



(51) International Patent Classification:

A23L 1/105 (2006.01) C12N 15/63 (2006.01)
C12N 9/62 (2006.01)

(21) International Application Number:

PCT/EP2015/061001

(22) International Filing Date:

19 May 2015 (19.05.2015)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

14168865.5 19 May 2014 (19.05.2014) EP

(71) Applicant: DSM IP ASSETS B.V. [NL/NL]; Het Overloon 1, NL-6411 TE Heerlen (NL).

(72) Inventors: LAAN, VAN DER, Jan, Metske; P.O. Box 4, NL-6100 AA Echt (NL). VONDERVOORT, VAN DE, Peter, Jozef, Ida; P.O. Box 4, NL-6100 AA Echt (NL). CHRISTIS, Chantal; P.O. Box 4, NL-6100 AA Echt (NL). SPAANS, Martine; P.O. Box 4, NL-6100 AA Echt (NL). BRUINE-PAULUS, DE, Angela; P.O. Box 4, NL-6100 AA Echt (NL).

(74) Agent: CAZEMIER, Anne, Engeline; DSM Intellectual Property, P.O. Box 4, NL-6100 AA Echt (NL).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: PROLINE-SPECIFIC ENDOPROTEASE

(57) Abstract: The present invention relates to a polypeptide having proline-specific endoprotease activity, selected from the group consisting of i. a polypeptide comprising a mature polypeptide sequence of SEQ ID NO: 2; ii. a polypeptide that has least 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the mature polypeptide sequence of SEQ ID NO: 2; iii. a polypeptide encoded by a nucleic acid that hybridizes under medium stringency, preferably under high stringency conditions to the complementary strand of the mature polypeptide coding sequence of SEQ ID NO:1; iv. a polypeptide encoded by a nucleic acid that has at least 80, 85%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to the mature polypeptide coding sequence of SEQ ID NO: 1; The invention further relates to a method of producing the polypeptide and a process for the preparation of a food or feed product wherein the polypeptide is used.



PROLINE-SPECIFIC ENDOPROTEASE

5

The present invention relates to a polypeptide having proline-specific endoprotease activity, a composition comprising the polypeptide, a nucleic acid encoding a proline-specific endoprotease, an expression vector comprising the nucleic acid encoding a proline-specific endoprotease, a recombinant host cell, a method for
10 preparing proline-specific endoprotease and a process for preparing a food or feed product wherein the proline-specific endoprotease is used.

Background

Proline-specific endoproteases are enzymes that hydrolyse a protein or peptide at a
15 position where there is a proline in the protein or peptide.

A proline-specific endoprotease may for instance be derived from *Aspergillus niger* or *Penicillium chrysogenum*, such as disclosed in WO2002/046381 and WO2009/144269 respectively.

Other proline-specific endoprotease are known from WO2012/174127.
20 WO2012/174127 discloses proline-specific protease from *Botryotinia fuckeliana*, *Aspergillus clavatus*, *Sclerotinia sclerotium*, *Mycosphaerella graminicola*, *Neurospora crasse*, *Talaromyces stipitatus* and *Gibberella zeae*.

Proline-specific endoprotease can be used in several applications, for instance in the degradation of gluten (see for instance WO2005/027953 or WO2003/068170).
25 Gluten is the insoluble protein fraction of cereals like wheat, rye, oat and barley. Gluten is a complex mixture of glutenin- and prolamine molecules which are thought to cause toxic effects, for instance in patients suffering from celiac disease. Celiac Sprue or celiac disease is considered to be an autoimmune disease. Patients suffering from Celiac Sprue need to follow a strict gluten-free diet, which is very
30 difficult to follow because gluten is so widely used. The use of proline-specific endoprotease as a medicament or dietary supplement may alleviate the need for a strict gluten free diet (WO2003/068170).

Proline-specific endoproteases are also used for reducing haze in beer, wherein the proline-specific protease may be added during several stages of a beer production process (WO 2002/046381).

It is desirable that enzymes in food and feed applications have a suitable pH optimum and preferably are not active in the final food or beverage.

The aim of the present invention is an alternative proline-specific endoprotease with improved characteristics.

Summary

In one aspect the present invention relates an isolated polypeptide having proline-specific endoprotease activity selected from the group consisting of

i.a polypeptide comprising a mature polypeptide sequence of SEQ ID NO: 2;

ii.a polypeptide that has least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to a mature polypeptide sequence of SEQ ID NO: 2;

iii.a polypeptide encoded by a nucleic acid that hybridizes under medium stringency, preferably under high stringency conditions to the complementary strand of a mature polypeptide coding sequence of SEQ ID NO:1;

iv.a polypeptide encoded by a nucleic acid that has at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to a mature polypeptide coding sequence of SEQ ID NO: 1;

In another aspect the present invention relates to a composition comprising a polypeptide as disclosed herein.

In one aspect the present invention relates to a nucleic acid that has at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to a mature polypeptide coding sequence of SEQ ID NO: 1.

The present invention also relates to an expression vector comprising a polynucleotide encoding a polypeptide according to the present invention.

In another aspect the present invention relates to a recombinant host cell comprising a polynucleotide sequence, or an expression vector as disclosed herein.

In yet another aspect the present invention relates to a method for the preparation of a polypeptide, comprising cultivating a host cell as disclosed herein, under conditions that allow expression of the polypeptide, and preparing the polypeptide.

5 In another aspect the present invention relates to a process for the preparation of a food or feed product comprising incubating an intermediate form of the food or feed product with a polypeptide as disclosed herein, and preparing the food product.

The present invention also relates to a food or feed product obtainable by a process as disclosed herein.

Definitions

10 The term "baked products" is herein defined as any product prepared from a dough or a batter. The product may have a soft or a crisp character and may be of a white, light or dark type. Baked products include, but are not limited to, bread such as for instance white, whole-meal or rye bread, French baguette-type bread, laminated
15 dough products such as (Danish) pastry, croissants or puff pastry, pita bread, tortillas, tacos, cakes, pancakes, biscuits, cookies, doughnuts, bagels, pie crusts, muffins, steamed bread, and crisp bread,. Types of baked products, methods to characterize and to produce them are known to those skilled in the art see for example "Baking
20 Science and Technology", by E.J. Pyler, L.A. Gorton, 2008, (2 volumes) Sosland Publishing Company, Kansas, USA, or "Baked Products: Science, Technology and Practice" by S.P. Cauvain, L.S. Young, 2006, Blackwell Publishing Ltd, Oxford, UK.

25 The term "complementary strand" can be used interchangeably with the term "complement". The complement of a nucleic acid strand can be the complement of a coding strand or the complement of a non-coding strand. When referring to double-stranded nucleic acids, the complement of a nucleic acid encoding a polypeptide refers to the complementary strand of the strand encoding the amino acid sequence or to any nucleic acid molecule containing the same.

30 The term "control sequence" can be used interchangeably with the term "expression-regulating nucleic acid sequence". The term as used herein refers to nucleic acid sequences necessary for and/or affecting the expression of an operably linked coding sequence in a particular host organism or *in vitro*. When two nucleic acid sequences

are operably linked, they usually will be in the same orientation and also in the same reading frame. They usually will be essentially contiguous, although this may not be required. The expression-regulating nucleic acid sequences, such as *inter alia* appropriate transcription initiation, termination, promoter, leader, signal peptide, propeptide, prepropeptide, or enhancer sequences; Shine-Dalgarno sequence, repressor or activator sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion, can be any nucleic acid sequence showing activity in the host organism of choice and can be derived from genes encoding proteins, which are either endogenous or heterologous to a host cell. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. When desired, the control sequence may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide. Control sequences may be optimized to their specific purpose.

A "dairy product" refers to any kind of milk-based product intended to be used as food, feed or beverage, including but not limited to cheese, milk, skimmed milk, acidified milk, butter milk, condensed milk, spreads, margarines, yoghurt, ice cream, milk powder, butter, EMC (Enzyme Modified Cheese), dulce de leche, coffee whitener; coffee creamer, cream, ghee, dairy analogue, etcetera. Cheese may be any kind of cheese, e.g. fresh cheese, hard cheese, curd cheese, cream cheese, white mould cheese, blue mould cheese and process cheese. Examples of fresh cheese are Ricotta, Cream cheese, Neufchatel or Cottage cheese. Examples of hard cheese are Chester, Danbo, Manchego, Saint Paulin, Cheddar, Monterey, Colby, Edam, Gouda, Muenster, Swiss type, Gruyere, Emmenthaler, Parmigiano Reggiano, Grana Padano, Parmesan, Pecorino, Provolone, and Romano. Examples of curd cheese such as Feta cheese, Quotija cheese, pasta filata cheese such as Mozzarella, and Queso fresco cheese. Examples of cream cheese are Philadelphia cheese. Examples of white mould cheese are Brie and Camembert cheese. Examples of blue mould cheese are Gorgonzola and Danish blue cheese.

As used herein, the term "endogenous" refers to a nucleic acid or amino acid sequence naturally occurring in a host.

Endopeptidases or endoproteinases are enzymes that are able to break peptide bonds of nonterminal amino acids (i.e. within the protein), in contrast to exopeptidases, which break peptide bonds either from the amino or the carboxyl terminus. Endopeptidases do not tend to break down peptides into monomers, but result in relatively large peptide fragments. The specific generation of relatively large fragments is highly preferred in many food and feed related applications. A particular case of endopeptidase is the oligopeptidase, whose substrates are oligopeptides instead of proteins.

The term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post transcriptional modification, translation, post-translational modification, and secretion.

An expression vector comprises a polynucleotide coding for a polypeptide, operably linked to the appropriate control sequences (such as a promoter, and transcriptional and translational stop signals) for expression and/or translation *in vitro*, or in the host cell of the polynucleotide.

The expression vector may be any vector (e.g., a plasmid or virus), which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the polynucleotide. The choice of the vector will typically depend on the compatibility of the vector with the cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids. The vector may be an autonomously replicating vector, i.e. a vector, which exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extra-chromosomal element, a mini-chromosome, or an artificial chromosome. Alternatively, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. The integrative cloning vector may integrate at random or at a predetermined target locus in the chromosomes of the host cell. The vector system may be a single vector or plasmid or two or more vectors or plasmids, which together contain the total DNA to be introduced into the genome of the host cell, or a transposon.

A host cell as defined herein is an organism suitable for genetic manipulation and one which may be cultured at cell densities useful for industrial production of a target product, such as a polypeptide according to the present invention. A host cell may be a host cell found in nature or a host cell derived from a parent host cell after genetic manipulation or classical mutagenesis. Advantageously, a host cell is a recombinant host cell.

A host cell may be a prokaryotic, archaeobacterial or eukaryotic host cell. A prokaryotic host cell may be, but is not limited to, a bacterial host cell. An eukaryotic host cell may be, but is not limited to, a yeast, a fungus, an amoeba, an algae, a plant, an animal, or an insect host cell.

The term "heterologous" as used herein refers to nucleic acid or amino acid sequences not naturally occurring in a host cell. In other words, the nucleic acid or amino acid sequence is not identical to that naturally found in the host cell.

The term "hybridization" means the pairing of substantially complementary strands of oligomeric compounds, such as nucleic acid compounds.

Hybridization may be performed under low, medium or high stringency conditions. Low stringency hybridization conditions comprise hybridizing in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by two washes in 0.2X SSC, 0.1% SDS at least at 50°C (the temperature of the washes can be increased to 55°C for low stringency conditions). Medium stringency hybridization conditions comprise hybridizing in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 60°C, and high stringency hybridization conditions comprise hybridizing in 6X SSC at about 45°C, followed by one or more washes in 0.2X SSC, 0.1% SDS at 65°C.

A nucleic acid or polynucleotide sequence is defined herein as a nucleotide polymer comprising at least 5 nucleotide or nucleic acid units. A nucleotide or nucleic acid refers to RNA and DNA. The terms "nucleic acid" and "polynucleotide sequence" are used interchangeably herein.

A "peptide" refers to a short chain of amino acid residues linked by a peptide (amide) bonds. The shortest peptide, a dipeptide, consists of 2 amino acids joined by single peptide bond.

The term "polypeptide" refers to a molecule comprising amino acid residues linked by peptide bonds and containing more than five amino acid residues. The term "protein" as used herein is synonymous with the term "polypeptide" and may also refer to two or more polypeptides. Thus, the terms "protein" and "polypeptide" can be used interchangeably. Polypeptides may optionally be modified (e.g., glycosylated, phosphorylated, acylated, farnesylated, prenylated, sulfonated, and the like) to add functionality. Polypeptides exhibiting activity in the presence of a specific substrate under certain conditions may be referred to as enzymes. It will be understood that, as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding a given polypeptide may be produced.

An "isolated nucleic acid fragment" is a nucleic acid fragment that is not naturally occurring as a fragment and would not be found in the natural state.

The term "isolated polypeptide" as used herein means a polypeptide that is removed from at least one component, e.g. other polypeptide material, with which it is naturally associated. The isolated polypeptide may be free of any other impurities. The isolated polypeptide may be at least 50% pure, e.g., at least 60% pure, at least 70% pure, at least 75% pure, at least 80% pure, at least 85% pure, at least 80% pure, at least 90% pure, or at least 95% pure, 96%, 97%, 98%, 99%, 99.5%, 99.9% as determined by SDS-PAGE or any other analytical method suitable for this purpose and known to the person skilled in the art. An isolated polypeptide may be produced by a recombinant host cell.

A "mature polypeptide" is defined herein as a polypeptide in its final form and is obtained after translation of a mRNA into polypeptide and post-translational modifications of said polypeptide. Post-translational modification include N-terminal processing, C-terminal truncation, glycosylation, phosphorylation and removal of leader sequences such as signal peptides, propeptides and/or prepropeptides by cleavage.

A "mature polypeptide coding sequence" means a polynucleotide that encodes a mature polypeptide.

The term "nucleic acid construct" is herein referred to as a nucleic acid molecule, either single-or double-stranded, which is isolated from a naturally occurring gene or

which has been modified to contain segments of nucleic acid which are combined and juxtaposed in a manner which would not otherwise exist in nature. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains all the control sequences required for expression of a coding sequence, wherein said control sequences are operably linked to said coding sequence.

A "proline-specific endoprotease" is a protease that hydrolyses a protein or peptide at a position where the protein or peptide contains a proline-residue. A proline-specific endoprotease may have proline-specific endoprotease and/or proline-specific oligopeptidase activity (EC3.4.21.26). A proline-specific endoprotease is preferably an enzyme that hydrolyses a peptide bond at the carboxy-terminal end of proline residues, resulting in a peptide and/or polypeptide fragment with a C-terminal proline..

The term "promoter" is defined herein as a DNA sequence that binds RNA polymerase and directs the polymerase to the correct downstream transcriptional start site of a nucleic acid sequence to initiate transcription.

The term "recombinant" when used in reference to a cell, nucleic acid, protein or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, underexpressed or not expressed at all. The term "recombinant" is synonymous with "genetically modified" and "transgenic".

Sequence identity. Sequence identity, or sequence homology are used interchangeably herein. For the purpose of this invention, it is defined here that in order to determine the percentage of sequence homology or sequence identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes. In order to optimize the alignment between the two sequences gaps may be introduced in any of the two sequences that are compared. Such alignment can be carried out over the full length of the sequences being compared. Alternatively, the alignment may be carried out over a shorter length, for

example over about 20, about 50, about 100 or more nucleic acids/bases or amino acids. The sequence identity is the percentage of identical matches between the two sequences over the reported aligned region. The percent sequence identity between two amino acid sequences or between two nucleotide sequences may be determined using the Needleman and Wunsch algorithm for the alignment of two sequences. (Needleman, S. B. and Wunsch, C. D. (1970) J. Mol. Biol. 48, 443-453). Both amino acid sequences and nucleotide sequences can be aligned by the algorithm. The Needleman-Wunsch algorithm has been implemented in the computer program NEEDLE. For the purpose of this invention the NEEDLE program from the EMBOSS package was used (version 2.8.0 or higher, EMBOSS: The European Molecular Biology Open Software Suite (2000) Rice, P. Longden, I. and Bleasby, A. Trends in Genetics 16, (6) pp276—277, <http://emboss.bioinformatics.nl/>). For protein sequences EBLOSUM62 is used for the substitution matrix. For nucleotide sequence, EDNAFULL is used. The optional parameters used are a gap-open penalty of 10 and a gap extension penalty of 0.5. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

After alignment by the program NEEDLE as described above the percentage of sequence identity between a query sequence and a sequence of the invention is calculated as follows: Number of corresponding positions in the alignment showing an identical amino acid or identical nucleotide in both sequences divided by the total length of the alignment after subtraction of the total number of gaps in the alignment. The identity as defined herein can be obtained from NEEDLE by using the NOBRIEF option and is labeled in the output of the program as “longest-identity”.

The nucleic acid and protein sequences of the present invention can further be used as a “query sequence” to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403—10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, word length = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped

alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See the homepage of the National Center for Biotechnology Information at <http://www.ncbi.nlm.nih.gov/>.

The term "substantially pure" with regard to polypeptides refers to a polypeptide preparation which contains at the most 50% by weight of other polypeptide material. The polypeptides disclosed herein are preferably in a substantially pure form. In particular, it is preferred that the polypeptides disclosed herein are in "essentially pure form", i.e. that the polypeptide preparation is essentially free of other polypeptide material. Optionally, the polypeptide may also be essentially free of non-polypeptide material such as nucleic acids, lipids, media components, and the like. Herein, the term "substantially pure polypeptide" is synonymous with the terms "isolated polypeptide" and "polypeptide in isolated form". The term "substantially pure" with regard to polynucleotide refers to a polynucleotide preparation which contains at the most 50% by weight of other polynucleotide material. The polynucleotides disclosed herein are preferably in a substantially pure form. In particular, it is preferred that the polynucleotide disclosed herein are in "essentially pure form", i.e. that the polynucleotide preparation is essentially free of other polynucleotide material. Optionally, the polynucleotide may also be essentially free of non-polynucleotide material such as polypeptides, lipids, media components, and the like. Herein, the term "substantially pure polynucleotide" is synonymous with the terms "isolated polynucleotide" and "polynucleotide in isolated form".

A "synthetic molecule", such as a synthetic nucleic acid or a synthetic polypeptide is produced by in vitro chemical or enzymatic synthesis. It includes, but is not limited to, variant nucleic acids made with optimal codon usage for host organisms of choice.

A synthetic nucleic acid may be optimized for codon use, preferably according to the methods described in WO2006/077258 and/or WO2008000632, which are herein incorporated by reference. WO2008/000632 addresses codon-pair optimization. Codon-pair optimization is a method wherein the nucleotide sequences encoding a polypeptide that have been modified with respect to their codon-usage, in particular the codon-pairs that are used, are optimized to obtain improved expression of the

nucleotide sequence encoding the polypeptide and/or improved production of the encoded polypeptide. Codon pairs are defined as a set of two subsequent triplets (codons) in a coding sequence. Those skilled in the art will know that the codon usage needs to be adapted depending on the host species, possibly resulting in variants with significant homology deviation from SEQ ID NO: 1, but still encoding the polypeptide according to the invention.

As used herein, the terms "variant", "derivative", "mutant" or "homologue" can be used interchangeably. They can refer to either polypeptides or nucleic acids. Variants include substitutions, insertions, deletions, truncations, transversions, and/or inversions, at one or more locations relative to a reference sequence. Variants can be made for example by site-saturation mutagenesis, scanning mutagenesis, insertional mutagenesis, random mutagenesis, site-directed mutagenesis, and directed-evolution, as well as various other recombination approaches known to a skilled person in the art. Variant genes of nucleic acids may be synthesized artificially by known techniques in the art.

FIGURES

Figure 1: pGBTOP-16 vector used for cloning of the GLA gene. The pGBTOP-16 vector is derived from the pGBTOP-12 vector described in WO 2011/009700. In addition to pGBTOP-12, it contains the *ccdB* gene from *E. coli* for positive selection for presence of an insert between the *EcoRI* and *PacI* cloning sites. The *PacI* restriction site replaces the *SnaBI* restriction site present in pGBTOP-12. This vector is linearized by *NotI* digestion prior to transformation.

Detailed description

In one aspect the present invention relates to a polypeptide having proline-specific endoprotease activity, selected from the group consisting of

- i. a polypeptide comprising a mature polypeptide sequence of SEQ ID NO: 2;
- ii. a polypeptide that has least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to a mature polypeptide sequence of SEQ ID NO: 2;

iii. a polypeptide encoded by a nucleic acid that hybridizes under medium stringency, preferably under high stringency conditions to the complementary strand of a mature polypeptide coding sequence of SEQ ID NO:1;

5 iv. a polypeptide encoded by a nucleic acid that has at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to a mature polypeptide coding sequence of SEQ ID NO: 1;

10 In one embodiment the present invention relates to a polypeptide that is an isolated, substantially pure, pure, recombinant, synthetic or variant polypeptide of the polypeptide as disclosed herein.

 In another aspect the present invention relates an isolated polypeptide having proline-specific endoprotease activity comprising the amino acid sequence of SEQ ID NO: 2. A proline-specific endoprotease as disclosed herein may be derived from
15 *Aspergillus carbonarius*. The wording "derived" or "derivable" from with respect to the origin of a polypeptide as disclosed herein, means that when carrying out a BLAST search with a polypeptide according to the present invention, the polypeptide according to the present invention may be derivable from a natural source, such as a microbial cell, of which an endogenous polypeptide shows the highest percentage
20 homology or identity with the polypeptide as disclosed herein.

 Advantageously, a polypeptide as disclosed herein has an improved property as compared to a proline-specific endoprotease known in the art. An improved property may be a reduced activity of a polypeptide provided by the invention after pasteurization conditions, such as pasteurization conditions that are applied in the
25 preparation of a food product. A reduced activity may be less than 80%, or less than 75%, 70%, 65%, 60%, 55%, 50%, 45%, 40%, 30%, 20%, 10% of the proline-specific endoprotease activity remains after pasteurizing the polypeptide as compared to a polypeptide of the invention that has not been submitted to pasteurization. Accordingly, a proline-specific endoprotease as disclosed herein can be readily used
30 in food applications, since there is a desire by the food industry that no active enzyme remains in the final food product. An improved property of a polypeptide having proline-specific endoprotease may also be an increased thermostability, or an increased specific activity at specific temperatures and pH values.

Preferably, a polypeptide provided by the invention may be a polypeptide that has least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the mature polypeptide sequence of SEQ ID NO: 2. A polypeptide according to SEQ ID NO: 2 comprises a pre-pro-sequence. Upon secretion of the polypeptide through a host cell's membrane, the pre-pro-sequence, such as amino acids 1 to 30, 1 to 31, 1 to 32, 1 to 33, 1 to 34, 1 to 35, 1 to 36, 1 to 37, 1 to 38, 1 to 39, 1 to 40, 1 to 41 of SEQ ID NO: 2 are removed. The mature polypeptide sequence having proline-specific endoprotease activity of SEQ ID NO:2 may comprise amino acids 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, or 45 to amino acid 512, 513, 514, 515, 516, 517, 518, 519, 520 or 521 of SEQ ID NO:2, wherein, the amino acid methionine at position 1 in SEQ ID NO: 2 is counted as number 1. Advantageously, the mature polypeptide sequence having proline-specific endoprotease activity comprises amino acids 37 to 521 of SEQ ID NO:2, wherein the amino acid methionine at position 1 in SEQ ID NO: 2 is counted as number 1.

A polypeptide according to the present invention may be encoded by any suitable polynucleotide sequence. Typically a polynucleotide sequence is codon optimized, or a codon pair optimized sequence for expression of a polypeptide as disclosed herein in a particular host cell.

A polypeptide as disclosed herein may be encoded by a polynucleotide that hybridizes under medium stringency, preferably under high stringency conditions to the complementary strand of the mature polypeptide coding sequence of SEQ ID NO:1. SEQ ID NO: 1 is a codon optimized polynucleotide sequence for expression of a polypeptide as disclosed herein in an *Aspergillus niger* host cell.

A polypeptide as disclosed herein may also be encoded by a nucleic acid that has at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to a mature polypeptide coding sequence of SEQ ID NO: 1.

A polypeptide having proline-specific endoprotease activity as disclosed herein may also be a variant of a mature polypeptide of SEQ ID NO: 2, comprising a substitution, deletion and/or insertion at one or more positions of the mature polypeptide SEQ ID NO: 2. A variant of the mature polypeptide of SEQ ID NO:2 may be an amino acid sequence that differs in 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 amino acids from the amino acids of of the mature polypeptide of SEQ ID NO:2.

In one embodiment the present invention features a biologically active fragment of a polypeptide as disclosed herein.

Biologically active fragments of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the proline-specific endoprotease protein (e.g., the mature amino acid sequence of SEQ ID NO:2), which include fewer amino acids than the full length protein but which exhibits at least one biological activity of the corresponding full-length protein. Typically, biologically active fragments comprise a domain or motif with at least one activity of the proline-specific endoprotease protein. A biologically active fragment may for instance comprise a catalytic domain. A biologically active fragment of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the biological activities of the native form of a polypeptide of the invention.

The invention also features nucleic acid fragments which encode the above biologically active fragments of the proline-specific endoprotease protein.

A polypeptide according to the present invention may be a fusion protein. Techniques for producing fusion polypeptides are known in the art, and include ligating the coding sequences encoding the polypeptides so that they are in frame. Expression of the fused polypeptide is under control of the same promoter (s) and terminator. The hybrid polypeptides may comprise a combination of partial or complete polypeptide sequences obtained from at least two different polypeptides wherein one or more may be heterologous to a host cell. Such fusion polypeptides from at least two different polypeptides may comprise a binding domain from one polypeptide, operably linked to a catalytic domain from a second polypeptide. Examples of fusion polypeptides and signal sequence fusions are for example as described in WO2010/121933, WO2013/007820 and WO2013/007821.

A polypeptide according to the present invention may be derived from any suitable eukaryotic cell. A eukaryotic cell may be a mammalian, insect, plant, fungal, or algal cell.

A polypeptide according to the present invention may also be derived from a filamentous fungal cell or thermophilic filamentous fungal cell. Preferred filamentous

5 fungal cells belong to a species of an *Acremonium*, *Aspergillus*, *Chrysosporium*,
Myceliophthora, *Penicillium*, *Talaromyces*, *Rasamsonia*, *Thielavia*, *Fusarium* or
Trichoderma, *Amorphotheca*, *Pseudocercospora*, *Trametes*, *Rhizomucor*,
Calcarisporiella, *Thermomyces*, *Thermoascus*, *Coryascus*, *Myricoccum*, *Scytalidium*,
10 *Chaetomium*, *Paecilomyces*, *Corynascus*, *Malbranchea*, *Stilbella*, *Thermomyces*,
Dactylomyces, *Humicola*, *Chaetomium*, *Melanocarpus*, *Rhizomucor*, *Lentinula*,
Anaeromyces genus, and most preferably belong to a species of *Aspergillus niger*,
Acremonium alabamense, *Aspergillus awamori*, *Aspergillus foetidus*, *Aspergillus*
sojae, *Aspergillus fumigatus*, *Talaromyces emersonii*, *Rasamsonia emersonii*,
15 *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*
20 *auratiacus*, *Coryascus thermophilus*, *Myricoccum thermophilum*, *Scytalidium*
thermophilum, *Myceliophthora henningsii*, *Chaetomium thermophilum*, *Paecilomyces*
byssochlamydoides, *Corynascus sepedonium*, *Malbranchea cinnamomea*, *Thielavia*
australiensis, *Stilbella thermophila*, *Thermomyces stellatus*, *Talaromyces emersonii*,
Dactylomyces thermophilus, *Humicola hyalothermophila*, *Acremonium thermophilum*,
30 *Chaetomium olivicolor*, *Melanocarpus albomyces*, *Rhizomucor miehei*, *Lentinula*
edodes or *Anaeromyces mucronatus*. A polypeptide preferably is derivable from
Aspergillus carbonarius.

A polypeptide according to the present invention may be a naturally occurring polypeptide or a genetically modified or recombinant polypeptide.

25 A polypeptide as disclosed herein may be purified. Purification of protein is known to a skilled person in the art

In one aspect the present invention relates to a composition comprising a polypeptide as disclosed herein.

30 A composition as disclosed herein, may comprise a carrier, an excipient, an auxiliary enzyme, or other compounds. Typically a composition, or a formulation, comprises a compound with which a proline-specific endoprotease may be formulated.

An excipient as used herein is an inactive substance formulated alongside with a polypeptide as disclosed herein, for instance sucrose or lactose, glycerol, sorbitol or sodium chloride. A composition comprising a polypeptide as disclosed herein may be a liquid composition or a solid composition. A liquid composition usually comprises water. When formulated as a liquid composition, the composition usually comprises components that lower the water activity, such as glycerol, sorbitol or sodium chloride (NaCl). A solid composition comprising a polypeptide as disclosed herein may comprise a granulate comprising the enzyme or the composition comprises an encapsulated polypeptide in liquid matrices like liposomes or gels like alginate or carrageenans. There are many techniques known in the art to encapsulate or granulate a polypeptide or enzyme (see for instance G.M.H. Meesters, "Encapsulation of Enzymes and Peptides", Chapter 9, in N.J. Zuidam and V.A. Nedović (eds.) "Encapsulation Technologies for Active Food Ingredients and food processing" 2010). A composition as disclosed herein may also comprise a carrier comprising a polypeptide as disclosed herein. A polypeptide as disclosed herein may be bound or immobilized to a carrier by known technologies in the art.

The present invention also relates to a process for preparing a composition comprising a polypeptide as disclosed herein, which may comprise spray drying a fermentation medium comprising the polypeptide, or granulating, or encapsulating a polypeptide as disclosed herein, and preparing the composition.

In another aspect the present invention relates to a polynucleotide sequence encoding a polypeptide as disclosed herein, which has at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 1, or to the mature polypeptide coding sequence of SEQ ID NO:1. A polynucleotide sequence as disclosed herein may comprise SEQ ID NO: 1, or may comprise the mature polypeptide coding sequence of SEQ ID NO:1.

In one other embodiment of the present invention a nucleic acid is disclosed that is an isolated, substantially pure, pure, recombinant, synthetic or variant nucleic acid of the nucleic acid of SEQ ID NO: 1. A variant nucleic acid sequence may for instance have at least 80% sequence identity to SEQ ID NO:1.

In another aspect, the present invention relates to an expression vector comprising a polynucleotide as disclosed herein operably linked to one or more

control sequence(s) that direct expression of the polypeptide in an expression host cell.

There are several ways of inserting a nucleic acid into a nucleic acid construct or an expression vector which are known to a skilled person in the art, see for instance Sambrook & Russell, *Molecular Cloning: A Laboratory Manual, 3rd Ed.*, CSHL Press, Cold Spring Harbor, NY, 2001. It may be desirable to manipulate a nucleic acid encoding a polypeptide of the present invention with control sequences, such as promoter and terminator sequences.

A promoter may be any appropriate promoter sequence suitable for a eukaryotic or prokaryotic host cell, which shows transcriptional activity, including mutant, truncated, and hybrid promoters, and may be obtained from polynucleotides encoding extracellular or intracellular polypeptides either endogenous (native) or heterologous (foreign) to the cell. The promoter may be a constitutive or inducible promoter. Preferably, the promoter is an inducible promoter, for instance a starch inducible promoter. Promoters suitable in filamentous fungi are promoters which may be selected from the group, which includes but is not limited to promoters obtained from the polynucleotides encoding *A. oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus gpdA* promoter, *A. niger* neutral alpha-amylase, *A. niger* acid stable alpha-amylase, *A. niger* or *A. awamori* glucoamylase (glaA), *A. niger* or *A. awamori* endoxylanase (xlnA) or beta-xylosidase (xlnD), *T. reesei* cellobiohydrolase I (CBHI), *R. miehei* lipase, *A. oryzae* alkaline protease, *A. oryzae* triose phosphate isomerase, *A. nidulans* acetamidase, *Fusarium venenatum* amyloglucosidase (WO 00/56900), *Fusarium venenatum* Dania (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), *Fusarium oxysporum* trypsin-like protease (WO 96/00787), *Trichoderma reesei* beta-glucosidase, *Trichoderma reesei* cellobiohydrolase I, *Trichoderma reesei* cellobiohydrolase II, *Trichoderma reesei* endoglucanase I, *Trichoderma reesei* endoglucanase II, *Trichoderma reesei* endoglucanase III, *Trichoderma reesei* endoglucanase IV, *Trichoderma reesei* endoglucanase V, *Trichoderma reesei* xylanase I, *Trichoderma reesei* xylanase II, *Trichoderma reesei* beta-xylosidase, as well as the NA2-tpi promoter (a hybrid of the promoters from the polynucleotides encoding *A. niger* neutral alpha-amylase and *A. oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof.

Any terminator which is functional in a cell as disclosed herein may be used, which are known to a skilled person in the art. Examples of suitable terminator sequences in filamentous fungi include terminator sequences of a filamentous fungal gene, such as from *Aspergillus* genes, for instance from the gene *A. oryzae* TAKA amylase, the genes encoding *A. niger* glucoamylase (*glaA*), *A. nidulans* anthranilate synthase, *A. niger* alpha-glucosidase, *trpC* and/or *Fusarium oxysporum* trypsin-like protease.

In another aspect the present invention relates to a host cell comprising a nucleic acid or a nucleic acid construct or an expression vector as disclosed herein.

A suitable host cell may be a mammalian, insect, plant, fungal, or algal cell, or a bacterial cell. A suitable host cell may be a fungal cell, for instance from the genus *Acremonium*, *Aspergillus*, *Chrysosporium*, *Fusarium*, *Myceliophthora*, *Penicillium*, *Rasamsonia*, *Talaromyces*, *Thielavia*, *Trichoderma*, *Saccaromyces*, *Kluyveromyces*, *Pichia*, for instance *Aspergillus niger*, *Aspergillus awamori*, *Aspergillus foetidus*, *A. oryzae*, *A. sojae*, *Talaromyces emersonii*, *Rasamsonia emersonii*, *Chrysosporium lucknowense*, *Fusarium oxysporum*, *Myceliophthora thermophila*, *Thielavia terrestris* or *Trichoderma reesei* or, *Saccharomyces cerevisiae*, *Kluyveromyces lactis*, *Pichia pastoris*

A host cell may be a recombinant or transgenic host cell. The host cell may be genetically modified with a nucleic acid construct or expression vector as disclosed herein with standard techniques known in the art, such as electroporation, protoplast transformation or conjugation for instance as disclosed in Sambrook & Russell, *Molecular Cloning: A Laboratory Manual*, 3rd Ed., CSHL Press, Cold Spring Harbor, NY, 2001.

In one aspect the present invention relates to a process for the production of a polypeptide as disclosed herein comprising cultivating a host cell in a suitable fermentation medium under conditions conducive to the production of the polypeptide and producing the polypeptide. A skilled person in the art understands how to perform a process for the production of a polypeptide as disclosed herein depending on a host cell used, such as pH, temperature and composition of a fermentation medium. Host cells can be cultivated in shake flasks, or in fermenters having a volume of 0.5 or 1 litre or larger to 10 to 100 or more cubic metres. Cultivation may be performed aerobically or anaerobically depending on the requirements of a host cell.

Advantageously a polypeptide as disclosed herein is recovered or isolated from the fermentation medium.

A polypeptide having proline-specific endoprotease activity or a composition comprising a polypeptide as disclosed herein may be used in a large variety of applications, for instance in the production of a food or feed product, such as in the production of a protein hydrolysate. Several food proteins contain highly allergenic subfractions which may be even toxic to specific individuals, such as gluten that contains prolamines with proline-rich peptide sequences. These proteins can be subjected to the new enzyme to alleviate their antigenicity or toxicity.

A group of people to which gluten is toxic are individuals suffering from Celiac Sprue. Celiac Sprue, also known as celiac disease, is an autoimmune disease of the small intestine caused by the ingestion of gluten proteins from cereals, such as alpha-gliadin from wheat, hordein from barley, secalin from rye and avenin from oats.

Accordingly, a polypeptide having proline-specific endoprotease activity or a composition comprising a polypeptide as disclosed herein may be used in the preparation of a dietary supplement or as a medicament in the treatment of a patient suffering from Celiac Sprue, or in the treatment of gluten intolerant people.

A polypeptide as disclosed herein may also be used as a processing aid to hydrolyse gluten in a food product.

Accordingly the present invention relates to a process for the preparation of a food or feed product comprising incubating an intermediate form of a food or feed product with a polypeptide or composition comprising the polypeptide as disclosed herein and preparing the food or feed product. A food product in a process as disclosed herein includes a beverage, such as beer, wine or fruit juice, or a baked product, or a dairy product, but is not limited thereto.

A food product and / or an intermediate form of a food product may comprise gluten.

It was found that a polypeptide having proline-specific endoprotease activity as disclosed herein was capable of hydrolysing the toxic epitopes in gluten into non-toxic fragments.

An intermediate form of a food product may be any suitable form of a food product during the preparation of the food product. For instance, an intermediate form

of beer, may be a mash and an intermediate form of bread may be a dough or a batter.

A process for the preparation of a food product according to the present disclosure may comprise a step of pasteurizing the food product. Pasteurization usually comprises heating a food product, or an intermediate form of a food product, for instance by bringing the food product or intermediate form of a food product to a temperature of between 60 to 68°C between 10 to 20 min, or between 12 and 18 min, or to a temperature of between 70-74°C, such as about 72 °C for at least 5, 10 or 15 seconds.

Advantageously, it was found that less than 70%, 60%, 50%, or less than 40% of the proline-specific endoprotease activity originally present was found in the pasteurized food product, or even no proline-specific activity was found after the food product was pasteurized.

A food product in a process as disclosed herein may also be a protein hydrolysate. Accordingly, the present disclosure relates to a process for the preparation of a protein hydrolysate, comprising contacting a protein substrate with a polypeptide or a composition as disclosed herein, and producing the protein hydrolysate. A protein hydrolysate may be prepared from any suitable a protein substrate, for instance a protein substrate that is rich in proline residues, such as gluten in cereals or caseins in bovine milk.

In one aspect the present invention relates to a food product obtainable by a process for the preparation of a food product as disclosed herein.

The following examples illustrate the invention.

EXAMPLES

MATERIALS and METHODS

Example 1. Cloning, Expression and Recovery proline-specific endoprotease (PEP) BC2G075

Example 1.1. Cloning and Expression

The protein sequence of the proline-specific endoprotease (PEP) according to present invention is shown in SEQ ID NO: 2. The first part of SEQ ID NO: 2 comprises 17 amino acids of a signal sequence of pectinmethylesterase of *A. niger* (PMeA ss; SEQ ID NO: 3) for efficient secretion in *Aspergillus niger*. The following part comprises
5 19 amino acids of a prosequence of *A. carbonarius* proline specific endoprotease (SEQ ID NO: 4).

A codon-adapted DNA sequence for expression of this protein in *Aspergillus niger* was designed containing additional restriction sites for subcloning in an *Aspergillus* expression vector. Codon adaptation was performed as described in WO
10 2008/000632. The codon optimized DNA sequence for *A. niger* of the gene encoding the PEP protein of SEQ ID: NO: 2 is shown in SEQ ID NO: 1.

The translational initiation sequence of the glucoamylase *glaA* promoter was modified into 5'-CACCGTCAAA ATG-3' and an optimal translational termination sequence 5'-TAAA-3' was used in the generation of the expression construct (as also
15 detailed in WO2006/077258). A DNA fragment containing a.o. part of the glucoamylase promoter and the PEP encoding gene was synthesized completely, purified and digested with EcoRI and PaeI. The pGBTOP-16 vector (Figure 1) was linearized by EcoRI/PaeI digestion and the linearized vector fragment was subsequently purified by gel-extraction. The DNA fragment containing the PEP coding
20 region was cloned into the pGBTOP-16 vector resulting in pGBTOP-PEP. Subsequently, *A. niger* GBA 306 ($\Delta glaA$, $\Delta pepA$, $\Delta hdfA$, adapted *Bam*HI amplicon, $\Delta amyBII$, $\Delta amyBI$, $\Delta amyA$ alpha-amylase and glucoamylase negative strain) was transformed with linearized pGBTOP-PEP vector by *NotI*-digestion, in a co-transformation protocol with linearized pGBAAS-4, with strain and methods as
25 described in WO 2011/009700 and references therein, and selected on acetamide containing media and colony purified according to standard procedures. Transformation and selection was performed as described in WO 98/46772 and WO 99/32617. Strains containing the PEP gene were selected via PCR with primers specific for the PEP gene to verify presence of the pGBTOP-PEP expression
30 cassette. A single transformant is selected, named PEP1, and further replica-plated to obtain a single strain inoculum.

Example 1.2. Production of PEP BC2G075 in *A. niger* PEP1 strain

Fresh *A. niger* PEP-1 spores were prepared. 4 shake flasks with 100 ml fermentation medium 1 (10 % w/v Corn Steep Solids, 1 % w/v glucose.H₂O, 0.1 % w/v NaH₂PO₄.H₂O, 0.05 % w/v MgSO₄.7H₂O, 0.025 % w/v Basildon, pH 5.8) in 500 ml shake flasks with baffle were inoculated with 10⁷ spores. These pre-cultures were incubated at 34 °C and 170 rpm for 16-24 hours. From the pre-cultures, 50 ml was used for inoculation of 1 shake flasks with 1 liter Fermentation medium 2 (15 % w/v maltose, 6 % w/v bacto-soytone, 1.5 % w/v (NH₄)₂SO₄, 0.1 % w/v NaH₂PO₄.H₂O, 0.1 % w/v MgSO₄.7H₂O, 0.1 % w/v L-arginine, 8 ‰ w/v Tween-80, 2 ‰ w/v Basildon, 2 % w/v MES pH 5.1) in a 5 liter shake flask size and shaken at 34 °C and 170 rpm. After 3, 4, 5, and 6 days incubation the pH of the culture was lowered to pH 5.0 using 2 N HCl and samples from each of these time points were analysed for PEP activity. 50 mL samples were taken and the supernatant was separated from the biomass by centrifugation and subsequent filtering. The sample with the highest activity was used to characterize the PEP produced.

Example 1.3. Production of the reference proline-specific endoprotease from *A. niger*

The proline-specific endoprotease from *A. niger* is known from WO2002/046381. The protein sequence of proline-specific endoprotease (PEP) from *A. niger* is shown in SEQ ID NO: 5, wherein the first 17 amino acids are a signal sequence of pectinemethylesterase of *A. niger* (PMeA ss; SEQ ID NO: 3) and the following part comprises 19 amino acids of the prosequence of *A. niger* proline specific endoprotease.

Expression and production of PEP from *A. niger* was performed in the same way as described under example 1.1. and 1.2. The *A. niger* PEP is used as the reference enzyme in Example 2.

Example 2. Proline-specific endoprotease (PEP) activity measurements BC2G075

100 µL of culture supernatant as produced in Example 1, diluted in 0.1 M sodium acetate buffer at pH4.5 with 50 mM NaCl, was incubated with 100 µL 6 mM Ac-AAP-pNA (acetyl-AlaAlaPro-paranitroaniline from Selleckchem or CPC Scientific; purity >95.0% based on HPLC analysis) in 0.1 M NaAc buffer at pH4.5 with 50 mM NaCl, in a Nunc 96 well flat bottom MTP (micro-titer plate). After 60 minutes at 20 °C the

reaction was stopped by adding 40 μ L of 1 M HCl. The pNA which had been liberated by PEP was measured in a Tecan MTP spectrophotometer at 405 nm (A405) (www.tecan.com). The blank was prepared by mixing the diluted culture supernatant with the substrate solution which had been mixed with the HCl solution beforehand.

5

The activity is expressed in pNASU's.

1pNASU is the amount of enzyme which liberates from Ac-AAP-pNA in 1 hour the amount of pNA that corresponds to an increase in absorption at 405 nm of 1 OD, using the conditions as described above. The A405 should not be below the blank value at the start of the reaction, or above 2.5 at the end of the reaction, nor may the A405 exceed the linear range of the spectrophotometer that is used.

10

Table 1. Relative activity of PEP BC2G075 compared to the reference PEP from *A. niger*.

Description	Activity (pNASU)
PEP reference	100%
PEP BC2G075	221%

15

Example 3. Thermal stability proline-specific endoprotease BC2G075

To assess the thermal stability of PEP BC2G075 the activity assay was preceded by an incubation of 100 μ L aliquots of a tenfold dilution of the culture supernatant produced in Example 1 in buffer (0.1 M NaAc pH 4.5, with 50 mM NaCl) at 55°C and 65°C for 15 min in a PCR plate in a PCR machine. After the 15 min incubation the samples were rapidly cooled to 25°C in the PCR machine. The pNASU/mL of every sample was measured. The initial activity measured before incubation at elevated temperature (0 minutes) is used as reference (100%) to determine the residual activity.

20
25

Table 2. Proline-specific endoprotease activity of BC2G075 after 15 minutes incubation at 55°C and 65°C

Description	Residual activity after 15' at 55°C (pNASU)	Residual activity at after 15' at 65°C (pNASU)
PEP BC2G075	95%	58%

The results in Table1 show that up to 55°C PEP BC2G075 is hardly prone to any inactivation while at 65°C the residual activity declined significantly.

Example 4. Determination of the molecular mass of the mature PEP BC2G075 with LC-MS

For LC-MS analysis of the mature BC2G075 PEP, 100 µl sample of the supernatant of a fermentation medium produced in Example 1.2. was mixed with 100 µl 20% TCA (Chem Lab NV, Belgium). This solution was put on ice for 1 hour. After precipitation of the proteins, the sample was centrifuged at 14000 rpm and 4°C, for 15 min. After centrifugation the supernatant was removed and the pellet was washed once with 500 µl acetone (- 20°C, Sigma-Aldrich, Netherlands) and centrifuged at 14000 rpm and 4°C for 10 min. The supernatant was removed and the pellet was dissolved in 50 µl 50 mM NaOH (Sigma-Aldrich, Netherlands) and then 350 µl 100 mM NH₄HCO₃ (Sigma-Aldrich, Netherlands) was added. Then 200 µl of the sample was reduced and alkylated. For reduction, 5 µl TCEP (Sigma-Aldrich, Netherlands) was added to the solution and incubated at room temperature in a thermomixer at 1000 rpm for 30 min. For the alkylation 5 µl of 550 mM IAA (Iodoacetamide, Sigma-Aldrich, Netherlands) was added and incubated in a thermomixer at 1000 rpm in the dark for 30 min. In order to deglycosylate the sample, 20 µl PNGase F (Promega, USA) was added. The sample was incubated in thermomixer at 1000 rpm at 37°C overnight. The next day 1% Formic Acid (Merk, Germany) was added in order to dilute the sample to a concentration of 50 µg/ml. The protein concentration was measured on the Qubit quantitative protein assay (Life technologies). The sample was analyzed on the Acquity I-class - Synapt G2-S (Waters, UK), with the following parameters: column: Waters Acquity UPLC BEH300 C4 1.7µm 300Å pore size 2.1×50 mm; column temperature: 75 °C; injection volume: 1 µl; mobile phase A: Formic Acid 0,1% in Mili-Q (Biosolve, Netherlands); mobile phase B: Formic Acid 0,1% in Acetonitrile (Biosolve,

Netherlands). A 15 minutes gradient was applied to the column by varying phase A and B (from 20 to 50 % B). The MS detector settings were: acquisition mass range 500-3500 m/z, scan time 1 sec, positive ESI, TOF MS Resolution mode, with data correction using Leu-Enk (Sigma-Aldrich, Netherlands) as lock mass on the fly during the run. Data spectral deconvolution, charge state stripping, was performed with a Waters MassLynx MaxEnt1 software tool: output mass resolution = 1 Da / channel; damage model: Gaussian (FWHH = 0.750 Da; minim intensity ratios = 33 % left and right); iterate to converge.

The mass spectrum shows a peak at 54885 Da. After correction for alkylation of the cysteines (7 cysteines, each IAA group contributes 57 Da) and deglycosylation (7 theoretical glycosylation sites contributing 1 Da each) the observed mass corresponds to amino acids 37-521 of SEQ ID NO 2. As a consequence the N-terminus of the major species in the mature PEP BC2G075 sample starts at ATTGEAY.

CLAIMS

1. A polypeptide having proline-specific endoprotease activity, selected from the group consisting of

i. a polypeptide comprising a mature polypeptide sequence of SEQ ID NO: 2;

ii. a polypeptide that has least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the mature polypeptide sequence of SEQ ID NO: 2;

iii. a polypeptide encoded by a nucleic acid that hybridizes under medium stringency, preferably under high stringency conditions to the complementary strand of the mature polypeptide coding sequence of SEQ ID NO:1;

iv. a polypeptide encoded by a nucleic acid that has at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identity to the mature polypeptide coding sequence of SEQ ID NO: 1.

2. A polypeptide that is an isolated, substantially pure, pure, recombinant, synthetic or variant polypeptide of the polypeptide of claim 1.

3. A polypeptide according to claim 1 or 2, wherein the mature polypeptide sequence comprises amino acids 37 to 521 of SEQ ID NO:2, wherein the amino acid methionine at position 1 in SEQ ID NO: 2 is counted as number 1.

4. A composition comprising a polypeptide according to any one of the claims 1 to 3.

5. A composition according to claim 4, comprising a carrier, an excipient, or an auxiliary enzyme.

6. A nucleic acid encoding a proline-specific endoprotease, which has at least 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% sequence identity to the mature polypeptide encoding sequence of SEQ ID NO: 1.

7. A nucleic acid that is an isolated, substantially pure, pure, recombinant, synthetic or variant nucleic acid of the nucleic acid of claim 6
- 5 8. An expression vector comprising a nucleic acid according to claim 7 operably linked to one or more control sequence(s) that direct expression of the polypeptide in a host cell.
9. A recombinant host cell comprising a nucleic acid according to claim 6 or 7, or
10 an expression vector according to claim 8.
10. A method for the preparation of a polypeptide according to claims 1 to 3, comprising cultivating a host cell according to claim 9 in a suitable fermentation medium, under conditions that allow expression of the polypeptide, and
15 preparing the polypeptide.
11. A method according to claim 10, further comprising recovering the polypeptide.
12. A process for the preparation of a food or feed product comprising bringing an
20 intermediate form of a food or feed product into contact with a polypeptide according to claim 1 to 3, or a composition according to claim 4 or 5, and preparing the food or feed product.
13. A process according to claim 12, wherein the food product is a beverage,
25 preferably a beer.
14. A process according to claims 12 or 13, wherein the food product comprises gluten.
- 30 15. A food or feed product obtainable by a process according to any one of the claims 12 to 14.

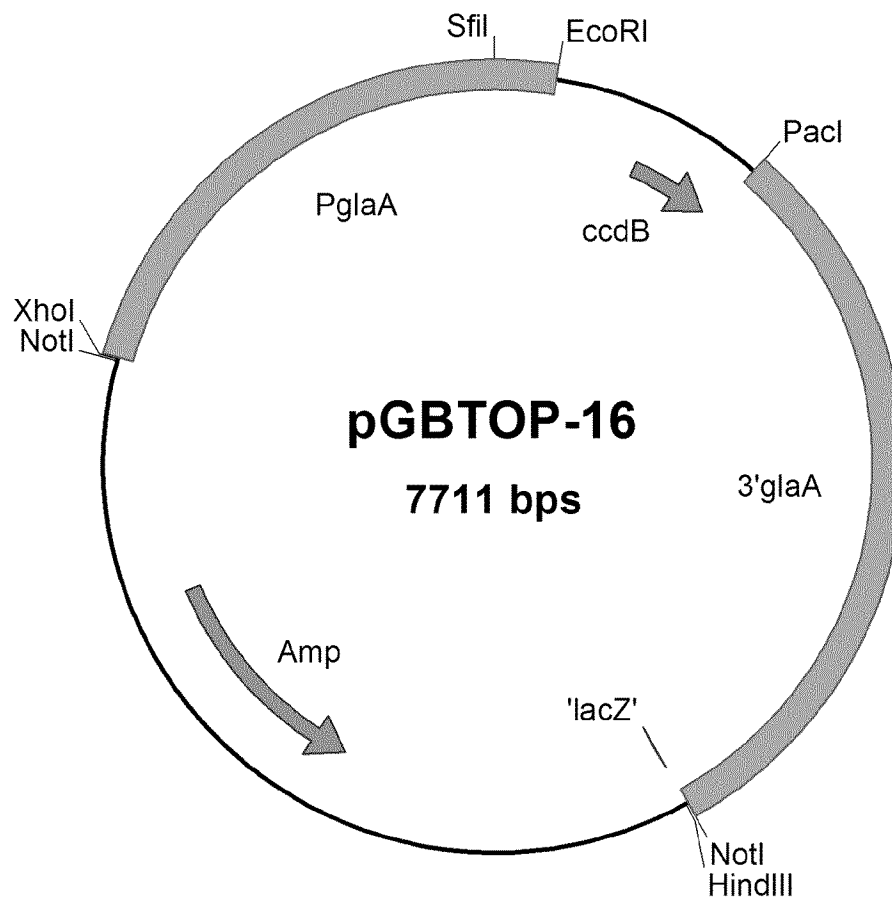


Fig 1

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/061001

A. CLASSIFICATION OF SUBJECT MATTER INV. A23L1/105 C12N9/62 C12N15/63 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) A23L C12N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 03/104382 A1 (DSM IP ASSETS BV [NL]; EDENS LUPPO [NL]; LOPEZ MICHEL [FR]) 18 December 2003 (2003-12-18) see SEQ ID 4 -----	1-15
X	WO 2005/027953 A2 (DSM IP ASSETS BV [NL]; EDENS LUPPO [NL]; VAN DER HOEVEN ROBERTUS ANTON) 31 March 2005 (2005-03-31) Aspergillus niger proline specific endoprotease; examples 1-11 -----	1-15
X	WO 2009/144269 A1 (DSM IP ASSETS BV [NL]; DEKKER PETRUS JACOBUS THEODORU [NL]; EDENS LUPP) 3 December 2009 (2009-12-03) claims 1-27 ----- -/-	1-15
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 17 July 2015		Date of mailing of the international search report 29/07/2015
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Vollbach, Silke

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2015/061001

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MITEA C ET AL: "Efficient degradation of gluten by a prolyl endoprotease in a gastrointestinal model: implications for coeliac disease", GUT, BRITISH MEDICAL ASSOCIATION, LONDON, UK, vol. 57, no. 1, 1 January 2008 (2008-01-01), pages 25-32, XP009113711, ISSN: 0017-5749 page 26 - page 29 -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2015/061001

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 03104382	A1	18-12-2003	AR 040950 A1 27-04-2005
			AU 2003228145 A1 22-12-2003
			BR 0311655 A 15-03-2005
			CA 2490020 A1 18-12-2003
			CN 1659271 A 24-08-2005
			DK 1511835 T3 05-01-2015
			EP 1511835 A1 09-03-2005
			ES 2525873 T3 30-12-2014
			JP 4713150 B2 29-06-2011
			JP 2005528906 A 29-09-2005
			US 2006257544 A1 16-11-2006
			WO 03104382 A1 18-12-2003
			ZA 200409447 A 14-10-2005
WO 2005027953	A2	31-03-2005	JP 5328100 B2 30-10-2013
			JP 2007505639 A 15-03-2007
			JP 2011120601 A 23-06-2011
			RU 2370279 C2 20-10-2009
			US 2007031399 A1 08-02-2007
			US 2011142987 A1 16-06-2011
			US 2014206036 A1 24-07-2014
			WO 2005027953 A2 31-03-2005
WO 2009144269	A1	03-12-2009	AR 071969 A1 28-07-2010
			AU 2009253110 A1 03-12-2009
			CA 2724459 A1 03-12-2009
			CN 102046784 A 04-05-2011
			EA 201001881 A1 30-06-2011
			EP 2283129 A1 16-02-2011
			JP 2011525349 A 22-09-2011
			TW 201000634 A 01-01-2010
			US 2011165306 A1 07-07-2011
			WO 2009144269 A1 03-12-2009