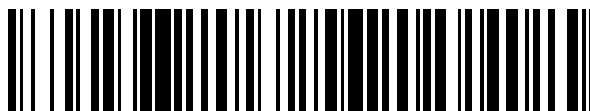


19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 668 202**

51 Int. Cl.:

C12N 9/58 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.12.2010 PCT/US2010/059814**

87 Fecha y número de publicación internacional: **16.06.2011 WO11072191**

96 Fecha de presentación y número de la solicitud europea: **10.12.2010 E 10803307 (7)**

97 Fecha y número de publicación de la concesión europea: **21.02.2018 EP 2510093**

54 Título: **Variantes de proteasa**

30 Prioridad:

11.12.2009 US 285601 P

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

17.05.2018

73 Titular/es:

**NOVOZYMES A/S (50.0%)
Krogshoejvej 36
2880 Bagsvaerd, DK y
NOVOZYMES NORTH AMERICA, INC. (50.0%)**

72 Inventor/es:

**MATSUI, TOMOKO;
NOERGAARD, ALLAN;
POULSEN, THOMAS, AGERSTEN y
MATTHEWS, JOHN**

74 Agente/Representante:

TOMAS GIL, Tesifonte Enrique

ES 2 668 202 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

Variantes de proteasa

5 Referencia a listado de secuencias

[0001] Esta solicitud contiene un listado de secuencias en formato legible por ordenador.

Campo de la invención

10

[0002] La presente descripción se refiere a una proteasa que tiene al menos el 85 % de identidad con una proteasa derivada de *Thermoascus aurantiacus*, cuya secuencia de aminoácidos se muestra en el listado de secuencias anexo como los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y comprende al menos una modificación en comparación con esta proteasa (es decir, es una variante de la misma). La invención también se refiere a ADN que codifica estas proteasas, constructos de ácido nucleico, vectores y células huésped que comprenden los polinucleótidos, así como a métodos de su producción, así como a su uso, por ejemplo, en la producción de productos de fermentación, por ejemplo, etanol.

15

Antecedentes de la invención

20

[0003] Las proteasas son enzimas bien conocidas, como lo son las ventajas de aplicarlas en la producción de productos de fermentación. Las proteasas se han aislado a partir de varias fuentes, que incluyen una serie de cepas bacterianas y fúngicas.

25

[0004] Es un objeto de la presente invención proporcionar polipéptidos alternativos con actividad de proteasa (variantes de proteasa) y polinucleótidos que codifican los polipéptidos. Es también un objeto de la invención proporcionar proteasa con propiedades mejoradas en comparación con la proteasa progenitora de la que se ha derivado. Las variantes de proteasa de la invención muestran termoestabilidad mejorada en comparación con la proteasa progenitora de tipo salvaje.

30

[0005] WO 2003/048353 divulga la proteasa de tipo salvaje de *Thermoascus aurantiacus* CGMCC N.º 0670.

Resumen de la invención

35

[0006] En un primer aspecto, la presente invención se refiere a un método de producción de una variante de proteasa con actividad de proteasa y una termoestabilidad mejorada en comparación con la proteasa progenitora, donde dicho método comprende el cultivo de una célula en la que se ha introducido un vector de expresión que comprende los siguientes elementos operativamente enlazados:

40

- (a) un promotor de transcripción,
- (b) una molécula de polinucleótido que codifica una variante de proteasa que tiene al menos el 85 % de identidad con la proteasa mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y que comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 en al menos una posición seleccionada de las siguientes: 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53,
- 45 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173,
- donde la molécula de polinucleótido se prepara introduciendo al menos una mutación en una molécula de ADN que codifica una proteasa, y
- (c) un terminador de transcripción,

50

con lo cual dicha célula expresa la variante de proteasa codificada por la molécula de polinucleótido; y la recuperación de la variante de proteasa.

55

[0007] En un segundo aspecto, la presente invención se refiere a una proteasa que tiene al menos el 85 % de identidad con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y que comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 en al menos una posición seleccionada de las siguientes: 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53, 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173, donde el polipéptido tiene termoestabilidad mejorada en comparación con la proteasa mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

60

[0008] La invención también se refiere a una secuencia de ácido nucleico que codifica estas proteasas, un constructo de ácido nucleico que comprende la secuencia de ácido nucleico, un vector de expresión que comprende el constructo de ácido nucleico, una célula huésped que comprende el vector de expresión y/o el constructo de ácido nucleico.

65

[0009] La invención también se refiere al uso de las proteasas, por ejemplo, en un proceso de licuefacción, sacarificación y/o fermentación de almidón.

Descripción detallada de la invención

- 5 [0010] La presente descripción se refiere a variantes aisladas de una proteasa progenitora, que comprenden una modificación en al menos una posición que corresponde con las posiciones 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53, 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173 del polipéptido maduro de la SEQ ID N.º: 2, donde la variante tiene actividad de proteasa. En particular las variantes según la invención tienen termoestabilidad mejorada.
- 10 [0011] **Actividad de proteasa:** El término "actividad de proteasa" significa un polipéptido de EC. 3.4.-.-. En una forma de realización preferida es un polipéptido que pertenece a las metaloendopeptidasas EC 3.4.24. La actividad de proteasa se puede determinar según el procedimiento descrito en el ejemplo 1, por ejemplo, usando un ensayo de azocaseína o ensayo de endoproteasa utilizando protazyme OL.
- 15 [0012] El sitio ENZYME en internet (<http://www.expasy.ch/enzyme/>) es un repositorio de información relativa a la nomenclatura de enzimas. Se basa principalmente en las recomendaciones del Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular (IUBBM) y describe cada tipo de enzima caracterizada para la que se proporciona un número EC (Comisión Enzimática) (Bairoch A. The ENZYME database, 2000, Nucleic Acids Res 28:304-305). Véase también el manual Enzyme Nomenclature de NC-IUBBM, 1992).
- 20 [0013] **Variante:** El término "variante" significa un polipéptido con actividad de proteasa que comprende al menos una modificación, es decir, una sustitución, inserción y/o delección, en comparación con la secuencia de aminoácidos mostrada como 1 a 177 de la SEQ ID N.º: 2. En el presente contexto, "al menos una" (por ejemplo, modificación) significa una o más, por ejemplo 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 modificaciones; o 12, 14, 15, 16, 18, 20, 22, 24, 25, 28 o 30 modificaciones; etcétera, hasta un número máximo de modificaciones de 44. Las variantes de proteasa de la invención, sin embargo, todavía deben ser al menos el 85 % idénticas a los aminoácidos 1 a 177 de la SEQ ID N.º: 2. En otra forma de realización, la variante tiene al menos el 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, o al menos el 99 % de identidad con los aminoácidos 1 a 177 de la SEQ ID N.º: 2. En otros ejemplos de realización particulares adicionales, el grado de identidad es al menos del 98,0 %, 98,2 %, 98,4 %, 98,6 %, 98,8 %, 99,0 %, 99,1 %, 99,2 %, 99,3 %, o al menos del 99,4 %, pero menos del 100 %, de identidad de secuencia con el polipéptido maduro de la SEQ ID N.º: 2.
- 30 [0014] **Mutante:** el término "mutante" significa un polinucleótido que codifica una variante.
- 35 [0015] **Enzima de tipo salvaje:** El término proteasa "de tipo salvaje" significa una proteasa expresada por un microorganismo de origen natural, tal como una bacteria, levadura u hongo filamentoso encontrado en la naturaleza. **Progenitor o proteasa progenitora:** El término "progenitor" o "proteasa progenitora" significa una proteasa a la que se hace una alteración para producir las variantes enzimáticas de la presente invención. El progenitor puede ser un polipéptido de origen natural (de tipo salvaje) o una variante del mismo.
- 40 [0016] **Variante progenitora:** El término variante progenitora significa que el punto de inicio para la fabricación de una proteasa variante fue en sí una variante en comparación con la proteasa de tipo salvaje.
- 45 [0017] **Variante aislada:** El término "variante aislada" significa una variante que se modifica por la acción del hombre. En un aspecto, la variante es al menos 1 % pura, por ejemplo, al menos 5 % pura, al menos 10 % pura, al menos 20 % pura, al menos 40 % pura, al menos 60 % pura, al menos 80 % pura, y al menos 90 % pura, como se determina por SDS-PAGE.
- 50 [0018] **Variante sustancialmente pura:** El término "variante sustancialmente pura" significa una preparación que contiene como mucho el 10 %, como mucho el 8 %, como mucho el 6 %, como mucho el 5 %, como mucho el 4 %, como mucho el 3 %, como mucho el 2 %, como mucho el 1 %, y como mucho el 0,5 % en peso de otro material de polipéptido con el cual se asocia originalmente o recombinantemente. Preferiblemente, la variante es al menos 92 % pura, por ejemplo, al menos 94 % pura, al menos 95 % pura, al menos 96 % pura, al menos 97 % pura, al menos 98 % pura, al menos 99 %, al menos 99,5 % pura, y 100 % pura en peso del material de polipéptido total presente en la preparación. Las variantes de la presente invención son preferiblemente en una forma sustancialmente pura. Esto se puede conseguir, por ejemplo, preparando la variante por métodos recombinantes bien conocidos o por métodos de purificación tradicionales.
- 55 [0019] **Polipéptido maduro:** El término "polipéptido maduro" significa un polipéptido en su forma final después de la traducción y de cualquier modificación postraduccion, tal como procesamiento del N-terminal, truncamiento del C-terminal, glicosilación, fosforilación, etc. En un aspecto, el polipéptido maduro son los aminoácidos 1 a 177 de la SEQ ID N.º: 2. La parte de péptido señal se puede predecir por programas conocidos en la técnica (SignalP (Nielsen *et al.*, 1997, Protein Engineering 10: 1-6)). Los aminoácidos 1 a 177 de SEQ ID N.º: 2 son la parte madura prevista. Generalmente, el primer aminoácido de la parte madura de una enzima se puede determinar por secuenciación del N-terminal de la enzima purificada. El N-terminal del polipéptido maduro
- 60
- 65

se confirmó mediante secuenciación del N-terminal. Cualquier diferencia entre la parte de péptido señal y la parte madura debe entonces deberse a la presencia de un propéptido.

[0020] **Secuencia codificante del polipéptido maduro:** El término "secuencia codificante del polipéptido maduro" significa un polinucleótido que codifica un polipéptido maduro con actividad de proteasa. En un aspecto, la secuencia codificante del polipéptido maduro son los nucleótidos 535 a 1065 de la SEQ ID N.º: 1 basándose en SignalP (Nielsen *et al.*, 1997, Protein Engineering 10: 1-6), que predice que los nucleótidos 535 a 1065 de la SEQ ID N.º: 1 codifican un péptido señal.

[0021] **Identidad de secuencia:** La relación entre dos secuencias de aminoácidos o entre dos secuencias de nucleótidos se describe por el parámetro "identidad de secuencia". Para los fines de la presente invención, el grado de identidad de secuencia entre dos secuencias de aminoácidos se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, J. Mol. Biol. 48: 443-453) como se implementa en el programa Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, Trends Genet. 16: 276-277), preferiblemente versión 3.0.0 o posterior. Los parámetros opcionales usados son penalización por apertura de espacio de 10, penalización por extensión de espacio de 0,5, y la matriz de sustitución EBLOSUM62 (versión EMBOSS de BLOSUM62). El resultado de Needle etiquetado como "mayor identidad" (obtenido utilizando la opción no resumida) se usa como el porcentaje de identidad y se calcula de la siguiente manera:

$$(\text{Residuos idénticos} \times 100) / (\text{Longitud del alineamiento} - \text{Número total de espacios en el alineamiento})$$

[0022] Para los fines de la presente invención, el grado de identidad de secuencia entre dos secuencias de desoxirribonucleótidos se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, *supra*) como se implementa en el programa Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *supra*), preferiblemente versión 3.0.0 o posterior. Los parámetros opcionales usados son penalización por apertura de espacio de 10, penalización por extensión de espacio de 0,5 y la matriz de sustitución EDNAFULL (versión EMBOSS de NCB1 NUC4.4). El resultado de Needle etiquetado como "mayor identidad" (obtenido utilizando la opción no resumida) se usa como el porcentaje de identidad y se calcula de la siguiente manera:

$$(\text{Desoxirribonucleótidos idénticos} \times 100) / (\text{Longitud del alineamiento} - \text{Número total de espacios en el alineamiento})$$

[0023] **Polinucleótido aislado:** El término "polinucleótido aislado" significa un polinucleótido que se modifica mediante la acción del hombre. En un aspecto, el polinucleótido aislado es al menos 1 % puro, por ejemplo, al menos 5 % puro, al menos 10 % puro, al menos 20 % puro, al menos 40 % puro, al menos 60 % puro, al menos 80 % puro, al menos 90 % puro, y al menos 95 % puro, determinado por electroforesis de agarosa. Los polinucleótidos pueden ser de origen genómico, de ADNc, de ARN, semisintético, sintético, o cualquier combinación de los mismos.

[0024] **Polinucleótido sustancialmente puro:** El término "polinucleótido sustancialmente puro" significa una preparación de polinucleótido libre de otros nucleótidos extraños o no deseados y en una forma adecuada para el uso en sistemas de producción de polipéptidos genéticamente modificados. Así, un polinucleótido sustancialmente puro contiene como mucho 10 %, como mucho 8 %, como mucho 6 %, como mucho 5 %, como mucho 4 %, como mucho 3 %, como mucho 2 %, como mucho 1 %, y como mucho 0,5 % en peso de otro material de polinucleótido con el cual se asocia originalmente o recombinantemente. Un polinucleótido sustancialmente puro puede, sin embargo, incluir regiones sin traducir 5'- y 3'- de origen natural, tales como promotores y terminadores. Se prefiere que el polinucleótido sustancialmente puro sea al menos 90 % puro, por ejemplo, al menos 92 % puro, al menos 94 % puro, al menos 95 % puro, al menos 96 % puro, al menos 97 % puro, al menos 98 % puro, al menos 99 % puro, y al menos 99,5 % puro en peso. Los polinucleótidos de la presente invención son preferiblemente en una forma sustancialmente pura.

[0025] **Secuencia codificante:** El término "secuencia codificante" significa un polinucleótido, que directamente especifica la secuencia de aminoácidos de su producto de polipéptido. Los límites de la secuencia codificante se determinan generalmente por un marco de lectura abierto, que empieza normalmente con el codón de inicio ATG o codones de inicio alternativos tales como GTG y TTG y termina con un codón de terminación tal como, TAA, TAG y TGA. La secuencia codificante puede ser un polinucleótido de ADN, ADNc, sintético, o recombinante. **Constructo de ácido nucleico:** El término "constructo de ácido nucleico" significa una molécula de ácido nucleico, ya sea mono o bicatenaria, que se aísla de un gen de origen natural o se modifica para contener segmentos de ácido nucleico de una manera que no existiría de otro modo en la naturaleza o que es sintética. El término constructo de ácido nucleico es sinónimo del término "casete de expresión" cuando el constructo de ácido nucleico contiene las secuencias de control requeridas para la expresión de una secuencia codificante de la presente invención.

[0026] **Secuencias de control:** El término "secuencias de control" significa todos los componentes necesarios para la expresión de un polinucleótido que codifica una variante de la presente invención. Cada secuencia de control puede ser nativa o foránea al polinucleótido que codifica la variante o nativas o foráneas entre sí. Tales secuencias de control incluyen, pero de forma no limitativa, un líder, una secuencia de poliadenilación, una secuencia de propéptidos, un promotor, una secuencia de péptido señal y un terminador de transcripción. Como mínimo, las secuencias de control incluyen un promotor y señales de parada de la transcripción y de la traducción. Las secuencias de control pueden proporcionarse con enlaces con el fin de introducir sitios de restricción específicos que faciliten la unión de las secuencias de control con la región codificante del polinucleótido que codifica una variante.

[0027] **Operativamente enlazado:** El término "operativamente enlazado" significa una configuración en la que una secuencia de control se coloca en una posición apropiada relativa a la secuencia de codificación de un polinucleótido de manera que la secuencia de control dirige la expresión de la secuencia codificante.

[0028] **Expresión:** El término "expresión" incluye cualquier paso implicado en la producción de la variante, entre los que se incluyen, pero de forma no limitativa, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.

[0029] **Vector de expresión:** El término "vector de expresión" significa una molécula de ADN lineal o circular que comprende un polinucleótido que codifica una variante y que está operativamente enlazada a nucleótidos adicionales que proporcionan su expresión.

[0030] **Célula huésped:** El término "célula huésped" significa cualquier tipo de célula que es susceptible de transformación, transfección, transducción y similares con un constructo de ácido nucleico o vector de expresión que comprende un polinucleótido de la presente invención. El término "célula huésped" abarca cualquier progenie de una célula progenitora que no sea idéntica a la célula progenitora debido a mutaciones que ocurren durante la replicación.

[0031] **Propiedad mejorada:** El término "propiedad mejorada" significa una característica asociada con una variante que se mejora en comparación con el progenitor. Tales propiedades mejoradas incluyen, pero de forma no limitativa, actividad térmica, termoestabilidad, actividad de pH, estabilidad de pH, especificidad de sustrato/cofactor, propiedades de superficie mejorada, especificidad de producto, estabilidad o solubilidad aumentada en presencia de biomasa pretratada, estabilidad mejorada bajo condiciones de almacenamiento y estabilidad química.

[0032] **Termoestabilidad mejorada:** El término "termoestabilidad mejorada" significa una variante que muestra retención de actividad de proteasa después de un periodo de incubación a temperatura elevada en relación con el progenitor, bien en un tampón o bien bajo condiciones tales como las que existen durante el almacenamiento/transporte del producto o condiciones similares a las que existen durante el uso industrial de la variante. Si una proteasa variante de la invención tiene una termoestabilidad mejorada o no en comparación con una proteasa progenitora se puede determinar como se describe en el ejemplo 1. La proteasa variante de la invención puede tener una termoestabilidad mejorada en comparación con una proteasa progenitora, donde la termoestabilidad mejorada se determina como actividad relativa aumentada. La proteasa variante de la invención puede tener una termoestabilidad mejorada en comparación con una proteasa progenitora, donde la termoestabilidad mejorada se determina como actividad restante aumentada.

[0033] En un aspecto, la termoestabilidad de la variante con actividad de proteasa es al menos 1,05 veces, por ejemplo, al menos 1,1 veces, al menos 1,5 veces, al menos 1,8 veces, al menos 2 veces, al menos 5 veces, al menos 10 veces, al menos 15 veces, al menos 20 veces, y al menos 25 veces más termoestable que el progenitor cuando la actividad residual se compara utilizando el ensayo para determinar la actividad restante (azocaseína) de los ejemplos.

Modificaciones, tales como sustituciones, deleciones, inserciones

[0034] Una variante de proteasa puede comprender varios tipos de modificaciones con respecto a un modelo (es decir, una proteasa progenitora o de referencia, o una secuencia de aminoácidos comparativa tal como los aminoácidos 1 a 177 de la SEQ ID N.º: 2): un aminoácido se puede sustituir con otro aminoácido; un aminoácido se puede eliminar; un aminoácido se puede insertar entre dos residuos; así como cualquier combinación de cualquier número de tales modificaciones.

[0035] Una sustitución o extensión sin ninguna indicación de qué sustituir o con qué extender se refiere a la inserción de cualquier aminoácido, natural o no natural, excepto el que ocupa esta posición en el modelo.

[0036] Una sustitución se refiere a un reemplazo de un aminoácido que ocupa una posición con un aminoácido diferente; una delección significa una eliminación de un aminoácido que ocupa una posición; y una inserción

significa añadir 1-3 aminoácidos adyacentes a un aminoácido que ocupa una posición. En el presente contexto, el término "inserción" se destina para cubrir también extensiones del N y/o C-terminal.

[0037] Para los fines de la presente invención, el polipéptido maduro comprendido en la SEQ ID N.º: 2 se usa para determinar el residuo de aminoácidos correspondiente en la proteasa variante. La secuencia de aminoácidos de la proteasa variante se alinea con el polipéptido maduro descrito en la SEQ ID N.º: 2, y basándose en el alineamiento, el número de posición del aminoácido que corresponde con cualquier residuo de aminoácidos en el polipéptido maduro descrito en la SEQ ID N.º: 2 se determina utilizando el algoritmo de Needleman-Wunsch (Needleman y Wunsch, 1970, J. Mol. Biol. 48: 443-453) como se implementa en el programa de Needle del paquete EMBOSS (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, Trends Genet. 16: 276-277), preferiblemente versión 3.0.0 o posterior. La identificación del residuo de aminoácidos correspondiente en otra proteasa se puede confirmar mediante un alineamiento de múltiples secuencias de polipéptido usando "ClustalW" (Larkin *et al.*, 2007, Bioinformatics 23: 2947-2948).

[0038] Sustituciones. Para una sustitución de aminoácidos, se usa la siguiente nomenclatura: aminoácido original, posición, aminoácido sustituido. Por consiguiente, la sustitución de treonina con alanina en la posición 226 se designa como "Thr226Ala" o "T226A". Las mutaciones múltiples se separan por marcas de barra ("/"), por ejemplo, "Gly205Arg / Ser411Phe" o "G205R / S411F" representan sustituciones en las posiciones 205 y 411 de glicina (G) con arginina (R) y de serina (S) con fenilalanina (F), respectivamente.

[0039] Delecciones. Para una delección de aminoácidos, se usa la siguiente nomenclatura: "Δ", aminoácido original, posición. Por consiguiente, la delección de glicina en la posición 195 se designa como "ΔGly195" o "ΔG195". Las delecciones múltiples se separan por marcas de barra ("/"), por ejemplo, "ΔGly195 / ΔSer411" o "ΔG195/ΔS411".

[0040] Inserciones. Para una inserción de aminoácidos, se usa la siguiente nomenclatura: aminoácido original, posición, aminoácido original, aminoácido insertado. Por consiguiente, la inserción de lisina después de glicina en la posición 195 se designa "Gly195GlyLys" o "G195GK".

[0041] Alteraciones múltiples. Las variantes que comprenden alteraciones múltiples se separan por marcas de barra ("/"), por ejemplo, "Arg170Tyr/Gly195Glu" o "R170Y/G195E" representan una sustitución de tirosina y ácido glutámico por arginina y glicina en las posiciones 170 y 195, respectivamente.

Numeración de posición

[0042] La nomenclatura que se usa aquí para definir las posiciones de los aminoácidos se basa en la secuencia de aminoácidos de la proteasa derivada de *Thermoascus aurantiacus* CGMCC N.º 0670, cuya secuencia del polipéptido maduro se da en el listado de secuencias como los aminoácidos 1 a 177 de la SEQ ID N.º: 2 (aminoácidos 1-177 de la SEQ ID N.º: 2). Por consiguiente, en el presente contexto, la base para la numeración de las posiciones es la SEQ ID N.º: 2 que empieza con T1 y termina con C177.

[0043] Cuando se usa aquí el término parte (o secuencia) "madura" se refiere a la parte del polipéptido que se segrega por una célula que contiene, como parte de su equipo genético, un polinucleótido que codifica el polipéptido. En otras palabras, la parte de polipéptido maduro se refiere a la parte del polipéptido que permanece después de que la parte de péptido señal, al igual que una parte de propéptido, si la hay, haya sido escindida. La parte de péptido señal se puede predecir mediante programas conocidos en la técnica (por ejemplo, SignalP). Los aminoácidos 1 a 177 de la SEQ ID N.º: 2 son la parte madura prevista. Generalmente, el primer aminoácido de la parte madura de una enzima se puede determinar mediante secuenciación del N-terminal de la enzima purificada. Cualquier diferencia entre la parte de péptido señal y la parte madura debe entonces deberse a la presencia de un propéptido.

Métodos de producción de una variante de proteasa

[0044] La presente invención se refiere a métodos de producción de una variante de proteasa con actividad de proteasa y una termoestabilidad mejorada en comparación con la proteasa progenitora, donde dicho método comprende el cultivo de una célula en la que se ha introducido un vector de expresión que comprende los siguientes elementos operativamente enlazados:

- (a) un promotor de transcripción,
 - (b) una molécula de polinucleótido que codifica una variante de proteasa que tiene al menos el 85 % de identidad con la proteasa original mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y que comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 en al menos una posición seleccionada de las siguientes: 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53, 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173,
- donde la molécula de polinucleótido se prepara introduciendo al menos una mutación en una molécula de ADN que codifica una proteasa, y

(c) un terminador de transcripción,

con lo cual dicha célula expresa la variante de proteasa codificada por la molécula de polinucleótido; y la recuperación de la variante de proteasa.

5

[0045] Preferiblemente la proteasa aplicada en el método de la invención que exhibe termoestabilidad mejorada comprende al menos una de las siguientes modificaciones: A27K, A27G, A27V, Q53K, Q53R, T54R, D79K, D79L, D79M, Y82F, S87G, S87P, A112P, D142L, R2P, ΔS5, C6R, ΔG8, N26R, S41R, Y43F, T46R, S49R, A73C, P81R, N88R, D104R, D104P, T114P, S115R, T116V, T124L, T124V, A126V, M152R, S157K, Q158W y I173V.

10

[0046] En un segundo aspecto, la presente invención se refiere a una variante de proteasa que tiene al menos el 85 % de identidad con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y que comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 en al menos una posición seleccionada de las siguientes: 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53, 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173, donde el polipéptido tiene termoestabilidad mejorada en comparación con la proteasa mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2. En una forma de realización más particular, las variantes de proteasa comprenden al menos una modificación en al menos una posición seleccionada del grupo que consiste en las posiciones: 27, 79, 82, 87, 104, 112, 126 y 142.

15

20

[0047] Preferiblemente la proteasa de la invención que exhibe termoestabilidad mejorada comprende al menos una de las siguientes modificaciones: 27K, A27G, A27V, Q53K, Q53R, T54R, D79K, D79L, D79M, Y82F, S87G, S87P, A112P, D142L, R2P, ΔS5, C6R, ΔG8, N26R, S41R, Y43F, T46R, S49R, A73C, P81R, N88R, D104P, D104R, T114P, S115R, T116V, T124L, T124V, A126V, M152R, S157K, Q158W y I173V.

25

[0048] Los números de posición se refieren a la numeración de posición de los aminoácidos 1 a 177 de la SEQ ID N.º: 2, como se describe en la sección "Numeración de posición". Las posiciones que corresponden a los números de posición de estos aminoácidos 1 a 177 de la SEQ ID N.º: 2 en otras proteasas se determinan como se ha descrito anteriormente.

30

[0049] La proteasa variante de la invención es una variante de la proteasa de los aminoácidos 1 a 177 de la SEQ ID N.º: 2, es decir, no es idéntica a los aminoácidos 1 a 177 de la SEQ ID N.º: 2, ya que esta comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

35

[0050] En ejemplos de realización preferidos adicionales, la proteasa comprende al menos una de las siguientes combinaciones de modificaciones:

40

ΔS5/D79L/S87P,
 ΔS5/D79L/S87P/A112P/D142L,
 ΔS5/N26R/D79L/S87P/A112P/D142L,
 C6R/D79L/S87P,
 ΔG8/D79L/S87P,
 N26R/D79L/S87P,
 N26R/T46R/D79L/S87P/A112P/D142L,
 A27G/D79L/S87P/A112P/D142L,
 45 A27K/D79L/S87P/A112P/D142L,
 A27K/D79L/S87P/A112P/T124V/D142L,
 A27V/D79L/S87P/A112P/D142L,
 S41R/D79L/S87P,
 S41R/D79L/S87P/A112P/D142L,
 50 S41R/D79L/S87P/A112P/D142L/S157K,
 Y43F/D79L/S87P/A112P/D142L,
 T46R/D79L/S87P,
 T46R/D79L/S87P/A112P/D142L,
 T46R/D79L/S87P/T116V/D142L,
 55 S49R/D79L/S87P,
 Q53K/D79L/S87P/I173V,
 Q53R/D79L/S87P,
 T54R/D79L/S87P,
 D79L/S87P/A112P,
 60 D79L/P81R/S87P/A112P/D142L,
 D79L/S87P,
 D79L/S87P/A112P,
 D79L/S87P/A112P/D142L,
 D79L/S87P/A112P/D142L/S157K,
 65 D79L/S87P/A112P/T124L/D142L,
 D79L/S87P/A112P/T124V/A126V/D142L,

5 D79L/S87P/A112P/T124V/D142L,
 D79L/S87P/D104R,
 D79L/S87P/D142L,
 D79L/S87P/N88R,
 D79L/S87P/Q158W,
 D79L/S87P/S115R,
 D79L/S87P/S157K,
 D79L/S87P/T114P,
 D79L/S87P/T116V/,
 10 D79L/Y82F/S87P/A112P/T124V/D142L,
 D79L/Y82F/S87P/A112P/T124V/D142L,
 A27K/D79L/Y82F/S87G/D104P/A112P/A126V/D142L,
 A27K/Y82F/S87G/D104P/A112P/A126V/D142L,
 A27K/D79L/Y82F/D104P/A112P/A126V/D142L,
 15 A27K/Y82F/D104P/A112P/A126V/D142L.

Estrategia para la preparación de variantes

20 [0051] Un modelo de homología de la estructura de la SEQ ID N.º: 2 se construyó usando EB6 de la entrada 1 de PDB como un modelo. El modelo se construyó utilizando el programa Yasara (YASARA Biosciences, Neue-Welt-Hoehe 13/b, 8042 Graz, Austria / Europa).

25 [0052] El modelo se sometió a simulaciones de dinámica molecular (MD), cálculos electrostáticos, reconstrucción de secuencia ancestral, cálculo de secuencia de consenso y predicción de sitio de autoescisión. Basándose en la estructura modelada y los resultados de simulación, se sugirieron modificaciones con énfasis particular en la mejora de las propiedades de termoestabilidad.

30 [0053] Basándose en el modelo estructural, se identificaron posiciones cercanas al ion Zn del sitio activo. D142 es un ejemplo de un residuo cercano al sitio activo. Asimismo, se identificaron posiciones expuestas a la superficie. Las posiciones en la superficie interactúan con el agua y son importantes para la estabilidad. Las posiciones 2, 5, 6, 8, 26, 27, 41, 49, 53, 73, 79, 87, 88, 104, 112, 114, 115, 116, 157, 158 están expuestas a la superficie. Las más importantes de estas son 112, 79 y 87.

35 [0054] Se realizaron simulaciones de dinámica molecular utilizando la versión 4 del programa GROMACS (<http://gromacs.org>. Hess, *et al.* (2008) J. Chem. Theory Comput. 4: 435-447). Las simulaciones se hicieron a 300 K, 400 K, 500 K en 3 repeticiones (utilizando diferentes semillas aleatorias). Basándose en las simulaciones, las posiciones que muestran alta movilidad (fluctuaciones en comparación con la estructura media) o las posiciones que muestran gran desplazamiento (en la estructura media en comparación con el modelo de inicio) se seleccionaron como candidatas para la estabilización, ya sea mediante la mutación de las mismas directamente
 40 o por mutación de las posiciones que interactúan con las mismas. Las posiciones con alto desplazamiento incluyen las posiciones en la región 144-163, en particular la posición 152. Las posiciones con alta movilidad incluyen las posiciones en las regiones 5-10, 86-89 y 109-120.

45 [0055] Basándose en el modelo, se calculó el potencial electrostático en la proteína y en su alrededor y las posiciones con aminoácidos cargados con un potencial electrostático desfavorable (por ejemplo, un aminoácido cargado negativamente con un potencial electrostático negativo) y las posiciones vecinas se marcaron como objetivos para las mutaciones. Las posiciones con aminoácidos cargados en posiciones desfavorables incluyen: 2, 19, 42, 52, 58, 60, 64, 80, 104, 121, 143 y en particular 142 y 79.

50 [0056] Secuencias homólogas conocidas se alinearon simultáneamente utilizando el programa MUSCLE (Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Research 32(5), 1792-97.). Basándose en el alineamiento de múltiples secuencias, se realizaron diferentes tipos de análisis basados en consenso. El programa PAUP se usó para encontrar posiciones que difieren del gen ancestral inferido por PAUP. Se identificaron las posiciones A37, A73, S86, F106, T108, A126, A145, M152,
 55 G153, S156, V160, M161 y T165 donde la SEQ ID N.º: 2 difiere del consenso basado en el alineamiento. La reconstrucción ancestral apuntó a las posiciones 73, 86, 126, 152, 153, 156, 38, 50, 94, 109, 135, 136, 142 y en particular a las posiciones 142, 73, 126, 152.

60 [0057] La pérdida de actividad de la proteasa se puede relacionar con la autoproteólisis. Basándose en el análisis de espectrometría de masa de proteína autodigerida de la SEQ ID N.º: 2 combinado con los datos de especificidad de enzimas homólogas, se sugirieron un número de posiciones: 59, 100, 105, 130, 131, 148 y en particular la posición 27.

65 [0058] Se conoce que las prolinas afectan la estabilidad y, utilizando la estructura modelo, se identificaron posiciones compatibles con la sustitución de una prolina utilizando la función SUGPRO del programa WHATIF (<http://www.cmbi.kun.nl/whatif>. WHAT IF Foundation / CMBI, Toernooiveld 1, 6525 ED Nijmegen, Países Bajos).

[0059] Puentes de cisteína pueden estabilizar la estructura y, aplicando la función SUGCYS del programa WHATIF en la estructura modelo, se identificaron posiciones comparables con sustituciones de cisteína.

- 5 [0060] Las variantes de proteasa correspondientes se prepararon por métodos conocidos en la técnica y se evaluaron como se describe en la parte experimental.

Termoestabilidad

- 10 [0061] Si una proteasa variante de la invención tiene una termoestabilidad mejorada o no en comparación con una proteasa progenitora se puede determinar como se describe en el ejemplo 1. La proteasa variante de la invención puede tener una termoestabilidad mejorada en comparación con una proteasa progenitora, donde la termoestabilidad mejorada se determina como actividad relativa aumentada. La proteasa variante de la invención puede tener una termoestabilidad mejorada en comparación con una proteasa progenitora, donde la termoestabilidad mejorada se determina como actividad restante aumentada.

[0062] La termoestabilidad también se puede determinar usando mediciones de DSC para determinar la temperatura de desnaturalización, Td, de la proteína de proteasa purificada. La Td es indicativa de la termoestabilidad de la proteína: a mayor Td, mayor termoestabilidad. Por consiguiente, en una forma de realización preferida, la proteasa de la invención tiene una Td que es superior a la Td de una proteasa de referencia, donde la Td se determina en las muestras de proteasa purificada después del intercambio de tampón a 20 mM de Na-acetato a pH 4,5 o 5,5 con o sin 2,5 mM de Zn²⁺ usando un dispositivo de filtrado centrífugo (10.000 MWCO) (usando calorimetría diferencial de barrido a un ritmo de barrido de 90 °C/h de 20-110 °C (tenemos unos pocos barridos hasta 120 °C) en 20 mM de tampón de Na-acetato.

20 En una forma de realización preferida, la Td de la proteasa de la invención es superior a la Td de la proteasa de los aminoácidos 1 a 177 de la SEQ ID N.º: 2, más preferiblemente al menos el 101 % de la misma, o al menos el 102 %, 103 %, 104 %, 105 %, 106 %, 107 %, 108 %, 109 %, o al menos el 110 % de la misma. Aún más preferiblemente, la Td de la proteasa de la invención es al menos el 120 %, 130 %, 140 %, 150 %, 160 %, 170 %, 180 %, o al menos el 190 % de la Td de la proteasa de los aminoácidos 1 a 177 de la SEQ ID N.º: 2. En otros ejemplos de realización particulares adicionales, la proteasa termoestable de la invención tiene una temperatura de fusión, Tm (o una temperatura de desnaturalización, Td), determinada usando calorimetría diferencial de barrido (DSC) como se describe en los ejemplos (es decir, en 20 mM de acetato sódico), de al menos 50 °C. En otros ejemplos de realización particulares adicionales, la Tm es al menos 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 o al menos 100 °C.

35 En formas de realización particulares, la invención se refiere a variantes de termoestabilidad mejorada de la proteasa progenitora de los aminoácidos 1 a 177 de la SEQ ID N.º: 2 que comprenden al menos una de las siguientes modificaciones: A27K, A27G, A27V, Q53K, Q53R, T54R, D79K, D79L, D79M, Y82F, S87P, S87G, A112P, D142L, R2P, ΔS5, C6R, ΔG8, N26R, S41R, Y43F, T46R, S49R, A73C, P81R, N88R, D104R, D104P, T114P, S115R, T116V, T124L, T124V, A126V, M152R, S157K, Q158W y I173V. En una forma de realización más particular, las modificaciones se seleccionan del grupo que consiste en: A27K, D79L, Y82F, S87G, S87P, D104P, A112P, A126V y D142L.

45 [0063] En formas de realización particulares, la invención se refiere a variantes con termoestabilidad mejorada de la proteasa progenitora de los aminoácidos 1 a 177 de la SEQ ID N.º: 2 que comprenden modificaciones seleccionadas de: ΔS5/D79L/S87P, ΔS5/D79L/S87P/A112P/D142L, ΔS5/N26R/D79L/S87P/A112P/D142L, C6R/D79L/S87P, ΔG8/D79L/S87P, N26R/D79L/S87P, N26R/T46R/D79L/S87P/A112P/D142L, A27G/D79L/S87P/A112P/D142L, A27K/D79L/S87P/A112P/D142L, A27K/D79L/S87P/A112P/T124V/D142L, A27V/D79L/S87P/A112P/D142L, S41R/D79L/S87P, S41R/D79L/S87P/A112P/D142L, S41R/D79L/S87P/A112P/D142L/S157K, Y43F/D79L/S87P/A112P/D142L, T46R/D79L/S87P, T46R/D79L/S87P/A112P/D142L, T46R/D79L/S87P/T116V/D142L, S49R/D79L/S87P, Q53K/D79L/S87P/I173V, Q53R/D79L/S87P, T54R/D79L/S87P, D79L/S87P/A112P, D79L/P81R/S87P/A112P/D142L, D79L/S87P, D79L/S87P/A112P, D79L/S87P/A112P/D142L, D79L/S87P/A112P/D142L/S157K, D79L/S87P/A112P/T124L/D142L, D79L/S87P/A112P/T124V/A126V/D142L, D79L/S87P/A112P/T124V/D142L, D79L/S87P/D104R, D79L/S87P/D142L, D79L/S87P/N88R, D79L/S87P/Q158W, D79L/S87P/S115R, D79L/S87P/S157K, D79L/S87P/T114P, D79L/S87P/T116V, D79L/Y82F/S87P/A112P/T124V/D142L, D79L/Y82F/S87P/A112P/T124V/D142L, A27K/D79L/Y82F/S87G/D104P/A112P/A126V/D142L, A27K/Y82F/S87G/D104P/A112P/A126V/D142L, A27K/D79L/Y82F/D104P/A112P/A126V/D142L, A27K/Y82F/D104P/A112P/A126V/D142L.

60 En una forma de realización particular, la invención se refiere a una variante con termoestabilidad mejorada de la proteasa progenitora de los aminoácidos 1 a 177 de la SEQ ID N.º: 2 que comprende las modificaciones: A27K/D79L/Y82F/S87G/D104P/A112P/A126V/D142L. En otra forma de realización particular, la invención se refiere a una variante con termoestabilidad mejorada de la proteasa progenitora de los aminoácidos 1 a 177 de la SEQ ID N.º: 2 que comprende las modificaciones: A27K/Y82F/S87G/D104P/A112P/A126V/D142L. En todavía otra forma de realización particular, la invención se refiere a una variante con termoestabilidad mejorada de la proteasa progenitora de los aminoácidos 1 a 177 de la SEQ ID N.º: 2 que comprende las modificaciones:

A27K/D79L/Y82F/D104P/A112P/A126V/D142L. En una otra forma de realización particular adicional, la invención se refiere a una variante con termoestabilidad mejorada de la proteasa progenitora de los aminoácidos 1 a 177 de la SEQ ID N.º: 2 que comprende las modificaciones: A27K/Y82F/D104P/A112P/A126V/D142L.

5 Secuencias y constructos de ácido nucleico

[0064] La presente invención también se refiere a secuencias de ácido nucleico que comprenden una secuencia de ácido nucleico que codifica una variante de proteasa de la invención.

10 [0065] El término "secuencia de ácido nucleico aislada" se refiere a una secuencia de ácido nucleico que está libre esencialmente de otras secuencias de ácido nucleico, por ejemplo, al menos aproximadamente 20 % pura, preferiblemente al menos aproximadamente 40 % pura, más preferiblemente al menos aproximadamente 60 % pura, aún más preferiblemente al menos aproximadamente 80 % pura, y de la forma más preferible al menos aproximadamente 90 % pura determinado por electroforesis de agarosa. Por ejemplo, una secuencia de ácido nucleico aislada se puede obtener por procedimientos de clonación estándar usados en ingeniería genética para
15 recolocar la secuencia de ácido nucleico de su ubicación natural a un sitio diferente donde se reproducirá. Los procedimientos de clonación pueden implicar escisión y aislamiento de un fragmento de ácido nucleico deseado que comprende la secuencia de ácido nucleico que codifica el polipéptido, la inserción del fragmento en una molécula del vector y la incorporación del vector recombinante en una célula huésped donde se replicarán las
20 múltiples copias o clones de la secuencia de ácido nucleico. La secuencia de ácido nucleico puede ser de origen genómico, de ADNc, de ARN, semisintético, sintético, o cualquier combinación de los mismos.

[0066] Las secuencias de ácido nucleico de la invención se pueden preparar introduciendo al menos una mutación en una secuencia codificante de proteasa modelo o una subsecuencia de la misma, donde la secuencia
25 de ácido nucleico mutante codifica una proteasa variante. La introducción de una mutación en la secuencia de ácido nucleico para intercambiar un nucleótido por otro nucleótido se puede realizar por cualquiera de los métodos conocidos en la técnica, por ejemplo, por mutagénesis dirigida al sitio, por mutagénesis aleatoria, o por mutagénesis dopada, adicionada o aleatoria localizada.

30 [0067] La mutagénesis aleatoria se realiza adecuadamente bien como mutagénesis aleatoria localizada o como específica de la región en al menos tres partes del gen que se traduce a la secuencia de aminoácidos mostrada en cuestión, o en el gen entero. Cuando la mutagénesis se realiza mediante el uso de un oligonucleótido, el oligonucleótido se puede dopar o adicionar con los tres nucleótidos no progenitores durante la síntesis del oligonucleótido en las posiciones que se van a cambiar. El dopaje o la adición se pueden realizar de modo que
35 se eviten los codones de aminoácidos no deseados. El oligonucleótido dopado o adicionado se puede incorporar en el ADN que codifica la enzima proteásica mediante cualquier técnica, usando, por ejemplo, PCR, LCR o cualquier polimerasa y ligasa de ADN que se considere apropiada.

[0068] Preferiblemente, el dopaje se realiza usando "dopaje aleatorio constante", donde el porcentaje de tipo salvaje y de mutación en cada posición está predefinido. Además, el dopaje se puede dirigir hacia una preferencia para la introducción de ciertos nucleótidos y, por lo tanto, una preferencia para la introducción de uno o más residuos de aminoácidos específicos. El dopaje se puede hacer, por ejemplo, para permitir la introducción del 90 % de tipo salvaje y el 10 % de mutaciones en cada posición. Una consideración adicional en la elección de un esquema de dopaje se basa en limitaciones genéticas y de estructuras de proteínas.
45

[0069] La mutagénesis aleatoria se puede localizar ventajosamente en una parte de la proteasa progenitora en cuestión. Esto puede, por ejemplo, ser ventajoso cuando ciertas regiones de la enzima se han identificado como que son de importancia particular para una propiedad determinada de la enzima.

50 [0070] Métodos alternativos para proporcionar variantes de la invención incluyen la transposición de genes, por ejemplo, como se describe en WO 95/22625 o en WO 96/00343, y el proceso de derivación de consenso como se describe en EP 897985.

55 Constructos de ácido nucleico

[0071] Un constructo de ácido nucleico comprende una secuencia de ácido nucleico de la presente invención operativamente enlazada a una o más secuencias de control que dirigen la expresión de la secuencia codificante en una célula huésped adecuada bajo condiciones compatibles con las secuencias de control. La expresión se entenderá que incluye cualquier paso implicado en la producción del polipéptido, entre los que se incluyen, pero de forma no limitativa, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.
60

[0072] El término "constructo de ácido nucleico" como se utiliza en este caso se refiere a una molécula de ácido nucleico, bien mono o bicatenaria, que se aísla a partir de un gen de origen natural o que se modifica para contener segmentos de ácido nucleico de una manera que no existiría de otro modo en la naturaleza. El término constructo de ácido nucleico es sinónimo del término "casete de expresión" cuando el constructo de ácido
65

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

5 [0073] El término "secuencias de control" se define aquí para incluir todos los componentes, que son necesarios o ventajosos para la expresión de un polinucleótido que codifica un polipéptido de la presente invención. Cada
 10 secuencia de control puede ser nativa o foránea a la secuencia de nucleótido que codifica el polipéptido. Tales secuencias de control incluyen, pero de forma no limitativa, un líder, una secuencia de poliadenilación, una secuencia de propéptidos, un promotor de transcripción, una secuencia de péptido señal y un terminador de transcripción. Como mínimo, las secuencias de control incluyen un promotor de transcripción y señales de
 15 parada de la transcripción y de la traducción. Las secuencias de control pueden proporcionarse con enlaces con el fin de introducir sitios de restricción específicos que faciliten la unión de las secuencias de control con la región codificante de la secuencia de nucleótido que codifica un polipéptido.

15 [0074] El término "enlazado operativamente" denota aquí una configuración donde una secuencia de control se coloca en una posición apropiada en relación con la secuencia codificante de la secuencia de polinucleótido de manera que la secuencia de control dirige la expresión de la secuencia codificante de un polipéptido.

20 [0075] Cuando se usa aquí, el término "secuencia codificante" (CDS) significa una secuencia de nucleótido, que especifica directamente la secuencia de aminoácidos de su producto proteínico. Los límites de la secuencia codificante se determinan generalmente por un marco de lectura abierto, que empieza normalmente con el codón de inicio ATG o codones de inicio alternativos tales como GTG y TTG. La secuencia codificante puede ser una secuencia de ADN, de ADNc o de nucleótido recombinante.

25 Vector de expresión

25 [0076] El término "expresión" incluye cualquier paso implicado en la producción del polipéptido entre los que se incluyen, pero de forma no limitativa, transcripción, modificación postranscripcional, traducción, modificación postraduccional y secreción.

30 [0077] El término "vector de expresión" se define aquí como una molécula de ADN lineal o circular que comprende un polinucleótido que codifica un polipéptido de la invención, y que está operativamente enlazada a nucleótidos adicionales que proporcionan su expresión.

35 [0078] Una secuencia de ácido nucleico que codifica una variante de proteasa de la invención se puede expresar utilizando un vector de expresión que incluye típicamente secuencias de control que codifican un promotor, operador, sitio de unión al ribosoma, señal de inicio de la traducción y, opcionalmente, un gen represor o varios genes activadores.

40 [0079] El vector de expresión recombinante que lleva la secuencia de ADN que codifica una variante de proteasa de la invención puede ser cualquier vector que pueda someterse convenientemente a procedimientos de ADN recombinante, y la elección del vector dependerá frecuentemente de la célula huésped en la que se vaya a introducir. El vector puede ser uno que, cuando se introduce en una célula huésped, se integra en el genoma de la célula huésped y se replica con el cromosoma (los cromosomas) donde se ha integrado.

45 [0080] La variante de proteasa también se puede coexpresar junto con al menos una otra enzima de interés, tal como una glucoamilasa, alfa-amilasa, fosfatasa, xilanasa, galactanasa, alfa-galactosidasa, proteasa, fosfolipasa y/o beta-glucanasa. Las enzimas se pueden coexpresar a partir de distintos vectores, de un vector, o utilizando una mezcla de ambas técnicas. Cuando se usan vectores diferentes, los vectores pueden tener diferentes
 50 marcadores seleccionables y diferentes orígenes de replicación. Cuando se usa solo un vector, los genes se pueden expresar a partir de uno o más promotores. Si se clonan bajo la regulación de un promotor (bi o multicistónico), el orden en el que se clonan los genes puede afectar a los niveles de expresión de las proteínas. La variante de proteasa también se puede expresar como una proteína de fusión, es decir, que el gen que codifica la variante de proteasa se ha fusionado en el marco al gen que codifica otra proteína. Esta proteína puede ser otra enzima o un dominio funcional de otra enzima.

55 Células huésped

60 [0081] El término "célula huésped", como se utiliza aquí, incluye cualquier tipo de célula que sea susceptible de transformación, transfección, transducción y similares con un constructo de ácido nucleico que comprende un polinucleótido de la presente invención.

65 [0082] La presente invención también se refiere a células huésped recombinantes, que comprenden un polinucleótido de la presente invención, que se usan ventajosamente en la producción recombinante de los polipéptidos. Un vector que comprende un polinucleótido de la presente invención se introduce en una célula huésped de modo que el vector se mantiene como un integrante cromosómico o como un vector extracromosómico autorreplicante como se ha descrito anteriormente. El término "célula huésped" abarca

cualquier progenie de una célula progenitora que no sea idéntica a la célula progenitora debido a mutaciones que ocurren durante la replicación. La elección de una célula huésped dependerá en gran medida del gen que codifica el polipéptido y su fuente.

5 [0083] La célula huésped puede ser un microorganismo unicelular, por ejemplo, una procarionta, o un microorganismo no unicelular, por ejemplo, una eucariota.

[0084] Los microorganismos unicelulares útiles son células bacterianas tales como bacterias grampositivas que incluyen, pero de forma no limitativa, una célula de *Bacillus*, por ejemplo, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus stearothermophilus*, *Bacillus subtilis* y *Bacillus thuringiensis*; o una célula de *Streptomyces*, por ejemplo, *Streptomyces lividans* y *Streptomyces murinus*, o bacterias gramnegativas tales como *E. coli* y *Pseudomonas sp.* En un aspecto preferido, la célula huésped bacteriana es una célula de *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* o *Bacillus subtilis*.
10
15 En otro aspecto preferido, la célula de *Bacillus* es un *Bacillus* alcalófilico.

[0085] La introducción de un vector en una célula huésped bacteriana se puede efectuar, por ejemplo, por transformación de protoplastos (véase, por ejemplo, Chang y Cohen, 1979, Molecular General Genetics 168: 111-115), usando células competentes (véase, por ejemplo, Young y Spizizin, 1961, Journal of Bacteriology 81: 823-829, o Dubnau y Davidoff-Abelson, 1971, Journal of Bacteriology 56: 209-221), electroporación (véase, por ejemplo, Shigekawa y Dower, 1988, Biotechniques 6: 742-751), o conjugación (véase, por ejemplo, Koehler y Thorne, 1987, Journal of Bacteriology 169: 5771-5278).
20

[0086] La célula huésped también puede ser una eucariota, tal como una célula de mamífero, de insecto, vegetal o fúngica.
25

[0087] En un aspecto preferido, la célula huésped es una célula fúngica. "Hongos" como se utiliza aquí incluye los filos *Ascomycota*, *Basidiomycota*, *Chytridiomycota* y *Zygomycota* (como se definen por Hawkswort *et al.*, en Ainsworth and Bisby's Dictionary of The Fungi, 8ª edición, 1995, CAB International, University Press, Cambridge, Reino Unido) al igual que el *Oomycota* (como se cita en Hawkswort *et al.*, 1995, *supra*, página 171) y todos los hongos mitospóricos (Hawkswort *et al.*, 1995, *supra*).
30

[0088] En un aspecto más preferido, la célula huésped fúngica es una célula de levadura. "Levadura" como se utiliza aquí incluye levadura ascoesporógena (*Endomycetales*), levadura basidioesporogénea y levadura de los hongos imperfectos (*Blastomycetes*). Dado que la clasificación de levadura puede cambiar en el futuro, para los fines de esta invención, la levadura se definirá como se describe en Biology and Activities of Yeast (Skinner, F.A., Passmore, S.M., y Davenport, R.R., eds, Soc. App. Bacteriol. Symposium Series No. 9, 1980).
35

[0089] En un aspecto más preferido, la célula huésped de levadura es una célula de *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces* o *Yarrowia*.
40

[0090] En un aspecto más preferido, la célula huésped de levadura es una célula de *Pichia pastoris*, *Pichia methanolica*, *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis* o *Saccharomyces oviformis*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Kluyveromyces lactis*. En otro aspecto más preferido, la célula huésped de levadura es una célula de *Yarrowia lipolytica*.
45

[0091] En otro aspecto más preferido, la célula huésped fúngica es una célula fúngica filamentosa. "Hongos filamentosos" incluyen todas las formas filamentosas de la subdivisión *Eumycota* y *Oomycota* (como se define por Hawkswort *et al.*, 1995, *supra*). Los hongos filamentosos se caracterizan generalmente por una pared micelial compuesta por quitina, celulosa, glucano, quitosano, manano y otros polisacáridos complejos. El crecimiento vegetativo es por alargamiento hifal y el catabolismo de carbono es estrictamente aeróbico. En cambio, el crecimiento vegetativo por levaduras tales como *Saccharomyces cerevisiae* es por injerto de un talo unicelular y el catabolismo de carbono puede ser fermentativo.
50
55

[0092] En un aspecto más preferido, la célula huésped fúngica filamentosa es una célula de *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filobasidium*, *Fusarium*, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomices*, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromices*, *Pleurotus*, *Schizofilum*, *Talaromyces*, *Teramoascus*, *Thielavia*, *Tolipocladium*, *Trametes* o *Trichoderma*.
60

[0093] En un aspecto más preferido, la célula huésped fúngica filamentosa es una célula de *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger* o *Aspergillus oryzae*. En otro aspecto más preferido, la célula huésped fúngica filamentosa es una célula de *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium*
65

reticulatum, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochrom*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides* o *Fusarium venenatum*. En otro aspecto más preferido, la célula huésped fúngica filamentosa es una célula de cepa de *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa* o *Ceriporiopsis subvermisporea*, *Coprinus cinereus*, *Coriolus hirsutus*, *Humicola insolens*, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*, *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei* o *Trichoderma viride*.

[0094] Las células fúngicas se pueden transformar mediante un proceso que implica la formación de protoplastos, la transformación de los protoplastos y la regeneración de la pared celular de una manera conocida *per se*. Los procedimientos adecuados para la transformación de células huésped de *Aspergillus* y *Trichoderma* se describen en EP 238 023 y Yelton *et al.*, 1984, Proc. Nat. Acad. Sci. USA 81: 1470-1474. Los métodos adecuados para la transformación de especies de *Fusarium* se describen por Malardier *et al.*, 1989, Gene 78: 147-156, y WO 96/00787. La levadura se puede transformar utilizando los procedimientos descritos por Becker y Guarente, en Abelson, J.N. y Simon, M.I., editores, Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Vol.194, págs. 182-187, Academic Press, Inc., Nueva York; Ito *et al.*, 1983, Journal of Bacteriology 153: 163; y Hinnen *et al.*, 1978, Proc. Nat. Acad. Sci. USA 75: 1920.

Métodos de producción

[0095] La presente invención también se refiere a métodos para la producción de una proteasa de la presente invención que comprenden (a) el cultivo de una célula huésped; y (b) la recuperación de la proteasa.

[0096] En los métodos de producción de la presente invención, las células se cultivan en un medio nutritivo adecuado para la producción del polipéptido utilizando métodos bien conocidos en la técnica. Por ejemplo, la célula se puede cultivar mediante cultivo en matraz de agitación y fermentación a pequeña escala o a gran escala (que incluyen fermentaciones continuas, por lotes, por lote alimentado o de estado sólido) en el laboratorio o fermentadores industriales realizadas en un medio adecuado y bajo condiciones que permitan que el polipéptido sea expresado y/o aislado. El cultivo tiene lugar en un medio nutritivo adecuado que comprende fuentes de nitrógeno y carbono y sales inorgánicas, usando procedimientos conocidos en la técnica. Medios adecuados están disponibles de proveedores comerciales o se pueden preparar según composiciones publicadas (por ejemplo, en catálogos de la American Type Culture Collection). Si el polipéptido se segrega en el medio nutritivo, el polipéptido se puede recuperar directamente a partir del medio. Si el polipéptido no se segrega, se puede recuperar a partir de lisados celulares.

[0097] El polipéptido resultante se puede recuperar utilizando métodos conocidos en la técnica. Por ejemplo, el polipéptido se puede recuperar a partir del medio nutritivo por procedimientos convencionales que incluyen, pero de forma no limitativa, centrifugación, filtración, extracción, secado por pulverización, evaporación o precipitación.

[0098] Los polipéptidos de la presente invención se pueden purificar mediante una variedad de procedimientos conocidos en la técnica que incluyen, pero de forma no limitativa, cromatografía (por ejemplo, de intercambio iónico, de afinidad, hidrofóbica, de cromatofoco y de exclusión de tamaño), procedimientos electroforéticos (por ejemplo, isoelectroenfoque preparativo), solubilidad diferencial (por ejemplo, precipitación de sulfato amónico), SDS-PAGE, o extracción (véase, por ejemplo, Protein Purification, J.-C. Janson y Lars Ryden, Eds, VCH Publishers, Nueva York, 1989).

Plantas transgénicas

[0099] La presente invención también se refiere a una planta transgénica, parte de planta o célula vegetal que haya sido transformada con una secuencia de nucleótido que codifica un polipéptido con actividad de proteasa de la presente invención para expresar y producir el polipéptido en cantidades recuperables. El polipéptido se puede recuperar a partir de la planta o la parte de planta. Alternativamente, la planta o parte de planta con el polipéptido recombinante se puede utilizar como tal para mejorar la calidad de un alimento o pienso, por ejemplo, mejorando el valor nutricional, la palatabilidad y las propiedades reológicas, o para destruir un factor antinutritivo.

[0100] En una forma de realización particular, el polipéptido se dirige a las vacuolas de almacenamiento de endospermo en semillas. Esto se puede obtener sintetizándolo como un precursor con un péptido señal adecuado, véase Horvath *et al.* en PNAS, Feb. 15, 2000, vol. 97, nº 4, págs. 1914-1919.

[0101] La planta transgénica puede ser dicotiledónea (una dicotiledónea) o monocotiledónea (una monocotiledónea) o variantes diseñadas de las mismas. Ejemplos de plantas monocotiledóneas son hierbas, tales como poa de prados (pasto azul, Poa), hierba forrajera tal como *Festuca*, *Lolium*, césped templado, tal como *Agrostis*, y cereales, por ejemplo, trigo, avena, centeno, cebada, arroz, sorgo, triticale (híbrido estabilizado de trigo (*Triticum*) y centeno (*Secale*), y maíz. Ejemplos de plantas dicotiledóneas son tabaco, leguminosas, tales

5 como girasol (*Helianthus*), algodón (*Gossypium*), altramuces, patata, remolacha azucarera, guisante, judía y soja, y plantas crucíferas (familia *Brassicaceae*), tales como coliflor, semilla de colza y el organismo de modelo cercanamente relacionado *Arabidopsis thaliana*. Las plantas con bajo contenido de fitato como se describen, por ejemplo, en la patente de EE.UU. N.º 5.689.054 y patente de EE.UU. N.º 6.111.168 son ejemplos de plantas manipuladas.

10 [0102] Ejemplos de partes de planta son tallo, callo, hojas, raíz, frutos, semillas y tubérculos, así como los tejidos individuales que comprenden estas partes, por ejemplo, epidermis, mesófilo, parénquima, tejidos vasculares, meristemos. Los compartimentos de células vegetales específicos, tales como cloroplasto, apoplasto, mitocondria, vacuola, peroxisomas y citoplasma también se consideran que son una parte de planta. Además, cualquier célula vegetal, sea cual sea el origen del tejido, se considera que es una parte de planta. Asimismo, partes de planta tales como tejidos específicos y células aislada para facilitar la utilización de la invención también se consideran partes de planta, por ejemplo, embriones, endospermas, aleurona y revestimientos de semillas.

15 [0103] También se incluye en el campo de la presente invención la progenie de tales plantas, partes de planta y células vegetales.

20 [0104] La planta transgénica o célula vegetal que expresa un polipéptido de la presente invención se puede construir conforme a métodos conocidos en la técnica. En resumen, la planta o célula vegetal se construye incorporando uno o más constructos de expresión que codifican un polipéptido de la presente invención en el genoma huésped vegetal y propagando la planta o célula vegetal modificada resultante en una planta o célula vegetal transgénica.

25 [0105] Convenientemente, el constructo de expresión es un constructo de ácido nucleico que comprende una secuencia de ácido nucleico que codifica un polipéptido de la presente invención operativamente enlazado con secuencias reguladoras apropiadas requeridas para la expresión de la secuencia de ácido nucleico en la planta o parte de planta de la elección. Además, el constructo de expresión puede comprender un marcador seleccionable útil para la identificación de células huésped en las que el constructo de expresión haya sido integrado y secuencias de ADN necesarias para la introducción del constructo en la planta en cuestión (lo último depende del método de introducción de ADN que se vaya a usar).

35 [0106] La elección de secuencias reguladoras, tales como secuencias promotoras y terminadoras y opcionalmente secuencias señal o de tránsito se determinan, por ejemplo, basándose en cuándo, dónde y cómo se desea que se exprese el polipéptido. Por ejemplo, la expresión del gen que codifica un polipéptido de la presente invención puede ser constitutiva o inducible, o puede ser de desarrollo, de fase o tejido específico, y el producto génico puede dirigirse a un compartimento celular, tejido o parte de planta específico tal como las semillas u hojas. Las secuencias reguladoras se describen, por ejemplo, por Tague *et al.*, 1988, *Plant Physiology* 86: 506.

40 [0107] Para la expresión constitutiva, se pueden utilizar los siguientes promotores: el promotor 35S-CaMV (Franck *et al.*, 1980, *Cell* 21: 285-294), la ubiquitina 1 de maíz (Christensen AH, Sharrock RA y Quail 1992. *Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation*), o el promotor de actina 1 de arroz (Plant. Mol. Biol. 18,675-689.; Zhang W, McElroy D. y Wu R 1991, *Analysis of rice Act1 5' region activity in transgenic rice plants*. *Plant Cell* 3, 1155-1165). Los promotores específicos de un órgano pueden ser, por ejemplo, un promotor de tejidos sumidero de almacenamiento tales como semillas, tubérculos de patata, y frutos (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), o de tejidos sumidero metabólicos tales como meristemos (Ito *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), un promotor específico de semilla como la glutelina, prolamina, globulina, o promotor de albúmina de arroz (Wu *et al.*, 1998, *Plant and Cell Physiology* 39: 885-889), un promotor de *Vicia faba* de la legúmina B4 y el gen de proteína de semilla desconocida de *Vicia faba* (Conrad *et al.*, 1998, *Journal of Plant Physiology* 152: 708-711), un promotor de una proteína del cuerpo oleaginoso de semilla (Chen *et al.*, 1998, *Plant and Cell Physiology* 39: 935-941), el promotor napA de proteína de almacenamiento de *Brassica napus*, o cualquier otro promotor específico de semilla conocido en la técnica, por ejemplo, como se describe en WO 91/14772. Además, el promotor puede ser un promotor específico de hoja tal como el promotor rbcS de arroz o tomate (Kozuka *et al.*, 1993, *Plant Physiology* 102: 991-1000, el promotor de gen de adenina metiltransferasa del virus de chlorella (Mitra y Higgins, 1994, *Plant Molecular Biology* 26: 85-93), o el promotor aldP del gen de arroz (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), o un promotor inducible por herida como el promotor pin2 de patata (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588). Asimismo, el promotor puede ser inducible por tratamientos abióticos tales como temperatura, sequía o modificaciones en la salinidad o inducible por sustancias aplicadas exógenamente que activan el promotor, por ejemplo, etanol, estrógenos, hormonas vegetales como etileno, ácido abscísico, ácido giberélico y/o metales pesados.

65 [0108] Un elemento intensificador del promotor también se puede usar para conseguir una mayor expresión del polipéptido en la planta. Por ejemplo, el elemento intensificador del promotor puede ser un intrón que se coloca

entre el promotor y la secuencia de nucleótidos que codifica un polipéptido de la presente invención. Por ejemplo, Xu *et al.*, 1993, *supra* revelan el uso del primer intrón del gen de actina 1 de arroz para mejorar la expresión.

5 [0109] Aún más, el uso del codón se puede optimizar para las especies vegetales en cuestión para mejorar la expresión (véase Horvath *et al.* al que se ha hecho referencia anteriormente).

[0110] El gen marcador seleccionable y cualquiera de las otras partes del constructo de expresión se pueden elegir de los disponibles en la técnica.

10 [0111] El constructo de ácido nucleico se incorpora en el genoma vegetal según técnicas convencionales conocidas en la técnica, que incluyen transformación mediada por *Agrobacterium*, transformación mediada por virus, microinyección, bombardeo de partículas, transformación biolística y electroporación (Gasser *et al.*, 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto *et al.*, 1989, *Nature* 338: 274).

15 [0112] Actualmente, la transferencia de gen mediada por *Agrobacterium* es el método de elección para generar dicotiledóneas transgénicas (para consulta, véase Hooykas y Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38), y este se puede usar también para la transformación de monocotiledóneas, aunque otros métodos de transformación se usan más frecuentemente para estas plantas. Actualmente, el método de elección para generar monocotiledóneas transgénicas, complementando el método del *Agrobacterium*, es el bombardeo de partículas (oro microscópico o partículas de tungsteno recubiertos con el ADN de transformación) de callos embriogénicos o embriones de desarrollo (Christou, 1992, *Plant Journal* 2: 275-281; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). Un método alternativo para la transformación de monocotiledóneas se basa en la transformación de protoplastos como se describe por Omirulleh *et al.*, 1993, *Plant Molecular Biology* 21: 415-428.

25 [0113] Después de la transformación, los transformantes que han incorporado en ellos el constructo de expresión se seleccionan y se regeneran en plantas enteras según métodos bien conocidos en la técnica. Frecuentemente el procedimiento de transformación se diseña para la eliminación selectiva de genes de selección bien durante la regeneración o bien en las siguientes generaciones usando, por ejemplo, cotransformación con dos constructos de ADN-T separados o escisión específica del sitio del gen de selección por una recombinasa específica.

30 [0114] La presente invención también se refiere a métodos de producción de un polipéptido de la presente invención que comprenden (a) el cultivo de una planta o una célula vegetal transgénica que comprende una secuencia de ácido nucleico que codifica un polipéptido con actividad de proteasa de la presente invención bajo condiciones propicias para la producción del polipéptido; y (b) la recuperación del polipéptido.

Composiciones y usos

40 [0115] En otros aspectos adicionales, la presente descripción se refiere a composiciones que comprenden un polipéptido de la presente invención, así como a métodos de utilización de los mismos.

45 [0116] Las composiciones de polipéptido se pueden preparar conforme a métodos conocidos en la técnica y pueden ser en forma de un líquido o una composición seca. Por ejemplo, la composición de polipéptido puede ser en forma de granulados o microgranulados. El polipéptido que se va a incluir en la composición se puede estabilizar conforme a métodos conocidos en la técnica.

50 [0117] El polipéptido y/o la composición de la invención se pueden utilizar para una composición detergente, una composición de pienso para animales, para licuefacción, sacarificación y/o fermentación de almidón, por ejemplo, como se describe en WO9220777.

55 [0118] El polipéptido y/o la composición de la invención se pueden utilizar en un proceso para la producción de un jarabe. Por ejemplo, el producto final puede ser glucosa, pero también puede convertirse, por ejemplo, mediante glucosa isomerasa en fructosa o una mezcla compuesta casi igualmente de glucosa y fructosa. Esta mezcla, o una mezcla adicionalmente enriquecida con fructosa, es el jarabe de maíz con alto contenido en fructosa (HFCS) más comúnmente usado comercializado en todo el mundo.

60 [0119] El polipéptido y/o la composición de la invención se pueden utilizar en un proceso para la producción de un producto de fermentación. En una forma particular de realización, la invención se refiere a un método para la producción de un producto de fermentación, que comprende (a) la fermentación utilizando un microorganismo fermentador, y un carbohidrato que contiene material en presencia de un polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 10, y (b) la producción del producto de fermentación a partir del carbohidrato fermentado que contiene material.

65 [0120] Los productos de fermentación incluyen alcoholes (por ejemplo, etanol, metanol, butanol, 1,3-propanediol); ácidos orgánicos (por ejemplo, ácido cítrico, ácido acético, ácido itacónico, ácido láctico, ácido glucónico, gluconato, ácido láctico, ácido succínico, ácido 2,5-diceto-D-glucónico); cetonas (por ejemplo,

acetona); aminoácidos (por ejemplo, ácido glutámico); gases (por ejemplo, H₂ y CO₂), y compuestos más complejos, que incluyen, por ejemplo, antibióticos (por ejemplo, penicilina y tetraciclina); enzimas; vitaminas (por ejemplo, riboflavina, B12, beta-caroteno); hormonas, y otros compuestos que son difíciles de producir sintéticamente. Los procesos de fermentación se usan también comúnmente para producir alcohol para el consumo (por ejemplo, cerveza y vino). Un producto de fermentación preferido es el etanol, por ejemplo, etanol industrial, etanol combustible y/o etanol potable. Otros productos de fermentación preferidos incluyen los coproductos del proceso de fermentación de etanol, por ejemplo, granos secos de destilería (DDG).

[0121] En una forma de realización preferida, se licua almidón en presencia de una alfa-amilasa, el triturado licuado se sacarifica en presencia de una glucoamilasa. El triturado sacarificado se fermenta con una levadura; y (d) el etanol producido se recupera. En el proceso, un polipéptido de la invención se añade al triturado antes de/durante la licuefacción o sacarificación y/o al almidón hidrolizado y los azúcares durante la fermentación.

[0122] Preferiblemente, las composiciones que comprenden un polipéptido de la presente invención comprenden al menos un otro polipéptido seleccionado de entre amilasa tal como, por ejemplo, alfa-amilasa (EC 3.2.1.1), glucoamilasa (EC 3.2.1.3) y pululanasa (EC 3.2.1.41); fitasa (EC 3.1.3.8 o 3.1.3.26); fosfatasa (EC 3.1.3.1; EC 3.1.3.2; EC 3.1.3.39); xilanasa (EC 3.2.1.8); galactanasa (EC 3.2.1.89); alfa-galactosidasa (EC 3.2.1.22); proteasa (EC 3.4.-.-), fosfolipasa A1 (EC 3.1.1.32); fosfolipasa A2 (EC 3.1.1.4); lisofosfolipasa (EC 3.1.1.5); fosfolipasa C (3.1.4.3); fosfolipasa D (EC 3.1.4.4) y/o beta-glucanasa (EC 3.2.1.4 o EC 3.2.1.6).

Ejemplos

[0123] Los productos químicos usados fueron productos comerciales de al menos grado reactivo.

Ejemplo 1: Preparación de variantes y pruebas de termoestabilidad

Cepas y plásmidos

[0124] *E. coli* DH12S (disponible de Gibco BRL) se usó para rescatar plásmido de levadura. pJTP000 es un vector transportador de *S. cerevisiae* y *E. coli* bajo el control del promotor TPI, construido a partir de pJC039 descrito en WO 01/92502, donde se ha insertado el gen de proteasa M35 de *Thermoascus aurantiacus* (WO 03048353).

[0125] Células competentes de *Saccharomyces cerevisiae* YNG318: MATa Dpep4[cir+] ura3-52, leu2-D2, his 4-539 se usaron para expresión de variantes de proteasa. Se describe en J. Biol. Chem. 272 (15), págs. 9720-9727, 1997.

Medios y sustratos

[0126] Solución basal 10X: 66,8 g/l de base nitrogenada de levadura sin aminoácidos (DIFCO), 100 g/l de succinato, 60 g/l de NaOH.

[0127] SC-glucosa: 20 % glucosa (es decir, una concentración final del 2 % = 2 g/100 ml) 100 ml/l, 5 % de treonina 4 ml/l, 1 % de triptófano 10 ml/l, 20 % de casamino ácidos 25 ml/l, solución basal 10 X 100 ml/l. La solución se esteriliza utilizando un filtro de un tamaño de poros de 0,20 micrómetros. Agar (2 %) y H₂O (aprox. 761 ml) se someten juntos a autoclave y la solución de SC-glucosa esterilizada separadamente se añade a la solución de agar.

[0128] YPD: 20 g/l de bacto peptona, 10 g/l de extracto de levadura, 20 % de glucosa 100 ml/l.

[0129] YPD+Zn: YPD + 0,25 mM de ZnSO₄.

[0130] PEG/solución de LiAc: 40 % de PEG4000 50 ml, 5 M de acetato de litio 1 ml.

[0131] Placa de microtitulación de zeína de 96 pocillos:

[0132] Cada pocillo contiene 200 microL de 0,05-0,1 % de zeína (Sigma), 0,25 mM de ZnSO₄ y 1 % de agar en 20 mM de tampón de acetato sódico, pH 4,5.

Manipulaciones de ADN

[0133] A menos que se indique lo contrario, las manipulaciones y transformaciones de ADN se realizaron utilizando métodos estándar de biología molecular como se describen en Sambrook *et al.* (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor lab. Cold Spring Harbor, NY; Ausubel, F. M. *et al.* (eds.) "Current protocols in Molecular Biology", John Wiley and Sons, 1995; Harwood, C. R. and Cutting, S. M. (Eds.).

Transformación de levadura

[0134] La transformación de levadura se realizó utilizando el método de acetato de litio. Se mezclan 0,5 microL de vector (digeridos por endonucleasas de restricción) y 1 microL de fragmentos de PCR. La mezcla de ADN, 100 microL de células competentes YNG318, y 10 microL de ADN (Clontech) portador de fabricante de levadura se añaden a un tubo de polipropileno de 12 ml (Falcon 2059). Se añaden 0,6 ml de PEG/solución LiAc y se mezcla suavemente. Se incuba durante 30 min a 30 °C, y 200 r.p.m. seguido de 30 min a 42 °C (choque térmico). Se transfiere a un tubo de Eppendorf y se centrifuga durante 5 s. Se elimina el sobrenadante y se resuelve en 3 ml de YPD. Se incuba la suspensión celular durante 45 min a 200 r.p.m. a 30 °C. Se vierte la suspensión en placas de SC-glucosa y se incuban a 30 °C durante 3 días para cultivar colonias. ADN total de levadura se extrae con Zymoprep Yeast Plasmid Miniprep Kit (ZYMO research).

Secuenciación de ADN

[0135] La transformación de *E. coli* para la secuenciación de ADN se efectuó por electroporación (BIO-RAD Gene Pulser). Plásmidos de ADN se prepararon por método alcalino (Molecular Cloning, Cold Spring Harbor) o con el Qiagen® Plasmid Kit. Los fragmentos de ADN se recuperaron a partir de gel de agarosa mediante el Qiagen Gel Extraction Kit. Se realizó PCR utilizando un motor de ADN PTC-200. Se utilizó el ABI PRISM™ 310 Genetic Analyzer para la determinación de todas las secuencias de ADN.

Construcción de vector de expresión de proteasa

[0136] El gen de proteasa M35 de *Thermoascus* se amplificó con el par de cebadores Prot F (SEQ ID N.º: 3) y Prot R (SEQ ID N.º: 4). Los fragmentos de PCR resultantes se introdujeron en *S. cerevisiae* YNG318 junto con el vector pJC039 (descrito en WO2001/92502) digerido con enzimas de restricción para eliminar el gen de cutinasa de *Humicola insolens*.

[0137] El plásmido en los clones de levadura en las placas de SC-glucosa se recuperó para confirmar la secuencia interna y se denominó como pJTP001.

Construcción de biblioteca de levadura y de variantes dirigidas al sitio

[0138] La biblioteca de levadura y de variantes dirigidas al sitio se construyó por el método de PCR SOE (empalme por extensión de superposición, véase "PCR: A practical approach", p. 207-209, Oxford University press, eds. McPherson, Quirke, Taylor), seguido de recombinación *in vivo* de levadura.

Cebadores generales para la amplificación y secuenciación

[0139] Los cebadores AM34 (SEQ ID N.º: 5) y AM35 (SEQ ID N.º: 6) se usaron para hacer fragmentos de ADN que contienen cualquier fragmento mutado por el método SOE junto con cebadores degenerados (AM34 + cebador inverso y AM35 + cebador directo) o solo para amplificar un gen de proteasa integral (AM34 + AM35).

Sistema de reacción de PCR:	Condiciones:	
48,5 microL de H ₂ O	1	94 °C 2 min
2 perlas de PCR Pure Taq Ready-To-Go (Amersham Biosciences)	2	94 °C 30 s
0,5 microL X 2 100 pmol/microL de cebadores	3	55 °C 30 s
0,5 microL ADN modelo	4	72 °C 90 s
	2-4	25 ciclos
	5	72 °C 10 min

[0140] Los fragmentos de ADN se recuperaron a partir de gel de agarosa mediante el Qiagen Gel Extraction Kit. Los fragmentos purificados resultantes se mezclaron con la digestión del vector. La solución mezclada se introdujo en *Saccharomyces cerevisiae* para construir bibliotecas o variantes dirigidas al sitio por recombinación *in vivo*.

Ensayo de actividad relativa

[0141] Los clones de levadura en SC-glucosa se inocularon a una placa de microtitulación de 96 pocillos que contiene medio de YPD+Zn y se cultivaron a 28 °C durante 3 días. Los sobrenadantes del cultivo se aplicaron a una placa de microtitulación de zeína de 96 pocillos y se incubaron a al menos 2 temperaturas (por ejemplo, 60 °C y 65 °C, 70 °C y 75 °C, 70 °C y 80 °C) durante más de 4 horas o durante toda la noche. La turbidez de la zeína en la placa se midió como A630 y la actividad relativa (temperaturas superiores/inferiores) se determinó como un indicador de mejora de la termoactividad. Los clones con actividad relativa más alta que la variante progenitora se seleccionaron y la secuencia se determinó.

Ensayo de actividad restante

[0142] Los clones de levadura en la SC-glucosa se inocularon a una placa de microtitulación de 96 pocillos y se cultivaron a 28 °C durante 3 días. La actividad de proteasa se midió a 65 °C usando azocaseína (Megazyme) después de la incubación del sobrenadante del cultivo en 20 mM de tampón de acetato sódico, pH 4,5, durante 10 min a una temperatura determinada (80 °C o 84 °C con 4 °C como una referencia) para determinar la actividad restante. Los clones con actividad restante más alta que la variante progenitora se seleccionaron y la secuencia se determinó.

Ensayo de azocaseína

[0143] 20 microL de muestras se mezclaron con 150 microL de solución de sustrato (4 ml de 12,5 % de azocaseína en el etanol en 96 ml de 20 mM de acetato sódico, pH 4,5, que contiene 0,01 % de triton-100 y 0,25 mM de ZnSO₄) y se incubaron durante 4 horas o más.

[0144] Después de añadir 20 microL/pocillo de 100 % de solución de ácido tricloroacético (ATC), la placa se centrifugó y 100 microL de sobrenadantes se sacaron con pipeta para medir A440.

Expresión de variantes de proteasa en *Aspergillus oryzae*

[0145] Los constructos que comprenden los genes de las variantes de proteasa se usaron para construir vectores de expresión para *Aspergillus*. Los vectores de expresión de *Aspergillus* consisten en un casete de expresión basado en el promotor de amilasa neutra II de *Aspergillus niger* II fusionado a la secuencia líder no traducida de triosa fosfato isomerasa de *Aspergillus nidulans* (Pna2/tpi) y el terminador de amiloglucosidasa (Tamg) de *Aspergillus niger*. También estaba presente en el plásmido el marcador selectivo de *Aspergillus amdS* de *Aspergillus nidulans* que permite el cultivo en acetamida como única fuente de nitrógeno. Los plásmidos de expresión para las variantes de proteasa se transformaron en *Aspergillus* como se describe en Lassen *et al.* (2001), Appl. Environ. Microbiol. 67, 4701-4707. Para cada uno de los constructos, se aislaron 10-20 cepas, se purificaron y se cultivaron en frascos de agitación.

Purificación de variantes expresadas

[0146]

1. Ajustar el pH de los 0,22 µm de las muestras de fermentación filtradas a 4,0.
2. Poner la muestra en un baño de hielo con agitación magnética. Añadir (NH₄)₂SO₄ en pequeñas partes alícuotas (correspondientes a aprox. 2,0-2,2 M de (NH₄)₂SO₄ sin tener en cuenta el aumento del volumen al añadir el compuesto).
3. Después de la adición final de (NH₄)₂SO₄, incubar la muestra en el baño de hielo con agitación magnética suave durante 45 min.
4. Centrifugación: centrifugador refrigerado de alta velocidad CR20G de Hitachi himac equipado con cabeza de rotor R20A2, 5 °C, 20.000 r.p.m., 30 min.
5. Disolver el precipitado formado en 200 ml de 50 mM de Na-acetato pH 4,0.
6. Filtrar la muestra por succión al vacío utilizando una membrana de 0,22 µm PES PLUS (IWAKI).
7. Desalar/cambiar de tampón la muestra a 50 mM de Na-acetato pH 4,0 usando ultrafiltración (Vivacell 250 de Vivascience equipado con membrana de 5 kDa MWCO PES) durante toda la noche en una cámara frigorífica. Diluir la muestra de retenido a 200 ml utilizando 50 mM de Na-acetato pH 4,0. La conductividad de muestra es preferiblemente inferior a 5 mS/cm.
8. Cargar la muestra sobre una columna de intercambio de cationes equilibrada con 50 mM de Na-acetato pH 4,0. Lavar la muestra sin unir de la columna utilizando 3 volúmenes de columna de tampón de unión (50 mM de Na-acetato pH 4,0), y eluir la muestra utilizando un gradiente lineal, 0-100 % de tampón de elución (50 mM de Na-acetato + 1 M de NaCl pH 4,0) en 10 volúmenes de columna.
9. Las fracciones recogidas se ensayan mediante un ensayo de endoproteasa (*cf.* a continuación) seguido de SDS-PAGE estándar (condiciones de reducción) en fracciones seleccionadas. Las fracciones se agrupan basándose en el ensayo de endoproteasa y SDS-PAGE.

Ensayo de endoproteasa

[0147]

1. Comprimido de Protazyme OL/5 ml 250 mM de Na-acetato pH 5,0 se disuelve por agitación magnética (sustrato: comprimido de Protazyme AK de endoproteasa de Megazyme - cat. # PRAK 11/08).
2. Con agitación, 250 microL de solución de sustrato se transfieren a un tubo de Eppendorf de 1,5 ml.
3. 25 microL de la muestra se añaden a cada tubo (la forma preliminar es tampón de muestra).
4. Los tubos se incuban en un termomezclador con agitación (1000 r.p.m.) a 50 °C durante 15 minutos.
5. 250 microL de 1 M de NaOH se añaden a cada tubo, seguido de mezcla en remolino.

6. Centrifugación durante 3 min a 16,100 x G y 25 °C.

7. 200 microL del sobrenadante se transfieren a una MTP y se registra la absorbancia a 590 nm.

Resultados

5

[0148]

Tabla 1. Actividad relativa de variantes de proteasa. La numeración de la sustitución (las sustituciones) comienza en el N-terminal del péptido maduro en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

Variante	Sustitución (sustituciones)	Actividad relativa 65 °C/60 °C
Tipo salvaje	Ninguno	31 %
JTP004	S87P	45 %
JTP005	A112P	43 %
JTP008	R2P	71 %
JTP009	D79K	69 %
JTP010	D79L	75 %
JTP011	D79M	73 %
JTP012	D79L/S87P	86 %
JTP013	D79L/S87P/A112P	90 %
JTP014	D79L/S87P/A112P	88 %
JTP016	A73C	52 %
JTP019	A126V	69 %
JTP021	M152R	59 %

Tabla 2. Actividad relativa de variantes de proteasa. La numeración de la sustitución (las sustituciones) comienza en el N-terminal del péptido maduro en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

Variante	Sustitución (sustituciones) y/o deleción (delecciones)	Actividad relativa		
		70 °C/65 °C	75 °C/65 °C	75 °C/70 °C
Tipo salvaje	Ninguna	59 %	17 %	
JTP036	D79L/S87P/D142L	73 %	73 %	
JTP040	T54R/D79L/S87P		71 %	
JTP042	Q53K/D79L/S87P/I173V		108 %	
JTP043	Q53R/D79L/S87P		80 %	
JTP045	S41R/D79L/S87P		82 %	
JTP046	D79L/S87P/Q158W		96 %	
JTP047	D79L/S87P/S157K		85 %	
JTP048	D79L/S87P/D104R		88 %	
JTP050	D79L/S87P/A112P/D142L		88 %	
JTP051	S41R/D79L/S87P/A112P/D142L			102 %
JTP052	D79L/S87P/A112P/D142L/S157K			111 %
JTP053	S41R/D79L/S87P/A112P/D142L/S157K			113 %
JTP054	ΔS5/D79L/S87P			92 %
JTP055	ΔG8/D79L/S87P			95 %
JTP059	C6R/D79L/S87P			92 %
JTP061	T46R/D79L/S87P			111 %
JTP063	S49R/D79L/S87P			94 %
JTP064	D79L/S87P/N88R			92 %
JTP068	D79L/S87P/T114P			99 %
JTP069	D79L/S87P/S115R			103 %
JTP071	D79L/S87P/T116V			105 %
JTP072	N26R/D79L/S87P		92 %	
JTP077	A27K/D79L/S87P/A112P/D142L		106 %	
JTP078	A27V/D79L/S87P/A112P/D142L		100 %	
JTP079	A27G/D79L/S87P/A112P/D142L		104 %	

10

Tabla 3. Actividad relativa de variantes de proteasa. La numeración de la sustitución (las sustituciones) comienza en el N-terminal del péptido maduro en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

Variante	Sustitución (sustituciones) y/o deleción (delecciones)	Actividad relativa 75 °C/65 °C	Actividad restante	
			80 °C	84 °C

JTP082	ΔS5/D79L/S87P/A112P/D142L	129 %		53 %
JTP083	T46R/D79L/S87P/A112P/D142L	126 %		
JTP088	Y43F/D79L/S87P/A112P/D142L	119 %		
JTP090	D79L/S87P/A112P/T124L/D142L	141 %		
JTP091	D79L/S87P/A112P/T124V/D142L	154 %	43 %	
JTP092	ΔS5/N26R/D79L/S87P/A112P/D142L			60 %
JTP095	N26R/T46R/D79L/S87P/A112P/D142L			62 %
JTP096	T46R/D79L/S87P/T116V/D142L			67 %
JTP099	D79L/P81R/S87P/A112P/D142L			80 %
JTP101	A27K/D79L/S87P/A112P/T124V/D142L		81 %	
JTP116	D79L/Y82F/S87P/A112P/T124V/D142L		59 %	
JTP117	D79L/Y82F/S87P/A112P/T124V/D142L		94 %	
JTP127	D79L/S87P/A112P/T124V/A126V/D142L		53 %	

Tabla 4. Actividad relativa de variantes de proteasa. La numeración de la sustitución (sustituciones) comienza en el N-terminal del péptido maduro en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

Variante	Sustituciones	Actividad relativa		
		75 °C/70 °C	80 °C/70 °C	85 °C/70 °C
JTP050	D79L S87P A112P D142L	55 %	23 %	9 %
JTP134	D79L Y82F S87P A112P D142L		40 %	
JTP135	S38T D79L S87P A112P A126V D142L		62 %	
JTP136	D79L Y82F S87P A112P A126V D142L		59 %	
JTP137	A27K D79L S87P A112P A126V D142L		54 %	
JTP140	D79L S87P N98C A112P G135C D142L	81 %		
JTP141	D79L S87P A112P D142L T141C M161C	68 %		
JTP143	S36P D79L S87P A112P D142L	69 %		
JTP144	A37P D79L S87P A112P D142L	57 %		
JTP145	S49P D79L S87P A112P D142L	82 %	59 %	
JTP146	S50P D79L S87P A112P D142L	83 %	63 %	
JTP148	D79L S87P D104P A112P D142L	76 %	64 %	
JTP161	D79L Y82F S87G A112P D142L		30 %	12 %
JTP180	S70V D79L Y82F S87G Y97W A112P D142L		52 %	
JTP181	D79L Y82F S87G Y97W D104P A112P D142L		45 %	
JTP187	S70V D79L Y82F S87G A112P D142L		45 %	
JTP188	D79L Y82F S87G D104P A112P D142L		43 %	
JTP189	D79L Y82F S87G A112P A126V D142L		46 %	
JTP193	Y82F S87G S70V D79L D104P A112P D142L			15 %
JTP194	Y82F S87G D79L D104P A112P A126V D142L			22 %
JTP196	A27K D79L Y82F S87G D104P A112P A126V D142L			18 %

Tabla 5. Actividad relativa de variantes de proteasa. La numeración de la sustitución (sustituciones) comienza en el N-terminal del péptido maduro en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

Variante	Sustituciones	Actividad relativa	
		75 °C/70 °C	80 °C/70 °C
JTP196	A27K D79L Y82F S87G D104P A112P A126V D142L	102 %	55 %
JTP210	A27K Y82F S87G D104P A112P A126V D142L	107 %	36 %
JTP211	A27KD79LY82F D104P A112P A126V D142L	94 %	44 %
JTP213	A27K Y82F D104P A112P A126V D142L	103 %	37 %

5

Ejemplo 2

Perfil de temperatura de variantes seleccionadas utilizando enzimas purificadas

- 10 [0149] Las variantes seleccionadas con buena termoestabilidad se purificaron y las enzimas purificadas se usaron en un ensayo de zeína-BCA como se describe a continuación. La actividad de proteasa restante se determinó a 60 °C tras la incubación de la enzima a temperaturas elevadas como se ha indicado durante 60 min.

Ensayo de zeína-BCA:

- 15 [0150] El ensayo de zeína-BCA se realizó para detectar cuantificación de proteína soluble liberada de la zeína por proteasas variantes a varias temperaturas.

Protocolo:

[0151]

- 5
- 1) Mezclar 10 ul de 10 ug/ml de soluciones enzimáticas y 100 ul de 0,025 % de solución de zeína en una placa de microtitulación (MTP).
 - 2) Incubar a varias temperaturas durante 60 min.
 - 3) Añadir 10 ul de 100 % de solución de ácido tricloroacético (ATC).
 - 4) Centrifugar la MTP a 3500 r.p.m. durante 5 min.
 - 5) Sacar 15 ul a una nueva MTP que contiene 100 ul de solución de ensayo BCA (Pierce Cat#: 23225, BCA Protein Assay Kit).
 - 6) Incubar durante 30 min. a 60 °C.
 - 7) Medir A562.
- 10
- 15

[0152] Los resultados se muestran en la tabla 6. Todas las variantes evaluadas mostraron una termoestabilidad mejorada en comparación con la proteasa de tipo salvaje.

Tabla 6. Ensayo de zeína-BCA

Tipo salvaje/ Variante	La muestra incubada 60 min a temperaturas indicadas (°C) (µg/ml de péptido equivalente de albúmina de suero bovino liberado)						
	60 °C	70 °C	75 °C	80 °C	85 °C	90 °C	95 °C
Tipo salvaje	94	103	107	93	58	38	
JTP050	86	101	107	107	104	63	36
JTP077	82	94	104	105	99	56	31
JTP188	71	83	86	93	100	75	53
JTP196	87	99	103	106	117	90	38

20 Listado de secuencias

[0153]

- 25 <110> Matsui, Tomoko
Noergaard, Allan
Poulsen, Thomas Agersten
Matthews, John
- 30 <120> Variantes de proteasa
- <130> 11665-WO-PCT
- <150> US 61/285,601
- 35 <151> 2009-12-11
- <160> 6
- <170> Versión de PatentIn 3.5
- 40 <210> 1
<211> 1068
<212> ADN
<213> Thermoascus aurantiacus
- 45 <220>
<221> Sig_peptide
<222> (1)..(57)
- 50 <220>
<221> CDS
<222> (1)..(1065)

ES 2 668 202 T3

```

<220>
<221> misc_feature
<222> (58)..(534)
5
<220>
<221> mat_peptide
<222> (535)..(1068)
10 <400> 1

atg cgg ctc gtt gct tcc cta acg gcc ttg gtg gcc ttg tcc gta
45
Met Arg Leu Val Ala Ser Leu Thr Ala Leu Val Ala Leu Ser Val
15 -175 -170 -165

cct gtc ttt ccc gct gct gtc aac gtg aag cgt gct tcg tcc tac
90
Pro Val Phe Pro Ala Ala Val Asn Val Lys Arg Ala Ser Ser Tyr
20 -160 -155 -150

ctg gag atc act ctg agc cag gtc agc aac act ctg atc aag gcc
135
Leu Glu Ile Thr Leu Ser Gln Val Ser Asn Thr Leu Ile Lys Ala
25 -145 -140 -135

gtg gtc cag aac act ggt agc gac gag ttg tcc ttc gtt cac ctg
180
Val Val Gln Asn Thr Gly Ser Asp Glu Leu Ser Phe Val His Leu
30 -130 -125 -120

aac ttc ttc aag gac ccc gct cct gtc aaa aag gta tcg gtc tat
225
Asn Phe Phe Lys Asp Pro Ala Pro Val Lys Lys Val Ser Val Tyr
35 -115 -110 -105

cgc gat ggg tct gaa gtg cag ttc gag ggc att ttg agc cgc tac aaa
273
Arg Asp Gly Ser Glu Val Gln Phe Glu Gly Ile Leu Ser Arg Tyr Lys
40 -100 -95 -90

tcg act ggc ctc tct cgt gac gcc ttt act tat ctg gct ccc gga gag
321
Ser Thr Gly Leu Ser Arg Asp Ala Phe Thr Tyr Leu Ala Pro Gly Glu
45 -85 -80 -75

tcc gtc gag gac gtt ttt gat att gct tcg act tac gat ctg acc agc
369
Ser Val Glu Asp Val Phe Asp Ile Ala Ser Thr Tyr Asp Leu Thr Ser
50 -70 -65 -60

ggc ggc cct gta act atc cgt act gag gga gtt gtt ccc tac gcc acg
417
Gly Gly Pro Val Thr Ile Arg Thr Glu Gly Val Val Pro Tyr Ala Thr
55 -55 -50 -45 -40

```

ES 2 668 202 T3

gct aac agc act gat att gcc ggc tac atc tca tac tcg tct aat gtg
465
Ala Asn Ser Thr Asp Ile Ala Gly Tyr Ile Ser Tyr Ser Ser Asn Val
-35 -30 -25

5
ttg acc att gat gtc gat ggc gcc gct gct gcc act gtc tcc aag gca
513
Leu Thr Ile Asp Val Asp Gly Ala Ala Ala Thr Val Ser Lys Ala
-20 -15 -10

10
atc act cct ttg gac cgc cgc act agg atc agt tcc tgc tcc ggc agc
561
Ile Thr Pro Leu Asp Arg Arg Thr Arg Ile Ser Ser Cys Ser Gly Ser
-5 -1 1 5

15
aga cag agc gct ctt act acg gct ctc aga aac gct gct tct ctt gcc
609
Arg Gln Ser Ala Leu Thr Thr Ala Leu Arg Asn Ala Ala Ser Leu Ala
10 15 20 25

20
aac gca gct gcc gac gcg gct cag tct gga tca gct tca aag ttc agc
657
Asn Ala Ala Ala Asp Ala Ala Gln Ser Gly Ser Ala Ser Lys Phe Ser
30 35 40

25
gag tac ttc aag act act tct agc tct acc cgc cag acc gtg gct gcg
705
Glu Tyr Phe Lys Thr Thr Ser Ser Ser Thr Arg Gln Thr Val Ala Ala
45 50 55

30
cgt ctt cgg gct gtt gcg cgg gag gca tct tcg tct tct tcg gga gcc
753
Arg Leu Arg Ala Val Ala Arg Glu Ala Ser Ser Ser Ser Ser Gly Ala
60 65 70

35
acc acg tac tac tgc gac gat ccc tac ggc tac tgt tcc tcc aac gtc
801
Thr Thr Tyr Tyr Cys Asp Asp Pro Tyr Gly Tyr Cys Ser Ser Asn Val
75 80 85

40
ctg gct tac acc ctg cct tca tac aac ata atc gcc aac tgt gac att
849
Leu Ala Tyr Thr Leu Pro Ser Tyr Asn Ile Ile Ala Asn Cys Asp Ile
90 95 100 105

45
ttc tat act tac ctg ccg gct ctg acc agt acc tgt cac gct cag gat
897
Phe Tyr Thr Tyr Leu Pro Ala Leu Thr Ser Thr Cys His Ala Gln Asp
110 115 120

50
caa gcg acc act gcc ctt cac gag ttc acc cat gcg cct ggc gtc tac
945
Gln Ala Thr Thr Ala Leu His Glu Phe Thr His Ala Pro Gly Val Tyr
125 130 135

55
agc cct ggc acg gac gac ctg gcg tat ggc tac cag gct gcg atg ggt
993

ES 2 668 202 T3

Ser Pro Gly Thr Asp Asp Leu Ala Tyr Gly Tyr Gln Ala Ala Met Gly
 140 145 150

5 ctc agc agc agc cag gct gtc atg aac gct gac acc tac gct ctc tat
 1041

Leu Ser Ser Ser Gln Ala Val Met Asn Ala Asp Thr Tyr Ala Leu Tyr
 155 160 165

10 gcg aat gcc ata tac ctt ggt tgc taa
 1068

Ala Asn Ala Ile Tyr Leu Gly Cys
 170 175

15 <210> 2
 <211> 355
 <212> PRT
 <213> *Thermoascus aurantiacus*

20 <400> 2

Met Arg Leu Val Ala Ser Leu Thr Ala Leu Val Ala Leu Ser Val
 -175 -170 -165

25 Pro Val Phe Pro Ala Ala Val Asn Val Lys Arg Ala Ser Ser Tyr
 -160 -155 -150

30 Leu Glu Ile Thr Leu Ser Gln Val Ser Asn Thr Leu Ile Lys Ala
 -145 -140 -135

35 Val Val Gln Asn Thr Gly Ser Asp Glu Leu Ser Phe Val His Leu
 -130 -125 -120

40 Asn Phe Phe Lys Asp Pro Ala Pro Val Lys Lys Val Ser Val Tyr
 -115 -110 -105

Arg Asp Gly Ser Glu Val Gln Phe Glu Gly Ile Leu Ser Arg Tyr Lys
 -100 -95 -90

45 Ser Thr Gly Leu Ser Arg Asp Ala Phe Thr Tyr Leu Ala Pro Gly Glu
 -85 -80 -75

50 Ser Val Glu Asp Val Phe Asp Ile Ala Ser Thr Tyr Asp Leu Thr Ser
 -70 -65 -60

55 Gly Gly Pro Val Thr Ile Arg Thr Glu Gly Val Val Pro Tyr Ala Thr
 -55 -50 -45 -40

ES 2 668 202 T3

Ala Asn Ser Thr Asp Ile Ala Gly Tyr Ile Ser Tyr Ser Ser Asn Val
 -35 -30 -25

5 Leu Thr Ile Asp Val Asp Gly Ala Ala Ala Thr Val Ser Lys Ala
 -20 -15 -10

10 Ile Thr Pro Leu Asp Arg Arg Thr Arg Ile Ser Ser Cys Ser Gly Ser
 -5 -1 1 5

15 Arg Gln Ser Ala Leu Thr Thr Ala Leu Arg Asn Ala Ala Ser Leu Ala
 10 15 20 25

Asn Ala Ala Ala Asp Ala Ala Gln Ser Gly Ser Ala Ser Lys Phe Ser
 30 35 40

20 Glu Tyr Phe Lys Thr Thr Ser Ser Ser Thr Arg Gln Thr Val Ala Ala
 45 50 55

25 Arg Leu Arg Ala Val Ala Arg Glu Ala Ser Ser Ser Ser Ser Gly Ala
 60 65 70

30 Thr Thr Tyr Tyr Cys Asp Asp Pro Tyr Gly Tyr Cys Ser Ser Asn Val
 75 80 85

35 Leu Ala Tyr Thr Leu Pro Ser Tyr Asn Ile Ile Ala Asn Cys Asp Ile
 90 95 100 105

40 Phe Tyr Thr Tyr Leu Pro Ala Leu Thr Ser Thr Cys His Ala Gln Asp
 110 115 120

45 Gln Ala Thr Thr Ala Leu His Glu Phe Thr His Ala Pro Gly Val Tyr
 125 130 135

50 Ser Pro Gly Thr Asp Asp Leu Ala Tyr Gly Tyr Gln Ala Ala Met Gly
 140 145 150

55 Leu Ser Ser Ser Gln Ala Val Met Asn Ala Asp Thr Tyr Ala Leu Tyr
 155 160 165

Ala Asn Ala Ile Tyr Leu Gly Cys
 170 175

<210> 3

ES 2 668 202 T3

<211> 49
<212> ADN
<213> Secuencia artificial

5 <220>
<223> Cebador artificial

<400> 3
aacgacggta cccggggatc ggatccatgc ggctcgttgc ttcctaac
10 49

<210> 4
<211> 48
15 <212> ADN
<213> Secuencia artificial

<220>
<223> Cebador artificial
20

<400> 4
ctaattacat gatgcggccc ttaattaatt agcaaccaag gtatatgg
48

25
<210> 5
<211> 20
<212> ADN
<213> Secuencia artificial

30
<220>
<223> Cebador artificial

<400> 5
35 taggagttta gtgaacttgc
20

<210> 6
40 <211> 18
<212> ADN
<213> Secuencia artificial

<220>
45 <223> Cebador artificial

<400> 6
ttcgagcgtc ccaaaacc
18

REIVINDICACIONES

1. Método de producción de una variante de proteasa con actividad de proteasa y una termoestabilidad mejorada en comparación con la proteasa progenitora, donde dicho método comprende el cultivo de una célula en la que se ha introducido un vector de expresión que comprende los siguientes elementos operativamente enlazados:

- (a) un promotor de transcripción,
 (b) una molécula de polinucleótido que codifica una variante de proteasa que tiene al menos el 85 % de identidad con la proteasa mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y que comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 en al menos una posición seleccionada de las siguientes: 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53, 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173, donde la molécula de polinucleótido se prepara introduciendo al menos una mutación en una molécula de ADN que codifica una proteasa, y
 (c) un terminador de transcripción,

con lo que dicha célula expresa la variante de proteasa codificada por la molécula de polinucleótido; y la recuperación de la variante de proteasa.

2. Método según la reivindicación 1, donde la variante de proteasa codificada por la molécula de polinucleótido comprende al menos una modificación seleccionada del grupo que consiste en A27K, A27G, A27V, Q53K, Q53R, T54R, D79K, D79L, D79M, Y82F, S87P, S87G, A112P, D142L, R2P, Δ S5, C6R, Δ G8, N26R, S41R, Y43F, T46R, S49R, A73C, P81R, N88R, D104R, D104P, T114P, S115R, T116V, T124L, T124V, A126V, M152R, S157K, Q158W y I173V.

3. Método según cualquiera de las reivindicaciones 1 o 2, donde la variante de proteasa codificada por la molécula de polinucleótido comprende al menos una de las siguientes combinaciones de modificaciones:

Δ S5/D79L/S87P, Δ S5/D79L/S87P/A112P/D142L, Δ S5/N26R/D79L/S87P/A112P/D142L, C6R/D79L/S87P, Δ G8/D79L/S87P, N26R/D79L/S87P, N26R/T46R/D79L/S87P/A112P/D142L, A27G/D79L/S87P/A112P/D142L, A27K/D79L/S87P/A112P/D142L, A27V/D79L/S87P/A112P/D142L, S41R/D79L/S87P, S41R/D79L/S87P/A112P/D142L, S41R/D79L/S87P/A112P/D142L/S157K, Y43F/D79L/S87P/A112P/D142L, T46R/D79L/S87P, T46R/D79L/S87P/A112P/D142L, T46R/D79L/S87P/T116V/D142L, S49R/D79L/S87P, Q53K/D79L/S87P/I173V, Q53R/D79L/S87P, T54R/D79L/S87P, D79L/S87P/A112P, D79L/P81R/S87P/A112P/D142L, D79L/S87P, D79L/S87P/A112P, D79L/S87P/A112P/D142L, D79L/S87P/A112P/D142L/S157K, D79L/S87P/A112P/T124L/D142L, D79L/S87P/A112P/T124V/A126V/D142L, D79L/S87P/A112P/T124V/D142L, D79L/S87P/D104R, D79L/S87P/D142L, D79L/S87P/N88R, D79L/S87P/Q158W, D79L/S87P/S115R, D79L/S87P/S157K, D79L/S87P/T114P, D79L/S87P/T116V, D79L/Y82F/S87P/A112P/T124V/D142L, D79L/Y82F/S87P/A112P/T124V/D142L, A27K/D79L/Y82F/S87G/D104P/A112P/A126V/D142L, A27K/Y82F/S87G/D104P/A112P/A126V/D142L, A27K/D79L/Y82F/D104P/A112P/A126V/D142L, A27K/Y82F/D104P/A112P/A126V/D142L.

4. Método según la reivindicación 1, donde la variante de proteasa comprende al menos una modificación en al menos una posición seleccionada del grupo que consiste en las posiciones: 27, 79, 82, 87, 104, 112, 126 y 142.

5. Método según la reivindicación 4, donde la variante de proteasa comprende al menos una de las modificaciones seleccionadas del grupo que consiste en: A27K, D79L, Y82F, S87G, S87P, D104P, A112P, A126V y D142L.

6. Polipéptido con actividad de proteasa que tiene al menos el 85 % de identidad con la proteasa mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2 y que comprende al menos una modificación en comparación con los aminoácidos 1 a 177 de la SEQ ID N.º: 2 en al menos una posición seleccionada de las siguientes: 27, 79, 82, 87, 112, 142, 2, 5, 6, 8, 26, 41, 43, 46, 49, 53, 54, 73, 88, 104, 114, 115, 116, 126, 152, 157, 158 y 173, donde el polipéptido tiene termoestabilidad mejorada en comparación con la proteasa mostrada en los aminoácidos 1 a 177 de la SEQ ID N.º: 2.

7. Polipéptido con actividad de proteasa según la reivindicación 6, que comprende al menos una modificación seleccionada de las siguientes: A27K, A27G, A27V, Q53K, Q53R, T54R, D79K, D79L, D79M, Y82F, S87P, S87G, A112P, D142L, R2P, Δ S5, C6R, Δ G8, N26R, S41R, Y43F, T46R, S49R, A73C, P81R, N88R, D104R, D104P, T114P, S115R, T116V, T124L, T124V, A126V, M152R, S157K, Q158W y I173V.

8. Polipéptido con actividad de proteasa según la reivindicación 6, donde el polipéptido comprende al menos una modificación en al menos una posición seleccionada del grupo que consiste en las posiciones: 27, 79, 82, 87, 104, 112, 126 y 142.

9. Polipéptido con actividad de proteasa según la reivindicación 8, donde el polipéptido comprende al menos una de las modificaciones seleccionadas del grupo que consiste en: A27K, D79L, Y82F, S87G, S87P, D104P, A112P, A126V y D142L.

5 10. Polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 9, donde el polipéptido comprende un conjunto de modificaciones seleccionadas de las siguientes:

10 Δ S5/D79L/S87P, Δ S5/D79L/S87P/A112P/D142L, Δ S5/N26R/D79L/S87P/A112P/D142L,
 C6R/D79L/S87P, Δ G8/D79L/S87P, N26R/D79L/S87P, N26R/T46R/D79L/S87P/A112P/D142L,
 A27G/D79L/S87P/A112P/D142L, A27K/D79L/S87P/A112P/D142L,
 A27K/D79L/S87P/A112P/T124V/D142L, A27V/D79L/S87P/A112P/D142L, S41R/D79L/S87P,
 S41R/D79L/S87P/A112P/D142L, S41R/D79L/S87P/A112P/D142L/S157K,
 Y43F/D79L/S87P/A112P/D142L, T46R/D79L/S87P, T46R/D79L/S87P/A112P/D142L,
 T46R/D79L/S87P/T116V/D142L, S49R/D79L/S87P, Q53K/D79L/S87P/I173V, Q53R/D79L/S87P,
 15 T54R/D79L/S87P, D79L/S87P/A112P, D79L/P81R/S87P/A112P/D142L, D79L/S87P,
 D79L/S87P/A112P, D79L/S87P/A112P/D142L, D79L/S87P/A112P/D142L/S157K,
 D79L/S87P/A112P/T124L/D142L, D79L/S87P/A112P/T124V/A126V/D142L,
 D79L/S87P/A112P/T124V/D142L, D79L/S87P/D104R, D79L/S87P/D142L, D79L/S87P/N88R,
 D79L/S87P/Q158W, D79L/S87P/S115R, D79L/S87P/S157K, D79L/S87P/T114P, D79L/S87PAT116V/
 20 D79LY82F/S87P/A112P/T124V/D142L, D79LY82F/S87P/A112P/T124V/D142L
 A27K/D79LY82F/S87G/D104P/A112P/A126V/D142L, A27K/Y82F/S87G/D104P/A112P/A126V/D142L,
 A27K/D79LY82F/D104P/A112P/A126V/D142L, A27K/Y82F/D104P/A112P/A126V/D142L.

25 11. Secuencia de ácido nucleico aislada que comprende una secuencia de ácido nucleico que codifica el polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 10.

30 12. Constructo de ácido nucleico que comprende la secuencia de ácido nucleico según la reivindicación 11 operativamente enlazada a una o más secuencias de control que dirigen la producción de la proteasa en un huésped de expresión adecuado.

13. Vector de expresión recombinante que comprende el constructo de ácido nucleico según la reivindicación 12.

35 14. Célula huésped recombinante que comprende el constructo de ácido nucleico según la reivindicación 12 y/o el vector de expresión según la reivindicación 13.

15. Método para la producción del polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 10, que comprende

40 (a) el cultivo de la célula huésped según la reivindicación 14 para producir un sobrenadante que comprende el polipéptido con actividad de proteasa; y (b) la recuperación del polipéptido con actividad de proteasa.

45 16. Planta transgénica, o parte de planta, capaz de expresar un polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 10.

50 17. Método para la producción de un producto de fermentación, que comprende (a) la fermentación utilizando un microorganismo fermentador y un carbohidrato que contiene material en presencia de un polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 10, y (b) la producción del producto de fermentación a partir del carbohidrato fermentado que contiene material.

18. Método según la reivindicación 17, donde el producto de fermentación es etanol y/o granos secos de destilería (DDG).

55 19. Uso de al menos un polipéptido con actividad de proteasa según cualquiera de las reivindicaciones 6 a 10 en un proceso para la producción de un producto de fermentación, preferiblemente etanol.