ciencia y Tecnología Industrial Avanzadas [(Nombre anterior) Instituto Nacional de Biocencia y
- 40 Agencia de Tecnología Humana de Ciencia y Tecnología Industrial (Dirección: AIST Tsukuba Central 6, 1-1, Higashi 1-chome Tukuba-shi, Ibaraki-ken 305-8566, Japón)] el 12 de julio de 1996 y se trasladó a un depósito internacional el 13 de junio de 1997. El número de depósito internacional (un número entre paréntesis [] después del número de depósito internacional es un número de depósito nacional) es FERM BP-5974 [FERM P-15730].
- 45

- Se depositó la cepa de *Escherichia coli* JM109 transformada con el vector de expresión pCB1-M2XR (FERM BP-6045) en el país en el Instituto Nacional Depositario de Organismos de Patentes Internacionales de Ciencia y Tecnología Industrial Avanzadas [(Nombre anterior) Instituto Nacional de Biocencia y Agencia de Tecnología Humana de Ciencia y Tecnología Industrial (Dirección: AIST Tsukuba Central 6, 1-1, Higashi 1-chome Tukuba-shi,
- 50 Ibaraki-ken 305-8566, Japón)] el 9 de septiembre de 1996 y se trasladó a un depósito internacional el 11 de agosto de 1997. El número de depósito internacional (un número entre paréntesis [] después del número de depósito internacional es un número de depósito nacional) es FERM BP-6045 [FERM P-15840].

- En relación con lo anterior, el plásmido pCB1-M2 de la presente invención se puede obtener no solo por el procedimiento descrito en los EJEMPLOS, sino también por auto-unión del plásmido pCB1-M2XR previamente
- 55 digerido con la enzima de restricción XbaI.

Se depositó la cepa de *Humicola insolens* MN200-1 (FERM BP-5977), que puede utilizarse como un huésped para el vector de expresión de la presente invención, en el país en el Instituto Nacional Depositario de Organismos de

Patentes Internacionales de Ciencia y Tecnología Industrial Avanzadas [(Nombre anterior) Instituto Nacional de Biocencia y Agencia de Tecnología Humana de Ciencia y Tecnología Industrial (Dirección: AIST Tsukuba Central 6, 1-1, Higashi 1-chome Tukuba-shi, Ibaraki-ken 305-8566, Japón)] el 15 de julio de 1996 y se trasladó a un depósito internacional el 13 de junio de 1997. El número de depósito internacional (un número entre paréntesis [] después del número de depósito internacional es un número de depósito nacional) es FERM BP-5977 [FERM P-15736].

Se depositó la cepa de *Trichoderma viride* MC300-1 (FERM BP-6047), que puede ser utilizada como huésped para el vector de expresión de la presente invención, en el país en el Instituto Nacional Depositario de Organismos de Patentes Internacionales de Ciencia y Tecnología Industrial Avanzadas [(Nombre anterior) Instituto Nacional de Biocencia y Agencia de Tecnología Humana de Ciencia y Tecnología Industrial (Dirección: AIST Tsukuba Central 6, 1-1, Higashi 1-chome Tukuba-shi, Ibaraki-ken 305-8566, Japón)] el 9 de septiembre de 1996 y se trasladó a un depósito internacional el 11 de agosto de 1997. El número de depósito internacional (un número entre paréntesis [] después del número de depósito internacional es un número de depósito nacional) es FERM BP-6047 [FERM P-15842].

Se depositó la cepa de *Escherichia coli* DH5 $\alpha$  transformada con el vector de expresión pUC118-STCEex (FERM BP-10127) en el país en el Instituto Nacional Depositario de Organismos de Patentes Internacionales de Ciencia y Tecnología Industrial Avanzadas (Dirección: AIST Tsukuba Central 6, 1-1, Higashi 1-chome Tukuba-shi, Ibaraki-ken 305-8566, Japón) el 1 de diciembre de 2003 y se trasladó a un depósito internacional el 15 de septiembre de 2004. El número de depósito internacional (un número entre paréntesis [] después del número de depósito internacional es un número de depósito nacional) es FERM BP-10127 [FERM P-19602]. En relación con lo anterior, el plásmido pUC118-STCEex es un plásmido obtenido por la inserción del gen de STCE en el sitio BamHI del plásmido pUC118. El gen de la endoglucanasa STCE1 presente en el fragmento BamHI contiene un intrón y tiene la secuencia de la SEQ ID NO: 5.

### Ejemplos

La presente invención ahora se ilustrará con más detalle por, pero de ningún medio está limitada a, los siguientes Ejemplos.

#### **Ejemplo A1: Construcción de vectores de expresión para expresar celulasa modificada en su extremo N-terminal y celulasa de tipo natural en *Humicola insolens***

(1) Construcción del vector de expresión pMKD01 para expresar celulasa modificada en su extremo N-terminal

El plásmido pUC118 (Takara Bio) se digirió con BamHI. Se generaron extremos romos en el fragmento resultante con un kit de generación de extremos romos de ADN (Takara Bio) y se auto-unió con un kit de unión de ADN (Takara Bio) para obtener pUC118-BN. El pUC118-BN se digirió con SphI, se generaron extremos romos y se auto-unió para obtener pUC118BSN. A continuación, se obtuvo un fragmento PstI-XbaI de 3,4 kb que comprendía el gen de celulasa NCE2 a partir de *Humicola insolens* MN200-1 (FERM BP-5977) según el procedimiento descrito en la Publicación de Patente Japonesa no examinada (Kokai) N.º: 8-126492. El fragmento se unió con pUC118BSN previamente digerido con la misma enzima de restricción para obtener pUC118BSN-PX. Se introdujo la mutación dirigida al sitio en pUC118BSN-PX con un sistema de mutagénesis in vitro Sculptor (Amersham Bioscience) para obtener el plásmido pM21. Como oligonucleótidos para la mutagénesis, se usaron los oligonucleótidos MNC-02 y MNC-03.

MNC-02: 5'-GAGCGCCAGAACTGTGGATCCACTTGGTGAGCAATG-3' (SEQ NO: 6)

MNC-03: 5'-TCCGCCGTTCTGAGCGGATCCAGGCGTTTGCGCG-3' (SEQ NO: 7)

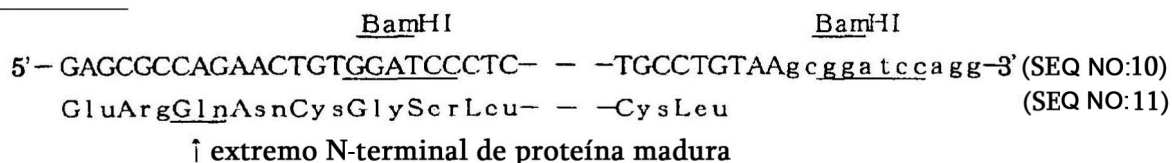
El plásmido pM21 se digirió utilizando con BamHI y se unió con un fragmento de PCR digerido con BamHI de un gen de celobiohidrolasa (NCE3) derivado de *Humicola insolens* MN200-1 para obtener el plásmido pKM04. El fragmento de PCR de NCE3 se amplificó utilizando ADN genómico (documento WO98/03640) de *Humicola insolens* MN200-1 (FERM BP-5977) como un molde y los siguientes cebadores MKA-05 y MKA-06.

MKA-05: 5'-GCCGCCAGCAGGCGGGATCCCTCACCACCGAGAGG-3' (SEQ NO: 8)

MKA-06: 5'-TGATCGTCGAGTCAGGGATCCAGAAATTACAGGCAC-3' (SEQ NO: 9)

A continuación, se utilizaron un promotor y un terminador (Mullaney, EJ y col., Mol. Gen. Genet., 199, 37-45, 1985) del gen *trpC* derivado de *Aspergillus nidulans* para preparar un gen capaz de expresar un gen de resistencia a la destomicina en *Humicola insolens*. El gen resultante se unió con el plásmido pKM04 previamente digerido con XbaI para construir el vector de expresión pMKD01 (FERM BP-5974) para expresar la celulasa modificada en el extremo N-terminal. En pMKD01, se introdujeron los sitios de reconocimiento de *BamHI* a 10 pb secuencia abajo del extremo N-terminal de la proteína madura del gen de NCE2 y a 3 pb secuencia abajo del codón de parada del mismo y por lo tanto se pueden añadir cinco residuos de aminoácidos al extremo N-terminal de una celulasa deseada perteneciente a la familia 45.

Tabla 1



(2) Construcción del vector de expresión pJD01 para expresar la celulasa de tipo natural

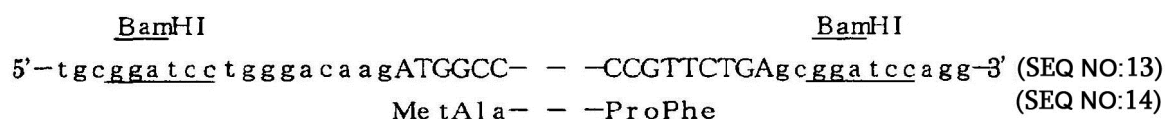
Se introdujo la mutación dirigida al sitio utilizando el oligonucleótido pMN-Bam en el plásmido pM21 obtenido en el punto anterior (1) para obtener el plásmido pM21-m-A1.

5 pMN-Bam:

5'-GGTCAAACAAGTCTGTGCGGATCCTGGGACAAGATGGCCAAGTTCTTCCTTAC-3' (SEQ NO: 12)

10 El plásmido pM21-m-A1 se digirió con las enzimas de restricción *Hind*III y *Bam*HI y se recuperó el fragmento resultante de aproximadamente 1 kb. A continuación, el pMKD01 obtenido en el Ejemplo A1(1) se digirió con *Hind*III y *Bam*HI y se recuperó un fragmento de aproximadamente 7 kb. Estos fragmentos se unieron para obtener el plásmido pJD01. En este vector, se introdujeron los sitios de reconocimiento de *Bam*HI a 15 pb secuencia arriba del sitio de iniciación de la traducción del gen de NCE2 y a 3 pb secuencia abajo del codón de parada del mismo y por consiguiente se expresó una celulasa deseada perteneciente a la familia 45 desde el punto de iniciación de la transcripción.

Tabla 2



15 **Ejemplo A2: Construcción de vectores de expresión para expresar NCE4 de tipo natural y NCE4 modificada en su extremo N-terminal, transformación de *Humicola insolens* con los mismos y evaluación de las actividades**

(1) Construcción de los vectores de expresión

20 Como un vector de expresión para NCE4 de tipo natural, se unió la región codificadora del gen de NCE4 con pJD01 obtenido en el Ejemplo A1(2) para construir pN2EX-N4.

En primer lugar, se amplificó la región codificadora del gen de NCE4 por PCR utilizando ADN genómico de *Humicola insolens* MN200-1 como un molde y NCE4-Ne y NCE4-Ce como cebadores. El producto de PCR resultante de aproximadamente 0,9 kpb se digirió con *Bam*HI y se unió con pJD01 previamente digerido con la misma enzima de restricción para construir pN2EX-N4.

25 NCE4-Ne: 5'-GGGGGGATCCTGGGACAAGATGCGTTCCTCCCCTCTC-3' (SEQ NO: 15)

NCE4-Ce: 5'-GGGGGGATCCCTGCGTTTACAGGCACTGATGG-3' (SEQ NO: 16)

30 Como un vector de expresión para NCE4 modificada en su extremo N-terminal, se unió la región codificadora de la proteína madura en la que se eliminó la secuencia señal para la secreción en el gen de NCE4 con pMKD01 (FERM BP-5974) obtenido en el Ejemplo A1(1) para construir pEGD01 (FERM BP-5973). En primer lugar, se amplificó la región codificadora de la parte de la proteína madura del gen de NCE4 llevando a cabo PCR utilizando ADN genómico de *Humicola insolens* MN200-1 como molde y NCE4-Ns y NCE4-Cs como cebadores. El producto de PCR resultante de aproximadamente 0,8 kpb se digirió con *Bam*HI y se unió con pMKD01 previamente digerido con la misma enzima de restricción para la construcción del pEGD01.

NCE4-Ns: 5'-CCGGTGTGGCCGGATCCGCTGATGGCAAG-3' (SEQ NO: 17)

35 NCE4-Cs: 5'-TAAGGCCCTCAAGGATCCCTGCGTCTACAG-3' (SEQ NO: 18)

(2) Transformantes que contienen NCE4 de tipo natural o NCE4 modificada en su extremo N-terminal

Se cultivó *Humicola insolens* MN200-1 en un medio S (glucosa al 3,0 %, extracto de levadura al 2,0 %, polipeptona al 0,1 %, cloruro de calcio al 0,03 % y sulfato de magnesio al 0,03 %; pH 6,8) a 37 °C durante 24 horas. Los micelios resultantes se recogieron por centrifugación, se lavaron con sacarosa 0,5 mol/l y se suspendieron en solución de

sacarosa 0,5 mol/l que contenía  $\beta$ -glucuronidasa 3 mg/ml (Sigma), quitinasa 1 mg/ml y zimoliasa 20T 1 mg/ml (Seikagaku Corp.). La suspensión se agitó suavemente a 30 °C durante 60 a 90 minutos para generar protoplastos. Los protoplastos resultantes se recogieron por filtración y centrifugación, se lavaron con un tampón SUTC (sacarosa 0,5 mol/l, cloruro de calcio 10 mmol/l y Tris-HCl 10 mmol/l, pH 7,5) y se resuspendieron en SUTC para preparar una suspensión de protoplastos.

A 100  $\mu$ l de la suspensión de protoplastos, se añadieron 10  $\mu$ l de cada solución de ADN (pEGD01 o pN2EX-N4). Tras dejar en reposo cada una en hielo durante 5 minutos, se añadieron 400  $\mu$ l de una solución de PEG (polietilenglicol 4000 al 60 %, cloruro de calcio 10 mmol/l y Tris-HCl 10 mmol/l, pH 7,5) y todo se dejó en reposo sobre hielo durante 20 minutos. Después de lavar la suspensión de protoplastos resultante con el tampón SUTC, se cubrió la suspensión de protoplastos con un medio YMG (glucosa al 1 %, extracto de levadura al 0,4 %, extracto de malta al 0,2 % y agar al 1 %, pH 6,8) que contenía higromicina B 200  $\mu$ g/ml junto con agar blando YMG. La placa se incubó a 37 °C durante 5 días para obtener los transformantes como colonias.

(3) Comparación de las actividades endoglucanasa de NCE4 de tipo natural y NCE4 modificada en su extremo N-terminal en los sobrenadantes de los cultivos

Los transformantes resultantes se cultivaron según el procedimiento descrito en el documento WO98/03667, para obtener los sobrenadantes de los cultivos. Los sobrenadantes de los cultivos se sometieron a SDS-PAGE para seleccionar los sobrenadantes de cultivo en los que se detectaba con claridad la banda de NCE4 de aproximadamente 43 kDa. Se midieron las actividades endoglucanasa en los sobrenadantes del cultivo a pH 10 en ausencia y con presencia de LAS y se compararon los grados de reducción en la actividad en presencia de LAS. Los resultados se muestran en la Tabla 3. En la tabla 3, se representan las actividades endoglucanasa por porcentajes cuando cada actividad endoglucanasa a pH 10 se considera como 100.

Tabla 3

| Vector de expresión                     |          | EGU (%) |         |
|-----------------------------------------|----------|---------|---------|
|                                         |          | Sin LAS | Con LAS |
| NCE de tipo natural                     | pN2EX-N4 | 100     | 27,1    |
| NCE modificada en su extremo N-terminal | pEGD01   | 100     | 55,6    |

Como resultado, el sobrenadante del cultivo obtenido a partir del transformante de NCE4 modificada en su extremo N-terminal mostró una pequeña reducción de la actividad en presencia de LAS. Tras una comparación de las actividades endoglucanasa en presencia de LAS, se encontró que la actividad endoglucanasa de NCE4 modificada en su extremo N-terminal fue 2,1 veces mayor que la de NCE4 de tipo natural. Además, la actividad endoglucanasa de *Humicola insolens* MN200-1 representó aproximadamente el 4 % de los transformantes, por consiguiente se encontró que casi todas las actividades endoglucanasa derivaron de las enzimas recombinantes expresadas por los vectores de expresión incorporados.

(4) Análisis de las secuencias de aminoácidos del extremo N-terminal de NCE4 de tipo natural y de NCE4 modificada en su extremo N-terminal

Los sobrenadantes de los cultivos de los transformantes anteriores se sometieron a SDS-PAGE y las proteínas sometidas a electroforesis se transfirieron eléctricamente a una membrana de PVDF (Immobilon-PSQ, Millipore). Las bandas de NCE4 de aproximadamente 43 kDa se recortaron de la mancha y se sometieron al secuenciador Protein Sequencer Model 492 (Applied Biosystems) para analizar las secuencias de aminoácidos.

Como resultado, se confirmó que la secuencia de NCE4 de tipo natural obtenida del transformante pN2EX-N4 era coherente con la secuencia de ADN de NCE4 derivada de *Humicola insolens* MN200-1 descrita en el documento WO98/03667.

Sin embargo, la secuencia de aminoácidos de NCE4 modificada en su extremo N-terminal obtenida a partir del transformante de pEGD01 no pudo determinarse en el procedimiento anterior. Se utilizó una piroglutamato aminopeptidasa Pfu (Takara Bio) para eliminar el extremo N-terminal modificado a partir de NCE4 modificada en su extremo N-terminal y a continuación se determinó el aminoácido (SEQ NO: 2).

Los resultados se resumieron en la Tabla 4, para comparar los aminoácidos del extremo N-terminal y las secuencias de aminoácidos de extremo N-terminal en NCE4 de tipo natural y en NCE4 con el extremo N-terminal añadido y se confirmó a partir de los resultados que cada NCE4 tenía la secuencia de aminoácidos del extremo N-terminal esperada del correspondiente vector de expresión construido.

Tabla 4

|                                          | Aminoácido en el extremo N-terminal | Secuencia de aminoácidos del extremo N-terminal |
|------------------------------------------|-------------------------------------|-------------------------------------------------|
| NCE4 de tipo natural                     | Alanina (no modificado)             | ADGKSTR (SEQ NO: 19)                            |
| NCE4 modificada en su extremo N-terminal | Ácido piroglutámico (modificado)    | pQNCGSADGKSTR (SEQ NO: 20)                      |

**Ejemplo B1: Construcción de vectores de expresión para *Trichoderma viride* para expresar celulasa modificada en su extremo N-terminal y celulasa de tipo natural**

(1) Construcción de los vectores de expresión

- 5 A partir de un fragmento de PstI de 7 kb que contenía el gen *cbh1* derivado de *Trichoderma viride* MC300-1 (obtenido de acuerdo con el documento WO98/11239) se clonaron un fragmento de *HindIII* de aproximadamente 3,1 kb que contenía una región promotora de *cbh1* y un fragmento de *Sall* de 2,7 kb que contenía una región terminadora de *cbh1* en el plásmido pUC119 para construir el plásmido pCB1-H4 y el plásmido pCB1-S1, respectivamente. Se infectó *E. coli* JM109 que contenía cada plásmido resultante con el fago ayudante M13KO7 para obtener cada ADN de hebra única. A los ADN de hebra única resultantes de los plásmidos pCB1-H4 y pCB1-S1 como moldes, se hibridaron oligonucleótidos fosforilados CBN-Stu y CBC-Xho y se introdujeron mutaciones con un sistema de mutagénesis in vitro Sculptor (Amersham Bioscience), para la construcción de los plásmidos pCB1H4-19 y pCB1S1-17, respectivamente. Se unió un fragmento de aproximadamente 6 kb, que se había obtenido por la digestión del plásmido pCB1H4-19 con XbaI y XhoI, con un fragmento de aproximadamente 1,2 kb, que se había sido obtenido por digestión del plásmido pCB1S1-17 con XbaI y posteriormente sometiénolo a digestión parcial con XhoI, para la construcción del pCB1-M.

CBN-Stu: 5'-GATACATGATGCGCAGGCCTTAGTCGACTAGAATGC-3' (SEQ NO:21)

CBC-Xho: 5'-GATCCTCAAGCTTTTGTCTCGAGTACCTTACAAGCAC-3' (SEQ NO: 22)

- 20 A continuación, se clonó un fragmento de aproximadamente 2,7 kb, obtenido por digestión del pCB1-M con *Sall*, en el plásmido pUC119 para construir el plásmido pCB1-Salm. Al ADN de hebra única obtenido del plásmido pCB1-Salm se le hibridaron los oligonucleótidos fosforilados CB1-SmSph, CB1-Bam y CB1-Pst y se introdujeron las mutaciones con un sistema de mutagénesis in vitro Sculptor.

CB1-SmSph: 5'-GGAGGGTGCATGCCGACTGAGCCCGGGCAGTAGCC-3' (SEQ NO: 23)

CB1-Bam: 5'-GCCGGGAGAGGATCCAGTGGAGG-3' (SEQ NO: 24)

- 25 CB1-Pst: 5'-GCTCGAGTACCTTACTGCAGGCACTGAGAG-3' (SEQ NO: 25)

- 30 A continuación, el plásmido pUC118 se digirió utilizando las enzimas XbaI y EcoRI, se generaron extremos romos con el kit de generación de extremos romos de ADN y se auto-unió, para construir el plásmido pUC118-SBN. Al pUC118-SBN digerido con *HindIII* y *Sall*, se le insertó el fragmento del promotor de *cbh1* de aproximadamente 1,4 kb previamente digerido con las mismas enzimas de restricción. El plásmido resultante se digirió con *Sall* y se unió con la región codificadora de *cbh1* con las mutaciones introducidas y la región terminadora de aproximadamente 2,8 kb, para la construcción del pCB1-M2. Las mutaciones introducidas (excluyendo la mutación de *BamHI*) se muestran en la Tabla 5.

Tabla 5

| <u>Sal I</u>                                       | <u>Stu I</u>                               |                                            |                                             |                 |             |
|----------------------------------------------------|--------------------------------------------|--------------------------------------------|---------------------------------------------|-----------------|-------------|
| 5'-c t a g t c g a c t a a g g c c t g c g c a t c | ATGTATCAAAAGTTGGCCCTCATCTCGGCCTTCTTGGCTACT | (SEC NO:26)                                |                                             |                 |             |
|                                                    | MetTyrGlnLysLeuAlaLeuIleSerAlaPheLeuAlaThr | (SEC NO:27)                                |                                             |                 |             |
| <u>Sma I</u>                                       | <u>Sph I</u>                               | <u>Pst I</u>                               | <u>Xho I</u>                                | <u>Hind III</u> | (SEC NO:28) |
| 5'-G C C C G G G C T C A G T C G G C A T G C A C C | - -                                        | - C A G T G C C T G C A G T A A            | g g t a c t c g a g c a a a a g c t t g a g | - 3'            | (SEC NO:29) |
| AlaArgAlaGlnSerAlaCysThr                           | - -                                        | GlnCysLeuGln                               |                                             |                 |             |
|                                                    |                                            | ↑ extremo N-terminal de la proteína madura |                                             |                 |             |

- 35 **Ejemplo B2: Construcción de vectores de expresión para expresar STCE1 modificada en su extremo N-terminal y STCE1 de tipo natural, transformación de *Trichoderma viride* con los mismos y evaluación de las actividades**

## (1) Construcción de los vectores de expresión

Como un vector de expresión para STCE1 de tipo natural, se unió la región codificadora del gen de STCE1 con el plásmido pCB1-M2 obtenido en el Ejemplo B1 para la construcción del pCB-Ste.

- 5 En primer lugar, se amplificó la región codificadora del gen de STCE1 por PCR utilizando el plásmido pUC118-STCEex (FERM BP-10127) como molde y la combinación de cebadores STCE1-TNERV y STCE1-TCET22I. El producto resultante de la PCR de aproximadamente 1 kpb se digirió con EcoRV y EcoT22I y se unió con pCB1-M2 previamente digerido con StuI y PstI para la construcción del PCB-Ste.

STCE1-TNERV: GGGGATATCGCGCATCATGCGTTCCTCCCCCGTCCTC (SEQ NO: 30)

STCE1-TCET22I: GGGATGCATTTAAAGGCACTGCGAGTACCAGTC (SEQ NO: 31)

- 10 Como un vector de expresión para STCE1 modificada en el extremo N-terminal, la región codificadora de la proteína madura en la que se había eliminado la secuencia señal para la secreción en el gen de STCE1, se unió con pCB1-M2 obtenido en el Ejemplo B1 para la construcción del PCB-Sts.

- 15 En primer lugar, se amplificó la región codificadora de la proteína madura del gen de STCE1 por PCR utilizando el plásmido pUC118-STCEex (FERM BP-10127) como un molde y STCE1-TmNSph y STCE1-TCET22I como cebadores. El producto de la PCR resultante de aproximadamente 1 kpb se digirió con SphI y EcoT22I y se unió con pCB1-M2 previamente digerido con SphI y PstI para la construcción del PCB-Sts.

STCE1-TmNSph: GGGGCATGCGATGGCAAGTCGACCCGCTAC (SEQ NO: 32)

(2) Generación de la cepa 2 de *Trichoderma viride*

- 20 Se irradió una suspensión de esporas (aproximadamente  $10^9$  UFC/ml) de *Trichoderma viride* MC300-1 (FERM BP-6047) con UV de dos lámparas UV colocadas a una altura de 30 cm por encima de la misma mientras que se agitaba suavemente. La suspensión se extendió en un medio de selección y se cultivó a 28 °C durante 7 días. Las cepas cultivadas se seleccionaron obteniéndose la cepa 2 de *Trichoderma viride* que mostraba requerimiento de uracilo.

- 25 Al respecto, el medio de selección era un medio mínimo [dihidrogenofosfato de potasio al 0,2 %, sulfato de amonio al 0,4 %, urea al 0,03 %, sulfato de magnesio heptahidratado al 0,03 %, cloruro de calcio al 0,03 %, glucosa al 0,5 %, agar al 2,5 % y oligoelementos al 0,01 % (preparados por la solución de 5 mg de sulfato de hierro heptahidratado, 1,56 mg de sulfato de manganeso heptahidratado, 1,4 mg de sulfato de cinc heptahidratado y 2 mg de cloruro de cobalto en 1 litro de agua)] suplementado con uridina 10 µg/ml y ácido 5-fluoroorótico 1 mg/ml.

## (3) Transformantes que contienen STCE1 de tipo natural o STCE1 modificada en su extremo N-terminal

- 30 Se transformó *Trichoderma viride* con pCB-Ste o pCB-Sts obtenidos en el Ejemplo B2(1), llevando a cabo un procedimiento de co-transformación utilizando la cepa 2 de *Trichoderma viride* obtenida en el Ejemplo B2(2) y el plásmido marcador pPYR4 que contenía el gen pyr4 derivado de *Neurospora crassa*, para obtener los transformantes como cepas cultivadas en el medio mínimo.

- 35 En primer lugar, se inoculó la cepa 2 de *Trichoderma viride* en un matraz cónico de 200 ml que contenía 50 ml de un medio de formación de micelios [extracto de levadura al 1 %, extracto de malta al 1 %, polipeptona al 2 %, glucosa al 2,5 %, hidrogenofosfato dipotásico al 0,1 %, sulfato de magnesio heptahidratado al 0,05 % y uridina al 0,0001 % (pH 7,0)] y se cultivó a 28 °C durante 2 días. Las células resultantes se recogieron por centrifugación, se lavaron con sacarosa 0,5 mol/l y se suspendieron en solución de sacarosa 0,5 mol/l que contenía β-glucuronidasa 3 mg/ml (Sigma), quitinasa 1 mg/ml y zimoliasa 20T 1 mg/ml (Seikagaku Corp.). La suspensión se agitó suavemente a 30 °C durante 60 a 90 minutos para generar protoplastos. Los protoplastos resultantes se recogieron por filtración y centrifugación, se lavaron con tampón SUTC y se resuspendieron en SUTC para preparar una suspensión de protoplastos.

- 40 A 100 µl de la suspensión de protoplastos, se le añadieron 10 µl de cada solución de ADN (pCB-Ste o pCB-Sts). Tras dejar cada una en reposo sobre hielo durante 5 minutos, se añadieron 400 µl de la solución de PEG y se dejó todo en reposo sobre hielo durante 20 minutos. Tras lavar la suspensión de protoplastos resultante con el tampón SUTC, se cubrió el medio mínimo que contenía 0,5 mol/l de sacarosa con la suspensión de protoplastos junto con agar blando. La placa se incubó a 28 °C durante 5 días y a continuación las colonias que se habían desarrollado se transfirieron al medio mínimo, para obtener los transformantes como colonias desarrolladas en el mismo.

- 50 (4) Comparación de las actividades endoglucanasa de STCE1 de tipo natural y STCE1 modificada en su extremo N-terminal en los sobrenadantes de los cultivos

Los transformantes resultantes se cultivaron según el procedimiento descrito en el documento WO98/11239, para obtener los sobrenadantes de cultivo. Los sobrenadantes de los cultivos se sometieron a SDS-PAGE, para seleccionar los sobrenadantes de cultivo en los que se detectaba con claridad la banda de STCE1 de

aproximadamente 45 kDa. Se midieron las actividades endoglucanasa en los sobrenadantes de los cultivos a pH 10 en ausencia y con presencia de LAS y se compararon los grados de reducción de la actividad en presencia de LAS. Los resultados se muestran en la Tabla 6. En la Tabla 6, se representan las actividades endoglucanasa por porcentajes cuando cada actividad endoglucanasa a pH 10 se considera como 100.

5

Tabla 6

|                                           | Vector de expresión | EGU (%) |         |
|-------------------------------------------|---------------------|---------|---------|
|                                           |                     | Sin LAS | con LAS |
| STCE1 de tipo natural                     | pCB-Ste             | 100     | 14,5    |
| STCE1 modificada en su extremo N-terminal | pCB-Sts             | 100     | 25,9    |

Como resultado, al comparar las actividades endoglucanasa en presencia de LAS, se encontró que la actividad endoglucanasa de STCE1 modificada en su extremo N-terminal fue aproximadamente 1,8 veces superior a la de STCE1 de tipo natural. Además, no se detectó la actividad endoglucanasa de la cepa 2 de *Trichoderma viride* y por lo tanto se encontró que las actividades endoglucanasa se derivaban de las enzimas recombinantes expresadas por los vectores de expresión incorporados.

10

(5) Análisis de secuencias de aminoácidos del extremo N-terminal de STCE1 de tipo natural y STCE1 modificada en su extremo N-terminal

A partir de cada sobrenadante de cultivo de los transformantes anteriores, se preparó una solución de los mismos que contenía Tris-HCl 0,05 mol/l (pH 8,0)/sulfato de sodio 1 mol/l que se adsorbió en un lecho de igual volumen (BV) de TOYOPEARL Butil-650S (Tosoh Corporation). Los vehículos se lavaron con tres BV de Tris-HCl 0,05 mol/l (pH 8,0)/sulfato de sodio 1 mol/l para eliminar los componentes no adsorbidos. La elución se realizó con tres BV de Tris-HCl 0,05 mol/l (pH 8,0)/sulfato de sodio 0,5 mol/l y la solución eluida se concentró por medio de Pellicon XL (valor de corte 5000, Millipore) y Ultrafree-15 (valor de corte 5000, Millipore). A partir del concentrado obtenido, se preparó una solución del mismo que contenía fosfato de sodio 0,1 mol/l (pH 5,8)/sulfato de amonio 1 mol/l. La solución resultante se sometió a una cromatografía hidrófoba utilizando Resource PHE (6 ml, Amersham Bioscience), para recuperar una fracción no adsorbida. La fracción no adsorbida se concentró y se desaló por medio de PD-10 (Amersham Bioscience) y a partir de la fracción, se preparó una solución de la misma que contenía solución de Tris-HCl 0,05 mol/l (pH 7,5). La solución se sometió a Resource Q (6 ml, Amersham Bioscience), para recuperar una fracción no adsorbida. La fracción no adsorbida se desaló y a partir de la fracción, se preparó una solución de la misma que contenía solución de acetato de sodio 0,05 mol/l (pH 5,0). La solución se sometió a Resource S (6 ml, Amersham Bioscience), para recuperar una fracción no adsorbida. La fracción no adsorbida se concentró y se sometió a una cromatografía en gel utilizando Superdex 75pg (16 X 60 mm; Amersham Bioscience) para recuperar una fracción con un peso molecular fraccionado de aproximadamente 45.000. Cada fracción (derivada de cada sobrenadante de cultivo de los transformantes anteriores) contenía solo la banda única de 45 kDa.

15

20

25

30

Las fracciones purificadas se sometieron a SDS-PAGE y las proteínas sometidas a electroforesis se transfirieron eléctricamente a una membrana de PVDF. Las bandas de STCE1 de aproximadamente 45 kDa se recortaron de la mancha y se sometieron al Secuenciador de Proteínas Modelo 492 (Applied Biosystems) para analizar las secuencias de aminoácidos.

35

Como resultado, se confirmó que la secuencia de STCE1 de tipo natural obtenida del transformante pCB-Ste era coherente con la secuencia de ADN (SEQ ID NO: 5) de STCE1 derivada de *Staphylotrichum coccosporum* IFO (nombre actual: NBRC) 31817.

Sin embargo, la secuencia de aminoácidos de STCE1 obtenida del transformante de STCE1 modificada en su extremo N-terminal no pudo determinarse en el procedimiento anterior. Se utilizó una piroglutamato aminopeptidasa Pfu (Takara Bio) para eliminar el extremo N-terminal de la STCE1 modificada en su extremo N-terminal y a continuación se determinó el aminoácido (SEQ NO: 4).

40

Los resultados se resumieron en la Tabla 7 para comparar los aminoácidos del extremo N-terminal y las secuencias de aminoácidos del extremo N-terminal de STCE1 de tipo natural y STCE1 modificada en su extremo N-terminal. A partir de los resultados se confirmó que cada STCE1 tenía la secuencia de aminoácidos del extremo N-terminal esperada del correspondiente vector de expresión construido.

45

Tabla 7

|                                           | Aminoácido del extremo N-terminal | Secuencia de aminoácidos del extremo N-terminal |
|-------------------------------------------|-----------------------------------|-------------------------------------------------|
| STCE1 de tipo natural                     | Alanina (sin modificar)           | ADGKSTR (SEQ NO: 33)                            |
| STCE1 modificada en su extremo N-terminal | Ácido piroglutámico (modificado)  | pQSACDGKSTR (SEQ NO: 34)                        |

(6) Comparación de las actividades endoglucanasa de STCE1 de tipo natural purificada y STCE1 modificada en su extremo N-terminal purificada

- 5 Se midieron las actividades endoglucanasa de STCE1 de tipo natural purificada y STCE1 sustituida en su extremo N-terminal purificada obtenidas en el Ejemplo B2(5) a pH 10 en ausencia y con presencia de LAS y se compararon los grados de reducción de la actividad en presencia de LAS. Los resultados se muestran en la Tabla 8. En la Tabla 8, se representan las actividades endoglucanasa por porcentajes cuando cada actividad endoglucanasa a pH 10 se considera como 100.

Tabla 8

|                                                      | EGU (%) |         |
|------------------------------------------------------|---------|---------|
|                                                      | Sin LAS | Con LAS |
| STCE1 de tipo natural purificada                     | 100     | 16,2    |
| STCE1 modificada en su extremo N-terminal purificada | 100     | 30,0    |

- 10 Como resultado, al comparar las actividades endoglucanasa en presencia de LAS, se encontró que la actividad endoglucanasa de STCE1 modificada en su extremo N-terminal purificada fue aproximadamente 1,9 veces superior a la de la STCE1 de tipo natural purificada.

**Ejemplo B3: Cambio de la secuencia añadida al extremo N-terminal de STCE1, construcción del vector, transformación de *Trichoderma viride* con el mismo y evaluación de las actividades**

- 15 (1) Construcción del vector de expresión para la adición de ácido piroglutámico solo al extremo N-terminal de proteína madura

Como un vector de expresión para STCE1 en la que solamente se añadió ácido piroglutámico, se unió la región codificadora de la proteína madura en la que se había eliminado la secuencia señal para la secreción en el gen de STCE1 con el plásmido pCB1-M2 obtenido en el Ejemplo B1 para la construcción del pCB-Stm11.

- 20 En primer lugar, se amplificó la región codificadora de la proteína madura del gen de STCE1 por PCR utilizando el plásmido pUC118-STCEex (FERM BP-10127) como molde y STCE-TmNSma y STCE1-TCET22I como cebadores. El producto de la PCR resultante de aproximadamente 1 kpb se digirió con *Sma*I y *Eco*T22I y se unió con pCB1-M2, previamente digerido con *Sma*I y *Pst*I para la construcción del pCB-Stm11. *Trichoderma viride* transformado con el vector pCB-Stm11 construido de la manera descrita anteriormente produjo la STCE1 que tiene la secuencia de aminoácidos N-terminal del pQADGKSTR (SEQ ID NO: 41) [es decir, la secuencia en la que se añadió un ácido piroglutámico (pQ) al extremo N-terminal de STCE1 de tipo natural].

STCE-TmNSma: GGCCCGGGCTCAGGCCGATGGCAAGTCGACCCG (SEQ NO: 35)

(2) Comparación de la actividad endoglucanasa de STCE1 a la que se añadió ácido piroglutámico

- 30 Se transformó la cepa 2 de *Trichoderma viride* con pCB-Stm11 según el procedimiento descrito en el Ejemplo B2(3). Los transformantes resultantes se cultivaron de acuerdo con el documento WO98/11239 para obtener los sobrenadantes de cultivo. Los sobrenadantes de los cultivos se sometieron a SDS-PAGE para seleccionar un sobrenadante de cultivo en el que se detectaba con claridad la banda de STCE1 de aproximadamente 45 kDa. Se midió la actividad endoglucanasa en el sobrenadante del cultivo a pH 10 en ausencia y con presencia de LAS y se comparó el grado de reducción de la actividad en presencia de LAS con el de la STCE1 de tipo natural en el Ejemplo B2(4). Los resultados se muestran en la Tabla 9. En la Tabla 9, se representan las actividades endoglucanasa por porcentajes cuando cada actividad endoglucanasa a pH 10 se considera como 100.

Tabla 9

| Vector expresión      |           | EGU (%) |         |
|-----------------------|-----------|---------|---------|
|                       |           | Sin LAS | Con LAS |
| STCE1 de tipo natural | pCB-Ste   | 100     | 14,5    |
| Vector expresión      |           | EGU (%) |         |
|                       |           | Sin LAS | Con LAS |
| STCE1 con pQ añadido  | pCB-Stm11 | 100     | 23,0    |

Como resultado, al comparar las actividades endoglucanasa en presencia de LAS, se encontró que la actividad endoglucanasa de la STCE1 a la que se añadió ácido piroglutámico fue aproximadamente 1,6 veces superior a la de la STCE1 de tipo natural.

#### 5 Ejemplo B4: Cambio de la secuencia añadida al extremo N-terminal de STCE1, construcción del vector, transformación de *Trichoderma viride* con el mismo y evaluación de las actividades

(1) Construcción del vector de expresión para la adición de 4 aminoácidos que contienen pQ al extremo N-terminal de la proteína madura

10 Como vector de expresión para la STCE1 a la que se añadieron cuatro aminoácidos que contenían ácido piroglutámico, se unió la región codificadora de proteína madura en la que se había suprimido la secuencia señal para la secreción en el gen de STCE1, con el plásmido pCB1-M2 obtenido en el Ejemplo B1 para construir pCB-Stm12.

15 En primer lugar, se amplificó la región codificadora de la proteína madura del gen de STCE1 llevando a cabo PCR utilizando el plásmido pUC118-STCEex (FERM BP-10127) como molde y STCE-TmNSph-2 y STCE1-TCET22I como cebadores. El producto de la PCR resultante de aproximadamente 1 kpb se digirió con *SphI* y *EcoT22I* y se unió con pCB1-M2 previamente digerido con *SphI* y *PstI* para construir pCB-Stm12. *Trichoderma viride* transformado con el vector pCB-Stm12 construido como se describió anteriormente produjo la STCE1 que tiene la secuencia de aminoácidos del extremo N-terminal del pQSACADGKSTR (SEQ ID NO: 42) [es decir, la secuencia en la que se añadió un péptido que consiste en 4 aminoácidos (el aminoácido N-terminal es ácido piroglutámico) al extremo N-terminal de la STCE1N de tipo natural].

STCE-TmNSph-2: GGGCATGCGCCGATGGCAAGTCGACCCGC (SEQ NO: 36)

(2) Comparación de la actividad endoglucanasa de la STCE1 con 4 aminoácidos añadidos que contienen pQ

25 La cepa 2 de *Trichoderma viride* se transformó con pCB-Stm12 según el procedimiento descrito en el Ejemplo B2(3). Los transformantes resultantes se cultivaron de acuerdo con el documento WO98/11239 para obtener los sobrenadantes del cultivo. Los sobrenadantes de los cultivos se sometieron a SDS-PAGE, para seleccionar un sobrenadante de cultivo en el que se detectaba con claridad la banda de STCE1 de aproximadamente 45 kDa. Se midió la actividad endoglucanasa en el sobrenadante del cultivo a pH 10 en ausencia y con presencia de LAS y se comparó el grado de reducción de la actividad en presencia de LAS con el de la STCE1 de tipo natural en el Ejemplo B2(4). Los resultados se muestran en la Tabla 10. En la Tabla 10, se representan las actividades endoglucanasa por porcentajes cuando cada actividad endoglucanasa a pH 10 se considera como 100.

Tabla 10

| Vector de expresión                       |           | EGU (%) |         |
|-------------------------------------------|-----------|---------|---------|
|                                           |           | Sin LAS | Con LAS |
| STCE1 de tipo natural                     | pCB-Ste   | 100     | 14,5    |
| STCE1 con 4-AAAs añadidos que contiene pQ | pCB-Stm12 | 100     | 22,4    |

Como resultado, al comparar las actividades endoglucanasa en presencia de LAS, se encontró que la actividad endoglucanasa de la STCE1 con 4 aminoácidos añadidos que contienen pQ (STCE1 con 4 AAs añadidos que contienen pQ) fue aproximadamente 1,5 veces superior a la de la STCE1 de tipo natural.

#### 35 Aplicabilidad industrial

La proteína de la presente invención presenta una pequeña reducción en la actividad endoglucanasa en presencia de un tensioactivo, en comparación con las celulasas de tipo natural y por consiguiente es útil como componente de

un detergente para la ropa.

Aunque la presente invención ha sido descrita con referencia a formas de realización concretas, es posible realizar diversos cambios y modificaciones obvias para los expertos en la técnica sin apartarse del alcance de las reivindicaciones adjuntas.

## 5 Breve descripción de los dibujos

La figura 1 muestra los resultados de la comparación entre la secuencia de aminoácidos de la endoglucanasa STCE1 [el péptido señal (SEQ ID NO: 43) y la proteína madura (SEQ ID NO: 44)] y las secuencias de aminoácidos de endoglucanasas conocidas pertenecientes a la familia de 45, NCE4 [el péptido señal (SEQ ID NO: 45) y la proteína madura (SEQ ID NO: 46)] y NCE5 [el péptido señal (SEQ ID NO: 47) y la proteína madura (SEQ ID NO: 48)], con respecto a las secuencias de aminoácidos de la mitad N-terminal.

La figura 2 muestra los resultados de la comparación que se describe en la figura 1, con respecto a las secuencias de aminoácidos de la mitad C terminal.

## Texto libre en el listado de secuencias

Las características de "secuencia artificial" se describen en el identificador numérico <223> en el Listado de secuencias. Cada una de las secuencias de nucleótidos de las SEQ ID NO: 6-10, 12-13, 15-18, 21-26, 28, 30-32 y 35-36 son: cebador MNC-02, cebador MNC-03, cebador MKA-05, cebador MKA-06, pMKD01, cebador pMN-Bam, pJD01, cebador NCE4-Ne, cebador NCE4-Ce, cebador NCE4-Ns, cebador NCE4-Cs, cebador CBN-Stu, cebador CBC-Xho, CB1-SmSph, cebador CB1-Bam, cebador CB1-Pst, pCB1-M2, pCB1-M2, cebador STCE1-TNERV, cebador STCE1-TCET22I, cebador STCE1-TmNSph, cebador STCE1-TmNSma y cebador STCE1-TmNSph-2. Cada una de las secuencias de aminoácidos de las SEQ ID NO: 11, 14, 27 y 29 son pMKD01, pJD01, pCB1-M2 y pCB1-M2.

## Listado de secuencias

<110> MEIJI SEIKA KAISHA, LTD.

<120> Celulosa tolerante a los detergentes y procedimiento para modificar la misma

25 <130> MEJ-719

<150> JP 2003-409692

<151> 8-12-2003

<160> 48

<170> PatentIn version 3.1

30 <210> 1

<211> 870

<212> ADN

<213> Humicola insolens MN200-1

<220>

35 <221> CDS

<221> (1) .. (870)

<223>

<220>

<221> característica miscelánea

40 <221> (1) .. (3)

<223> Ácido piroglutámico

<220>

<221> fuente

# ES 2 539 955 T3

<222> (16) .. (870)

<223> Humicola insolens MN200-1

<400> 1

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| cag | aac | tgt | gga | tcc | gct | gat | ggc | aag | tcc | acc | cgc | tac | tgg | gac | tgc | 48 |
| Gln | Asn | Cys | Gly | Ser | Ala | Asp | Gly | Lys | Ser | Thr | Arg | Tyr | Trp | Asp | Cys |    |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     |     | 15  |     |    |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |    |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| tgc | aag | cct | tcg | tgc | ggc | tgg | gcc | aag | aag | gct | ccc | gtg | aac | cag | cct | 96 |
| Cys | Lys | Pro | Ser | Cys | Gly | Trp | Ala | Lys | Lys | Ala | Pro | Val | Asn | Gln | Pro |    |

ES 2 539 955 T3

| 20                                                              | 25  | 30  |     |
|-----------------------------------------------------------------|-----|-----|-----|
| gtc ttc tcc tgc aac gcc aac ttc cag cgt ctc act gac ttc gac gcc |     |     | 144 |
| Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Leu Thr Asp Phe Asp Ala |     |     |     |
| 35                                                              | 40  | 45  |     |
| aag tcc ggc tgc gag ccg ggc ggt gtc gcc tac tcg tgc gcc gac cag |     |     | 192 |
| Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln |     |     |     |
| 50                                                              | 55  | 60  |     |
| acc cca tgg gct gtg aac gac gac ttc gcg ttc ggt ttt gct gcc acc |     |     | 240 |
| Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Phe Gly Phe Ala Ala Thr |     |     |     |
| 65                                                              | 70  | 75  | 80  |
| tct att gcc ggc agc aat gag gcg ggc tgg tgc tgc gcc tgc tac gag |     |     | 288 |
| Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu |     |     |     |
| 85                                                              | 90  | 95  |     |
| ctc acc ttc aca tcc ggt cct gtt gct ggc aag aag atg gtc gtc cag |     |     | 336 |
| Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln |     |     |     |
| 100                                                             | 105 | 110 |     |
| tcc acc agc act ggc ggt gat ctt ggc agc aac cac ttc gat ctc aac |     |     | 384 |
| Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn |     |     |     |
| 115                                                             | 120 | 125 |     |
| atc ccc ggc ggc ggc gtc ggc atc ttc gac gga tgc act ccc cag ttc |     |     | 432 |
| Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe |     |     |     |
| 130                                                             | 135 | 140 |     |
| ggc ggt ctg ccc ggc cag cgc tac ggc ggc atc tcg tcc cgc aac gag |     |     | 480 |
| Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu |     |     |     |
| 145                                                             | 150 | 155 | 160 |
| tgc gat cgg ttc ccc gac gcc ctc aag ccc ggc tgc tac tgg cgc ttc |     |     | 528 |
| Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe |     |     |     |
| 165                                                             | 170 | 175 |     |
| gac tgg ttc aag aac gcc gac aac ccg agc ttc agc ttc cgt cag gtc |     |     | 576 |
| Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val |     |     |     |
| 180                                                             | 185 | 190 |     |
| caa tgc cca gcc gag ctc gtc gct cgc acc gga tgc cgc cgc aac gac |     |     | 624 |
| Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp |     |     |     |

# ES 2 539 955 T3

| 195                                                             | 200 | 205 |     |
|-----------------------------------------------------------------|-----|-----|-----|
| gac ggc aac ttc cct gcc gtc cag atc ccc tcc agc agc acc agc tct |     |     | 672 |
| Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser |     |     |     |
| 210                                                             | 215 | 220 |     |
| ccg gtc ggc cag cct acc agt acc agc acc acc tcc acc tcc acc acc |     |     | 720 |
| Pro Val Gly Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr |     |     |     |
| 225                                                             | 230 | 235 | 240 |
| tcg agc ccg ccc gtc cag cct acg act ccc agc ggc tgc act gct gag |     |     | 768 |
| Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu |     |     |     |
| 245                                                             | 250 | 255 |     |
| agg tgg gct cag tgc ggc ggc aat ggc tgg agc ggc tgc acc acc tgc |     |     | 816 |
| Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys |     |     |     |
| 260                                                             | 265 | 270 |     |
| gtc gct ggc agc acc tgc acg aag att aat gac tgg tac cat cag tgc |     |     | 864 |
| Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys |     |     |     |
| 275                                                             | 280 | 285 |     |
| ctg taa                                                         |     |     | 870 |
| Leu                                                             |     |     |     |

<210> 2

<211> 289

<212> PRT

5 <213> Humicola insolens MN200-1

<220>

<221> característica miscelánea

<221> (1) .. (3)

<223> Ácido piroglutámico

10 <400> 2

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Asn | Cys | Gly | Ser | Ala | Asp | Gly | Lys | Ser | Thr | Arg | Tyr | Trp | Asp | Cys |
| 1   |     |     |     | 5   |     |     |     | 10  |     |     |     |     |     | 15  |     |

Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro

# ES 2 539 955 T3

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     | 20  |     | 25  |     | 30  |     |     |     |     |     |     |     |     |     |     |
| Val | Phe | Ser | Cys | Asn | Ala | Asn | Phe | Gln | Arg | Leu | Thr | Asp | Phe | Asp | Ala |
|     | 35  |     |     |     |     | 40  |     |     |     |     |     | 45  |     |     |     |
| Lys | Ser | Gly | Cys | Glu | Pro | Gly | Gly | Val | Ala | Tyr | Ser | Cys | Ala | Asp | Gln |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Thr | Pro | Trp | Ala | Val | Asn | Asp | Asp | Phe | Ala | Phe | Gly | Phe | Ala | Ala | Thr |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |
| Ser | Ile | Ala | Gly | Ser | Asn | Glu | Ala | Gly | Trp | Cys | Cys | Ala | Cys | Tyr | Glu |
|     |     |     |     | 85  |     |     |     |     | 90  |     |     |     |     | 95  |     |
| Leu | Thr | Phe | Thr | Ser | Gly | Pro | Val | Ala | Gly | Lys | Lys | Met | Val | Val | Gln |
|     |     |     |     | 100 |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Ser | Thr | Ser | Thr | Gly | Gly | Asp | Leu | Gly | Ser | Asn | His | Phe | Asp | Leu | Asn |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Ile | Pro | Gly | Gly | Gly | Val | Gly | Ile | Phe | Asp | Gly | Cys | Thr | Pro | Gln | Phe |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Gly | Gly | Leu | Pro | Gly | Gln | Arg | Tyr | Gly | Gly | Ile | Ser | Ser | Arg | Asn | Glu |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     | 160 |     |
| Cys | Asp | Arg | Phe | Pro | Asp | Ala | Leu | Lys | Pro | Gly | Cys | Tyr | Trp | Arg | Phe |
|     |     |     |     | 165 |     |     |     |     | 170 |     |     |     |     | 175 |     |
| Asp | Trp | Phe | Lys | Asn | Ala | Asp | Asn | Pro | Ser | Phe | Ser | Phe | Arg | Gln | Val |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Gln | Cys | Pro | Ala | Glu | Leu | Val | Ala | Arg | Thr | Gly | Cys | Arg | Arg | Asn | Asp |

|  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
|  | 195 |     | 200 |     | 205 |     |     |     |     |     |     |     |     |     |     |     |  |
|  | Asp | Gly | Asn | Phe | Pro | Ala | Val | Gln | Ile | Pro | Ser | Ser | Ser | Thr | Ser | Ser |  |
|  | 210 |     |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |  |
|  | Pro | Val | Gly | Gln | Pro | Thr | Ser | Thr | Ser | Thr | Thr | Ser | Thr | Ser | Thr | Thr |  |
|  | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |  |
|  | Ser | Ser | Pro | Pro | Val | Gln | Pro | Thr | Thr | Pro | Ser | Gly | Cys | Thr | Ala | Glu |  |
|  |     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |  |
|  | Arg | Trp | Ala | Gln | Cys | Gly | Gly | Asn | Gly | Trp | Ser | Gly | Cys | Thr | Thr | Cys |  |
|  |     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |  |
|  | Val | Ala | Gly | Ser | Thr | Cys | Thr | Lys | Ile | Asn | Asp | Trp | Tyr | His | Gln | Cys |  |
|  |     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |  |

Leu

<210> 3

<211> 897

<212> ADN

5 <213> Staphylotrichum coccosporum IF0 31817

<220>

<221> CDS

<221> (1) .. (897)

<223>

10 <220>

<221> característica miscelánea

<221> (1) .. (3)

<223> Ácido piroglutámico

<220>

15 <221> fuente

<222> (13) .. (897)

<223> Staphylotrichum coccosporum IF0 31817

<400> 3

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

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser |     |
| 145 150 155 160                                                 |     |
| ttc ccc gcc gcc ctc aag ccc ggc tgc tac tgg cgc ttc gac tgg ttc | 528 |
| Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe |     |
| 165 170 175                                                     |     |
| aag aac gcc gac aac ccg acc ttc acc ttc cgc cag gtc cag tgc ccg | 576 |
| Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro |     |
| 180 185 190                                                     |     |
| tcg gag ctc gtc gcc cgc acc ggc tgc cgc cgc aac gac gac ggc aac | 624 |
| Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn |     |
| 195 200 205                                                     |     |
| ttc ccc gtc ttc acc cct ccc tcg ggc ggt cag tcc tcc tcg tct tcc | 672 |
| Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser |     |
| 210 215 220                                                     |     |
| tcc tcc agc agc gcc aag ccc acc tcc acc tcc acc tcg acc acc tcc | 720 |
| Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr Ser |     |
| 225 230 235 240                                                 |     |
| acc aag gct acc tcc acc acc tcg acc gcc tcc agc cag acc tcg tcg | 768 |
| Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser |     |
| 245 250 255                                                     |     |
| tcc acc ggc ggc ggc tgc gcc gcc cag cgc tgg gcg cag tgc ggc ggc | 816 |
| Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly |     |
| 260 265 270                                                     |     |
| atc ggg ttc tcg ggc tgc acc acg tgc gtc agc ggc acc acc tgc aac | 864 |
| Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn |     |
| 275 280 285                                                     |     |
| aag cag aac gac tgg tac tcg cag tgc ctt taa                     | 897 |
| Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu                         |     |
| 290 295                                                         |     |

<210> 4

<211> 298

<212> PRT

5 <213> Staphylotrichum coccosporum IF0 31817

<220>

<221> característica miscelánea

<221> (1) .. (3)

<223> Ácido piroglutámico

5 <400> 4

Gln Ser Ala Cys Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys  
1 5 10 15

Pro Ser Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe  
20 25 30

Ala Cys Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser  
35 40 45

Gly Cys Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp  
50 55 60

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

Gly Gly Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe  
85 90 95

Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser  
100 105 110

Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly  
115 120 125

Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu  
130 135 140

ES 2 539 955 T3

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

Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe  
165 170 175

Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro  
180 185 190

Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn  
195 200 205

Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser  
210 215 220

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

Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser  
245 250 255

Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly  
260 265 270

Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn  
275 280 285

Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu  
290 295

<210> 5

<211> 1037

<212> ADN

5 <213> Staphylotrichum coccosporum IF0 31817

<220>

<221> péptido señal

<221> (1) .. (63)

<223>

5 <220>

<221> exón

<221> (64) .. (333)

<223>

<220>

10 <221> Intrón

<221> (334) .. (419)

<223>

<220>

<221> exón

15 <221> (420) .. (1037)

<223>

<400> 5

atgogttcct ccccgctcct ccgcacggcc ctggccgctg cccctccccct ggccgocctc 60

gct gcc gat ggc aag tcg acc cgc tac tgg gac tgt tgc aag ccg tcg 108  
Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser  
1 5 10 15

tgc tcg tgg ccc ggc aag gcc tcg gtg aac cag ccc gtc ttc gcc tgc 156  
Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe Ala Cys  
20 25 30

agc gcc aac ttc cag cgc atc agc gac ccc aac gtc aag tcg ggc tgc 204  
Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser Gly Cys  
35 40 45

gac ggc ggc tcc gcc tac gcc tgc gcc gac cag acc ccg tgg gcc gtc 252  
Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp Ala Val  
50 55 60

aac gac aac ttc tcg tac ggc ttc gcc gcc acg tcc atc tcg ggc ggc 300  
Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser Gly Gly  
65 70 75

|                                                                   |     |
|-------------------------------------------------------------------|-----|
| aac gag gcc tgg tgg tgc tgt ggc tgc tac gag tgagtgttc ccccccccc   | 353 |
| Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu                       |     |
| 80 85 90                                                          |     |
| ccccccccac ccccggttcg gtcccttgcc gtgccttctt cataactaacc gctaccccc | 413 |
| tccagg ctg acc ttc acc tgg ggc ccc gtc gct ggc aag acc atg gtt    | 461 |
| Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val           |     |
| 95 100                                                            |     |
| gtc cag tcc acc tgg acc ggc ggc gac ctc ggc acc aac cac ttc gac   | 509 |
| Val Gln Ser Thr Ser Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp   |     |
| 105 110 115 120                                                   |     |
| ctg gcc atg ccc ggt ggt ggt gtc ggc atc ttc gac ggc tgc tgg ccc   | 557 |
| Leu Ala Met Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro   |     |
| 125 130 135                                                       |     |
| cag ttc ggc ggc ctc gcc ggc gac cgc tac ggc ggc gtc tgg tgg cgc   | 605 |
| Gln Phe Gly Gly Leu Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg   |     |
| 140 145 150                                                       |     |
| agc cag tgc gac tgg ttc ccc gcc gcc ctc aag ccc ggc tgc tac tgg   | 653 |
| Ser Gln Cys Asp Ser Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp   |     |
| 155 160 165                                                       |     |
| cgc ttc gac tgg ttc aag aac gcc gac aac ccg acc ttc acc ttc cgc   | 701 |
| Arg Phe Asp Trp Phe Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg   |     |
| 170 175 180                                                       |     |
| cag gtc cag tgc ccg tgg gag ctc gtc gcc cgc acc ggc tgc cgc cgc   | 749 |
| Gln Val Gln Cys Pro Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg   |     |
| 185 190 195 200                                                   |     |
| aac gac gac ggc aac ttc ccc gtc ttc acc cct ccc tgg ggc ggt cag   | 797 |
| Asn Asp Asp Gly Asn Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln   |     |
| 205 210 215                                                       |     |
| tcc tcc tgg tct tcc tcc tcc agc agc gcc aag ccc acc tcc acc tcc   | 845 |
| Ser Ser Ser Ser Ser Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser   |     |
| 220 225 230                                                       |     |
| acc tgg acc acc tcc acc aag gct acc tcc acc acc tgg acc gcc tcc   | 893 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |  |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|------|
| Thr | Ser | Thr | Thr | Ser | Thr | Lys | Ala | Thr | Ser | Thr | Thr | Ser | Thr | Ala | Ser |  |      |
|     |     | 235 |     |     |     |     | 240 |     |     |     |     | 245 |     |     |     |  |      |
| agc | cag | acc | tgc | tgc | tcc | acc | ggc | ggc | ggc | tgc | gcc | gcc | cag | cgc | tgg |  | 941  |
| Ser | Gln | Thr | Ser | Ser | Ser | Thr | Gly | Gly | Gly | Cys | Ala | Ala | Gln | Arg | Trp |  |      |
|     |     | 250 |     |     |     |     | 255 |     |     |     | 260 |     |     |     |     |  |      |
| gcg | cag | tgc | ggc | ggc | atc | ggg | ttc | tgc | ggc | tgc | acc | acg | tgc | gtc | agc |  | 989  |
| Ala | Gln | Cys | Gly | Gly | Ile | Gly | Phe | Ser | Gly | Cys | Thr | Thr | Cys | Val | Ser |  |      |
| 265 |     |     |     |     | 270 |     |     |     |     | 275 |     |     |     |     | 280 |  |      |
| ggc | acc | acc | tgc | aac | aag | cag | aac | gac | tgg | tac | tgc | cag | tgc | ctt | tga |  | 1037 |
| Gly | Thr | Thr | Cys | Asn | Lys | Gln | Asn | Asp | Trp | Tyr | Ser | Gln | Cys | Leu |     |  |      |
|     |     |     |     | 285 |     |     |     |     | 290 |     |     |     |     | 295 |     |  |      |

<210> 6

<211> 36

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Cebador MNC-02

<400> 6

gagcgccaga actgtggatc cacttggtga gcaatg 36

10 <210> 7

<211> 35

<212> ADN

<213> Secuencia artificial

<220>

15 <223> Descripción de la secuencia artificial: Cebador MNC-03

<400> 7

tccgccgttc tgagcggatc caggcgtttg gcgcg 35

<210> 8

<211> 36

20 <212> ADN

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Cebador MKA-05

<400> 8

25 gccgcccagc aggcgggatc cctcaccacc gagagg 36

<210> 9  
 <211> 36  
 <212> ADN  
 <213> Secuencia artificial  
 5 <220>  
 <223> Descripción de la secuencia artificial: Cebador MKA-06  
 <400> 9  
 tgatcgctga gtcagggatc cagaatttac aggcac 36  
 <210> 10  
 10 <211> 44  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Descripción de la secuencia artificial: pMKD01  
 15 <220>  
 <221> característica miscelánea  
 <221> (24) .. (25)  
 <223> bases no consecutivas  
 <400> 10  
 20 gagcgccaga actgtggatc cctctgcctg taagcggatc cagg 44  
 <210> 11  
 <211> 10  
 <212> PRT  
 <213> Secuencia artificial  
 25 <220>  
 <223> Descripción de la secuencia artificial: pMKD01  
 <220>  
 <221> NON\_CONS  
 <221> (8) .. (9)  
 30 <223>  
 <220>  
 <221> MOD\_RES  
 <221> (1) .. (1)  
 <223> ÁCIDO PIRROLIDONA CARBOXÍLICO  
 35 <400> 11

Glu Arg Gln Asn Cys Gly Ser Leu Cys Leu  
1 5 10

<210> 12

<211> 53

<212> ADN

5 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Cebador pMN-Bam

<400> 12

ggcacaacaa gtctgtcgg atcctgggac aagatggcca agttcttct tac 53

10 <210> 13

<211> 44

<212> ADN

<213> Secuencia artificial

<220>

15 <223> Descripción de la secuencia artificial: pJD01

<220>

<221> característica miscelánea

<221> (24) .. (25)

<223> bases no consecutivas

20 <400> 13

tgcgatcct gggacaagat ggccccgttc tgagcggatc cagg 44

<210> 14

<211> 4

<212> PRT

25 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: pJD01

<220>

<221> NON\_CONS

30 <221> (2) .. (3)

<223>

<400> 14

Met Ala Pro Phe  
1

<210> 15  
 <211> 37  
 <212> ADN  
 <213> Secuencia artificial  
 5 <220>  
 <223> Descripción de la secuencia artificial: Cebador NCE4-Ne  
 <400> 15  
 ggggggatcc tgggacaaga tgcgttcctc ccctctc 37  
 <210> 16  
 10 <211> 32  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Descripción de la secuencia artificial: Cebador NCE4-Ce  
 15 <400> 16  
 ggggggatcc ctgcgtttac aggactgat gg 32  
 <210> 17  
 <211> 30  
 <212> ADN  
 20 <213> Secuencia artificial  
 <220>  
 <223> Descripción de la secuencia artificial: Cebador NCE4-Ns  
 <400> 17  
 ccggtgttgg ccggatccgc tgatggcaag 30  
 25 <210> 18  
 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 30 <223> Descripción de la secuencia artificial: Cebador NCE4-Cs  
 <400> 18  
 taaggccctc aaggatccct gcgctctacag 30  
 <210> 19  
 <211> 7  
 35 <212> PRT  
 <213> Humicola insolens MN200-1  
 <400> 19

Ala Asp Gly Lys Ser Thr Arg  
1 5

<210> 20

<211> 12

<212> PRT

5 <213> Humicola insolens MN200-1

<220>

<221> MOD\_RES

<221> (1) .. (1)

<223> ÁCIDO PIRROLIDONA CARBOXÍLICO

10 <220>

<221> MUTÁGENO

<221> (1) .. (5)

<223>

<400> 20

Gln Asn Cys Gly Ser Ala Asp Gly Lys Ser Thr Arg  
1 5 10

15

<210> 21

<211> 36

<212> ADN

<213> Secuencia artificial

20 <220>

<223> Descripción de la secuencia artificial: Cebador CBN-Stu

<400> 21

gatacatgat ggcagggcct tagtcgacta gaatgc 36

<210> 22

25 <211> 36

<212> ADN

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Cebador CBc-Xho

30 <400> 22

gatcctcaag ctttgctcg agtaccttac aagcac 36

<210> 23

<211> 35

<212> ADN

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Cebador CB1-SmSph

5 <400> 23

ggaggggtgca tgccgactga gcccgggcag tagcc 35

<210> 24

<211> 23

<212> ADN

10 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: Cebador CB1-Bam

<400> 24

gccgggagag gatccagtgg agg 23

15 <210> 25

<211> 30

<212> ADN

<213> Secuencia artificial

<220>

20 <223> Descripción de la secuencia artificial: Cebador CB1-Pst

<400> 25

gctcgagtac ctactgcag gcactgagag 30

<210> 26

<211> 66

25 <212> ADN

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: pCB1-M2

<400> 26

ctagtcgact aaggcctgcg catcatgtat caaaagttgg ccctcatctc ggccttcttg 60

30 gctact 66

<210> 27

<211> 14

<212> PRT

<213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: pCB1-M2

<400> 27

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Tyr | Gln | Lys | Leu | Ala | Leu | Ile | Ser | Ala | Phe | Leu | Ala | Thr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     |

5 <210> 28

<211> 61

<212> ADN

<213> Secuencia artificial

<220>

10 <223> Descripción de la secuencia artificial: pCB1-M2

<210>

<211> característica miscelánea

<212> (24) .. (25)

<213> bases no consecutivas

15 <400> 28

gcccgggctc agtcggcatg caccagtgctc ctgcagtaag gtactcgagc aaaagcttga 60

g 61

<210> 29

<211> 12

<212> PRT

20 <213> Secuencia artificial

<220>

<223> Descripción de la secuencia artificial: pCB1-M2

<220>

<221> NON\_CONS

25 <221> (8) .. (9)

<223>

<400> 29

|     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Arg | Ala | Gln | Ser | Ala | Cys | Thr | Gln | Cys | Leu | Gln |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |

<210> 30

30 <211> 37

<212> ADN

<213> Secuencia artificial  
 <220>  
 <223> Descripción de la secuencia artificial: Cebador STCE1-TNERV  
 <400> 30  
 5 ggggatatcg cgcatcatgc gttcctcccc cgtcctc 37  
 <210> 31  
 <211> 33  
 <212> ADN  
 <213> Secuencia artificial  
 10 <220>  
 <223> Descripción de la secuencia artificial: Cebador STCE1-TCET221  
 <400> 31  
 gggatgcatt taaaggcact gcgagtacca gtc 33  
 <210> 32  
 15 <211> 30  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <223> Descripción de la secuencia artificial: Cebador STCE1-TmNSph  
 20 <400> 32  
 ggggcatgcg atggcaagtc gacccgctac 30  
 <210> 33  
 <211> 7  
 <212> PRT  
 25 <213> Staphylotrichum coccosporum IFO 31817  
 <400> 33

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| Ala | Asp | Gly | Lys | Ser | Thr | Arg |
| 1   |     |     |     | 5   |     |     |

<210> 34  
 <211> 10  
 30 <212> PRT  
 <213> Staphylotrichum coccosporum IFO 31817  
 <220>  
 <221> MOD\_RES  
 <221> (1) .. (1)  
 35 <223> ÁCIDO PIRROLIDONA CARBOXÍLICO

<220>  
 <221> MUTÁGENO  
 <221> (1) .. (4)  
 <223>  
 5 <400> 34

|  |     |     |     |     |     |     |     |     |     |     |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|  | Gln | Ser | Ala | Cys | Asp | Gly | Lys | Ser | Thr | Arg |
|  | 1   |     |     |     | 5   |     |     |     |     | 10  |

<210> 35  
 <211> 33  
 <212> ADN  
 10 <213> Secuencia artificial  
 <220>  
 <223> Descripción de la secuencia artificial: Cebador STCE1-TmNSma  
 <400> 35  
 ggcccgggct caggccgatg gcaagtcgac ccg 33

15 <210> 36  
 <211> 29  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 20 <223> Descripción de la secuencia artificial: Cebador STCE1-TmNSph-2  
 <400> 36  
 ggccatgctgc ccatggcaag tcgaccgc 29  
 <210> 37  
 <211> 891

25 <212> ADN  
 <213> Staphylotrichum coccosporum IF0 31817  
 <220>  
 <221> CDS  
 <221> (1) .. (891)

30 <223>  
 <220>  
 <221> característica miscelánea  
 <221> (1) .. (3)  
 <223> Ácido piroglutámico

35 <220>

<221> fuente

<222> (4) .. (891)

<223> Staphylotrichum coccosporum IF0 31817

<400> 37

cag gcc gat ggc aag tcc acc cgc tac tgg gac tgc tgc aag cct tcg 48

5

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Gln | Ala | Asp | Gly | Lys | Ser | Thr | Arg | Tyr | Trp | Asp | Cys | Cys | Lys | Pro | Ser |     |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |     |
| tgc | tgc | tgg | ccc | ggc | aag | gcc | tgc | gtg | aac | cag | ccc | gtc | ttc | goc | tgc | 96  |
| Cys | Ser | Trp | Pro | Gly | Lys | Ala | Ser | Val | Asn | Gln | Pro | Val | Phe | Ala | Cys |     |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |     |
| agc | gcc | aac | ttc | cag | cgc | atc | agc | gac | ccc | aac | gtc | aag | tgc | ggc | tgc | 144 |
| Ser | Ala | Asn | Phe | Gln | Arg | Ile | Ser | Asp | Pro | Asn | Val | Lys | Ser | Gly | Cys |     |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |     |
| gac | ggc | ggc | tcc | gcc | tac | gcc | tgc | gcc | gac | cag | acc | ccg | tgg | gcc | gtc | 192 |
| Asp | Gly | Gly | Ser | Ala | Tyr | Ala | Cys | Ala | Asp | Gln | Thr | Pro | Trp | Ala | Val |     |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |     |
| aac | gac | aac | ttc | tgc | tac | ggc | ttc | gcc | gcc | acg | tcc | atc | tgc | ggc | ggc | 240 |
| Asn | Asp | Asn | Phe | Ser | Tyr | Gly | Phe | Ala | Ala | Thr | Ser | Ile | Ser | Gly | Gly |     |
| 65  |     |     |     |     | 70  |     |     |     |     | 75  |     |     |     |     | 80  |     |
| aac | gag | gcc | tgc | tgg | tgc | tgt | ggc | tgc | tac | gag | ctg | acc | ttc | acc | tgc | 288 |
| Asn | Glu | Ala | Ser | Trp | Cys | Cys | Gly | Cys | Tyr | Glu | Leu | Thr | Phe | Thr | Ser |     |
|     |     |     | 85  |     |     |     |     |     | 90  |     |     |     |     | 95  |     |     |
| ggc | ccc | gtc | gct | ggc | aag | acc | atg | gtt | gtc | cag | tcc | acc | tgc | acc | ggc | 336 |
| Gly | Pro | Val | Ala | Gly | Lys | Thr | Met | Val | Val | Gln | Ser | Thr | Ser | Thr | Gly |     |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |     |
| ggc | gac | ctc | ggc | acc | aac | cac | ttc | gac | ctg | gcc | atg | ccc | ggt | ggt | ggt | 384 |
| Gly | Asp | Leu | Gly | Thr | Asn | His | Phe | Asp | Leu | Ala | Met | Pro | Gly | Gly | Gly |     |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |     |
| gtc | ggc | atc | ttc | gac | ggc | tgc | tgc | ccc | cag | ttc | ggc | ggc | ctc | gcc | ggc | 432 |
| Val | Gly | Ile | Phe | Asp | Gly | Cys | Ser | Pro | Gln | Phe | Gly | Gly | Leu | Ala | Gly |     |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |     |
| gac | cgc | tac | ggc | ggc | gtc | tgc | tgc | cgc | agc | cag | tgc | gac | tgc | ttc | ccc | 480 |
| Asp | Arg | Tyr | Gly | Gly | Val | Ser | Ser | Arg | Ser | Gln | Cys | Asp | Ser | Phe | Pro |     |
| 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |     |     |     | 160 |     |
| gcc | gcc | ctc | aag | ccc | ggc | tgc | tac | tgg | cgc | ttc | gac | tgg | ttc | aag | aac | 528 |
| Ala | Ala | Leu | Lys | Pro | Gly | Cys | Tyr | Trp | Arg | Phe | Asp | Trp | Phe | Lys | Asn |     |
|     |     |     | 165 |     |     |     |     | 170 |     |     |     |     |     | 175 |     |     |
| gcc | gac | aac | ccg | acc | ttc | acc | ttc | cgc | cag | gtc | cag | tgc | ccg | tgc | gag | 576 |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Asp | Asn | Pro | Thr | Phe | Thr | Phe | Arg | Gln | Val | Gln | Cys | Pro | Ser | Glu |     |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |     |
| ctc | gtc | gcc | cgc | acc | ggc | tgc | cgc | cgc | aac | gac | gac | ggc | aac | ttc | ccc | 624 |
| Leu | Val | Ala | Arg | Thr | Gly | Cys | Arg | Arg | Asn | Asp | Asp | Gly | Asn | Phe | Pro |     |
|     |     | 195 |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |     |
| gtc | ttc | acc | cct | ccc | tgc | ggc | ggt | cag | tcc | tcc | tgc | tct | tcc | tcc | tcc | 672 |
| Val | Phe | Thr | Pro | Pro | Ser | Gly | Gly | Gln | Ser | Ser | Ser | Ser | Ser | Ser | Ser |     |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |     |
| agc | agc | gcc | aag | ccc | acc | tcc | acc | tcc | acc | tgc | acc | acc | tcc | acc | aag | 720 |
| Ser | Ser | Ala | Lys | Pro | Thr | Ser | Thr | Ser | Thr | Ser | Thr | Thr | Ser | Thr | Lys |     |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |     |
| gct | acc | tcc | acc | acc | tgc | acc | gcc | tcc | agc | cag | acc | tgc | tgc | tcc | acc | 768 |
| Ala | Thr | Ser | Thr | Thr | Ser | Thr | Ala | Ser | Ser | Gln | Thr | Ser | Ser | Ser | Thr |     |
|     |     |     |     | 245 |     |     |     |     | 250 |     |     |     |     | 255 |     |     |
| ggc | ggc | ggc | tgc | gcc | gcc | cag | cgc | tgg | gcg | cag | tgc | ggc | ggc | atc | ggg | 816 |
| Gly | Gly | Gly | Cys | Ala | Ala | Gln | Arg | Trp | Ala | Gln | Cys | Gly | Gly | Ile | Gly |     |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |     |
| ttc | tgc | ggc | tgc | acc | acg | tgc | gtc | agc | ggc | acc | acc | tgc | aac | aag | cag | 864 |
| Phe | Ser | Gly | Cys | Thr | Thr | Cys | Val | Ser | Gly | Thr | Thr | Cys | Asn | Lys | Gln |     |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |     |
| aac | gac | tgg | tac | tgc | cag | tgc | ctt | taa |     |     |     |     |     |     |     | 891 |
| Asn | Asp | Trp | Tyr | Ser | Gln | Cys | Leu |     |     |     |     |     |     |     |     |     |
|     | 290 |     |     |     |     | 295 |     |     |     |     |     |     |     |     |     |     |

<210> 38

<211> 296

<212> PRT

5 <213> Staphylotrichum coccosporum IF0 31817

<220>

<221> característica miscelánea

<221> (1) .. (3)

<223> Ácido piroglutámico

10 <400> 38

# ES 2 539 955 T3

Gln Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser  
1 5 10 15

Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe Ala Cys  
20 25 30

Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser Gly Cys  
35 40 45

Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro Trp Ala Val  
50 55 60

Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser Gly Gly  
65 70 75 80

Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe Thr Ser  
85 90 95

Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser Thr Gly  
100 105 110

Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly Gly Gly  
115 120 125

Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu Ala Gly  
130 135 140

Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser Phe Pro  
145 150 155 160

Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe Lys Asn  
165 170 175

Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro Ser Glu  
180 185 190

Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn Phe Pro  
195 200 205

Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser Ser Ser  
210 215 220

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

Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser Ser Thr  
245 250 255

Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly Ile Gly  
260 265 270

Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn Lys Gln  
275 280 285

Asn Asp Trp Tyr Ser Gln Cys Leu  
290 295

<210> 39

<211> 900

<212> ADN

5 <213> Staphylotrichum coccosporum IF0 31817

<220>

<221> CDS

<221> (1) .. (900)

<223>

10 <220>

<221> fuente

<222> (13) .. (90)

<223> Ácido piroglutámico

<220>

<221> característica miscelánea

<221> (1) .. (3)

5 <223> Ácido piroglutámico

<400> 39

|                                                                 |     |
|-----------------------------------------------------------------|-----|
| cag tgc gca tgc gcc gat ggc aag tcc acc cgc tac tgg gac tgc tgc | 48  |
| Gln Ser Ala Cys Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys |     |
| 1 5 10 15                                                       |     |
| aag cct tgc tgc tgc tgg ccc ggc aag gcc tgc gtg aac cag ccc gtc | 96  |
| Lys Pro Ser Cys Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val |     |
| 20 25 30                                                        |     |
| ttc gcc tgc agc gcc aac ttc cag cgc atc agc gac ccc aac gtc aag | 144 |
| Phe Ala Cys Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys |     |
| 35 40 45                                                        |     |
| tgc ggc tgc gac ggc ggc tcc gcc tac gcc tgc gcc gac cag acc ccg | 192 |
| Ser Gly Cys Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro |     |
| 50 55 60                                                        |     |
| tgg gcc gtc aac gac aac ttc tgc tac ggc ttc gcc gcc acg tcc atc | 240 |
| Trp Ala Val Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile |     |
| 65 70 75 80                                                     |     |
| tgc ggc ggc aac gag gcc tgc tgg tgc tgt ggc tgc tac gag ctg acc | 288 |
| Ser Gly Gly Asn Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr |     |
| 85 90 95                                                        |     |
| ttc acc tgc ggc ccc gtc gct ggc aag acc atg gtt gtc cag tcc acc | 336 |
| Phe Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr |     |
| 100 105 110                                                     |     |
| tgc acc ggc ggc gac ctc ggc acc aac cac ttc gac ctg gcc atg ccc | 384 |
| Ser Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro |     |
| 115 120 125                                                     |     |

|                                                                                                                                                       |     |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| ggt ggt ggt gtc ggc atc ttc gac ggc tgc tcg ccc cag ttc ggc ggc<br>Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly<br>130 135 140     | 432 |
| ctc gcc ggc gac cgc tac ggc ggc gtc tcg tcg cgc agc cag tgc gac<br>Leu Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp<br>145 150 155 160 | 480 |
| tcg ttc ccc gcc gcc ctc aag ccc ggc tgc tac tgg cgc ttc gac tgg<br>Ser Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp<br>165 170 175     | 528 |
| ttc aag aac gcc gac aac ccg acc ttc acc ttc cgc cag gtc cag tgc<br>Phe Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys<br>180 185 190     | 576 |
| ccg tcg gag ctc gtc gcc cgc acc ggc tgc cgc cgc aac gac gac ggc<br>Pro Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly<br>195 200 205     | 624 |
| aac ttc ccc gtc ttc acc cct ccc tcg ggc ggt cag tcc tcc tcg tct<br>Asn Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser<br>210 215 220     | 672 |
| tcc tcc tcc agc agc gcc aag ccc acc tcc acc tcc acc tcg acc acc<br>Ser Ser Ser Ser Ser Ala Lys Pro Thr Ser Thr Ser Thr Ser Thr Thr<br>225 230 235 240 | 720 |
| tcc acc aag gct acc tcc acc acc tog acc gcc tcc agc cag acc tcg<br>Ser Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser<br>245 250 255     | 768 |
| tcg tcc acc ggc ggc ggc tgc gcc gcc cag cgc tgg gcg cag tgc ggc<br>Ser Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly<br>260 265 270     | 816 |
| ggc atc ggg ttc tcg ggc tgc acc acg tgc gtc agc ggc acc acc tgc<br>Gly Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys<br>275 280 285     | 864 |
| aac aag cag aac gac tgg tac tcg cag tgc ctt taa<br>Asn Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu<br>290 295                                             | 900 |

<210> 40

<211> 299

<212> PRT

<213> Staphylotrichum coccosporum IF0 31817

<220>

<221> característica miscelánea

5 <221> (1) .. (3)

<223> Ácido piroglutámico

<400> 40

Gln Ser Ala Cys Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys  
1 5 10 15

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

Phe Ala Cys Ser Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys  
35 40 45

Ser Gly Cys Asp Gly Gly Ser Ala Tyr Ala Cys Ala Asp Gln Thr Pro  
50 55 60

Trp Ala Val Asn Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile  
65 70 75 80

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

Phe Thr Ser Gly Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr  
100 105 110

Ser Thr Gly Gly Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro  
115 120 125

ES 2 539 955 T3

Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly  
130 135 140

Leu Ala Gly Asp Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp  
145 150 155 160

Ser Phe Pro Ala Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp  
165 170 175

Phe Lys Asn Ala Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys  
180 185 190

Pro Ser Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly  
195 200 205

Asn Phe Pro Val Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser  
210 215 220

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

Ser Thr Lys Ala Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser  
245 250 255

Ser Ser Thr Gly Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly  
260 265 270

Gly Ile Gly Phe Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys  
275 280 285

Asn Lys Gln Asn Asp Trp Tyr Ser Gln Cys Leu  
290 295

<211> 8

<212> PRT

<213> Staphylotrichum coccosporum IFO 31817

<220>

5 <221> MOD\_RES

<221> (1) .. (1)

<223> ÁCIDO PIRROLIDONA CARBOXÍLICO

<220>

<221> MUTÁGENO

10 <221> (1) .. (1)

<223>

<400> 41

Gln Ala Asp Gly Lys Ser Thr Arg  
1 5

<210> 42

15 <211> 11

<212> PRT

<213> Staphylotrichum coccosporum IFO 31817

<220>

<221> MOD\_RES

20 <221> (1) .. (1)

<223> ÁCIDO PIRROLIDONA CARBOXÍLICO

<220>

<221> MUTÁGENO

<221> (1) .. (4)

25 <223>

<400> 42

Gln Ser Ala Cys Ala Asp Gly Lys Ser Thr Arg

1

5

10

<210> 43

<211> 21

30 <212> PRT

<213> Staphylotrichum coccosporum IFO 31817

# ES 2 539 955 T3

<400> 43

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

Leu Ala Ala Leu Ala  
20

<210> 44

<211> 295

5 <212> PRT

<213> Staphylotrichum coccosporum IFO 31817

<400> 44

Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser Cys  
1 5 10 15

Ser Trp Pro Gly Lys Ala Ser Val Asn Gln Pro Val Phe Ala Cys Ser  
20 25 30

Ala Asn Phe Gln Arg Ile Ser Asp Pro Asn Val Lys Ser Gly Cys Asp  
35 40 45

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

Asp Asn Phe Ser Tyr Gly Phe Ala Ala Thr Ser Ile Ser Gly Gly Asn  
65 70 75 80

ES 2 539 955 T3

Glu Ala Ser Trp Cys Cys Gly Cys Tyr Glu Leu Thr Phe Thr Ser Gly  
85 90 95

Pro Val Ala Gly Lys Thr Met Val Val Gln Ser Thr Ser Thr Gly Gly  
100 105 110

Asp Leu Gly Thr Asn His Phe Asp Leu Ala Met Pro Gly Gly Gly Val  
115 120 125

Gly Ile Phe Asp Gly Cys Ser Pro Gln Phe Gly Gly Leu Ala Gly Asp  
130 135 140

Arg Tyr Gly Gly Val Ser Ser Arg Ser Gln Cys Asp Ser Phe Pro Ala  
145 150 155 160

Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe Lys Asn Ala  
165 170 175

Asp Asn Pro Thr Phe Thr Phe Arg Gln Val Gln Cys Pro Ser Glu Leu  
180 185 190

Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn Phe Pro Val  
195 200 205

Phe Thr Pro Pro Ser Gly Gly Gln Ser Ser Ser Ser Ser Ser Ser  
210 215 220

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

Thr Ser Thr Thr Ser Thr Ala Ser Ser Gln Thr Ser Ser Ser Thr Gly  
245 250 255

# ES 2 539 955 T3

Gly Gly Cys Ala Ala Gln Arg Trp Ala Gln Cys Gly Gly Ile Gly Phe  
260 265 270

Ser Gly Cys Thr Thr Cys Val Ser Gly Thr Thr Cys Asn Lys Gln Asn  
275 280 285

Asp Trp Tyr Ser Gln Cys Leu  
290 295

<210> 45

<211> 20

<212> PRT

5 <213> Humicola insolens

<400> 45

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

Val Leu Ala Leu  
20

<210> 46

<211> 286

10 <212> PRT

<213> Humicola insolens

<400> 46

Ala Ala Asp Gly Lys Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser  
1 5 10 15

Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro Val Phe Ser Cys  
20 25 30

Asn Ala Asn Phe Gln Arg Leu Thr Asp Phe Asp Ala Lys Ser Gly Cys

|                                                                 |     |     |
|-----------------------------------------------------------------|-----|-----|
| 35                                                              | 40  | 45  |
| Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln Thr Pro Trp Ala |     |     |
| 50                                                              | 55  | 60  |
| Val Asn Asp Asp Phe Ala Phe Gly Phe Ala Ala Thr Ser Ile Ala Gly |     |     |
| 65                                                              | 70  | 75  |
| Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu Leu Thr Phe Thr |     |     |
| 85                                                              | 90  | 95  |
| Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln Ser Thr Ser Thr |     |     |
| 100                                                             | 105 | 110 |
| Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn Ile Pro Gly Gly |     |     |
| 115                                                             | 120 | 125 |
| Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe Gly Gly Leu Pro |     |     |
| 130                                                             | 135 | 140 |
| Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu Cys Asp Arg Phe |     |     |
| 145                                                             | 150 | 155 |
| Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe Asp Trp Phe Lys |     |     |
| 165                                                             | 170 | 175 |
| Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val Gln Cys Pro Ala |     |     |
| 180                                                             | 185 | 190 |
| Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp Asp Gly Asn Phe |     |     |
| 195                                                             | 200 | 205 |
| Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser Pro Val Gly Gln |     |     |

210

215

220

Pro Thr Ser Thr Ser Thr Thr Ser Thr Ser Thr Thr Ser Ser Pro Pro  
225 230 235 240

Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu Arg Trp Ala Cys  
245 250 255

Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys Val Ala Gly  
260 265 270

Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys Leu  
275 280 285

<210> 47

<211> 17

<212> PRT

5 <213> Humicola insolens

<400> 47

Met Gln Leu Pro Leu Thr Thr Leu Leu Thr Leu Leu Pro Ala Leu Ala  
1 5 10 15

Ala

<210> 48

<211> 206

10 <212> PRT

<213> Humicola insolens

<400> 48

Ala Gln Ser Gly Ser Gly Arg Thr Thr Arg Tyr Trp Asp Cys Cys Lys  
1 5 10 15

ES 2 539 955 T3

Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ala Pro Val Arg Thr Cys  
20 25 30

Asp Arg Trp Asp Asn Pro Leu Phe Asp Gly Gly Asn Thr Arg Ser Gly  
35 40 45

Cys Asp Ala Gly Gly Gly Ala Tyr Met Cys Ser Asp Gln Ser Pro Trp  
50 55 60

Ala Val Ser Asp Asp Leu Ala Tyr Gly Trp Ala Ala Val Asn Ile Ala  
65 70 75 80

Gly Ser Asn Glu Arg Gln Trp Cys Cys Ala Cys Tyr Glu Leu Thr Phe  
85 90 95

Thr Ser Gly Pro Val Ala Gly Lys Arg Met Ile Val Gln Ala Ser Asn  
100 105 110

Thr Gly Gly Asp Leu Gly Asn Asn His Phe Asp Ile Ala Met Pro Gly  
115 120 125

Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr Gly Ala Pro  
130 135 140

Pro Asn Gly Trp Gly Gln Arg Tyr Gly Gly Ile Ser Gln Arg His Glu  
145 150 155 160

Cys Asp Ala Phe Pro Glu Lys Leu Lys Pro Gly Cys Tyr Trp Arg Phe  
165 170 175

Asp Trp Phe Leu Asn Ala Asp Asn Pro Ser Val Asn Trp Arg Gln Val  
180 185 190

Ser Cys Pro Ala Glu Ile Val Ala Lys Ser Gly Cys Ser Arg  
195 200 205

## REIVINDICACIONES

1. Una combinación de un polinucleótido y una célula huésped, en la que el polinucleótido está seleccionado del grupo que consiste en:

- (a) un polinucleótido que comprende la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39;
- (b) un polinucleótido que comprende una secuencia de nucleótidos en la que de 1 a 90 nucleótidos han sido eliminados, sustituidos, insertados o añadidos en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y que codifica una proteína en la que el extremo N-terminal es ácido glutámico o glutamina y que tiene una actividad endoglucanasa; y
- (c) un polinucleótido que hibrida en condiciones rigurosas con un polinucleótido que consiste en la secuencia de nucleótidos de la SEQ ID NO: 1, 3, 37 o 39 y que codifica una proteína en la que el extremo N-terminal es ácido glutámico o glutamina y que tiene una actividad endoglucanasa, y

en la que la célula huésped es un huésped en el que se lleva a cabo la formación de piroglutamato del aminoácido N-terminal.

2. Una combinación de la célula huésped de acuerdo con la reivindicación 1 y un vector de expresión que comprende el polinucleótido de acuerdo con la reivindicación 1.

3. Una célula huésped transformada con el vector de expresión que comprende el polinucleótido de acuerdo con la reivindicación 1, en la que la célula huésped es un huésped en el que se lleva a cabo la formación de piroglutamato del aminoácido N-terminal.

4. La célula huésped de acuerdo con la reivindicación 3, en la que la célula huésped es una levadura u hongo filamentoso.

5. La célula huésped de acuerdo con la reivindicación 4, en la que el hongo filamentoso es un microorganismo que pertenece al género *Humicola* o *Trichoderma*.

6. La célula huésped de acuerdo con la reivindicación 5, en la que el hongo filamentoso es *Humicola insolens* o *Trichoderma viride*.

7. Un procedimiento de producción de una proteína seleccionada del grupo que consiste en:

- (a) una proteína que comprende la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40, en la que el extremo N-terminal es ácido piroglutámico;
- (b) una proteína modificada que comprende una secuencia de aminoácidos en la que el extremo N-terminal es ácido piroglutámico y en la que han sido eliminados, sustituidos, insertados o añadidos de 1 a 30 aminoácidos en la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40 y que tiene una actividad endoglucanasa; y
- (c) una proteína que comprende una secuencia de aminoácidos que tiene al menos el 85 % de identidad con una proteína que comprende la secuencia de aminoácidos de la SEQ ID NO: 2, 4, 38 o 40, en la que el extremo N-terminal es ácido piroglutámico y que tiene una actividad endoglucanasa, que comprende:

cultivar la célula huésped de acuerdo con una cualquiera de las reivindicaciones 3 a 6 y recuperar la proteína a partir de la célula huésped o del cultivo obtenido por medio del cultivo.

**Figura 1**

|       |                                                        |     |
|-------|--------------------------------------------------------|-----|
| NCE4  | MRSSPLLRSVVAALPVLAL—————AADGKSTRYWDCKPSCGWAK           | 21  |
|       | *****                                                  |     |
| STCE1 | MRSSPVLRTALAAALPLAALA—————ADGKSTRYWDCKPSCSWPG          | 20  |
|       | * ***** *                                              |     |
| NCE5  | MQLPLTLLTLLPALAA—————AQSGSGRTTRYWDCKPSCAWPG            | 23  |
|       | <u>Péptido señal</u> → <u>Dominio catalítico</u> →     |     |
|       |                                                        |     |
| NCE4  | KAPVNQPVFSCNANFQRLTDF—DAKSGCEPGGVAYSCADQTPWAVNDDFA     | 70  |
|       | ** ***** * ***** *      ***** ** ** ***** *            |     |
| STCE1 | KASVNQPVFACSANFQRI SDP—NVKSGCD—GGSAYACADQTPWAVNDNFS    | 68  |
|       | *      ** *      * * ***** ** ** * ** ***** *          |     |
| NCE5  | KGPA—PVRTCDRWDNPLFDGGNTRSGCDAGGGAYMCSQSPWAVSDDL—       | 71  |
|       |                                                        |     |
| NCE4  | FGFAATS I AGSNEAGWCCACYELTFTSGPVAGKKMVVQSTSTGGDLGSNH   | 120 |
|       | ***** * *** *** ***** ***** *                          |     |
| STCE1 | YGFAATS I SGGNEASWCCGCYELTFTSGPVAGKTMVVQSTSTGGDLGTNH   | 118 |
|       | ** ** * * ** *** ***** * ** ***** *                    |     |
| NCE5  | YGWAAVNI AGSNERQWCCACYELTFTSGPVAGKRM I VQASNTGGDLGNNH  | 121 |
|       |                                                        |     |
| NCE4  | FDLNI PGGGVG I FDGCTPQFGGLP—GQRYGG I SSRNECDRFPDALKPG  | 167 |
|       | *** ***** ***** * ***** ** ** *****                    |     |
| STCE1 | FDLAMPGGGVG I FDGCSPPQFGGLA—GDRYGGVSSRSQCDSFPAALKPG    | 165 |
|       | ** ***** * * *      * ***** * * ** ** *****            |     |
| NCE5  | FDIAMPGGGVG I FNACTDQYGAPPNGWGQRYGG I SQRHECDAFPEKLKPG | 171 |

**Figura 2**

|       |                                                   |     |
|-------|---------------------------------------------------|-----|
| NCE4  | CYWRFDWFKNADNPSFSFRQVQCPAELVARTGCCRNDGDFPAVQIPSSS | 217 |
|       | ***** * ***** ***** **                            |     |
| STCE1 | CYWRFDWFKNADNPTFTFRQVQCPSELVARTGCCRNDGDFPVFTPPSGG | 215 |
|       | ***** ***** *** ** * ** ** *                      |     |
| NCE5  | CYWRFDWFLNADNPSVNRQVSCPAEIVAKSGCSR-----           | 206 |
|       | <b>Conector →</b>                                 |     |
| NCE4  | TSSPVGQPTSTSTTSTTSSPPVQPTTPS-----GCTAERWA         | 255 |
|       | ** * ***** * * ** * ***                           |     |
| STCE1 | QSSSSSSSSSAKPTSTSTTSTKATSTTSTASSQTSSSTGGGCAAQRWA  | 265 |
| NCE5  | -----                                             |     |
|       | <b>DUC →</b>                                      |     |
| NCE4  | CQCGGNGWSGCTTCVAGSTCTKINDWYHQCL                   | 286 |
|       | ***** * ***** * ** * *****                        |     |
| STCE1 | CQCGGIGFSGCTTCVSGTTCNKQNDWYSQCL                   | 295 |
| NCE5  | -----                                             |     |