

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 750 753**

51 Int. Cl.:

C12N 9/62 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **03.06.2015 PCT/EP2015/062325**

87 Fecha y número de publicación internacional: **10.12.2015 WO15185590**

96 Fecha de presentación y número de la solicitud europea: **03.06.2015 E 15725383 (2)**

97 Fecha y número de publicación de la concesión europea: **24.07.2019 EP 3152304**

54 Título: **Endoproteasa específica para prolina y su uso**

30 Prioridad:

03.06.2014 EP 14170879
17.06.2014 EP 14172644

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
27.03.2020

73 Titular/es:

DSM IP ASSETS B.V. (100.0%)
Het Overloon 1
6411 TE Heerlen , NL

72 Inventor/es:

LAAN, VAN DER, JAN METSKE;
BRUINE-PAULUS, DE, ANGELA;
CHRISTIS, CHANTAL;
SPAANS, MARTINE y
VONDERVOORT, VAN DE, PETER JOZEF IDA

74 Agente/Representante:

LEHMANN NOVO, María Isabel

ES 2 750 753 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

Endoproteasa específica para prolina y su uso

5 La presente invención se refiere a un polipéptido que tiene actividad de endoproteasa específica para prolina, una composición que comprende el polipéptido, un ácido nucleico que codifica una endoproteasa específica para prolina, un vector de expresión que comprende el ácido nucleico que codifica una endoproteasa específica para prolina, una célula hospedante recombinante, un método para preparar una endoproteasa específica para prolina y un proceso para preparar un producto alimentario o pienso en donde se usa la endoproteasa específica para prolina.

Antecedentes

10 Las endoproteasas específicas para prolina son enzimas que hidrolizan una proteína o péptido en una posición donde hay una prolina en la proteína o péptido.

Una endoproteasa específica para prolina puede derivar, por ejemplo, de *Aspergillus niger* o *Penicillium chrysogenum*, tal como se describe en WO2002/046381 y WO2009/144269, respectivamente.

15 Otra endoproteasa específica para prolina se conoce a partir de WO2012/174127. El documento WO2012/174127 describe una proteasa específica para prolina de *Botryotinia fuckeliana*, *Aspergillus clavatus*, *Sclerotinia sclerotium*, *Mycosphaerella graminicola*, *Neurospora crasse*, *Talaromyces stipitatus* y *Gibberella zeae*.

20 Una endoproteasa específica para prolina se puede usar en diversas aplicaciones, por ejemplo, en la degradación del gluten (véase, por ejemplo, WO2005/027953 o WO2003/068170). El gluten es la fracción de proteína insoluble de cereales como el trigo, centeno, avena y cebada. El gluten es una mezcla compleja de moléculas de glutenina y prolamina que se cree que causan efectos tóxicos, por ejemplo, en pacientes que padecen enfermedad celíaca. Se considera que la celiaquía o enfermedad celíaca es una enfermedad autoinmunitaria. Los pacientes que padecen celiaquía deben seguir una dieta estricta sin gluten, que es muy difícil de seguir porque el gluten se usa tan ampliamente. El uso de endoproteasa específica para prolina como medicamento o suplemento alimentario puede aliviar la necesidad de una dieta estricta sin gluten (WO2003/068170).

25 Las endoproteasas específicas para prolina también se usan para reducir la turbidez en la cerveza, en donde la proteasa específica para prolina se puede agregar durante varias etapas de un proceso de producción de cerveza (WO 2002/046381).

Es deseable que las enzimas en aplicaciones alimentarias y de pienso tengan un pH óptimo adecuado y que preferiblemente no sean activas en el alimento o bebida final.

30 El objetivo de la presente invención es una endoproteasa específica para prolina alternativa con características mejoradas.

Compendio

35 En un aspecto, la presente descripción se refiere a un polipéptido que tiene actividad de endoproteasa específica para prolina, en donde el polipéptido tiene menos de 70 % de actividad residual sobre acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min. La actividad residual de un polipéptido que tiene actividad de endoproteasa específica para prolina se determina, de manera ventajosa, mediante el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) como sustrato a una temperatura de 20 °C y a pH 4,5, por ejemplo, en un tampón a pH 4,5, por ejemplo, un tampón de acetato de sodio (NaAc), que puede comprender una sal adicional tal como NaCl. La actividad residual se puede determinar al incubar un polipéptido, según se describe en la presente memoria, a una temperatura de 20 °C, y a un pH 4,5 durante 60 min.

45 En otro aspecto, la presente descripción se refiere a un polipéptido que tiene actividad de endoproteasa específica para prolina, en donde el polipéptido se selecciona del grupo que consiste en un polipéptido que tiene actividad de endoproteasa específica para prolina, opcionalmente que tiene menos de 50 % de actividad residual después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min, en donde el polipéptido se selecciona del grupo que consiste en:

50 i. un polipéptido, que, cuando se alinea con una secuencia de aminoácidos según SEQ ID NO:1, comprende al menos una combinación de aminoácidos que corresponde a la combinación seleccionada del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición 204, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición

304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), en el que las posiciones se definen con referencia a SEQ ID NO:1;

5 ii. un polipéptido que comprende una secuencia de aminoácidos según la SEQ ID NO: 1, en donde la SEQ ID NO: 1 comprende al menos una combinación de sustituciones de aminoácidos seleccionada del grupo que consiste en las combinaciones (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) en donde las sustituciones se definen con respecto a la SEQ ID NO: 1;

iii. un polipéptido según i) a iii), pero que carece de una secuencia señal y/o una secuencia propeína;

10 iv. un polipéptido según i) a iv) que tiene al menos 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad con respecto a la secuencia de aminoácidos de la SEQ ID NO: 1;

15 v. un polipéptido codificado por un ácido nucleico que tiene al menos 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 %, 100 % de identidad con respecto a la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácido seleccionada del grupo que consiste en (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) en donde las sustituciones de aminoácidos se definen con respecto a la SEQ ID NO: 1.

20 La invención también proporciona una composición que comprende un polipéptido que tiene endoproteasa específica para prolina según se describe en la presente memoria.

En otro aspecto, la presente invención proporciona un método para generar un polipéptido variante que tiene actividad de endoproteasa específica para prolina según se describe en la presente memoria.

25 La descripción también proporciona un ácido nucleico que codifica una endoproteasa específica para prolina que tiene al menos 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad secuencial con respecto a la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácidos seleccionada del grupo que consiste en (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) en donde las sustituciones de aminoácidos se definen con respecto a la SEQ ID NO: 1.

30 En otro aspecto, la presente invención se refiere a un vector de expresión que comprende un polinucleótido que codifica un polipéptido según se describe en la presente memoria.

En otro aspecto, la presente invención se refiere a una célula hospedante recombinante que comprende una secuencia polinucleotídica o un vector de expresión según se describe en la presente memoria.

35 En otro aspecto adicional, la presente invención se refiere a un método para la preparación de un polipéptido, que comprende cultivar una célula hospedante según se describe en la presente memoria, en condiciones que permiten la expresión del polipéptido y preparar el polipéptido.

40 En otro aspecto, la presente invención se refiere a un proceso para la preparación de un producto alimentario o pienso que comprende incubar una forma intermediaria del producto alimentario o pienso con un polipéptido, o una composición que comprende un polipéptido según se describe en la presente memoria, y preparar el producto alimentario.

La presente descripción también se refiere a un producto alimentario o pienso obtenible mediante un proceso según se describe en la presente memoria.

Definiciones

45 El término "productos horneados" se define en la presente memoria como cualquier producto preparado a partir de una masa o una mezcla. El producto puede tener un carácter suave o crujiente y puede ser de tipo blanco, claro u oscuro. Los productos horneados incluyen, pero no se limitan a, pan tal como, por ejemplo, pan blanco, pan integral o de centeno, pan tipo baguette francés, productos de masa laminados tal como pastel (danés), medialuna u hojaldre, pan de pita, tortillas, tacos, tortas, panqueques, biscochos, galletas, donas, roscas, tapas de masa, bollos, pan al vapor y galleta al agua. Los tipos de productos horneados, los métodos para caracterizarlos y producirlos son conocidos para los expertos en la técnica, véase, por ejemplo, "Baking Science and Technology", por E.J. Pyler, L.A. Gorton, 2008, (2 tomos) Sosland Publishing Company, Kansas, EE. UU., o "Baked Products: Science, Technology and Practice" por S.P. Cauvain, L.S. Young, 2006, Blackwell Publishing Ltd, Oxford, RU.

El término “cadena complementaria” se puede usar de manera intercambiable con el término “complemento”. El complemento de una cadena de ácido nucleico puede ser el complemento de una cadena codificante o el complemento de una cadena no codificante. Cuando se hace referencia a ácidos nucleicos bicatenarios, el complemento de un ácido nucleico que codifica un polipéptido hace referencia a la cadena complementaria de la cadena que codifica la secuencia de aminoácidos o a cualquier molécula de ácido nucleico que la contiene.

El término “secuencia de control” se puede usar de manera intercambiable con el término “secuencia de ácido nucleico reguladora de la expresión”. El término, según se usa en la presente memoria, se refiere a secuencias de ácido nucleico necesarias para y/o que afectan la expresión de una secuencia codificante enlazada funcionalmente en un organismo hospedante particular o *in vitro*. Cuando dos secuencias de ácido nucleico están enlazadas funcionalmente, normalmente estarán en la misma orientación y también en el mismo marco de lectura. Normalmente serán esencialmente contiguas, aunque esto puede no ser necesario. Las secuencias de ácido nucleico que regulan la expresión, tal como *inter alia*, secuencias de iniciación de transcripción, de terminación, promotoras, líderes, péptido señal, propéptido, prepropéptido o potenciadoras adecuadas; secuencia de Shine-Dalgarno, secuencias represoras o activadoras; señales de procesamiento de ARN eficaces tales como señales de empalme y poliadenilación; secuencias que estabilizan el ARNm citoplasmático; secuencias que potencian la eficacia de traducción (p. ej., sitios de unión a ribosoma); secuencias que potencian la estabilidad de la proteína; y, cuando se desea, secuencias que potencian la secreción de la proteína, pueden ser cualquier secuencia de ácido nucleico que exhibe actividad en el organismo hospedante de elección y que se puede derivar de genes que codifican proteínas, que son endógenos o heterólogos con respecto a una célula hospedante. Cada secuencia de control puede ser natural o extraña para la secuencia de ácido nucleico que codifica el polipéptido. Cuando se desea, la secuencia de control se puede proporcionar con enlazadores con el fin de introducir sitios de restricción específicos que facilitan la ligadura de las secuencias de control con la región codificante de la secuencia de ácido nucleico que codifica un polipéptido. Las secuencias de control se pueden optimizar para su fin específico.

Un “producto lácteo” se refiere a cualquier tipo de producto a base de leche previsto para su uso como alimento, pienso o bebida que incluyen, pero no se limitan a, queso, leche, leche desnatada, leche acidificada, suero de mantequilla, leche condensada, pastas para untar, margarinas, yogur, helado, leche en polvo, mantequilla, EMC (siglas en inglés para “queso modificado enzimáticamente”), dulce de leche, blanqueador para café; crema para café, nata, ghee, análogo lácteo, etc. El queso puede ser cualquier tipo de queso, p. ej., queso fresco, queso duro, queso en grano, queso crema, queso con moho blanco, queso con moho azul y queso procesado. Los ejemplos de queso fresco son la ricota, el queso crema, el Neufchatel o el requesón. Los ejemplos de queso duro son Chester, Danbo, Manchego, Saint Paulin, Cheddar, Monterey, Colby, Edam, Gouda, Muenster, de tipo suizo, Gruyere, Emmental, Parmigiano Reggiano, Grana Padano, Parmesano, Pecorino, Provolone y Romano. Los ejemplos de queso en grano son el queso Feta, el queso Quotija, el queso pasta filata tal como la Mozzarella y el Queso fresco. Los ejemplos de queso crema son el queso Filadelfia. Los ejemplos de queso de moho blanco son el queso Brie y el Camembert. Los ejemplos de queso de moho azul son el queso Gorgonzola y el danés azul.

Según se usa en la presente memoria, el término “endógeno” se refiere a una secuencia de ácido nucleico o de aminoácidos que se origina naturalmente en un hospedante.

Las endopeptidasas o endoproteinasas son capaces de descomponer los enlaces peptídicos de aminoácidos no terminales (es decir, dentro de la proteína), a diferencia de las exopeptidasas, que descomponen los enlaces peptídicos del extremo amínico o el carboxílico. Las endopeptidasas no tienden a descomponer los péptidos en monómeros, sino que resultan en fragmentos peptídicos relativamente grandes. La generación específica de fragmentos relativamente grandes se prefiere en gran medida en muchas aplicaciones relacionadas con alimentos y piensos. Un caso específico de endopeptidasa es la oligopeptidasa, cuyos sustratos son oligopéptidos en lugar de proteínas.

El término “expresión” incluye cualquier etapa relacionada con la producción del polipéptido que incluye, pero no se limita a, transcripción, modificación postranscripción, traducción, modificación postraducción y secreción.

Los polinucleótidos de la presente invención, según se describen en la presente memoria, se pueden sobreexpresar en una célula hospedante de la invención en comparación con una célula genitora en la que dicho gen no se sobreexpresa. La sobreexpresión de una secuencia polinucleotídica se define en la presente memoria como la expresión de dicho gen de secuencia que resulta en una actividad del polipéptido codificado por la dicha secuencia en una célula hospedante que es al menos 1,1, al menos 1,25 o al menos 1,5 veces mayor que la actividad del polipéptido en la célula hospedante; preferiblemente, la actividad de dicho polipéptido es al menos 2 veces, más preferiblemente, al menos 3 veces, más preferiblemente, al menos 4 veces, más preferiblemente, al menos 5 veces, incluso más preferiblemente, al menos 10 veces y, lo más preferiblemente, al menos 20 veces mayor que la actividad del polipéptido en la célula genitora.

Un vector de expresión comprende un polinucleótido que codifica un polipéptido, tal como un polipéptido según la presente invención, enlazado funcionalmente con las secuencias de control adecuadas (tal como un promotor, y

señales de terminación de la transcripción y traducción) para la expresión y/o traducción *in vitro*, o en una célula hospedante del polinucleótido.

5 El vector de expresión puede ser cualquier vector (p. ej., un plásmido o virus), que se puede someter convenientemente a procedimientos de ADN recombinante y provocar la expresión del polinucleótido. La elección del vector típicamente dependerá de la compatibilidad del vector con la célula en la que el vector se va a introducir. Los vectores pueden ser plásmidos lineales o circulares cerrados. El vector puede ser un vector que se replica de manera autónoma, es decir, un vector, que existe como una entidad extracromosómica, cuya replicación es independiente de la replicación cromosómica, p. ej., un plásmido, un elemento extracromosómico, un minicromosoma o un cromosoma artificial. Alternativamente, el vector puede ser uno que, cuando se introduce en la célula hospedante, se integra en el genoma y se replica junto con el(los) cromosoma(s) en el(los) que se ha integrado. El vector de clonación integradora se puede integrar en un locus aleatorio o uno diana predeterminado en los cromosomas de la célula hospedante. El sistema de vector puede ser un vector o plásmido simple o dos o más vectores o plásmidos, que juntos contienen el ADN total que se va a introducir en el genoma de la célula hospedante, o un transposón. Un vector de la invención puede comprender uno, dos o más, por ejemplo, tres, cuatro o cinco polinucleótidos de la invención, por ejemplo, para la sobreexpresión.

15 El término "gen", según se usa en la presente memoria, hace referencia a un segmento de una molécula de ácido nucleico que codifica una cadena polipeptídica, que puede o no incluir secuencias reguladoras para el gen antes y después de la secuencia codificante, p. ej., promotores, potenciadores, etc., así como secuencias intercaladas (intrones) entre segmentos codificantes individuales (exones). Se apreciará, además, que la definición de gen puede incluir ácidos nucleicos que no codifican un polipéptido, sino que proporcionan plantillas para la transcripción de moléculas de ARN funcionales tales como ARNt, ARNr, etc.

20 Una célula hospedante, según se define en la presente memoria, es un organismo adecuado para manipulación genética y que se puede cultivar a densidades celulares útiles para producción industrial de un producto diana, tal como un polipéptido según la presente invención. Una célula hospedante puede ser una célula hospedante que se encuentra en la naturaleza o una célula hospedante derivada de una célula hospedante genitora después de la manipulación genética o mutagénesis clásica. De manera ventajosa, una célula hospedante es una célula hospedante recombinante. Una célula hospedante puede ser una célula hospedante procariota, arqueobacteriana o eucariota. Una célula hospedante procariota puede ser, pero no se limita a, una célula hospedante bacteriana. Una célula hospedante eucariota puede ser, pero no se limita a, una célula hospedante de levadura, de hongo, de ameba, de alga, de planta, de animal o de insecto.

25 El término "heterólogo", según se usa en la presente memoria, se refiere a secuencias de ácido nucleico o de aminoácidos que no se originan naturalmente en una célula hospedante. En otras palabras, la secuencia de ácido nucleico o de aminoácidos no es idéntica a la que se encuentra naturalmente en la célula hospedante.

30 El término "hibridación" significa el apareamiento de cadenas sustancialmente complementarias de compuestos oligoméricos, tales como compuestos de ácido nucleico. La hibridación se puede llevar a cabo en condiciones de rigurosidad baja, media o alta. Las condiciones de hibridación de rigurosidad baja comprenden hibridación en 6X cloruro de sodio/citrato de sodio (SSC, por sus siglas en inglés) a aproximadamente 45 °C, posteriormente dos lavados en 0,2 X de SSC, SDS al 0,1 % a al menos 50 °C (la temperatura de los lavados se puede aumentar hasta 55 °C para condiciones de rigurosidad baja). Las condiciones de hibridación de rigurosidad media comprenden hibridación en 6X de SSC a aproximadamente 45 °C, posteriormente uno o más lavados en 0,2 X de SSC, SDS al 0,1 % a 60 °C y las condiciones de hibridación de rigurosidad alta comprenden hibridación en 6X de SSC a aproximadamente 45 °C, posteriormente uno o más lavados en 0,2X de SSC, SDS al 0,1 % a 65 °C.

35 Una secuencia de ácido nucleico o polinucleotídica se define en la presente memoria como un polímero nucleotídico que comprende al menos 5 unidades de nucleótido o ácido nucleico. Un nucleótido o ácido nucleico se refiere a ARN y ADN. Los términos "ácido nucleico" y "secuencia polinucleotídica" se usan de manera intercambiable en la presente memoria.

40 Un "péptido" se refiere a una cadena corta de residuos aminoácidos enlazados mediante enlaces peptídicos (amida). El péptido más corto, un dipéptido, consiste en 2 aminoácidos unidos mediante un enlace peptídico simple.

45 El término "polipéptido" se refiere a una molécula que comprende residuos aminoácidos enlazados mediante enlaces peptídicos y que contiene más de cinco residuos aminoácidos. El término "proteína", según se usa en la presente memoria, es sinónimo del término "polipéptido" y también puede hacer referencia a dos o más polipéptidos. Por lo tanto, los términos "proteína" y "polipéptido" se pueden usar de manera intercambiable. Los polipéptidos opcionalmente se pueden modificar (p. ej., glucosilar, fosforilar, acilar, farnesilar, prenilar, sulfonar y similares) para agregar funcionalidad. Los polipéptidos que exhiben actividad en presencia de un sustrato específico en ciertas condiciones pueden denominarse enzimas. Se entenderá que, como resultado de la degeneración del código genético, se puede producir una multitud de secuencias nucleotídicas que codifican un polipéptido dado.

Un "fragmento de ácido nucleico aislado" es un fragmento de ácido nucleico que no se origina naturalmente como un fragmento y no se encontraría en estado natural.

5 El término "polipéptido aislado", según se usa en la presente memoria, significa un polipéptido que se retira de al menos un componente, p. ej., otro material polipeptídico, con el cual se asocia naturalmente. El polipéptido aislado puede estar libre de cualesquiera otras impurezas. El polipéptido aislado puede ser al menos 50 % puro, p. ej., al menos 60 % puro, al menos 70 % puro, al menos 75 % puro, al menos 80 % puro, al menos 85 % puro, al menos 80 % puro, al menos 90 % puro, o al menos 95 % puro, 96 %, 97 %, 98 %, 99 %, 99,5 %, 99,9 %, según se determina mediante SDS-PAGE o cualquier otro método analítico adecuado con este fin y conocido para el experto en la técnica. Un polipéptido aislado se puede producir mediante una célula hospedante recombinante.

10 Un "polipéptido maduro" se define en la presente memoria como un polipéptido en su forma final y se obtiene después de la traducción de un ARNm en polipéptido y las modificaciones postraducción de dicho polipéptido. La modificación postraducción incluye procesamiento del extremo N, truncamiento del extremo C, glucosilación, fosforilación y extracción de secuencias líderes tales como péptidos señal, propéptidos y/o prepropéptidos mediante escisión.

15 Una "secuencia codificante de polipéptido maduro" significa un polinucleótido que codifica un polipéptido maduro.

20 El término "construcción de ácido nucleico" se refiere en la presente memoria a una molécula de ácido nucleico, ya sea mono o bicatenaria, que se aísla de un gen de origen natural o que ha sido modificada para que contenga segmentos de ácido nucleico que se combinan y yuxtaponen de una manera que no existiría de cualquier otra manera en la naturaleza. El término construcción de ácido nucleico es sinónimo del término "casete de expresión" o "vector de expresión" cuando la construcción de ácido nucleico contiene todas las secuencias de control necesarias para la expresión de una secuencia codificante, en donde dichas secuencias de control se enlazan funcionalmente con dicha secuencia codificante.

25 Una "endoproteasa específica para prolina" es una proteasa que hidroliza una proteína o péptido en una posición donde la proteína o péptido contiene un residuo prolina. Una endoproteasa específica para prolina puede tener actividad de endoproteasa específica para prolina y/o de oligopeptidasa específica para prolina (EC3.4.21.26). Una endoproteasa específica para prolina es preferiblemente una enzima que hidroliza un enlace peptídico en el extremo carboxílico de residuos prolina, que resulta en un fragmento de péptido y/o polipéptido con una prolina en el extremo C.

30 El término "promotor" se define en la presente memoria como una secuencia de ADN que se une a ARN polimerasa y dirige la polimerasa al sitio de inicio de la transcripción posterior correcto de una secuencia de ácido nucleico para iniciar la transcripción.

35 El término "recombinante" cuando se usa en referencia a una célula, ácido nucleico, proteína o vector, indica que la célula, ácido nucleico, proteína o vector se ha modificado mediante la introducción de un ácido nucleico, aminoácido o proteína heterólogo o la alteración de un ácido nucleico o proteína natural, o que la célula deriva de una célula modificada de esta forma. Por lo tanto, por ejemplo, las células recombinantes expresan genes que no se encuentran dentro de la forma natural (no recombinante) de la célula o expresan genes naturales que de cualquier otra manera se expresan de manera anormal, se subexpresan o no se expresan en absoluto. El término "recombinante" es sinónimo de "genéticamente modificado" y "transgénico".

40 "Identidad secuencial", u homología secuencial se usan de manera intercambiable en la presente memoria. A los efectos de la presente invención, se define en la presente que para determinar el porcentaje de homología secuencial o identidad secuencial de dos secuencias de aminoácidos o de dos secuencias de ácido nucleico, las secuencias se alinean para una comparación óptima. Para optimizar la alineación entre las dos secuencias, se pueden introducir espacios en cualquiera de las dos secuencias que se comparan. Dicha alineación se puede llevar a cabo a lo largo de la longitud completa de las secuencias que se comparan. Alternativamente, la alineación se puede llevar a cabo a lo largo de una longitud más corta, por ejemplo, a lo largo de aproximadamente 20, aproximadamente 50, aproximadamente 100 o más ácidos nucleicos/bases o aminoácidos. La identidad secuencial es el porcentaje de coincidencias idénticas entre las dos secuencias a lo largo de la región alineada informada. El porcentaje de identidad secuencial entre dos secuencias de aminoácidos o entre dos secuencias nucleotídicas se puede determinar usando el algoritmo de Needleman y Wunsch para la alineación de dos secuencias. (Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). Las secuencias de aminoácidos y las secuencias nucleotídicas se pueden alinear mediante el algoritmo. El algoritmo de Needleman-Wunsch se ha implementado en el programa informático NEEDLE. A los efectos de la presente invención, se usó el programa NEEDLE del paquete EMBOSS (versión 2.8.0 o más alta, EMBOSS: The European Molecular Biology Open Software Suite (2000) Rice, P. Longden, I. y Bleasby, A. Trends in Genetics 16, (6) págs. 276-277, <http://emboss.bioinformatics.nl/>). Para las secuencias proteicas, se usa EBLOSUM62 para la matriz de sustitución. Para la secuencia nucleotídica se usa EDNAFULL. Los parámetros opcionales que se usan son una penalización por espacio abierto de 10 y una penalización por extensión de espacio de 0,5. El experto apreciará que todos estos parámetros diferentes

proporcionarán resultados ligeramente diferentes, pero que el porcentaje de identidad global de dos secuencias no se altera significativamente cuando se usan algoritmos diferentes.

Después de la alineación mediante el programa NEEDLE, según se describió anteriormente, se calcula el porcentaje de identidad secuencial entre una secuencia de consulta y una secuencia de la invención de la siguiente manera:
 5 Cantidad de posiciones correspondientes en la alineación que exhiben un aminoácido idéntico o nucleótido idéntico en ambas secuencias dividida entre la longitud total de la alineación después de restar la cantidad total de espacios en la alineación. La identidad, según se define en la presente memoria, se puede obtener a partir de NEEDLE mediante el uso de la opción NOBRIEF y se etiqueta en el resultado del programa como "identidad más larga".

Las secuencias de ácido nucleico y proteína de la presente invención se pueden usar adicionalmente como una
 10 "secuencia de consulta" para llevar a cabo una búsqueda en bases de datos públicas para, por ejemplo, identificar otros miembros de la familia o secuencias relacionadas. Dichas búsquedas se pueden llevar a cabo mediante el uso de los programas NBLAST y XBLAST (versión 2.0) de Altschul, et al. (1990) J. Mol. Biol. 215:403-10. Las búsquedas de nucleótidos por BLAST se pueden llevar a cabo con el programa NBLAST, puntuación = 100, longitud de palabra= 12 para obtener secuencias nucleotídicas homólogas a las moléculas de ácido nucleico de la invención.
 15 Las búsquedas de proteína por BLAST se pueden llevar a cabo con el programa XBLAST, puntuación = 50, longitud de palabra= 3 para obtener secuencias de aminoácidos homólogas a las moléculas de proteína de la invención. Para obtener alineaciones especiadas para hacer comparaciones, se puede usar Gapped BLAST según se describe en Altschul et al., (1997) Nucleic Acids Res. 25(17): 3389-3402. Cuando se usan los programas BLAST y Gapped BLAST, se pueden usar los parámetros predeterminados de los respectivos programas (p. ej., XBLAST y NBLAST).
 20 Véase la página de inicio del National Center for Biotechnology Information en <http://www.ncbi.nlm.nih.gov/>.

El término "sustancialmente puro" con respecto a polipéptidos se refiere a una preparación polipeptídica que contiene como mucho 50 % en peso de otro material polipeptídico. Los polipéptidos descritos en la presente memoria están preferiblemente en una forma sustancialmente pura. En particular, se prefiere que los polipéptidos descritos en la presente memoria estén en "forma esencialmente pura", es decir, que la preparación polipeptídica
 25 esté esencialmente libre de otro material polipeptídico. Opcionalmente, el polipéptido también puede estar esencialmente libre de material no polipeptídico, tal como ácidos nucleicos, lípidos, componentes de medios y similares. En la presente memoria, el término "polipéptido sustancialmente puro" es sinónimo de los términos "polipéptido aislado" y "polipéptido en forma aislada". El término "sustancialmente puro" con respecto a un polinucleótido se refiere a una preparación polinucleotídica que contiene como mucho 50 % en peso de otro material polinucleotídico. Los polinucleótidos descritos en la presente memoria están preferiblemente en una forma sustancialmente pura. En particular, se prefiere que el polinucleótido descrito en la presente memoria esté en "forma esencialmente pura", es decir, que la preparación polinucleotídica esté esencialmente libre de otro material polinucleotídico. Opcionalmente, el polinucleótido también puede estar esencialmente libre de material no polinucleotídico, tal como polipéptidos, lípidos, componentes de medios y similares. En la presente memoria, el
 30 término "polinucleótido sustancialmente puro" es sinónimo de los términos "polinucleótido aislado" y "polinucleótido en forma aislada".

Una "sustitución", según se usa en la presente memoria, en relación con polipéptidos o ácidos nucleicos, denota el reemplazo de uno o más aminoácidos en una secuencia polipeptídica o de uno o más nucleótidos en una secuencia polinucleotídica, respectivamente, por diferentes aminoácidos o nucleótidos, respectivamente. Por ejemplo, una
 40 sustitución indica que una posición en un polipéptido, según se describe en la presente memoria, tal como un polipéptido variante, que corresponde a al menos una posición establecida anteriormente en la SEQ ID NO: 1, comprende un residuo aminoacídico que no aparece en dicha posición en el polipéptido genitor (por ejemplo, la secuencia genitora de la SEQ ID NO: 1).

Una "molécula sintética", tal como un ácido nucleico sintético o un polipéptido sintético se produce mediante síntesis química o enzimática in vitro. Incluye, pero no se limita a, ácidos nucleicos variantes producidos con uso óptimo de
 45 codones para organismos hospedantes de elección.

Un ácido nucleico sintético se puede optimizar para el uso de codones, preferiblemente, según los métodos descritos en WO2006/077258 y/o WO2008000632, que se incorporan en la presente memoria por referencia. WO2008/000632 se refiere a la optimización de un par de codones. La optimización de un par de
 50 codones es un método en donde las secuencias nucleotídicas que codifican un polipéptido que se han modificado con respecto a su uso de condones, en particular, los pares de codones que se usan, se optimizan para obtener expresión mejorada de la secuencia nucleotídica que codifica el polipéptido y/o producción mejorada del polipéptido codificado. Los pares de codones se definen como un conjunto de dos tripletes posteriores (codones) en una secuencia codificante. Los expertos en la técnica sabrán que el uso de codones debe adaptarse dependiendo de la especie hospedante y posiblemente resulta en variantes con desviación de homología significativa con respecto a la
 55 SEQ ID NO: 2, pero que todavía codifican el polipéptido según la invención.

Según se usan en la presente memoria, los términos “variante”, “derivado”, “mutante” u “homólogo” se pueden usar de manera intercambiable. Pueden hacer referencia a polipéptidos o ácidos nucleicos. Las variantes incluyen sustituciones, inserciones, eliminaciones, truncamientos, transversiones e/o inversiones, en una o más ubicaciones con respecto a una secuencia de referencia. Las variantes se pueden producir, por ejemplo, mediante mutagénesis de saturación de sitio, mutagénesis de exploración, mutagénesis de inserción, mutagénesis aleatoria, mutagénesis dirigida al sitio y evolución dirigida, así como diversas otras estrategias de recombinación conocidas para un experto en la técnica. Los genes variantes de ácidos nucleicos se pueden sintetizar artificialmente mediante técnicas conocidas en la técnica.

Figuras

10 Figura 1: vector pGBTOP-16 usado para clonar el gen GLA. El vector pGBTOP-16 deriva del vector pGBTOP-12 descrito en WO 2011/009700. Además de pGBTOP-12, contiene el gen *ccdB* de *E. coli* para selección positiva para la presencia de un inserto entre los sitios de clonación *EcoRI* y *Pacl*. El sitio de restricción *Pacl* reemplaza el sitio de restricción *SnaBI* presente en pGBTOP-12. Este vector se linealiza mediante digestión por *NotI* antes de la transformación.

15 Secuencias

SEQ ID NO: 1: Secuencia de aminoácidos de la endoproteasa específica para prolina de *Aspergillus niger* que contiene una secuencia señal de pectinametilsterasa.

SEQ ID NO: 2: Secuencia de ácido nucleico de la endoproteasa específica para prolina de *Aspergillus niger* que contiene una secuencia señal de pectinametilsterasa.

20 SEQ ID NO: 3 Secuencia de aminoácidos de citocromo C de corazón de caballo.

SEQ ID NO: 4: Fragmento de citocromo C digerido con una PEP según la presente invención

SEQ ID NO:5: Fragmento de citocromo C digerido con una PEP según la presente invención

SEQ ID NO: 6: Fragmento de citocromo C digerido con una PEP según la presente invención

SEQ ID NO:7: Fragmento de citocromo C digerido con una PEP según la presente invención

25 SEQ ID NO: 8: Fragmento de citocromo C digerido con una PEP según la presente invención

Descripción detallada

En un aspecto, la presente descripción se refiere a un polipéptido que tiene actividad de endoproteasa específica para prolina, en donde el polipéptido tiene menos de 70 % de actividad residual con el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min. La actividad de endoproteasa específica para prolina residual se mide mediante el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) a pH 4,5, por ejemplo, en un tampón de acetato de sodio a pH 4,5 a 20 grados Celsius. De manera sorprendente, un polipéptido que tiene menos de 50 % de actividad residual después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min, se puede usar de manera ventajosa en aplicaciones tales como en alimentos o piensos, en donde se desea escasa o nula actividad residual. Preferiblemente, un polipéptido proporcionado por la invención tiene menos de 45 %, 40 %, 30 %, 20 %, 15 %, 10 %, tal como menos de 5 % de actividad residual después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min. Según se define en la presente memoria, menos de 70%, 60 %, 50 %, o menos de 45 %, 40 %, 30 %, 20 %, 15 %, 10 % o 5 % de actividad residual significa que el polipéptido exhibe menos de 70%, 60%, 50 %, o menos de 45 %, 40 %, 30 %, 20 %, 15 %, 10 % o 5 %, respectivamente, de actividad en comparación con la actividad del polipéptido antes de mantener al polipéptido a 65 °C durante 15 min. Preferiblemente, un polipéptido según la presente invención no exhibe ninguna actividad residual después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min.

En una realización, un polipéptido según se describe en la presente memoria es un polipéptido que tiene actividad de endoproteasa específica para prolina, en donde el polipéptido tiene menos de 90 % de actividad residual con el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 60 °C durante 15 min. La actividad de endoproteasa específica para prolina residual se mide mediante el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) a pH 4,5, por ejemplo, en un tampón de acetato de sodio a pH 4,5 a 20 grados Celsius. De manera sorprendente, un polipéptido que tiene menos de 70 % de actividad residual después de que el polipéptido se ha mantenido a una temperatura de 60 °C durante 15 min, se puede usar de manera ventajosa en aplicaciones tales como en alimentos o piensos, en donde se desea escasa o nula actividad residual. Preferiblemente, un polipéptido proporcionado en la presente memoria tiene menos de 80 %, 70 %, 60 %, 50 %, 40 %, 30 %, 20 %, 15 %, 10 %, tal como menos de 5 % de actividad residual después de que el polipéptido se ha mantenido a una temperatura de 60 °C durante 15 min. Según se define en la presente

memoria, menos de 70 %, o menos de 60 %, 50 %, 45 %, 40 %, 30 %, 20 %, 15 %, 10 % o 5 % de actividad residual significa que el polipéptido exhibe menos de 70 %, 60 %, 50 %, o menos de 40 %, 30 %, 20 %, 15 %, 10 % o 5 %, respectivamente, de actividad en comparación con la actividad del polipéptido antes de mantener al polipéptido a 60 °C durante 15 min. Preferiblemente, un polipéptido según la presente invención no exhibe ninguna actividad residual después de que el polipéptido se ha mantenido a una temperatura de 60 °C durante 15 min.

La invención también proporciona un polipéptido que tiene actividad de endoproteasa específica para prolina, opcionalmente que tiene menos de 70 % de actividad residual con el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAp-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min, u opcionalmente que tiene menos de 90 % de actividad residual con el uso de acetil-AlaAlaPro-paranitroanilina (Ac-AAp-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 60 °C durante 15 min, en donde el polipéptido se selecciona del grupo que consiste en

i. comprende al menos una combinación de aminoácidos que corresponde a la combinación seleccionada del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición 204, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición 304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), y opcionalmente al menos un aminoácido adicional de Phe (F) en la posición 204, Asp (N) en la posición 205 o Ala (A) en la posición 477, en el que el polipéptido no comprende el aminoácido Phe (F) en la posición 204, cuando el polipéptido comprende los aminoácidos Glu (E) en la posición 238, Val (V) en la posición 204, y Ala (A) en la posición 460, en el que las posiciones se definen con referencia a SEQ ID NO:1;

ii. un polipéptido que comprende una secuencia de aminoácidos según la SEQ ID NO: 1, en donde la SEQ ID NO: 1 comprende al menos una combinación de sustituciones de aminoácidos seleccionada del grupo que consiste en las combinaciones (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) y opcionalmente al menos una sustitución de aminoácido adicional I204F, Y205N o P477A, en la que la sustitución de aminoácido I204F no está presente en un polipéptido que tiene las sustituciones de aminoácidos K238E, I204V y V460A en las que las sustituciones se definen con referencia a SEQ ID NO:1;

iii. un polipéptido según i) a iii), pero que carece de una secuencia señal y/o una secuencia propeptina;

iv. un polipéptido según i) a iv), en donde el polipéptido tiene al menos 90%, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100% de identidad con respecto a la secuencia de aminoácidos según la SEQ ID NO: 1;

v. un polipéptido codificado por un ácido nucleico que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % o 100% de identidad con respecto a la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácidos seleccionada del grupo (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) y opcionalmente al menos una sustitución de aminoácido adicional I204F, Y205N o P477A, en la que la sustitución del aminoácido I204F no está presente en un polipéptido que tiene las sustituciones de aminoácidos K238E, I204V y V460A, en el que las posiciones de los aminoácidos se definen con referencia a SEQ ID NO:1.

Como se usa en la presente memoria, cuando se alinea un polipéptido con una secuencia de endoproteasa específica de prolina de SEQ ID NO:1, un polipéptido de la presente invención comprenderá al menos una combinación de aminoácidos seleccionados del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición 204, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición 304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), y opcionalmente al menos un aminoácido adicional de Phe (F) en la posición 204, Asp (N) en la posición 205 o Ala (A) en la posición 477, en el que el polipéptido no comprende el aminoácido Phe (F) en la posición 204, cuando el polipéptido comprende los aminoácidos Glu (E) en la posición 238, Val (V) en la posición 204, y Ala (A) en la posición 460, en el que las posiciones se definen con referencia a SEQ ID NO:1.

Aquellas posiciones en un polipéptido de la invención, que puede ser un polipéptido recombinante, sintético o variante, que corresponden a las posiciones establecidas anteriormente en la SEQ ID NO: 1 se puede identificar al alinear la secuencia del polipéptido de la presente invención con la de la SEQ ID NO: 1 mediante el uso de, por ejemplo, la alineación mediante el programa Needle. Las posiciones en el polipéptido de la presente invención que

corresponden a las posiciones en la SEQ ID NO: 1, según se establecieron anteriormente, se pueden identificar, por lo tanto, y se hace referencia a ellas como las posiciones definidas con respecto a la SEQ ID NO: 1.

La presente invención también proporciona un polipéptido que es un polipéptido aislado, sustancialmente puro, puro, recombinante, sintético o variante del polipéptido según se describe en la presente memoria.

5 De forma ventajosa un polipéptido proporcionado por la invención comprende al menos una combinación de aminoácidos que corresponde a la combinación seleccionada del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición 304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), y opcionalmente al menos un aminoácido adicional de Phe (F) en la posición 204, Asp (N) en la posición 205 o Ala (A) en la posición 477, en el que el polipéptido no comprende el aminoácido Phe (F) en la posición 204, cuando el polipéptido comprende los aminoácidos Glu (E) en la posición 238, Val (V) en la posición 204, y Ala (A) en la posición 460, en el que las posiciones se definen con referencia a SEQ ID NO:1.

En una realización, la invención proporciona un polipéptido que tiene actividad endoproteasa específica de prolina que comprende una secuencia de aminoácidos según SEQ ID NO:1, en la que SEQ ID NO:1 comprende al menos una combinación de sustituciones de aminoácidos seleccionados del grupo que consiste en las combinaciones (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A), (P466T, P469Q), (L470H, E387G, Y205N), (P466T, P469Q, I204F) y (P466T, P469Q, P477A), en el que las sustituciones se definen con referencia a SEQ ID NO:1.

Preferiblemente, la invención proporciona un polipéptido que comprende al menos una combinación de sustituciones de aminoácidos como se define en la presente invención anterior, en la que el polipéptido tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad a la secuencia de aminoácidos según SEQ ID NO:1. Por consiguiente, la presente invención proporciona un polipéptido que tiene al menos 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% o 100% de identidad a la secuencia de aminoácidos según la SEQ ID NO:1 o al polipéptido maduro de SEQ ID NO:1, en la que SEQ ID NO:1 comprende al menos una combinación de sustituciones de aminoácidos como se define en la presente memoria anterior.

Un polipéptido que tiene actividad endoproteasa específica de prolina que, cuando se alinea con una secuencia de aminoácidos según SEQ ID NO:1 comprende al menos una combinación de sustituciones de aminoácidos como se define en la presente memoria anterior, puede comprender sustituciones, supresiones y/o inserciones adicionales en una o más posiciones de aminoácidos adicionales. Por ejemplo un polipéptido como se describe en la presente memoria puede ser una variante del polipéptido o el polipéptido maduro de SEQ ID NO:1 que comprende una sustitución, supresión o inserción en una posición seleccionada del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición 204, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición 304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), y opcionalmente al menos un aminoácido adicional de Phe (F) en la posición 204, Asp (N) en la posición 205 o Ala (A) en la posición 477, en el que el polipéptido no comprende el aminoácido Phe (F) en la posición 204, cuando el polipéptido comprende los aminoácidos Glu (E) en la posición 238, Val (V) en la posición 204, y Ala (A) en la posición 460, como se define con referencia a SEQ ID NO:1, y que tiene además 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 o 12 o más sustituciones, supresiones y/o inserciones amino adicionales, por lo que el polipéptido aún tiene la actividad o función del polipéptido de la invención. El experto apreciará que estos cambios de aminoácidos menores en el polipéptido de la invención pueden estar presentes (por ejemplo, mutaciones de origen natural) o producirse (por ejemplo, usando tecnología de r-ADN) sin la pérdida de la función o actividad de la proteína. En el caso de que estas mutaciones estén presentes en un dominio de unión, sitio activo u otro dominio funcional del polipéptido, se puede cambiar una propiedad del polipéptido, pero el polipéptido puede mantener su actividad. En el caso de que una mutación esté presente y no esté cerca del sitio activo, dominio de unión u otro dominio funcional, se puede esperar un efecto menor.

También se pueden identificar equivalentes funcionales de un polipéptido según la invención, p. ej., al someter a barrido bibliotecas combinatorias de mutantes, p. ej., mutantes de truncamiento, del polipéptido de la invención para determinar la actividad biológica del polipéptido de la invención. En una realización, se genera una biblioteca heterogénea de variantes mediante mutagénesis combinatoria a nivel de ácido nucleico. Se puede producir una biblioteca heterogénea de variantes, por ejemplo, al ligar enzimáticamente una mezcla de oligonucleótidos sintéticos con secuencias de gen de manera que un conjunto degenerado de posibles secuencias proteicas se pueda expresar

como polipéptidos individuales, o alternativamente, como un conjunto de proteínas de fusión más grandes (p. ej., para visualización en fagos). Existe una variedad de métodos que se pueden usar para producir bibliotecas de posibles variantes de los polipéptidos de la invención a partir de una secuencia de oligonucleótidos degenerada. Los métodos para sintetizar oligonucleótidos degenerados se conocen en la técnica (véase, p. ej., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477). El término "secuencia de ácido nucleico degenerada" o "secuencia de (oligo)nucleótido degenerada" denota una secuencia de ácidos nucleicos que incluye uno o más codones degenerados (en comparación con una molécula de ácido nucleico de referencia que codifica un polipéptido). Los codones degenerados contienen tripletes de ácidos nucleicos diferentes, pero codifican el mismo residuo aminoacídico (es decir, los tripletes GAU y GAC codifican cada uno Asp). La degeneración del codón se refiere a la naturaleza del código genético que permite la variación de la secuencia de ácido nucleico sin afectar la secuencia de aminoácidos de un polipéptido codificado. El experto conoce sobre la "preferencia codónica" que exhibe una célula hospedante específica en el uso de codones de ácido nucleico para especificar un aminoácido dado.

Además, se pueden usar bibliotecas de fragmentos de la secuencia codificante de un polipéptido de la invención para generar una población heterogénea de polipéptidos para someter a barrido una selección posterior de variantes. Por ejemplo, se puede generar una biblioteca de fragmentos de la secuencia codificante al tratar un fragmento de PCR bicatenario de la secuencia codificante de interés con una nucleasa en condiciones en donde el mellado se produce solo aproximadamente una vez por molécula, desnaturalizar el ADN bicatenario, renaturalizar el ADN para formar un ADN bicatenario que puede incluir pares codificantes/no codificantes de productos mellados diferentes, retirar porciones monocatenarias de las estructuras dobles reformadas mediante tratamiento con S1 nucleasa y ligar la biblioteca de fragmentos resultante con un vector de expresión. Mediante este método, se puede derivar una biblioteca de expresión que codifica fragmentos del extremo N e internos de diversos tamaños de la proteína de interés.

En la técnica, se conocen diversas técnicas para barrer productos génicos de bibliotecas combinatorias creadas por mutaciones de truncamiento puntuales y para barrer bibliotecas de ADNc para productos génicos que tienen una propiedad seleccionada. Las técnicas más ampliamente usadas, que son adecuadas para análisis de alto rendimiento, para barrer grandes bibliotecas de genes típicamente incluyen clonar la biblioteca de genes en vectores de expresión, transformar células adecuadas con la biblioteca de vectores resultante y expresar los genes combinatorios en condiciones en las que la detección de una actividad deseada facilita el aislamiento del vector que codifica el gen cuyo producto se detectó. Se puede usar mutagénesis de conjunto recursivo (REM, por sus siglas en inglés), una técnica que potencia la frecuencia de mutantes funcionales en las bibliotecas, en combinación con los ensayos de barrido para identificar variantes de una proteína de la invención (Arkin and Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3): 327-331).

Un polipéptido proporcionado por la invención puede carecer de una secuencia señal y/o una secuencia proproteína. Por ejemplo, un polipéptido como se proporciona en la presente memoria puede ser un polipéptido que comprende una secuencia de aminoácidos según la SEQ ID NO: 1 que comprende al menos una combinación de sustituciones de aminoácidos según se definieron anteriormente en la presente memoria y que carece de los primeros 17 aminoácidos que codifican una secuencia señal y/o que carece de los siguientes 19 aminoácidos que codifican una prosequencia. Por consiguiente, un polipéptido proporcionado por la invención puede comprender un polipéptido maduro de la SEQ ID NO: 1, tal como los aminoácidos 37 a 521 de la SEQ ID NO: 1 y que comprende al menos una combinación de sustituciones de aminoácidos según se definieron en la presente memoria, en donde en aminoácido metionina en la posición 1 en la SEQ ID NO: 1 se cuenta como el número 1.

Un polipéptido proporcionado por la invención se puede codificar mediante cualquier ácido nucleico adecuado, tal como un ácido nucleico que tiene al menos 90%, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad con respecto al ácido nucleico según la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácidos seleccionada del grupo (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) en donde las posiciones de aminoácidos se definen con respecto a la SEQ ID NO: 1.

En una realización, un polipéptido codificado por un ácido nucleico de SEQ ID NO:2 o una secuencia de codificación madura de SEQ ID NO:2 como se describe en la presente memoria puede codificarse mediante una secuencia de ácido nucleico, que cuando se alinea con SEQ ID NO:2 comprende al menos una mutación adicional que codifica al menos una sustitución de aminoácido adicional I204F, Y205N o P477A, en la que en la que la sustitución de aminoácido I204F no está presente en un polipéptido que tiene las sustituciones de aminoácido K238E, I204V y V460A, y en la que las sustituciones de aminoácidos se definen con referencia a SEQ IN NO:1.

Típicamente, una secuencia polinucleotídica según se describe en la presente memoria está optimizada por codón, o es una secuencia optimizada por un par de codones para la expresión óptima de un polipéptido según se describe en la presente memoria en una célula hospedante particular.

En una realización, la presente descripción incluye un fragmento biológicamente activo de un polipéptido según se describe en la presente memoria.

Los fragmentos biológicamente activos de un polipéptido de la invención incluyen polipéptidos que comprenden secuencias de aminoácidos suficientemente idénticas a o derivadas de la secuencia de aminoácidos de la proteína endoproteasa específica para prolina (p. ej., la secuencia de aminoácidos madura de SEQ ID NO:1), que incluyen menos aminoácidos que la proteína de longitud completa, pero que exhiben al menos una actividad biológica de la correspondiente proteína de longitud completa. Típicamente, los fragmentos biológicamente activos comprenden un dominio o motivo con al menos una actividad de la proteína endoproteasa específica para prolina. Un fragmento biológicamente activo puede comprender, por ejemplo, un dominio catalítico. Un fragmento biológicamente activo de una proteína de la invención puede ser un polipéptido que tiene, por ejemplo, 10, 25, 50, 100 o más aminoácidos de longitud. Además, se pueden preparar otras porciones biológicamente activas, en las que se eliminan otras regiones de la proteína mediante técnicas recombinantes y se evalúan para determinar una o más de las actividades biológicas de la forma natural de un polipéptido de la invención.

La descripción también incluye fragmentos de ácido nucleico que codifican los fragmentos biológicamente activos mencionados anteriormente de la proteína endoproteasa específica para prolina.

Un polipéptido según la presente invención puede ser una proteína de fusión. Las técnicas para producir polipéptidos de fusión se conocen en la técnica e incluyen ligar las secuencias codificantes que codifican los polipéptidos de manera que están dentro del mismo marco. La expresión del polipéptido fusionado está bajo el control del(de los) mismo(s) promotor(es) y terminador. Los polipéptidos híbridos pueden comprender una combinación de secuencias polipeptídicas parciales o completas obtenidas de al menos dos polipéptidos diferentes en donde una o más pueden ser heterólogas con respecto a una célula hospedante. Dichos polipéptidos de fusión de al menos dos polipéptidos diferentes pueden comprender un dominio de unión de un polipéptido, enlazado funcionalmente con un dominio catalítico de un segundo polipéptido. Se describen ejemplos de polipéptidos de fusión y fusiones de secuencia señal, por ejemplo, en WO2010/121933, WO2013/007820 y WO2013/007821.

Un polipéptido según la presente invención puede derivar de cualquier célula eucariota adecuada. Una célula eucariota puede ser una célula de mamífero, insecto, planta, hongo o alga. La expresión "derivado/a" o "derivable de" con respecto al origen de un polipéptido según se describe en la presente memoria, significa que cuando se lleva a cabo una búsqueda BLAST con un polipéptido según la presente invención, el polipéptido según la presente invención puede ser derivable de una fuente natural, tal como una célula microbiana, de la cual un polipéptido endógeno exhibe el porcentaje de homología o identidad más alto con el polipéptido según se describe en la presente memoria.

Un polipéptido según la presente invención puede derivar de una célula fúngica filamentosa o de una célula fúngica filamentosa termófila. Las células fúngicas filamentosas preferidas pertenecen a una especie de un género *Acremonium*, *Aspergillus*, *Chrysosporium*, *Myceliophthora*, *Penicillium*, *Talaromyces*, *Rasamsonia*, *Thielavia*, *Fusarium* o *Trichoderma*, *Amorphotheca*, *Pseudocercospora*, *Trametes*, *Rhizomucor*, *Calcarisporiella*, *Thermomyces*, *Thermoascus*, *Cornyascus*, *Myricoccum*, *Scytalidium*, *Chaetomium*, *Paecilomyces*, *Corynascus*, *Malbranchea*, *Stilbella*, *Thermomyces*, *Dactylomyces*, *Humicola*, *Chaetomium*, *Melanocarpus*, *Rhizomucor*, *Lentinula*, *Anaeromyces*, y lo más preferiblemente, pertenecen a una especie de *Aspergillus niger*, *Acremonium alabamense*, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus sojae*, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Rasamsonia emersonii*, *Aspergillus oryzae*, *Chrysosporium lucknowense*, *Fusarium oxysporum*, *Myceliophthora thermophila*, *Trichoderma reesei*, *Thielavia terrestris*, *Penicillium chrysogenum*, *Amorphotheca resinae*, *Aureobasidium pullulans*, *Pseudocercospora herpotrichoides*, *Trametes versicolor* 52J, *Rhizomucor pusillus*, *Calcarisporiella thermophila*, *Talaromyces thermophilus*, *Thermomyces lanuginosus*, *Thermoascus auratiacus*, *Cornyascus thermophilus*, *Myricoccum thermophilum*, *Scytalidium thermophilum*, *Myceliophthora hinnulea*, *Chaetomium thermophilum*, *Paecilomyces byssochlamydoides*, *Corynascus sepedonium*, *Malbranchea cinnamomea*, *Thielavia australiensis*, *Stilbella thermophila*, *Thermomyces stellatus*, *Talaromyces emersonii*, *Dactylomyces thermophilus*, *Humicola hyalothermophila*, *Acremonium thermophilum*, *Chaetomium olivicolor*, *Melanocarpus albomyces*, *Rhizomucor miehei*, *Lentinula edodes* o *Anaeromyces mucronatus*. Un polipéptido según la presente invención puede derivar de *Aspergillus niger*, *Aspergillus flavus*, *Aspergillus carbonari*, o *Aspergillus aculeatus*, *Rasamsonia emersonii*.

Un polipéptido según la presente invención puede ser un polipéptido de origen natural o un polipéptido genéticamente modificado o recombinante.

Un polipéptido según se describe en la presente memoria puede estar purificado. La purificación de proteínas es conocida para un experto en la técnica. Un método conocido para la purificación de proteínas es la cromatografía líquida de alto rendimiento.

Los polipéptidos según la presente invención pueden tener de manera ventajosa una propiedad mejorada. Una propiedad mejorada puede ser una actividad específica mejorada y/o una sensibilidad a temperatura aumentada en comparación con un polipéptido que no comprende una sustitución de aminoácido según se define en la presente

memoria, o cualquier otra propiedad mejorada, por ejemplo, deseable en el procesamiento de un alimento o pienso. De manera ventajosa, un polipéptido según se describe en la presente memoria tiene menos de 70 % de actividad residual sobre acetil-AlaAlaPro-paranitroanilina (Ac-AAp-pNA) como sustrato cuando el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min.

5 Los polipéptidos de la invención se pueden obtener mediante diversos procedimientos conocidos para un experto en la técnica, tales como:

1. PCR propensa a error para introducir mutaciones aleatorias y posteriormente un barrido de los polipéptidos (variantes) obtenidos y aislamiento de polipéptido(s) (variante(s)) con propiedades mejoradas

10 2. Redistribución por familia de variantes relacionadas de los genes que codifican el polipéptido según la invención y posteriormente un barrido de las variantes obtenidas y aislamiento de las variantes con propiedades mejoradas

Se pueden obtener variantes de genes que codifican un polipéptido de la presente invención que conducen a un nivel aumentado de ARNm y/o proteína, que resultan en mayor actividad, al modificar las secuencias polinucleotídicas de dichos genes. Entre dichas modificaciones se incluyen:

15 1. Mejorar el uso de codones de tal manera que los codones estén (óptimamente) adaptados al hospedante microbiano genitor.

2. Mejorar el uso de pares de codones de tal manera que los codones estén (óptimamente) adaptados al hospedante microbiano genitor.

3. Adición de secuencias estabilizantes a la información genómica que codifica un polipéptido según la invención que resulta en moléculas de ARNm con una semivida aumentada

20 Se describen métodos para aislar variantes con propiedades catalíticas mejoradas o niveles aumentados de ARNm o proteína en WO03/010183 y WO03/01311. Se describen métodos para optimizar el uso de codones en cepas microbianas genitoras, por ejemplo, en WO2008/000632. Se describen métodos para la adición de elementos estabilizantes en los genes que codifican el polipéptido de la invención en WO2005/059149.

25 Por consiguiente, en un aspecto, la invención proporciona un método para generar un polipéptido variante, en donde el método comprende

i. seleccionar un polipéptido genitor que comprende al menos 90%, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % o 100 % de identidad con respecto a la secuencia de aminoácidos según la SEQ ID NO: 1, o con respecto a un polipéptido maduro de la SEQ ID NO: 1; y,

30 ii. sustituir al menos una combinación de aminoácidos que dan por resultado una combinación de aminoácidos seleccionados del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición 304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), y opcionalmente al menos un aminoácido adicional de Phe (F) en la posición 204, Asp (N) en la posición 205 o Ala (A) en la posición 477, en el que el polipéptido no comprende el aminoácido Phe (F) en la posición 204, cuando el polipéptido comprende los aminoácidos Glu (E) en la posición 238, Val (V) en la posición 204, y Ala (A) en la posición 460, en el que las posiciones se definen con referencia a SEQ ID NO:1; y

40 iii. generar el polipéptido variante,

en donde, opcionalmente, el polipéptido variante tiene menos de 70 % de actividad residual con el uso de Ac-AAp-pNA como sustrato después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min.

45 Generar un polipéptido variante según se describe en la presente memoria puede incluir expresar un gen que codifica el polipéptido variante en una célula hospedante adecuada (recombinante) y cultivar la célula hospedante para generar el polipéptido variante.

En otro aspecto, la presente invención proporciona una composición que comprende un polipéptido según se describe en la presente memoria.

50 Una composición según se describe en la presente memoria puede comprender un vehículo, un excipiente, una enzima auxiliar u otros compuestos. Típicamente, una composición o una formulación comprende un compuesto con el cual se puede formar una endoproteasa específica para prolina. Un excipiente, según se usa en la presente memoria, es una sustancia inactiva formulada junto con a polipéptido según se describe en la presente memoria, por

- ejemplo, sacarosa o lactosa, glicerol, sorbitol o cloruro de sodio. Una composición que comprende un polipéptido según se describe en la presente memoria puede ser una composición líquida o una composición sólida. Una composición líquida normalmente comprende agua. Cuando se formula como una composición líquida, la composición normalmente comprende componentes que bajan la actividad del agua, tal como glicerol, sorbitol o cloruro de sodio (NaCl). Una composición sólida que comprende un polipéptido según se describe en la presente memoria puede comprender un granulado que comprende la enzima o la composición comprende un polipéptido encapsulado en matrices líquidas como liposomas o geles como alginato o carragenanos. Existen muchas técnicas conocidas en la técnica para encapsular o granular un polipéptido o enzima (véase, por ejemplo, G.M.H. Meesters, "Encapsulation of Enzymes and Peptides", Capítulo 9, en N.J. Zuidam and V.A. Nedović (eds.) "Encapsulation Technologies for Active Food Ingredients and food processing" 2010). Una composición según se describe en la presente memoria puede comprender también un vehículo que comprende un polipéptido según se describe en la presente memoria. Un polipéptido según se describe en la presente memoria se puede unir o inmovilizar en un vehículo mediante tecnologías conocidas en la técnica.
- La presente descripción también se refiere a un envase, tal como una lata, un barril o un tonel que comprende un polipéptido o una composición que comprende un polipéptido según se describe en la presente memoria.
- La presente descripción también se refiere a un proceso para preparar una composición que comprende un polipéptido según se describe en la presente memoria, que puede comprender secar por pulverización un medio de fermentación que comprende el polipéptido, o granular, o encapsular un polipéptido según se describe en la presente memoria, y preparar la composición.
- En otro aspecto, la presente invención se refiere a un ácido nucleico que codifica una endoproteasa específica para prolina, que tiene al menos 90 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad secuencial con respecto a la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácidos seleccionada del grupo que consiste en (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) y opcionalmente al menos una sustitución de aminoácido adicional I204F, Y205N o P477A, en el que polipéptido no comprende la sustitución de aminoácido I204F cuando el polipéptido tiene las sustituciones de aminoácidos K238E, I204V y V460A, en donde las sustituciones de aminoácidos se definen con respecto a la SEQ ID NO: 1.
- En otra realización, una molécula de ácido nucleico de la invención comprende una molécula de ácido nucleico que es el complemento inverso de la secuencia nucleotídica que se muestra en la SEQ ID NO: 2, o el complemento inverso de la secuencia codificante madura de la SEQ ID NO: 2, que comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácidos tal como se define anteriormente o una variante de cualquiera de dicha secuencia nucleotídica.
- También se describe un ácido nucleico que se hibrida en condiciones de rigurosidad media, preferiblemente, en condiciones de rigurosidad alta con la cadena complementaria de la secuencia codificante del polipéptido maduro de la SEQ ID NO:2, que comprende al menos una mutación que codifica al menos una combinación de sustituciones de aminoácidos tal como se define anteriormente.
- Una molécula de ácido nucleico que es complementaria a otra secuencia nucleotídica es una que es suficientemente complementaria a la otra secuencia nucleotídica para que pueda hibridarse con la otra secuencia nucleotídica y formar así una estructura doble estable. El término "ADNc" (ADN complementario) se define en la presente memoria como una molécula de ADN que se puede preparar mediante transcripción inversa a partir de una molécula de ARNm. En procariotas, la molécula de ARNm se obtiene a partir de la transcripción del ADN genómico de un gen presente en una célula. En células eucariotas, los genes contienen ambos exones, es decir, secuencias codificantes e intrones, es decir, secuencias intercaladas ubicadas entre los exones. Por lo tanto, en células eucariotas, el ARN primario, inicial obtenido a partir de la transcripción del ADN genómico de un gen se procesa a través de una serie de etapas antes de aparecer como ARNm. Estas etapas incluyen la eliminación de las secuencias intrónicas mediante un proceso denominado empalme. El ADNc derivado de ARNm solo contiene secuencias codificantes y se puede traducir directamente en el producto polipeptídico correspondiente.
- En una realización, se describe un ácido nucleico que es un ácido nucleico aislado, sustancialmente puro, puro, recombinante, sintético o variante del ácido nucleico descrito en la presente memoria.
- En otro aspecto, la presente invención se refiere a un vector de expresión que comprende un polinucleótido según se describe en la presente memoria enlazado funcionalmente con al menos una secuencia de control que dirige la expresión del polipéptido en una célula hospedante para expresión.
- Existen diversos modos de insertar un ácido nucleico en una construcción de ácido nucleico o un vector de expresión que son conocidos para un experto en la técnica, véase por ejemplo Sambrook & Russell, Molecular Cloning: A Laboratory Manual, 3a Ed., CSHL Press, Cold Spring Harbor, NY, 2001. Puede ser deseable manipular

un ácido nucleico que codifica un polipéptido de la presente invención con secuencias de control, tal como secuencias promotoras y terminadoras.

Un promotor puede ser cualquier secuencia promotora apropiada adecuada para una célula hospedante eucariota o procariota, que exhibe actividad de transcripción, incluidos promotores mutantes, truncados e híbridos, y se puede obtener a partir de polinucleótidos que codifican polipéptidos extracelulares o intracelulares ya sean endógenos (natural) o heterólogos (extraños) con respecto a la célula. El promotor puede ser un promotor constitutivo o inducible. Preferiblemente, el promotor es un promotor inducible, por ejemplo, un promotor inducible por almidón. Los promotores adecuados en hongos filamentosos son promotores que se pueden seleccionar del grupo que incluye, pero no se limita a, promotores obtenidos a partir de polinucleótidos que codifican TAKA amilasa de *A. oryzae*, proteinasa aspártica de *Rhizomucor miehei*, promotor gdpA de *Aspergillus*, alfa-amilasa neutra de *A. niger*, alfa-amilasa estable en ácido de *A. niger*, glucoamilasa (glaA) de *A. niger* o *A. awamori*, endoxilanas (xlnA) o beta-xilosidasa (xlnD) de *A. niger* o *A. awamori*, cellobiohidrolasa I (CBHI) de *T. reesei*, lipasa de *R. miehei*, proteasa alcalina de *A. oryzae*, triosa fosfato isomerasa de *A. oryzae*, acetamidasa de *A. nidulans*, camiloglucosidasa de *Fusarium venenatum* (WO 00/56900), Dania de *Fusarium venenatum* (WO 00/56900), Quinn de *Fusarium venenatum* (WO 00/56900), proteasa similar a tripsina de *Fusarium oxysporum* (WO 96/00787), beta-glucosidasa de *Trichoderma reesei*, cellobiohidrolasa I de *Trichoderma reesei*, cellobiohidrolasa II de *Trichoderma reesei*, endoglucanasa I de *Trichoderma reesei*, endoglucanasa II de *Trichoderma reesei*, endoglucanasa III de *Trichoderma reesei*, endoglucanasa IV de *Trichoderma reesei*, endoglucanasa V de *Trichoderma reesei*, xilanasas I de *Trichoderma reesei*, xilanasas II de *Trichoderma reesei*, beta-xilosidasa de *Trichoderma reesei*, así como el promotor NA2-tpi (un híbrido de los promotores de los polinucleótidos que codifican la alfa-amilasa neutra de *A. niger* y la triosa fosfato isomerasa de *A. oryzae*), y promotores mutantes, truncados e híbridos de estos.

Se puede usar cualquier terminador que es funcional en una célula según se describe en la presente memoria, que son conocidos para un experto en la técnica. Los ejemplos de secuencias terminadoras adecuadas en hongos filamentosos incluyen secuencias terminadoras de un gen fúngico filamentosos, tal como de genes de *Aspergillus*, por ejemplo, del gen de TAKA amilasa de *A. oryzae*, los genes que codifican glucoamilasa (glaA) de *A. niger*, antranilato sintasa de *A. nidulans*, alfa-glucosidasa de *A. niger*, trpC y/o proteasa similar a tripsina de *Fusarium oxysporum*.

En otro aspecto, la presente invención se refiere a una célula hospedante que comprende una construcción de ácido nucleico o un vector de expresión según se describen en la presente memoria. Una célula hospedante adecuada puede ser una célula de mamífero, insecto, planta, hongo o alga, o una célula bacteriana. Una célula hospedante adecuada puede ser una célula fúngica, por ejemplo, del género *Acremonium*, *Aspergillus*, *Chrysosporium*, *Fusarium*, *Myceliophthora*, *Penicillium*, *Rasamsonia*, *Talaromyces*, *Thielavia*, *Trichoderma*, *Saccaromyces*, *Kluyveromyces*, *Pichia*, por ejemplo, *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *A. oryzae*, *A. sojae*, *Talaromyces emersonii*, *Rasamsonia emersonii*, *Chrysosporium lucknowense*, *Fusarium oxysporum*, *Myceliophthora thermophila*, *Thielavia terrestris* o *Trichoderma reesei* o *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Pichia pastoris*.

Una célula hospedante puede ser una célula hospedante recombinante o transgénica. La célula hospedante se puede modificar genéticamente con una construcción de ácido nucleico o vector de expresión según se describen en la presente memoria, con técnicas estándares conocidas en la técnica, tales como electroporación, transformación de protoplasto o conjugación, por ejemplo, según se describe en Sambrook & Russell, *Molecular Cloning: A Laboratory Manual*, 3ª Ed., CSHL Press, Cold Spring Harbor, NY, 2001. Un hospedante recombinante puede sobreexpresar un polipéptido según la presente descripción mediante técnicas conocidas en la técnica.

En un aspecto, la presente invención se refiere a un proceso para la producción de un polipéptido según se describe en la presente memoria, que comprende cultivar una célula hospedante recombinante en un medio de fermentación adecuado en condiciones que conducen a la producción del polipéptido y producir el polipéptido. Un experto en la técnica entiende cómo llevar a cabo un proceso para la producción de un polipéptido según se describe en la presente memoria dependiendo de la célula hospedante usada, tal como el pH, la temperatura y la composición de un medio de fermentación. Las células hospedantes se pueden cultivar en matraces de agitación o en fermentadores que tienen un volumen de 0,5 o 1 litro o mayor hasta 10 a 100 o más metros cúbicos. El cultivo se puede llevar a cabo de manera aerobia o anaerobia, dependiendo de los requisitos de una célula hospedante.

De manera ventajosa, un polipéptido según se describe en la presente memoria se recupera o aísla del medio de fermentación. Recuperar o aislar un polipéptido de un medio de fermentación se puede llevar a cabo, por ejemplo, mediante centrifugación, filtración y/o ultrafiltración.

Un polipéptido que tiene actividad de endoproteasa específica para prolina o una composición que comprende un polipéptido según se describe en la presente memoria se puede usar en una gran variedad de aplicaciones, por ejemplo, en la producción de un producto alimentario o pienso, tal como en la producción de un hidrolizado proteico. Diversas proteínas alimentarias contienen subfracciones altamente alergénicas que incluso pueden ser tóxicas para individuos específicos, tales como gluten que contiene prolaminas con secuencias peptídicas ricas en prolina. Estas proteínas se pueden someter a la nueva enzima para aliviar su antigenicidad o toxicidad.

Un grupo de personas para las cuales el gluten es tóxico son los individuales que padecen celiacía. La celiacía, también conocida como enfermedad celíaca, es una enfermedad autoinmunitaria del intestino delgado causada por la ingesta de proteínas de gluten de cereales, tales como alfa-gliadina del trigo, hordeína de la cebada, secalina del centeno y avenina de la avena.

5 Por consiguiente, un polipéptido que tiene actividad de endoproteasa específica para prolina o una composición que comprende un polipéptido según se describe en la presente memoria se puede usar en la preparación de un suplemento alimentario o como un medicamento en el tratamiento de un paciente que padece celiacía, o en el tratamiento de personas intolerantes al gluten.

10 Un polipéptido según se describe en la presente memoria también se puede usar como un auxiliar de procesamiento para hidrolizar el gluten en un producto alimentario.

15 Por consiguiente, la presente invención se refiere a un proceso para la preparación de un producto alimentario o pienso que comprende incubar una forma intermedia de un producto alimentario o pienso con un polipéptido o una composición que comprende un polipéptido según se describe en la presente memoria, y preparar el producto alimentario o pienso. Un producto alimentario en un proceso según se describe en la presente memoria incluye una bebida, tal como cerveza, vino o jugo de fruta, o un producto horneado, o un producto lácteo, pero no se limita a estos.

Un producto alimentario y/o una forma intermedia de un producto alimentario puede comprender gluten.

Se halló que un polipéptido que tiene actividad de endoproteasa específica para prolina según se describe en la presente memoria fue capaz de hidrolizar los epítopos tóxicos en el gluten en fragmentos no tóxicos.

20 Una forma intermedia de un producto alimentario puede ser cualquier forma adecuada de un producto alimentario durante la preparación del producto alimentario. Por ejemplo, una forma intermedia de cerveza puede ser un mosto y una forma intermedia de pan puede ser una masa o una mezcla.

25 Un proceso para la preparación de un producto alimentario según la presente descripción puede comprender una etapa de pasteurizar el producto alimentario. La pasteurización normalmente comprende calentar un producto alimentario, o una forma intermedia de un producto alimentario, por ejemplo, al llevar el producto alimentario o forma intermedia de un producto alimentario hasta una temperatura de entre 60 a 68 °C entre 10 a 20 min, o entre 12 y 18 min, o hasta una temperatura de entre 70-74 °C, tal como aproximadamente 72 °C durante al menos 5, 10 o 15 segundos.

30 Un producto alimentario en un proceso según se describe en la presente memoria puede ser también un hidrolizado proteico. Por consiguiente, la presente descripción se refiere a un proceso para la preparación de un hidrolizado proteico que comprende poner en contacto un sustrato proteico con un polipéptido o una composición, según se describen en la presente memoria, y producir el hidrolizado proteico. Un hidrolizado proteico se puede preparar a partir de cualquier sustrato proteico adecuado, por ejemplo, un sustrato proteico que es rico en residuos prolina, tal como gluten en cereales o caseínas en leche bovina.

35 En un aspecto, la presente descripción se refiere a un producto alimentario obtenible mediante un proceso para la preparación de un producto alimentario según se describe en la presente memoria.

Ejemplos

Materiales y métodos

40 Se llevaron a cabo procedimientos de ADN estándares según se describen en otro lugar (Sambrook & Russell, 2001, Molecular cloning: a laboratory manual, 3ª Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, Nueva York) a menos que se indique lo contrario. Las secuencias de ADN se ordenaron en DNA 2.0.

Ejemplo 1. Clonación, expresión y recuperación de endoproteasa específica para prolina (PEP) (mutante)

Ejemplo 1.1. Clonación y expresión

45 La secuencia proteica de la endoproteasa específica para prolina (PEP) de *A. niger* se muestra en la SEQ ID NO: 1, en donde los primeros 17 aminoácidos son una secuencia señal de pectinametiltransferasa de *A. niger* (PMeA ss; SEQ ID NO: 2) y la parte siguiente comprende 19 aminoácidos de la prosequencia de endoproteasa específica para prolina de *A. niger* (SEQ ID NO: 4).

50 Se diseñó una secuencia de ADN adaptada con codones para la expresión de esta proteína en *Aspergillus niger* que contiene sitios de restricción adicionales para la subclonación en un vector de expresión de *Aspergillus*. La adaptación con codones se llevó a cabo según se describe en WO 2008/000632. La secuencia de ADN optimizada con codones para *A. niger* del gen que codifica la proteína PEP de la SEQ ID: NO: 1 se muestra en la SEQ ID NO: 2.

De forma similar la endoproteasa específica de prolina mutante de SEQ ID NO:1, PEP13-PEP23, que tiene la combinación de sustituciones de aminoácidos como se enumera en la Tabla 1, la endoproteasa específica de prolina mutante que tiene una única mutación I204F, Y205N o P477A, y la endoproteasa específica de prolina mutante PEP-5_70, PEP-5_73, PEP-5_74 (Tabla 3) tuvieron una optimización en los codones para la expresión en *Aspergillus niger*.

La secuencia de iniciación de la traducción del promotor de glucoamilasa *glaA* se modificó en 5'-CACCGTCAAAA ATG-3' y se usó una secuencia de terminación de la traducción óptima 5'-TAAA-3' en la generación de la construcción para expresión (según se detalla también en WO2006/077258). Un fragmento de ADN que contenía a.o. parte del promotor de glucoamilasa y el gen codificante de PEP se sintetizó completamente, se purificó y digirió con EcoRI y Pacl. El vector pGBTOP-16 (Figura 1) se linealizó mediante digestión con EcoRI/Pacl y el fragmento de vector linealizado posteriormente se purificó mediante extracción en gel. El fragmento de ADN que contenía la región codificante de PEP se clonó en el vector pGBTOP-16 que resultó en pGBTOP-PEP. Posteriormente, se transformó GBA 306 de *A. niger* (cepa $\Delta glaA$, $\Delta pepA$, $\Delta hdfA$, amplicón *Bam*HI adaptado, $\Delta amyBI$, $\Delta amyBI$, $\Delta amyA$ alfa-amilasa y glucoamilasa negativa) con el vector pGBTOP-PEP linealizado mediante digestión por *Not*I, en un protocolo de cotransformación con pGBAAS-4 linealizado, con cepa y métodos según se describen en WO 2011/009700 y las referencias en él, y se seleccionó en medios que contenían acetamida y se purificó la colonia según procedimientos estándares. La transformación y selección se llevan a cabo según se describe en WO 98/46772 y WO 99/32617. Las cepas que contenían el gen de PEP se seleccionaron a través de PCR con cebadores específicos para el gen de PEP para verificar la presencia del casete de expresión de pGBTOP-PEP. Los transformantes se seleccionaron, denominados PEP13-PEP23, I204F, Y205N, P477A y EDP-5_70, EDP-5_73, EDP-5_74, y se replicaron en placas adicionalmente para obtener un inóculo de cepa único.

Ejemplo 1.2. Producción de PEP (mutante) en cepa PEP de *A. niger*

Para cada endoproteasa específica para prolina PEP (mutante) del Ejemplo 1.1 se prepararon esporas de PEP de *A. niger* nuevas. Se inocularon con 10^7 de esporas 4 matraces de agitación con 100 ml de medio de fermentación 1 (10 % p/v de sólidos de maíz fermentado, 1 % p/v de glucosa.H₂O, 0,1 % p/v de NaH₂PO₄.H₂O, 0,05 % p/v de MgSO₄.7H₂O, 0,025 % p/v Basildon, pH 5,8) en matraces de agitación de 500 ml con deflector. Estos precultivos se incubaron a 34 °C y 170 rpm durante 16-24 horas. De los precultivos, se usaron 50 ml para la inoculación de 1 matraz de agitación con 1 litro de medio de fermentación 2 (15 % p/v de maltosa, 6 % de p/v bacto-soitona, 1,5 % p/v de (NH₄)₂SO₄, 0,1 % p/v de NaH₂PO₄.H₂O, 0,1 % p/v de MgSO₄.7H₂O, 0,1 % p/v de L-arginina, 8 ‰ p/v de Tween-80, 2 ‰ p/v de Basildon, 2 % p/v de MES pH 5,1) en un tamaño de matraz de agitación de 5 litros y se agitó a 34 °C y 170 rpm. Después de 3, 4, 5 y 6 días de incubación, el pH del cultivo se redujo hasta pH 5,0 usando HCl 2 N y se analizaron muestras de cada uno de estos puntos de tiempo para determinar la actividad de PEP. Se tomaron muestras de 50 mL y el sobrenadante se separó de la biomasa mediante centrifugación y posteriormente filtración. La muestra con la actividad más alta se usó para caracterizar el mutante de PEP producido.

Ejemplo 2. Mediciones de la actividad de endoproteasa específica para prolina (PEP)

Se incubaron 100 μ L de sobrenadante de cultivo según se produjo en el Ejemplo 1, diluidos en 0,1 M tampón de acetato de sodio a pH 4,5 con 50 mM de NaCl, con 100 μ L de 6 mM Ac-AAP-pNA (acetil-AlaAlaPro-paranitroanilina de Selleckchem o CPC Scientific; pureza >95,0 % en base al análisis por HPLC) en 0,1 M tampón de NaAc a pH4,5 con 50 mM de NaCl, en una MTC (placa de microtitulación) Nunc de 96 pocillos con fondo plano. Después de 60 minutos a 20 °C, la reacción se detuvo mediante la adición de 40 μ L de HCl 1 M. El pNA que se había liberado mediante PEP se midió en un espectrofotómetro Tecan MTP a 405 nm (A405) (www.tecan.com). El blanco se preparó al mezclar el sobrenadante de cultivo diluido con la disolución de sustrato que se había mezclado con la disolución de HCl de antemano. La actividad se expresa en pNASU. 1pNASU es la cantidad de enzima que se libera de Ac-AAP-pNA en 1 hora la cantidad de pNA que corresponde a un aumento en la absorción a 405 nm de 1 OD, usando las condiciones descritas anteriormente. La A405 no debería estar por debajo del valor blanco en el inicio de la reacción ni por encima de 2,5 al finalizar la reacción, ni puede la A405 exceder el rango lineal del espectrofotómetro que se usa.

Ejemplo 3. Estabilidad térmica de la endoproteasa específica para prolina

Para evaluar la estabilidad térmica de la PEP genitora y de los mutantes PEP13-PEP23, antes del ensayo de actividad se procedió a una incubación de alícuotas de 100 μ L de una dilución de diez veces del sobrenadante de cultivo producido en el Ejemplo 1 en tampón (0,1 M NaAc pH 4,5, con 50 mM de NaCl) a 55 °C y 65 °C durante 15 min en una placa de PCR en una máquina de PCR. Después de los 15 min de incubación, las muestras se enfriaron rápidamente hasta 25 °C en la máquina de PCR. Se midió el pNASU/mL de cada muestra. La actividad inicial medida antes de la incubación a temperatura elevada (0 minutos) se usó como referencia (100 %) para determinar la actividad residual. Todas las actividades se midieron cuatro veces.

La Tabla 1 muestra que todas las endoproteasas específicas para prolina mutantes tienen una actividad residual significativamente reducida en comparación con la endoproteasa específica para prolina genitora después de mantener la enzima a 65 °C y 60°C durante 15 min.

Tabla 1. Actividad residual de mutantes de endoproteasa específica para prolina PEP13-PEP23 en comparación con la endoproteasa específica para prolina genitora.

Clon	Sustituciones con respecto a la SEQ ID NO genitora: 1	Actividad residual (pNASU) en el T indicado después de 15'		
		55 °C	60 °C	65 °C
PEP	Genitor	100 %	93 %	80 %
PEP13	K238E+I204V+V460A	61 %	14 %	0 %
PEP14	F279S+A242V+N507I	86 %	40 %	1 %
PEP15	T145A+K424M	99 %	63 %	9 %
PEP16	T359A+F379S	96 %	55 %	0 %
PEP17	M170I+A421T	88 %	53 %	4 %
PEP18	N441S+G484S	92 %	50 %	3 %
PEP19	L470H+Q288R	92 %	48 %	5 %
PEP20	L470H+E387G	89 %	37 %	1 %
PEP21	T281S+L373I	64 %	59 %	14 %
PEP22	P304A+P469A	95 %	77 %	5 %
PEP23	P466T+P469Q	93 %	40 %	0 %

Ejemplo 4. Mutación adicional para reducir más la termoestabilidad en endoproteasas específicas de prolina

5 Las endoproteasas específicas de prolina mutante PEP20 que tiene sustituciones L470H+E387G y PEP23, que tiene sustituciones P466T+P469Q de la Tabla 1 se combinaron adicionalmente con las mutaciones sencillas Y205N y o P477A o I204F, respectivamente. El clonado, expresión y recuperación de los mutantes sencillos y triples se realizó como se describe en el Ejemplo 1. Para evaluar la estabilidad térmica de los mutantes sencillos y triples el ensayo de actividad se precedió por una incubación de alícuotas de 100 µL de una dilución de diez veces del sobrenadante del cultivo en el tampón (NaAc 0,1 M pH 4,5, con NaCl 50 mM) en una placa de PCR en una máquina de PCR en las temperaturas como se indican en la Tabla 2 y 3. Después de los 15 min de incubación las muestras se enfriaron rápidamente a 25°C. Se midió la pNASU/mL de cada muestra. La actividad inicial medida al comienzo de la incubación se usó como referencia (100%) para determinar la actividad residual.

La termoestabilidad de la endoproteasa específica de prolina de *Aspergillus niger* que contiene las mutaciones sencillas Y205N, P477A o I204F y que contiene las mutaciones triples se muestra en la Tabla 2 y 3 posterior.

15 Tabla 2. Actividad residual de endoproteasas específicas de prolina mutante de *Aspergillus niger* que contiene una única sustitución

Sustituciones con respecto a SEQ ID NO:1 parental	Actividad residual (pNASU) a la T indicada después de 15'					
	66,5°C	63,3°C	58,6°C	53,0°C	48,5°C	45,4°C
I204F	0,0%	4,4%	53,7%	86,4%	123,2%	130,5%

Y205N	0,3%	5,0%	42,2%	83,1%	101,3%	117,4%
P477A	0,4%	0,6%	25,6%	84,1%	108,8%	125,7%

Tabla 3. Actividad residual de mutantes triples de endoproteasa específica de prolina de *Aspergillus niger*

		Actividad residual (pNASU) a la T indicada después de 15'					
Sustituciones con respecto a SEQ ID NO:1 parental		60,4°C	57,3°C	52,5°C	47,0°C	42,4°C	39,4°C
PEP-5_70	L470H+ E387G+ Y205N	0%	1%	5%	52%	90%	99%
PEP-5_73	P466T+ P469Q+ I204F	1%	1%	12%	66%	91%	98%
PEP-5_74	P466T+ P469Q+ P477A	0%	0%	3%	46%	80%	96%

Los resultados muestran que la termoestabilidad de la endoproteasa específica de prolina mutante PEP20 y PEP23 puede reducirse más mediante sustituciones de aminoácidos adicionales tal como mediante la introducción de las sustituciones de aminoácidos I204F, Y205 o P477A.

Ejemplo 5. Especificidad para sustrato de las endoproteasas específicas para prolina

Para confirmar que los mutantes PEP13 a PEP23 son endoproteasas específicas para prolina, se evaluó la especificidad para el sustrato de diferentes mutantes de PEP PEP13-PEP23 usando citocromo c de corazón de caballo. Se incuban diluciones del sobrenadante de cultivo preparado en el Ejemplo 1 con citocromo c de corazón de caballo (Sigma) que tenía la secuencia de aminoácidos de la SEQ ID NO: 3. El sustrato se preparó al disolver 1 mg/mL de citocromo c en 100 mM de tampón de acetato de sodio, pH 4,5 y calentar a 95 °C durante 15 min. El sobrenadante del cultivo se diluye en 100 mM de tampón NaAc pH 4,5 y se incubó con la disolución de sustrato de citocromo c a 50 °C durante 3 horas. La reacción se detiene mediante dilución en agua y adición de 0,4M NaOH para subir el pH hasta 10. Las mezclas de reacción incubadas se analizan en un Accela UHPLC (Thermo Electron, Breda, Países Bajos) acoplado a un LTQ-Orbitrap Fourier Transform Mass Spectrometer (Thermo Electron, Bremen, Alemania). La separación cromatográfica se logró con una columna C-18 Eclipse XDB Zorbax de 2,1 x 50 mm, tamaño de partícula 1,8 µm, tamaño de poro 80Å (Agilent Santa Clara, CA, EE. UU.), usando una elución con gradiente con (A) agua de grado LC-MS que contenía ácido fórmico al 0,1 % y (B) acetonitrilo de grado LC-MS que contenía disolución de ácido fórmico al 0,1 % (Biosolve BV, Países Bajos) como fases móviles. El gradiente de 25 min se inicia a partir de 0 %, se mantuvo así durante 1 minuto y después se aumentó linealmente hasta 40 % (B) en 14 min, después se lavó con 80 % de (B) durante 4 min y se reequilibró con 0 % de (B) durante 5 min. La velocidad de flujo se mantiene a 0,4 ml/min, usando un volumen de inyección de 5 µl y la temperatura de la columna fue de 50 °C. La adquisición de los datos de espectrometría de masas se logra con adquisición dependiente de datos Top 3 usando las opciones "Cromatografía" y "Exclusión dinámica" en las que se incluyeron solo y estados de carga 2 y 3. La resolución para la exploración de FT MS fue de 15000 y se barrió para un intervalo m/z de 210-2000, mientras que los experimentos de MS/MS se llevaron a cabo en la trampa de iones. El ancho de aislamiento se fija en 3,0 m/z y la energía de colisión normalizada se fijó en 35. La identificación de péptidos se lleva a cabo usando datos secuenciados con masa precisa y MS/MS de novo.

Cuando los principales péptidos SEQ ID NO: 4 a SEQ ID NO: 8, en particular SEQ ID NO: 4 a SEQ ID NO: 6, se forman, esto muestra que las endoproteasas específicas para prolina mutantes tienen actividad de endoproteasa específica para prolina, es decir, tienen una preferencia por un corte en una posición donde hay un residuo prolina en el péptido.

ES 2 750 753 T3

Listado de secuencias

<110> DSM IP Assets B.V.
 <120> Endoproteasa específica para prolina y su uso
 <130> 30377-WO-PCT
 <160> 8
 5 <170> BiSSAP 1.2
 <210> 1
 <211> 521
 <212> PRT
 <213> Secuencia artificial
 10 <220>
 <223> Endoproteasa específica para prolina de A. niger, con secuencia señal de Pectinametilsterasa
 <400> 1
 Met Val Lys Ser Ile Leu Ala Ser Val Phe Phe Ala Ala Thr Ala Leu
 1 5 10 15
 Ala Ala Arg Pro Arg Leu Val Pro Lys Pro Val Ser Arg Pro Ala Ser
 20 25 30
 Ser Lys Ser Ala Ala Thr Thr Gly Glu Ala Tyr Phe Glu Gln Leu Leu
 35 40 45
 Asp His His Asn Pro Glu Lys Gly Thr Phe Ser Gln Arg Tyr Trp Trp
 50 55 60
 Ser Thr Glu Tyr Trp Gly Gly Pro Gly Ser Pro Val Val Leu Phe Thr
 65 70 75 80
 Pro Gly Glu Val Ser Ala Asp Gly Tyr Glu Gly Tyr Leu Thr Asn Glu
 85 90 95
 Thr Leu Thr Gly Val Tyr Ala Gln Glu Ile Gln Gly Ala Val Ile Leu
 100 105 110
 Ile Glu His Arg Tyr Trp Gly Asp Ser Ser Pro Tyr Glu Val Leu Asn
 115 120 125
 Ala Glu Thr Leu Gln Tyr Leu Thr Leu Asp Gln Ala Ile Leu Asp Met
 130 135 140
 Thr Tyr Phe Ala Glu Thr Val Lys Leu Gln Phe Asp Asn Ser Thr Arg
 145 150 155 160
 Ser Asn Ala Gln Asn Ala Pro Trp Val Met Val Gly Gly Ser Tyr Ser
 165 170 175
 Gly Ala Leu Thr Ala Trp Thr Glu Ser Val Ala Pro Gly Thr Phe Trp
 180 185 190
 Ala Tyr His Ala Thr Ser Ala Pro Val Glu Ala Ile Tyr Asp Tyr Trp
 195 200 205
 Gln Tyr Phe Tyr Pro Ile Gln Gln Gly Met Ala Gln Asn Cys Ser Lys
 210 215 220
 Asp Val Ser Leu Val Ala Glu Tyr Val Asp Lys Ile Gly Lys Asn Gly
 225 230 235 240
 Thr Ala Lys Glu Gln Gln Ala Leu Lys Glu Leu Phe Gly Leu Gly Ala
 245 250 255
 Val Glu His Phe Asp Asp Phe Ala Ala Val Leu Pro Asn Gly Pro Tyr
 260 265 270
 Leu Trp Gln Asp Asn Asp Phe Ala Thr Gly Tyr Ser Ser Phe Phe Gln
 275 280 285
 Phe Cys Asp Ala Val Glu Gly Val Glu Ala Gly Ala Ala Val Thr Pro
 290 295 300
 Gly Pro Glu Gly Val Gly Leu Glu Lys Ala Leu Ala Asn Tyr Ala Asn
 305 310 315 320
 Trp Phe Asn Ser Thr Ile Leu Pro Asp Tyr Cys Ala Ser Tyr Gly Tyr

ES 2 750 753 T3

```

                325                330                335
Trp Thr Asp Glu Trp Ser Val Ala Cys Phe Asp Ser Tyr Asn Ala Ser
                340                345                350
Ser Pro Ile Tyr Thr Asp Thr Ser Val Gly Asn Ala Val Asp Arg Gln
                355                360                365
Trp Glu Trp Phe Leu Cys Asn Glu Pro Phe Phe Tyr Trp Gln Asp Gly
                370                375                380
Ala Pro Glu Gly Thr Ser Thr Ile Val Pro Arg Leu Val Ser Ala Ser
385                390                395                400
Tyr Trp Gln Arg Gln Cys Pro Leu Tyr Phe Pro Glu Thr Asn Gly Tyr
                405                410                415
Thr Tyr Gly Ser Ala Lys Gly Lys Asn Ala Ala Thr Val Asn Ser Trp
                420                425                430
Thr Gly Gly Trp Asp Met Thr Arg Asn Thr Thr Arg Leu Ile Trp Thr
                435                440                445
Asn Gly Gln Tyr Asp Pro Trp Arg Asp Ser Gly Val Ser Ser Thr Phe
                450                455                460
Arg Pro Gly Gly Pro Leu Ala Ser Thr Ala Asn Glu Pro Val Gln Ile
465                470                475                480
Ile Pro Gly Gly Phe His Cys Ser Asp Leu Tyr Met Ala Asp Tyr Tyr
                485                490                495
Ala Asn Glu Gly Val Lys Lys Val Val Asp Asn Glu Val Lys Gln Ile
                500                505                510
Lys Glu Trp Val Glu Glu Tyr Tyr Ala
                515                520

```

<210> 2

<211> 1566

<212> ADN

5

<213> Secuencia artificial

<220>

<221> fuente

<222> 1..1566

10

<223> /organismo="Secuencia artificial" /nota=" Endoproteasa específica para prolina de A. niger, con secuencia señal de Pectinametilsterasa" /mol_tipo="AND no asignado"

<400> 2

```

atgggtcaagt ccatcctggc ctccgtcttc ttcgctgcc a ctgctcttgc tgcaaggcct      60
cgtctcgttc ccaagcccgt ttctcgtccc gccagctcca agtccgctgc tactactggt      120
gaggcctact ttgaacagct gttggaccac cacaacctg agaagggtag tttctcgtcaa      180
agatactggt ggagcaccga gtactggggt ggtcccgat ccccgttgt cctgttcaact      240
cccggtgagg tcagcgtga tggctacgag ggttatctga ccaacgagac tctcaccggt      300
gtctacgccc aggagattca gggtgctgtc atcctgatcg aacaccgata ctggggtgac      360
tcgtctccct acgaggtgct gaacgccgag actctccagt acttgaccct cgaccaggct      420
atccttgata tgacctactt cgccgaaacc gtcaagctcc agtttgaaa ctccaccggc      480
tccaacgctc agaacgctcc ttgggttatg gtcggcggca gctacagcgg tgctctgact      540
gcttgaccgg agtccggtgc tcccggcacc ttctgggctt accacgccac ctctgctcct      600
gttgaggcca tctacgacta ctggcaatac ttctacccca ttcagcaggg tatggctcag      660
aactgctcca aagatgtctc tctttagtca gaatacgtcg acaagatcgg caagaacggc      720

```

ES 2 750 753 T3

actgccaaagg agcaacaggc tctgaaggag cttttcggcc taggagcagt ggagcacttc 780
gacgacttgc cgcgtgttct gcccaacggt ccttacctct ggcaagacaa cgactttgcc 840
accggttact cttctttctt ccagttctgt gatgccgctg aggggtgctga ggctgggtgct 900
gccgtcacc cccgtcctga aggtgttggt ctggaaaagg cccttgctaa ctacgcgaac 960
tggttcaact ctaccatcct ccccgattac tgcgccagct acggctactg gactgacgag 1020
tgggtccgctg cctgcttoga ctactacaac gcctcctctc ctatatacac cgacaccagc 1080
gttggttaacg ccgtcgaccg tcagtgggag tggttcctct gcaatgagcc cttcttttac 1140
tggcaggacg gtgccccga ggggtacttca acgatagtagt cccgcttagt gtccgcctcc 1200
tactggcagc gtcaatgtcc gttgtacttc cccgagacta acggttacac ctacggctcc 1260
gccaagggaa agaacgccgc caccgtcaac agctggaccg gtggctggga catgaccctg 1320
aacaccacc cgtctgatctg gacgaacggc caatacgacc cctggcgtga ctccggtgct 1380
tcttccacct tccgccccgg tggteccctc gcttcgaccg ccaacgagcc cgtccagata 1440
ataccgggtg gtttccattg ctccgacctc tacatggcag actactacgc caacgagggc 1500
gtcaagaagg ttgtcgacaa cgaagtcaaa caaatcaagg agtgggttga ggaatactac 1560
gcgtaa 1566

5

<210> 3
<211> 104
<212> PRT
<213> Secuencia artificial
<220>
<223> Secuencia de aminoácidos de citocromo C de corazón de caballo

<400> 3
Gly Asp Val Glu Lys Gly Lys Lys Ile Phe Val Gln Lys Cys Ala Gln
1 5 10 15
Cys His Thr Val Glu Lys Gly Gly Lys His Lys Thr Gly Pro Asn Leu
20 25 30
His Gly Leu Phe Gly Arg Lys Thr Gly Gln Ala Pro Gly Phe Thr Tyr
35 40 45
Thr Asp Ala Asn Lys Asn Lys Gly Ile Thr Trp Lys Glu Glu Thr Leu
50 55 60
Met Glu Tyr Leu Glu Asn Pro Lys Lys Tyr Ile Pro Gly Thr Lys Met
65 70 75 80
Ile Phe Ala Gly Ile Lys Lys Lys Thr Glu Arg Glu Asp Leu Ile Ala
85 90 95
Tyr Leu Lys Lys Ala Thr Asn Glu
100

10

<210> 4
<211> 30
<212> PRT
<213> Secuencia artificial
<220>

15

<223> Fragmento de citocromo C digerido con PEP
<400> 4
Gly Asp Val Glu Lys Gly Lys Lys Ile Phe Val Gln Lys Cys Ala Gln
1 5 10 15
Cys His Thr Val Glu Lys Gly Gly Lys His Lys Thr Gly Pro
20 25 30

20

<210> 5
<211> 14
<212> PRT
<213> Secuencia artificial
<220>
<223> Fragmento de citocromo C digerido con PEP
<400> 5

ES 2 750 753 T3

Asn Leu His Gly Leu Phe Gly Arg Lys Thr Gly Gln Ala Pro
 1 5 10
 <210> 6
 <211> 32
 <212> PRT
 5 <213> Secuencia artificial
 <220>
 <223> Fragmento de citocromo C digerido con PEP
 <400> 6
 Gly Phe Thr Tyr Thr Asp Ala Asn Lys Asn Lys Gly Ile Thr Trp Lys
 1 5 10 15
 Glu Glu Thr Leu Met Glu Tyr Leu Glu Asn Pro Lys Lys Tyr Ile Pro
 20 25 30
 10 <210> 7
 <211> 7
 <212> PRT
 <213> Secuencia artificial
 <220>
 15 <223> Fragmento de citocromo C digerido con PEP
 <400> 7
 Gly Thr Lys Met Ile Phe Ala
 1 5
 <210> 8
 <211> 21
 20 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Fragmento de citocromo C digerido con PEP
 <400> 8
 Gly Ile Lys Lys Lys Thr Glu Arg Glu Asp Leu Ile Ala Tyr Leu Lys
 1 5 10 15
 Lys Ala Thr Asn Glu
 20
 25

REIVINDICACIONES

1. Un polipéptido que tiene actividad de endoproteasa específica para prolina, en donde el polipéptido es un polipéptido variante que tiene menos de 70 % de actividad residual sobre acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min en comparación con el polipéptido de la SEQ ID NO:1 y se selecciona del grupo que consiste en
- 5
- i. un polipéptido que comprende secuencia de aminoácidos según la SEQ ID NO: 1, en donde la SEQ ID NO: 1 comprende al menos una combinación de una sustituciones de aminoácidos seleccionada del grupo que consiste en las combinaciones (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) y (P466T, P469Q), y opcionalmente una sustitución de aminoácido adicional I204F, Y205N o P477A en la que el polipéptido no comprende la sustitución del aminoácido I204F cuando el polipéptido tiene las sustituciones de aminoácidos K238E, I204V y V460A, en las que las posiciones se definen con referencia a SEQ ID NO:1;
- 10
- ii. un polipéptido según i), pero que carece de una secuencia señal y/o una secuencia proproteína;
- iii. un polipéptido según i) a ii), en donde el polipéptido tiene al menos 90%, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad con respecto a la secuencia de aminoácidos según la SEQ ID NO: 1; y,
- 15
- iv. un polipéptido codificado por un ácido nucleico que tiene al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad con respecto a la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende al menos una mutación que codifica al menos una sustitución de aminoácidos seleccionada del grupo que consiste en (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R), (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) y opcionalmente una sustitución de aminoácido adicional I204F, Y205N o P477A en la que la sustitución de aminoácidos I204F no está presente, cuando el polipéptido tiene las sustituciones de aminoácidos K238E, I204V y V460A, en la que las sustituciones de aminoácidos se definen con referencia a SEQ ID NO:1.
- 20
2. Un polipéptido que es un polipéptido aislado, sustancialmente puro, puro, recombinante, sintético o variante del polipéptido de la reivindicación 1.
- 25
3. Un método para generar un polipéptido variante, en donde el método comprende
- i. seleccionar un polipéptido genitor que comprende al menos 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad con respecto a la secuencia de aminoácidos según la SEQ ID NO: 1, o con respecto a un polipéptido maduro de la SEQ ID NO: 1; y,
- 30
- ii. sustituir al menos una combinación de aminoácidos que da por resultado una combinación de aminoácidos seleccionados del grupo que consiste en (Glu (E) en la posición 238, Val (V) en la posición 204, Ala (A) en la posición 460), (Ser (S) en la posición 279, Val (V) en la posición 242, Ile (I) en la posición 507), (Thr (T) en la posición 145, Met (M) en la posición 424), (Ala (A) en la posición 359, Ser (S) en la posición 379), (Ile (I) en la posición 170, Thr (T) en la posición 421), (Ser (S) en la posición 441, Ser (S) en la posición 484), (His (H) en la posición 470, Arg (R) en la posición 288), (His (H) en la posición 470, Gly (G) en la posición 387), (Ser (S) en la posición 281, Ile (I) en la posición 373), (Ala en la posición 304, Ala en la posición 469), y (Thr (T) en la posición 466, Gln (Q) en la posición 469), y opcionalmente sustituir al menos un aminoácido adicional que da por resultado Phe (F) en la posición 204, Asp (N) en la posición 205 o Ala (A) en la posición 477, en el que la sustitución de aminoácido Phe (F) no está presente en el polipéptido que tiene las sustituciones de aminoácidos Glu (E) en la posición 238, Val (V) en la posición 204, y Ala (A) en la posición 460); en el que las posiciones se definen con referencia a SEQ ID NO:1; y
- 35
- iii. generar el polipéptido variante,
- 40
- en donde el polipéptido variante tiene menos de 70 % de actividad residual sobre acetil-AlaAlaPro-paranitroanilina (Ac-AAP-pNA) como sustrato después de que el polipéptido se ha mantenido a una temperatura de 65 °C durante 15 min.
- 45
4. Una composición que comprende un polipéptido según la reivindicación 1 o 2.
5. Una composición según la reivindicación 4, que comprende un vehículo, un excipiente o una enzima auxiliar.
6. Un ácido nucleico que codifica una endoproteasa específica para prolina, que tiene al menos 90 %, 95 %, 96 %, 97 %, 98 % o 99 % de identidad secuencial con respecto a la SEQ ID NO: 2, o con respecto a una secuencia codificante madura de la SEQ ID NO: 2, en donde la SEQ ID NO: 2 comprende mutaciones que codifican al menos una combinación de sustituciones de aminoácidos seleccionada del grupo que consiste en (K238E, I204V, V460A), (F279S, A242V, N507I), (T145A, K424M), (T359A, F379S), (M170I, A421T), (N441S, G484S), (L470H, Q288R),
- 50

- (L470H, E387G), (T281S, L373I), (P304A, P469A) y (P466T, P469Q) sustitución de aminoácido seleccionada del grupo que consiste en y (P466T, P469Q), y opcionalmente una sustitución de aminoácido adicional I204F, Y205N o P477A, en la que el polipéptido no comprende la sustitución de aminoácido I204F cuando el polipéptido tiene las sustituciones de aminoácidos K238E, I204V y V460A, en donde las sustituciones de aminoácidos se definen con respecto a la SEQ ID NO: 1.
- 5
7. Un ácido nucleico que es un ácido nucleico aislado, sustancialmente puro, puro, recombinante o sintético del ácido nucleico de la reivindicación 6.
8. Un vector de expresión que comprende un ácido nucleico según la reivindicación 6 o 7 enlazado funcionalmente con al menos una secuencia de control que dirige la expresión del polipéptido en una célula hospedante.
- 10
9. Una célula hospedante recombinante que comprende un ácido nucleico según la reivindicación 6 o 7, o un vector de expresión según la reivindicación 9.
10. Un método para la preparación de un polipéptido según las reivindicaciones 1 o 2, que comprende cultivar una célula hospedante según la reivindicación 9 en un medio de fermentación adecuado, en condiciones que permiten la expresión del polipéptido y preparar el polipéptido y, opcionalmente, que comprende además recuperar el polipéptido.
- 15
11. Un proceso para la preparación de un producto alimentario o pienso que comprende poner en contacto una forma intermediaria de un producto alimentario o pienso con un polipéptido según la reivindicación 1 o 2, o una composición según la reivindicación 4 o 5, y preparar el producto alimentario o pienso.
- 20
12. Un proceso según la reivindicación 11, en donde el producto alimentario es una bebida, preferiblemente, una cerveza.
13. Un proceso según las reivindicaciones 11 o 12, en donde el producto alimentario comprende gluten.

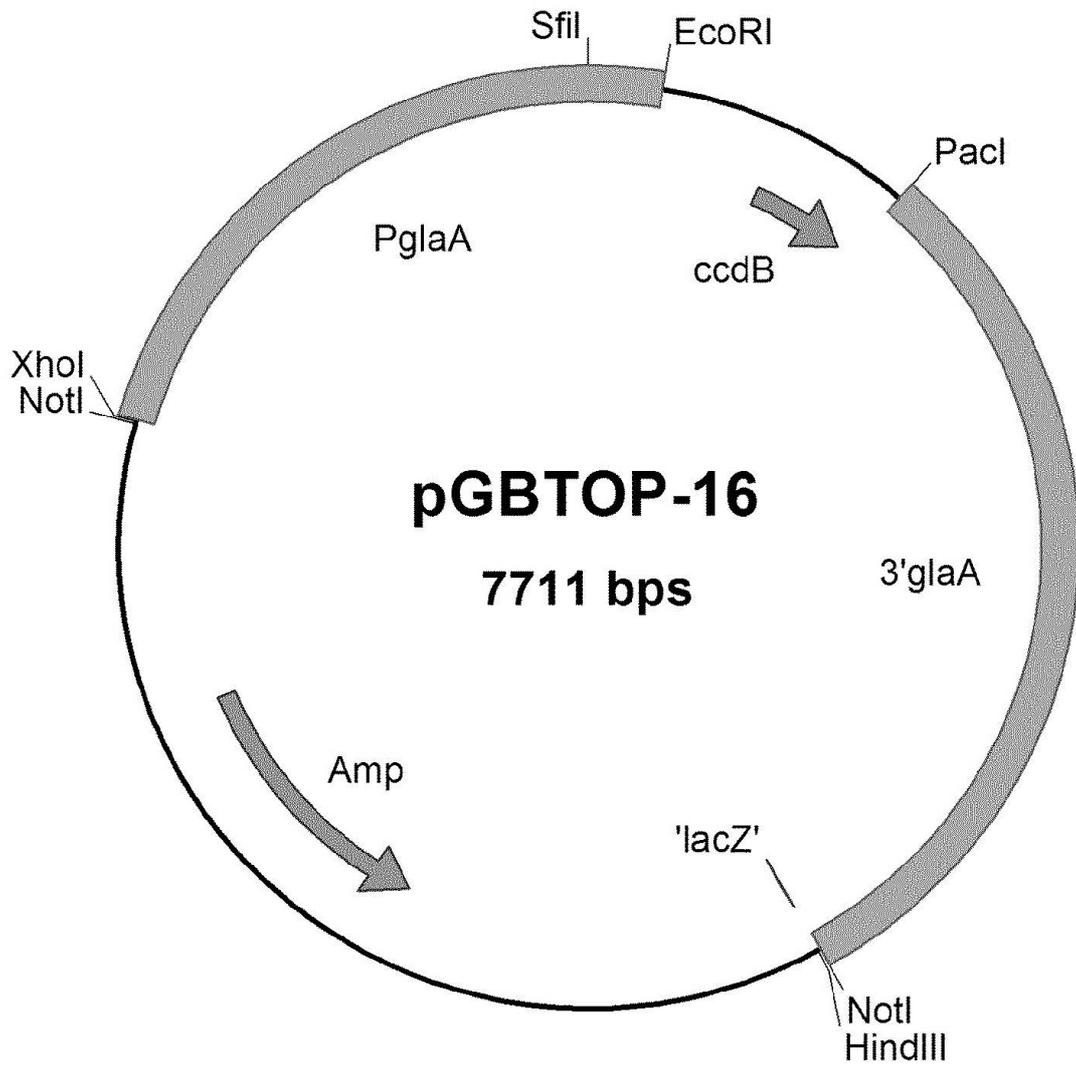


Fig. 1