



- (51) **International Patent Classification:**
C12N 9/24 (2006.01)
- (21) **International Application Number:**
PCT/US2016/037224
- (22) **International Filing Date:**
13 June 2016 (13.06.2016)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
62/181,538 18 June 2015 (18.06.2015) US
- (71) **Applicant:** NOVOZYMES A/S [DK/DK]; Krogshoejvej 36, DK-2880 Bagsvaerd (DK).
- (72) **Inventor; and**
- (71) **Applicant (for BW only):** KANG, Zhengfang [CN/US]; 77 Perry Chapel Church Rd., P.O. Box 576, Franklinton, North Carolina 27525 (US).
- (72) **Inventors:** SCHNORR, Kirk Matthew; Krogshoejvej 36, DK-2880 Bagsvaerd (DK). SKOV, Lars Kobberø; Krogshoejvej 36, DK-2880 Bagsvaerd (DK).
- (74) **Agents:** KRENICKY, Michael W. et al.; Novozymes North America, Inc., 60 E. 42nd St., Suite 700, New York, New York 10165 (US).
- (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).

Declarations under Rule 4.17:

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** POLYPEPTIDES HAVING TREHALASE ACTIVITY AND THE USE THEREOF IN PROCESS OF PRODUCING FERMENTATION PRODUCTS

(57) **Abstract:** The present invention relates to polypeptides having trehalase activity. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the polynucleotides as well as methods of producing and using a trehalase of the invention, in particular a process of producing a fermentation product, such as ethanol.



POLYPEPTIDES HAVING TREHALASE ACTIVITY AND THE USE THEREOF IN PROCESS OF PRODUCING FERMENTATION PRODUCTS

Reference to a Sequence Listing

5 This application contains a Sequence Listing in computer readable form, which is incorporated herein by reference.

Background of the Invention

Field of the Invention

10 The present invention relates to polypeptides having trehalase activity and polynucleotides encoding the polypeptides. The invention also relates to nucleic acid constructs, vectors, and host cells comprising the polynucleotides as well as methods of producing the polypeptides. The invention also relates to processes of producing fermentation products using a trehalase of the invention.

Description of the Related Art

15 Trehalose is a stable disaccharide sugar consisting of two sugar monomers (glucose). Trehalose is accumulated in yeast as a response to stress in up to 10-15% of cell dry weight (GrBa et al. (1975) Eur. J. Appl. Microbiol. 2:29-37). Trehalose cannot be metabolized by the yeast. The enzyme trehalase cleaves trehalose into two glucose units.

20 Trehalases are classified in EC 3.2.1.28 (alpha,alpha-trehalase) and EC. 3.2.1.93 (alpha,alpha-phosphotrehalase). The EC classes are based on recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). Description of EC classes can be found on the internet, e.g., on "<http://www.expasy.org/enzyme/>". The two enzyme classes are both referred to as "trehalases". Examples of neutral trehalases include trehalases from *Saccharomyces cerevisiae* (Londesborouh et al. (1984) Characterization of two trehalases from baker's yeast" Biochem J 219, 511-518; Mucor roxii (Dewerchin et al (1984), "Trehalase activity and cyclic AMP content during early development of Mucor roxii spores", J. Bacteriol. 158, 575-579); *Phycomyces blakesleeana* (Thevelein et al (1983), "Glucose-induced trehalase activation and trehalose mobilization during early germination of *Phycomyces blakesleeana* spores" J. Gen
25 Microbiol. 129, 719-726); *Fusarium oxysporium* (Amaral et al (1996), "Comparative study of two trehalase activities from *Fusarium oxysporium* var *Linii*" Can. J Microbiol. 41, 1057-1062). Examples of neutral trehalases include, but are not limited to, trehalases from *Saccharomyces cerevisiae* (Parvaeh et al. (1996) Purification and biochemical characterization of the ATH1 gene product, vacuolar acid trehalase from *Saccharomyces cerevisiae*" FEBS Lett. 391, 273-
30

278); *Neorospora crassa* (Hecker et al (1973), "Location of trehalase in the ascospores of *Neorospora*: Relation to ascospore dormancy and germination". J. Bacteriol. 115, 592-599); *Chaetomium aureum* (Sumida et al. (1989), "Purification and some properties of trehalase from *Chaetomium aureum* MS-27. J. Ferment. Bioeng. 67, 83-86); *Aspergillus nidulans* (d'Enfert et al. (1997), "Molecular characterization of the *Aspergillus nidulans* *treA* gene encoding an acid trehalase required for growth on trehalose. Mol. Microbiol. 24, 203-216); *Humicola grisea* (Zimmermann et al. (1990) ." Purification and properties of an extracellular conidial trehalase from *Humicola grisea* var. *thermoidea*", Biochim. Acta 1036, 41-46); *Humicola grisea* (Cardello et al. (1994), "A cytosolic trehalase from the thermophilic fungus *Humicola grisea* var. *thermoidea*", Microbiology UK 140, 1671-1677; *Scytalidium thermophilum* (Kadowaki et al. (1996), "Characterization of the trehalose system from the thermophilic fungus *Scytalidium thermophilum*" Biochim. Biophys. Acta 1291, 199-205); and *Fusarium oxysporum* (Amaral et al (1996), "Comparative study of two trehalase activities from *Fusarium oxysporum* var *Linii*" Can. J Microbiol. 41, 1057-1062).

A trehalase is also known from soybean (Aeschbacher et al (1999) "Purification of the trehalase GmTRE1 from soybean nodules and cloning of its cDNA", Plant Physiol 119, 489-496).

Trehalases are also present in small intestine and kidney of mammals.

WO 2009/121058 (Novozymes) concerns a method of fermenting sugars derived from plant material into a fermentation product, such as ethanol, using a fermenting organism by adding one or more trehalase into in the fermentation medium.

WO 2012/027374 (Dyadic) discloses a trehalase from *Myceliophthora thermophila* which can be used in an enzyme mixture for degrading lignocellulosic biomass to fermentable sugars.

WO 2013/148993 (Novozymes) discloses a process of producing a fermentation product, such as ethanol, from starch-containing material by liquefying, saccharifying and fermenting the starch-containing material wherein wherein a carbohydrate-source generating enzyme, a cellulolytic composition and a trehalase is present in fermentation. A trehalase from *Trichoderma reesei* is disclosed.

WO 2015/065978 (Danisco US Inc.) discloses a method of increasing the production of ethanol from a liquefact in a fermentation reaction including fermenting the liquefact with a glucoamylase, a fermenting organism and a trehalase and recovering the ethanol and other fermentation products at the end of the fermentation.

There is still a need for providing enzymes or enzyme composition suitable for use in processes for producing fermentation products, such as ethanol, in increased yields.

Summary of the Invention

The present invention relates to polypeptides having trehalase activity and polynucleotides encoding the polypeptides. The invention also relates to processes of producing fermentation products using a trehalase of the invention. The trehalases concerned are E.C. 3.2.1.28 and further belong to Family 37 Glucoside Hydrolases ("GH37") as defined by CAZY (see www.cazy.org).

The inventors have found, purified and characterized two trehalases (shown in SEQ ID NO: 30 herein and 4 herein, respectively) having high thermostability and a broad pH range. It was also found that an increased ethanol yield can be obtained when adding a trehalase of the invention in fermentation in a process of the invention.

A trehalase of the invention can be used in any yeast fermentation product production process, in particular ethanol production process. The trehalase can be used as an exogenous enzyme or can be expressed in a fermentation product producing organism, in particular a yeast strain producing ethanol, especially a strain of *Sacchariomyces*, in particular a *Saccharomyces cerevisiae* strain.

Accordingly, in the first aspect the present invention relates to polypeptides having trehalase activity selected from the group consisting of:

(a) a polypeptide having at least 90% sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 or at least 80% sequence identity to the mature polypeptide of SEQ ID NO: 4;

(b) a polypeptide encoded by a polynucleotide that hybridizes under very high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 29, SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii);

(c) a polypeptide encoded by a polynucleotide having at least 90% sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 29 or SEQ ID NO: 1 or at least 80% sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 3 or the cDNA sequence thereof;

(d) a variant of the mature polypeptide of SEQ ID NO: 30, SEQ ID NO: 2 or SEQ ID NO: 4 comprising a substitution, deletion, and/or insertion at one or more (e.g., several) positions; and

(e) a fragment of the polypeptide of (a), (b), (c), or (d) that has trehalase activity.

The mature polypeptide of SEQ ID NO: 30 is shown as amino acids 21-694. The mature polypeptide of SEQ ID NO: 2 is shown as amino acids 21-697. The mature polypeptide of SEQ ID NO: 4 is shown as amino acids 21-690. The signal of SEQ ID NO: 30 is shown as amino acids 1-20. The signal of SEQ ID NO: 2 is shown as amino acids 1-20. The signal of SEQ ID NO: 4 is shown as amino acids 21-690.

The present invention also relates to polynucleotides encoding the polypeptides having trehalase activity of the present invention; nucleic acid constructs; recombinant expression vectors; recombinant host cells comprising the polynucleotides; and methods of producing the polypeptides.

5 In an aspect the present invention also relates to processes of producing a fermentation product, in particular ethanol, comprising

- (a) liquefying a starch-containing material with an alpha-amylase;
optionally pre-saccharifying the liquefied material before step (b);
 - (b) saccharifying the liquefied material;
 - 10 (c) fermenting using a fermentation organism;
- wherein
- i) a glucoamylase;
 - ii) a trehalase of the invention;
 - iii) optionally a cellulolytic enzyme composition and/or a protease;

15 are present and/or added during

- saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally the presaccharification step before step (b).

20 Liquefaction in step (a) may be carried out at a temperature above the initial gelatinization temperature, in particular between 80°C-90°C.

In an embodiment the invention relates to processes of producing fermentation products from starch-containing material comprising:

25 (i) saccharifying a starch-containing material at a temperature below the initial gelatinization temperature; and

(ii) fermenting using a fermentation organism;

wherein saccharification and/or fermentation is done in the presence of the following enzymes: glucoamylase, alpha-amylase, trehalase of the invention, and optionally a cellulolytic enzyme composition and/or a protease.

30 In an embodiment the invention relates to processes of producing a fermentation product from pretreated cellulosic material, comprising:

(a) hydrolyzing said pretreated cellulosic material with a cellulolytic enzyme composition;

(b) fermenting using a fermenting organism; and

35 (c) optionally recovering the fermentation product;

wherein a trehalase of the invention is added and/or present in hydrolysis step (a) and/or fermentation step (b).

In an embodiment the trehalase may be added to the thin stillage, i.e., at the backend of a starch based or a cellulosic material based process.

A trehalase of the invention may be present and/or added in saccharification/hydrolysis and/or fermentation in a fermentation product producing process of the invention in any suitable amount. In a preferred embodiment the trehalase is present and/or added in an amount between 0.01-20 ug EP trehalase/g DS, such as between 0.05-15 ug EP terhalase/g DS, such as between 0.5 and 10 ug EP trehalase/g DS.

Brief Description of the Figures

Figure 1 shows the glucose production by trehalase addition in supernatant of ferm drop.

Figure 2 shows the glucose production by Ms trehalase (SEQ ID NO: 30) and GH65 Tr trehalase (SEQ ID NO: 31) addition in supernatant of ferm drop.

Definitions

Trehalase: The term “trehalase” means an enzyme which degrades trehalose into its unit monosaccharides (i.e., glucose). Trehalases are classified in EC 3.2.1.28 (alpha,alpha-trehalase) and EC. 3.2.1.93 (alpha,alpha-phosphotrehalase). The EC classes are based on recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB). Description of EC classes can be found on the internet, e.g., on “<http://www.expasy.org/enzyme/>”. Trehalases are enzymes that catalyze the following reactions:

EC 3.2.1.28:

Alpha,alpha-trehalose + H₂O ⇌ 2 D-glucose;

EC 3.2.1. 93:

Alpha,alpha-trehalose 6-phosphate + H₂O ⇌ D-glucose + D-glucose 6-phosphate.

For purposes of the present invention, trehalase activity may be determined according to “Trehalase Assay” procedure described in the “Materials & Methods”-section. In one aspect, the polypeptides of the present invention have at least 20%, e.g., at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 100% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30, SEQ ID NO: 2 or SEQ ID NO: 4, respectively. In a preferred embodiment a trehalase of the invention is a Family 37 Glycoside Hydrolase (“GH37 trehalase”).

Allelic variant: The term “allelic variant” means any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid

sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

Catalytic domain: The term “catalytic domain” means the region of an enzyme containing the catalytic machinery of the enzyme.

5 **cDNA:** The term "cDNA" means a DNA molecule that can be prepared by reverse transcription from a mature, spliced, mRNA molecule obtained from a eukaryotic or prokaryotic cell. cDNA lacks intron sequences that may be present in the corresponding genomic DNA. The initial, primary RNA transcript is a precursor to mRNA that is processed through a series of steps, including splicing, before appearing as mature spliced mRNA.

10 **Coding sequence:** The term “coding sequence” means a polynucleotide, which directly specifies the amino acid sequence of a polypeptide. The boundaries of the coding sequence are generally determined by an open reading frame, which begins with a start codon such as ATG, GTG, or TTG and ends with a stop codon such as TAA, TAG, or TGA. The coding sequence may be a genomic DNA, cDNA, synthetic DNA, or a combination thereof.

15 **Control sequences:** The term “control sequences” means nucleic acid sequences necessary for expression of a polynucleotide encoding a mature polypeptide of the present invention. Each control sequence may be native (*i.e.*, from the same gene) or foreign (*i.e.*, from a different gene) to the polynucleotide encoding the polypeptide or native or foreign to each other. Such control sequences include, but are not limited to, a leader, polyadenylation
20 sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences 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 polynucleotide encoding a polypeptide.

25 **Expression:** The term “expression” includes any step involved in the production of a polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

30 **Expression vector:** The term “expression vector” means a linear or circular DNA molecule that comprises a polynucleotide encoding a polypeptide and is operably linked to control sequences that provide for its expression.

35 **Fragment:** The term “fragment” means a polypeptide having one or more (*e.g.*, several) amino acids absent from the amino and/or carboxyl terminus of a mature polypeptide or domain; wherein the fragment has trehalase activity. In one aspect, a fragment contains at least 592 amino acid residues , at least 627 amino acid residues , or at least 662 amino acid residues of SEQ ID NO: 2. In another aspect, a fragment contains at least 587 amino acid residues, at least 621 amino acid residues, or at least 655 amino acid residues of SEQ ID NO: 4.

Host cell: The term "host cell" means any cell type that is susceptible to transformation, transfection, transduction, or the like with a nucleic acid construct or expression vector comprising a polynucleotide of the present invention. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication.

Isolated: The term "isolated" means a substance in a form or environment that does not occur in nature. Non-limiting examples of isolated substances include (1) any non-naturally occurring substance, (2) any substance including, but not limited to, any enzyme, variant, nucleic acid, protein, peptide or cofactor, that is at least partially removed from one or more or all of the naturally occurring constituents with which it is associated in nature; (3) any substance modified by the hand of man relative to that substance found in nature; or (4) any substance modified by increasing the amount of the substance relative to other components with which it is naturally associated (e.g., recombinant production in a host cell; multiple copies of a gene encoding the substance; and use of a stronger promoter than the promoter naturally associated with the gene encoding the substance). An isolated substance may be present in a fermentation broth sample; e.g. a host cell may be genetically modified to express the polypeptide of the invention. The fermentation broth from that host cell will comprise the isolated polypeptide.

Mature polypeptide: The term "mature polypeptide" means a polypeptide in its final form following translation and any post-translational modifications, such as N-terminal processing, C-terminal truncation, glycosylation, phosphorylation, etc. In one aspect, the mature polypeptide is amino acids 21 to 694 of SEQ ID NO: 30. Amino acids 1 to 20 of SEQ ID NO: 30 are the signal peptide. In one aspect, the mature polypeptide is amino acids 21 to 697 of SEQ ID NO: 2. Amino acids 1 to 20 of SEQ ID NO: 2 are a signal peptide. In another aspect, the mature polypeptide is amino acids 21 to 690 of SEQ ID NO: 4. Amino acids 1 to 20 of SEQ ID NO: 4 are a signal peptide. The signal peptides are determined using the SignalP program (Nielsen et al., 1997, Protein Engineering 10: 1-6). It is known in the art that a host cell may produce a mixture of two or more different mature polypeptides (*i.e.*, with a different C-terminal and/or N-terminal amino acid) expressed by the same polynucleotide. It is also known in the art that different host cells process polypeptides differently, and thus, one host cell expressing a polynucleotide may produce a different mature polypeptide (e.g., having a different C-terminal and/or N-terminal amino acid) as compared to another host cell expressing the same polynucleotide.

Mature polypeptide coding sequence: The term "mature polypeptide coding sequence" means a polynucleotide that encodes a mature polypeptide having trehalase activity.

In one aspect, the mature polypeptide coding sequence is nucleotides (101)...(2326) of

SEQ ID NO: 29 or the cDNA sequence thereof shown as nucleotides 501-679, 366-1025, 1084-2326.

In one aspect, the mature polypeptide coding sequence is nucleotides (501)...(2814) of SEQ ID NO: 1 or the cDNA sequence thereof shown as nucleotides 501-679, 766-1425, 1484-2676 and 2756-2814. Nucleotides 501 to 560 of SEQ ID NO: 1 encode a signal peptide. In another aspect, the mature polypeptide coding sequence is nucleotides 501-2701 of SEQ ID NO: 3 or the cDNA sequence thereof shown as nucleotides 501-679, 752-1411, and 1471-2701. Nucleotides 501 to 560 of SEQ ID NO: 3 encode a signal peptide.

Nucleic acid construct: The term "nucleic acid construct" means a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature or which is synthetic, which comprises one or more control sequences.

Operably linked: The term "operably linked" means a configuration in which a control sequence is placed at an appropriate position relative to the coding sequence of a polynucleotide such that the control sequence directs expression of the coding sequence.

Sequence identity: The relatedness between two amino acid sequences or between two nucleotide sequences is described by the parameter "sequence identity".

For purposes of the present invention, the sequence identity between two amino acid sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *Trends Genet.* 16: 276-277), preferably version 5.0.0 or later. The parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled "longest identity" (obtained using the -nobrief option) is used as the percent identity and is calculated as follows:

$$(\text{Identical Residues} \times 100) / (\text{Length of Alignment} - \text{Total Number of Gaps in Alignment})$$

Stringency conditions: The term "high stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 50% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS at 65°C.

The term "very high stringency conditions" means for probes of at least 100 nucleotides in length, prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 micrograms/ml sheared and denatured salmon sperm DNA, and 50% formamide, following standard Southern blotting procedures for 12 to 24 hours. The carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS at 70°C.]

Subsequence: The term “subsequence” means a polynucleotide having one or more (e.g., several) nucleotides absent from the 5' and/or 3' end of a mature polypeptide coding sequence; wherein the subsequence encodes a fragment having trehalase activity. In one aspect, a subsequence contains at least 1776 nucleotides, at least 1881 nucleotides, or at least 1962 nucleotides of SEQ ID NO: 29 or SEQ ID NO: 1. In one aspect, a subsequence contains at least 1761 nucleotides, at least 1863 nucleotides, or at least 1965 nucleotides of SEQ ID NO: 3.

Variant: The term “variant” means a polypeptide having trehalase activity comprising an alteration, *i.e.*, a substitution, insertion, and/or deletion, at one or more (e.g., several) positions. A substitution means replacement of the amino acid occupying a position with a different amino acid; a deletion means removal of the amino acid occupying a position; and an insertion means adding an amino acid adjacent to and immediately following the amino acid occupying a position.

Detailed Description of the Invention

Polypeptides Having Trehalase Activity

In an embodiment, the present invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, which have trehalase activity. In one aspect, the polypeptides differ by up to 10 amino acids, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, from the mature polypeptide of SEQ ID NO: 30.

In an embodiment, the present invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 2 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, which have trehalase activity. In one aspect, the polypeptides differ by up to 10 amino acids, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, from the mature polypeptide of SEQ ID NO: 2.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 70% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 2 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%,

at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 70% of the trehalase activity of the mature polypeptide of SEQ ID NO: 2.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 2 of at least 80%, at least 85%, at least 90%,
5 at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least 75% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 of at least 80%, at least
10 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 80% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 of at least 80%, at least
15 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 85% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 of at least 80%, at least
20 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 90% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.
25

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 of at least 80%, at least
30 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least 95% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 of at least 80%, at least
35 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least 100% of the trehalase activity of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.

In an embodiment, the present invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, which have trehalase activity. In one aspect, the polypeptides differ by up to 10 amino acids, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, from the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 70% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 75% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 80% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 85% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least at least 90% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least 95% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In a particular embodiment the invention relates to polypeptides having a sequence identity to the mature polypeptide of SEQ ID NO: 4 of at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%, and wherein the polypeptide has at least 100% of the trehalase activity of the mature polypeptide of SEQ ID NO: 4.

In an embodiment, the polypeptide has been isolated. A polypeptide of the present invention preferably comprises or consists of the amino acid sequence of SEQ ID NO: 30 or SEQ ID NO: 2 or an allelic variant thereof; or is a fragment thereof having trehalase activity. In another aspect, the polypeptide comprises or consists of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2. In another aspect, the polypeptide comprises or consists of amino acids 21 to 694 of SEQ ID NO: 30 or amino acids 21 to 697 of SEQ ID NO: 2.

In an embodiment, the polypeptide has been isolated. A polypeptide of the present invention preferably comprises or consists of the amino acid sequence of SEQ ID NO: 4 or an allelic variant thereof; or is a fragment thereof having trehalase activity. In another aspect, the polypeptide comprises or consists of the mature polypeptide of SEQ ID NO: 4. In another aspect, the polypeptide comprises or consists of amino acids 21 to 690 of SEQ ID NO: 4.

In another embodiment, the present invention relates to a polypeptide having trehalase activity encoded by a polynucleotide that hybridizes under high stringency conditions or very high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 29 or SEQ ID NO: 1, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii) (Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York). In an embodiment, the polypeptide has been isolated.

In another embodiment, the present invention relates to a polypeptide having trehalase activity encoded by a polynucleotide that hybridizes under high stringency conditions or very high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 3, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii) (Sambrook *et al.*, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York). In an embodiment, the polypeptide has been isolated.

The polynucleotide of SEQ ID NO: 29 or SEQ ID NO: 1 or a subsequence thereof, as well as the polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 or a fragment thereof may be used to design nucleic acid probes to identify and clone DNA encoding polypeptides having trehalase activity from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic DNA or cDNA of a cell of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 15, e.g., at least 25, at least 35, or at least 70 nucleotides in length. Preferably, the nucleic acid probe is at least 100 nucleotides in length, e.g., at least 200

nucleotides, at least 300 nucleotides, at least 400 nucleotides, at least 500 nucleotides, at least 600 nucleotides, at least 700 nucleotides, at least 800 nucleotides, or at least 900 nucleotides in length. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with ^{32}P , ^3H , ^{35}S , biotin, or avidin). Such probes are encompassed by the present invention.

The polynucleotide of SEQ ID NO: 3 or a subsequence thereof, as well as the polypeptide of SEQ ID NO: 4 or a fragment thereof may be used to design nucleic acid probes to identify and clone DNA encoding polypeptides having trehalase activity from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic DNA or cDNA of a cell of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 15, e.g., at least 25, at least 35, or at least 70 nucleotides in length. Preferably, the nucleic acid probe is at least 100 nucleotides in length, e.g., at least 200 nucleotides, at least 300 nucleotides, at least 400 nucleotides, at least 500 nucleotides, at least 600 nucleotides, at least 700 nucleotides, at least 800 nucleotides, or at least 900 nucleotides in length. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with ^{32}P , ^3H , ^{35}S , biotin, or avidin). Such probes are encompassed by the present invention.

A genomic DNA or cDNA library prepared from such other strains may be screened for DNA that hybridizes with the probes described above and encodes a polypeptide having trehalase activity. Genomic or other DNA from such other strains may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA that hybridizes with SEQ ID NO: 1 or a subsequence thereof or SEQ ID NO: 3 or a subsequence thereof, the carrier material is used in a Southern blot.

For purposes of the present invention, hybridization indicates that the polynucleotide hybridizes to a labeled nucleic acid probe corresponding to (i) SEQ ID NO: 29 or SEQ ID NO: 1; (ii) the mature polypeptide coding sequence of SEQ ID NO: 29 or SEQ ID NO: 1; (iii) the cDNA sequence thereof; (iv) the full-length complement thereof; or (v) a subsequence thereof; under high to very high stringency conditions.

For purposes of the present invention, hybridization indicates that the polynucleotide hybridizes to a labeled nucleic acid probe corresponding to (i) SEQ ID NO: 3; (ii) the mature polypeptide coding sequence of SEQ ID NO: 3; (iii) the cDNA sequence thereof; (iv) the full-length complement thereof; or (v) a subsequence thereof; under high to very high stringency conditions.

Molecules to which the nucleic acid probe hybridizes under these conditions can be detected using, for example, X-ray film or any other detection means known in the art.

In another embodiment, the present invention relates to a polypeptide having trehalase activity encoded by a polynucleotide having a sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 29 or the cDNA sequence thereof of at least 60%, *e.g.*, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%. In a further embodiment, the polypeptide has been isolated.

In another embodiment, the present invention relates to a polypeptide having trehalase activity encoded by a polynucleotide having a sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or the cDNA sequence thereof of at least 60%, *e.g.*, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%. In a further embodiment, the polypeptide has been isolated.

In another embodiment, the present invention relates to a polypeptide having trehalase activity encoded by a polynucleotide having a sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 3 or the cDNA sequence thereof of at least 60%, *e.g.*, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%. In a further embodiment, the polypeptide has been isolated.

In another embodiment, the present invention relates to variants of the mature polypeptide of SEQ ID NO: 30 comprising a substitution, deletion, and/or insertion at one or more (*e.g.*, several) positions. In an embodiment, the number of amino acid substitutions, deletions and/or insertions introduced into the mature polypeptide of SEQ ID NO: 30 is up to 10, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

In another embodiment, the present invention relates to variants of the mature polypeptide of SEQ ID NO: 2 comprising a substitution, deletion, and/or insertion at one or more (*e.g.*, several) positions. In an embodiment, the number of amino acid substitutions, deletions and/or insertions introduced into the mature polypeptide of SEQ ID NO: 2 is up to 10, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

In another embodiment, the present invention relates to variants of the mature polypeptide of SEQ ID NO: 4 comprising a substitution, deletion, and/or insertion at one or more (*e.g.*, several) positions. In an embodiment, the number of amino acid substitutions, deletions and/or insertions introduced into the mature polypeptide of SEQ ID NO: 4 is up to 10, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

The amino acid changes may be of a minor nature, that is conservative amino acid substitutions or insertions that do not significantly affect the folding and/or activity of the protein;

small deletions, typically of 1-30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to 20-25 residues; or a small extension that facilitates purification by changing net charge or another function, such as a poly-histidine tract, an antigenic epitope or a binding domain.

5 Examples of conservative substitutions are within the groups of basic amino acids (arginine, lysine and histidine), acidic amino acids (glutamic acid and aspartic acid), polar amino acids (glutamine and asparagine), hydrophobic amino acids (leucine, isoleucine and valine), aromatic amino acids (phenylalanine, tryptophan and tyrosine), and small amino acids (glycine, alanine, serine, threonine and methionine). Amino acid substitutions that do not
10 generally alter specific activity are known in the art and are described, for example, by H. Neurath and R.L. Hill, 1979, *In, The Proteins*, Academic Press, New York. Common substitutions are Ala/Ser, Val/Ile, Asp/Glu, Thr/Ser, Ala/Gly, Ala/Thr, Ser/Asn, Ala/Val, Ser/Gly, Tyr/Phe, Ala/Pro, Lys/Arg, Asp/Asn, Leu/Ile, Leu/Val, Ala/Glu, and Asp/Gly.

 Alternatively, the amino acid changes are of such a nature that the physico-chemical
15 properties of the polypeptides are altered. For example, amino acid changes may improve the thermal stability of the polypeptide, alter the substrate specificity, change the pH optimum, and the like.

 Essential amino acids in a polypeptide can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham
20 and Wells, 1989, *Science* 244: 1081-1085). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant molecules are tested for trehalase activity to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton *et al.*, 1996, *J. Biol. Chem.* 271: 4699-4708. The active site of the enzyme or other biological interaction can also be determined by physical analysis of structure, as
25 determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction, or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos *et al.*, 1992, *Science* 255: 306-312; Smith *et al.*, 1992, *J. Mol. Biol.* 224: 899-904; Wlodaver *et al.*, 1992, *FEBS Lett.* 309: 59-64. The identity of essential amino acids can also be inferred from an alignment with a related polypeptide.

30 Single or multiple amino acid substitutions, deletions, and/or insertions can be made and tested using known methods of mutagenesis, recombination, and/or shuffling, followed by a relevant screening procedure, such as those disclosed by Reidhaar-Olson and Sauer, 1988, *Science* 241: 53-57; Bowie and Sauer, 1989, *Proc. Natl. Acad. Sci. USA* 86: 2152-2156; WO 95/17413; or WO 95/22625. Other methods that can be used include error-prone PCR,
35 phage display (e.g., Lowman *et al.*, 1991, *Biochemistry* 30: 10832-10837; U.S. Patent No. 5,223,409; WO 92/06204), and region-directed mutagenesis (Derbyshire *et al.*, 1986, *Gene* 46: 145; Ner *et al.*, 1988, *DNA* 7: 127).

Mutagenesis/shuffling methods can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides expressed by host cells (Ness *et al.*, 1999, *Nature Biotechnology* 17: 893-896). Mutagenized DNA molecules that encode active polypeptides can be recovered from the host cells and rapidly sequenced using standard methods in the art. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide.

The polypeptide may be a hybrid polypeptide in which a region of one polypeptide is fused at the N-terminus or the C-terminus of a region of another polypeptide.

The polypeptide may be a fusion polypeptide or cleavable fusion polypeptide in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide of the present invention. A fusion polypeptide is produced by fusing a polynucleotide encoding another polypeptide to a polynucleotide of the present invention. 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 and that expression of the fusion polypeptide is under control of the same promoter(s) and terminator. Fusion polypeptides may also be constructed using intein technology in which fusion polypeptides are created post-translationally (Cooper *et al.*, 1993, *EMBO J.* 12: 2575-2583; Dawson *et al.*, 1994, *Science* 266: 776-779).

A fusion polypeptide can further comprise a cleavage site between the two polypeptides. Upon secretion of the fusion protein, the site is cleaved releasing the two polypeptides. Examples of cleavage sites include, but are not limited to, the sites disclosed in Martin *et al.*, 2003, *J. Ind. Microbiol. Biotechnol.* 3: 568-576; Svetina *et al.*, 2000, *J. Biotechnol.* 76: 245-251; Rasmussen-Wilson *et al.*, 1997, *Appl. Environ. Microbiol.* 63: 3488-3493; Ward *et al.*, 1995, *Biotechnology* 13: 498-503; and Contreras *et al.*, 1991, *Biotechnology* 9: 378-381; Eaton *et al.*, 1986, *Biochemistry* 25: 505-512; Collins-Racie *et al.*, 1995, *Biotechnology* 13: 982-987; Carter *et al.*, 1989, *Proteins: Structure, Function, and Genetics* 6: 240-248; and Stevens, 2003, *Drug Discovery World* 4: 35-48.

Sources of Polypeptides Having Trehalase Activity

A polypeptide having trehalase activity of the present invention may be obtained from microorganisms of any genus. For purposes of the present invention, the term "obtained from" as used herein in connection with a given source shall mean that the polypeptide encoded by a polynucleotide is produced by the source or by a strain in which the polynucleotide from the source has been inserted. In one aspect, the polypeptide obtained from a given source is secreted extracellularly.

In another aspect, the polypeptide is a *Myceliophthora sepedonium* polypeptide having trehalase or a *Chaetomium virescens* polypeptide having trehalase activity.

It will be understood that for the aforementioned species, the invention encompasses both the perfect and imperfect states, and other taxonomic equivalents, e.g., anamorphs, regardless of the species name by which they are known. Those skilled in the art will readily recognize the identity of appropriate equivalents.

Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

The polypeptide may be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) or DNA samples obtained directly from natural materials (e.g., soil, composts, water, etc.) using the above-mentioned probes. Techniques for isolating microorganisms and DNA directly from natural habitats are well known in the art. A polynucleotide encoding the polypeptide may then be obtained by similarly screening a genomic DNA or cDNA library of another microorganism or mixed DNA sample. Once a polynucleotide encoding a polypeptide has been detected with the probe(s), the polynucleotide can be isolated or cloned by utilizing techniques that are known to those of ordinary skill in the art (see, e.g., Sambrook *et al.*, 1989, *supra*).

Polynucleotides

The present invention also relates to polynucleotides encoding a polypeptide of the present invention, as described herein. In an embodiment, the polynucleotide encoding the polypeptide of the present invention has been isolated.

The techniques used to isolate or clone a polynucleotide are known in the art and include isolation from genomic DNA or cDNA, or a combination thereof. The cloning of the polynucleotides from genomic DNA can be effected, e.g., by using the well-known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis *et al.*, 1990, *PCR: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligation activated transcription (LAT) and polynucleotide-based amplification (NASBA) may be used. The polynucleotides may be cloned from a strain of *Myceliophthora*, or a related organism or *Chaetomium*, or a related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the polynucleotide.

Nucleic Acid Constructs

The present invention also relates to nucleic acid constructs comprising a polynucleotide of the present invention operably linked to one or more control sequences that direct the

expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences.

The polynucleotide may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides utilizing recombinant DNA methods are well known in the art.

The control sequence may be a promoter, a polynucleotide that is recognized by a host cell for expression of a polynucleotide encoding a polypeptide of the present invention. The promoter contains transcriptional control sequences that mediate the expression of the polypeptide. The promoter may be any polynucleotide that shows transcriptional activity in the host cell including variant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

Examples of suitable promoters for directing transcription of the nucleic acid constructs of the present invention in a bacterial host cell are the promoters obtained from the *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus licheniformis* penicillinase gene (*penP*), *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), *Bacillus subtilis* levansucrase gene (*sacB*), *Bacillus subtilis* *xylA* and *xylB* genes, *Bacillus thuringiensis* *cryIIIA* gene (Agaisse and Lereclus, 1994, *Molecular Microbiology* 13: 97-107), *E. coli* *lac* operon, *E. coli* *trc* promoter (Egon *et al.*, 1988, *Gene* 69: 301-315), *Streptomyces coelicolor* agarase gene (*dagA*), and prokaryotic beta-lactamase gene (Villa-Kamaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. USA* 75: 3727-3731), as well as the *tac* promoter (DeBoer *et al.*, 1983, *Proc. Natl. Acad. Sci. USA* 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in Gilbert *et al.*, 1980, *Scientific American* 242: 74-94; and in Sambrook *et al.*, 1989, *supra*. Examples of tandem promoters are disclosed in WO 99/43835.

Examples of suitable promoters for directing transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes for *Aspergillus nidulans* acetamidase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Aspergillus oryzae* TAKA amylase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Fusarium oxysporum* trypsin-like protease (WO 96/00787), *Fusarium venenatum* amyloglucosidase (WO 00/56900), *Fusarium venenatum* Daria (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), *Rhizomucor miehei* lipase, *Rhizomucor miehei* aspartic proteinase, *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 V, *Trichoderma reesei* xylanase I, *Trichoderma reesei* xylanase II, *Trichoderma reesei* xylanase III, *Trichoderma reesei* beta-xylosidase, and *Trichoderma reesei* translation elongation factor, as well as the NA2-tpi promoter (a modified promoter from an *Aspergillus* neutral alpha-amylase gene in which the untranslated leader has been replaced by an untranslated leader from an *Aspergillus* triose phosphate isomerase gene; non-limiting examples include modified promoters from an *Aspergillus niger* neutral alpha-amylase gene in which the untranslated leader has been replaced by an untranslated leader from an *Aspergillus nidulans* or *Aspergillus oryzae* triose phosphate isomerase gene); and variant, truncated, and hybrid promoters thereof. Other promoters are described in U.S. Patent No. 6,011,147.

In a yeast host, useful promoters are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH1, ADH2/GAP), *Saccharomyces cerevisiae* triose phosphate isomerase (TPI), *Saccharomyces cerevisiae* metallothionein (CUP1), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are described by Romanos *et al.*, 1992, *Yeast* 8: 423-488.

The control sequence may also be a transcription terminator, which is recognized by a host cell to terminate transcription. The terminator is operably linked to the 3'-terminus of the polynucleotide encoding the polypeptide. Any terminator that is functional in the host cell may be used in the present invention.

Preferred terminators for bacterial host cells are obtained from the genes for *Bacillus clausii* alkaline protease (*aprH*), *Bacillus licheniformis* alpha-amylase (*amyL*), and *Escherichia coli* ribosomal RNA (*rrnB*).

Preferred terminators for filamentous fungal host cells are obtained from the genes for *Aspergillus nidulans* acetamidase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* glucoamylase, *Aspergillus niger* alpha-glucosidase, *Aspergillus oryzae* TAKA amylase, *Fusarium oxysporum* trypsin-like protease, *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 V, *Trichoderma reesei* xylanase I, *Trichoderma reesei* xylanase II, *Trichoderma reesei* xylanase III, *Trichoderma reesei* beta-xylosidase, and *Trichoderma reesei* translation elongation factor.

Preferred terminators for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be an mRNA stabilizer region downstream of a promoter and upstream of the coding sequence of a gene which increases expression of the gene.

Examples of suitable mRNA stabilizer regions are obtained from a *Bacillus thuringiensis* cryIIIA gene (WO 94/25612) and a *Bacillus subtilis* SP82 gene (Hue *et al.*, 1995, *Journal of Bacteriology* 177: 3465-3471).

The control sequence may also be a leader, a nontranslated region of an mRNA that is important for translation by the host cell. The leader is operably linked to the 5'-terminus of the polynucleotide encoding the polypeptide. Any leader that is functional in the host cell may be used.

Preferred leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase.

Suitable leaders for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3'-terminus of the polynucleotide and, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence that is functional in the host cell may be used.

Preferred polyadenylation sequences for filamentous fungal host cells are obtained from the genes for *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* glucoamylase, *Aspergillus niger* alpha-glucosidase *Aspergillus oryzae* TAKA amylase, and *Fusarium oxysporum* trypsin-like protease.

Useful polyadenylation sequences for yeast host cells are described by Guo and Sherman, 1995, *Mol. Cellular Biol.* 15: 5983-5990.

The control sequence may also be a signal peptide coding region that encodes a signal peptide linked to the N-terminus of a polypeptide and directs the polypeptide into the cell's secretory pathway. The 5'-end of the coding sequence of the polynucleotide may inherently contain a signal peptide coding sequence naturally linked in translation reading frame with the segment of the coding sequence that encodes the polypeptide. Alternatively, the 5'-end of the coding sequence may contain a signal peptide coding sequence that is foreign to the coding sequence. A foreign signal peptide coding sequence may be required where the coding sequence does not naturally contain a signal peptide coding sequence. Alternatively, a foreign signal peptide coding sequence may simply replace the natural signal peptide coding sequence in order to enhance secretion of the polypeptide. However, any signal peptide coding sequence that directs the expressed polypeptide into the secretory pathway of a host cell may be used.

Effective signal peptide coding sequences for bacterial host cells are the signal peptide coding sequences obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus stearothermophilus* neutral proteases (*nprT*, *nprS*, *nprM*), and *Bacillus subtilis* *prsA*. Further signal peptides are described by Simonen and Palva, 1993, *Microbiological Reviews* 57: 109-137.

Effective signal peptide coding sequences for filamentous fungal host cells are the signal peptide coding sequences obtained from the genes for *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Aspergillus oryzae* TAKA amylase, *Humicola insolens* cellulase, *Humicola insolens* endoglucanase V, *Humicola lanuginosa* lipase, and *Rhizomucor miehei* aspartic proteinase.

Useful signal peptides for yeast host cells are obtained from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase. Other useful signal peptide coding sequences are described by Romanos *et al.*, 1992, *supra*.

The control sequence may also be a propeptide coding sequence that encodes a propeptide positioned at the N-terminus of a polypeptide. The resultant polypeptide is known as a proenzyme or propolypeptide (or a zymogen in some cases). A propolypeptide is generally inactive and can be converted to an active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding sequence may be obtained from the genes for *Bacillus subtilis* alkaline protease (*aprE*), *Bacillus subtilis* neutral protease (*nprT*), *Myceliophthora thermophila* laccase (WO 95/33836), *Rhizomucor miehei* aspartic proteinase, and *Saccharomyces cerevisiae* alpha-factor.

Where both signal peptide and propeptide sequences are present, the propeptide sequence is positioned next to the N-terminus of a polypeptide and the signal peptide sequence is positioned next to the N-terminus of the propeptide sequence.

It may also be desirable to add regulatory sequences that regulate expression of the polypeptide relative to the growth of the host cell. Examples of regulatory sequences are those that cause expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. Regulatory sequences in prokaryotic systems include the *lac*, *tac*, and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the *Aspergillus niger* glucoamylase promoter, *Aspergillus oryzae* TAKA alpha-amylase promoter, and *Aspergillus oryzae* glucoamylase promoter, *Trichoderma reesei* cellobiohydrolase I promoter, and *Trichoderma reesei* cellobiohydrolase II promoter may be used. Other examples of regulatory sequences are those that allow for gene amplification. In eukaryotic systems, these regulatory sequences include the dihydrofolate reductase gene that is amplified in the presence of methotrexate, and

the metallothionein genes that are amplified with heavy metals. In these cases, the polynucleotide encoding the polypeptide would be operably linked to the regulatory sequence.

Expression Vectors

The present invention also relates to recombinant expression vectors comprising a polynucleotide of the present invention, a promoter, and transcriptional and translational stop signals. The various nucleotide and control sequences may be joined together to produce a recombinant expression vector that may include one or more convenient restriction sites to allow for insertion or substitution of the polynucleotide encoding the polypeptide at such sites. Alternatively, the polynucleotide may be expressed by inserting the polynucleotide or a nucleic acid construct comprising the polynucleotide into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (e.g., a plasmid or virus) that can be conveniently subjected to recombinant DNA procedures and can bring about expression of the polynucleotide. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vector may be a linear or closed circular plasmid.

The vector may be an autonomously replicating vector, i.e., a vector that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g., a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one that, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids that together contain the total DNA to be introduced into the genome of the host cell, or a transposon, may be used.

The vector preferably contains one or more selectable markers that permit easy selection of transformed, transfected, transduced, or the like cells. A selectable marker is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like.

Examples of bacterial selectable markers are *Bacillus licheniformis* or *Bacillus subtilis* *dal* genes, or markers that confer antibiotic resistance such as ampicillin, chloramphenicol, kanamycin, neomycin, spectinomycin, or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to, ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *adeA* (phosphoribosylaminoimidazole-succinocarboxamide synthase), *adeB* (phosphoribosylaminoimidazole synthase), *amdS* (acetamidase), *argB* (ornithine carbamoyltransferase), *bar*

(phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. Preferred for use in an *Aspergillus* cell are *Aspergillus nidulans* or *Aspergillus oryzae* *amdS* and *pyrG* genes and a *Streptomyces hygroscopicus* *bar* gene. Preferred for use in a *Trichoderma* cell are *adeA*, *adeB*, *amdS*, *hph*, and *pyrG* genes.

The selectable marker may be a dual selectable marker system as described in WO 2010/039889. In one aspect, the dual selectable marker is an *hph-tk* dual selectable marker system.

The vector preferably contains an element(s) that permits integration of the vector into the host cell's genome or autonomous replication of the vector in the cell independent of the genome.

For integration into the host cell genome, the vector may rely on the polynucleotide's sequence encoding the polypeptide or any other element of the vector for integration into the genome by homologous or non-homologous recombination. Alternatively, the vector may contain additional polynucleotides for directing integration by homologous recombination into the genome of the host cell at a precise location(s) in the chromosome(s). To increase the likelihood of integration at a precise location, the integrational elements should contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, 400 to 10,000 base pairs, and 800 to 10,000 base pairs, which have a high degree of sequence identity to the corresponding target sequence to enhance the probability of homologous recombination. The integrational elements may be any sequence that is homologous with the target sequence in the genome of the host cell. Furthermore, the integrational elements may be non-encoding or encoding polynucleotides. On the other hand, the vector may be integrated into the genome of the host cell by non-homologous recombination.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the host cell in question. The origin of replication may be any plasmid replicator mediating autonomous replication that functions in a cell. The term "origin of replication" or "plasmid replicator" means a polynucleotide that enables a plasmid or vector to replicate *in vivo*.

Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAM β 1 permitting replication in *Bacillus*.

Examples of origins of replication for use in a yeast host cell are the 2 micron origin of replication, ARS1, ARS4, the combination of ARS1 and CEN3, and the combination of ARS4 and CEN6.

Examples of origins of replication useful in a filamentous fungal cell are AMA1 and ANS1 (Gems *et al.*, 1991, *Gene* 98: 61-67; Cullen *et al.*, 1987, *Nucleic Acids Res.* 15: 9163-9175; WO 00/24883). Isolation of the AMA1 gene and construction of plasmids or vectors comprising the gene can be accomplished according to the methods disclosed in

More than one copy of a polynucleotide of the present invention may be inserted into a host cell to increase production of a polypeptide. An increase in the copy number of the polynucleotide can be obtained by integrating at least one additional copy of the sequence into the host cell genome or by including an amplifiable selectable marker gene with the polynucleotide where cells containing amplified copies of the selectable marker gene, and thereby additional copies of the polynucleotide, can be selected for by cultivating the cells in the presence of the appropriate selectable agent.

The procedures used to ligate the elements described above to construct the recombinant expression vectors of the present invention are well known to one skilled in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

Host Cells

The present invention also relates to recombinant host cells, comprising a polynucleotide of the present invention operably linked to one or more control sequences that direct the production of a polypeptide of the present invention. A construct or vector comprising a polynucleotide is introduced into a host cell so that the construct or vector is maintained as a chromosomal integrant or as a self-replicating extra-chromosomal vector as described earlier. The term "host cell" encompasses any progeny of a parent cell that is not identical to the parent cell due to mutations that occur during replication. The choice of a host cell will to a large extent depend upon the gene encoding the polypeptide and its source.

The host cell may be any cell useful in the recombinant production of a polypeptide of the present invention, *e.g.*, a prokaryote or a eukaryote.

The prokaryotic host cell may be any Gram-positive or Gram-negative bacterium. Gram-positive bacteria include, but are not limited to, *Bacillus*, *Clostridium*, *Enterococcus*, *Geobacillus*, *Lactobacillus*, *Lactococcus*, *Oceanobacillus*, *Staphylococcus*, *Streptococcus*, and *Streptomyces*. Gram-negative bacteria include, but are not limited to, *Campylobacter*, *E. coli*, *Flavobacterium*, *Fusobacterium*, *Helicobacter*, *Ilyobacter*, *Neisseria*, *Pseudomonas*, *Salmonella*, and *Ureaplasma*.

The bacterial host cell may be any *Bacillus* cell including, but not limited to, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*,

Bacillus megaterium, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis* cells.

The bacterial host cell may also be any *Streptococcus* cell including, but not limited to, *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, and *Streptococcus equi* subsp. *Zooepidemicus* cells.

The bacterial host cell may also be any *Streptomyces* cell including, but not limited to, *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, and *Streptomyces lividans* cells.

The introduction of DNA into a *Bacillus* cell may be effected by protoplast transformation (see, e.g., Chang and Cohen, 1979, *Mol. Gen. Genet.* 168: 111-115), competent cell transformation (see, e.g., Young and Spizizen, 1961, *J. Bacteriol.* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *J. Mol. Biol.* 56: 209-221), electroporation (see, e.g., Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or conjugation (see, e.g., Koehler and Thorne, 1987, *J. Bacteriol.* 169: 5271-5278). The introduction of DNA into an *E. coli* cell may be effected by protoplast transformation (see, e.g., Hanahan, 1983, *J. Mol. Biol.* 166: 557-580) or electroporation (see, e.g., Dower et al., 1988, *Nucleic Acids Res.* 16: 6127-6145). The introduction of DNA into a *Streptomyces* cell may be effected by protoplast transformation, electroporation (see, e.g., Gong et al., 2004, *Folia Microbiol. (Praha)* 49: 399-405), conjugation (see, e.g., Mazodier et al., 1989, *J. Bacteriol.* 171: 3583-3585), or transduction (see, e.g., Burke et al., 2001, *Proc. Natl. Acad. Sci. USA* 98: 6289-6294). The introduction of DNA into a *Pseudomonas* cell may be effected by electroporation (see, e.g., Choi et al., 2006, *J. Microbiol. Methods* 64: 391-397) or conjugation (see, e.g., Pinedo and Smets, 2005, *Appl. Environ. Microbiol.* 71: 51-57). The introduction of DNA into a *Streptococcus* cell may be effected by natural competence (see, e.g., Perry and Kuramitsu, 1981, *Infect. Immun.* 32: 1295-1297), protoplast transformation (see, e.g., Catt and Jollick, 1991, *Microbios* 68: 189-207), electroporation (see, e.g., Buckley et al., 1999, *Appl. Environ. Microbiol.* 65: 3800-3804), or conjugation (see, e.g., Clewell, 1981, *Microbiol. Rev.* 45: 409-436). However, any method known in the art for introducing DNA into a host cell can be used.

The host cell may also be a eukaryote, such as a mammalian, insect, plant, or fungal cell.

The host cell may be a fungal cell. "Fungi" as used herein includes the phyla *Ascomycota*, *Basidiomycota*, *Chytridiomycota*, and *Zygomycota* as well as the *Oomycota* and all mitosporic fungi (as defined by Hawksworth et al., In, *Ainsworth and Bisby's Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK).

The fungal host cell may be a yeast cell. "Yeast" as used herein includes ascosporogenous yeast (*Endomycetales*), basidiosporogenous yeast, and yeast belonging to the *Fungi Imperfecti* (*Blastomycetes*). Since the classification of yeast may change in the future,

for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, Passmore, and Davenport, editors, Soc. App. Bacteriol. Symposium Series No. 9, 1980).

The yeast host cell may be a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*,
5 *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* cell, such as a *Kluyveromyces lactis*,
Saccharomyces carlsbergensis, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*,
Saccharomyces douglasii, *Saccharomyces kluyveri*, *Saccharomyces norbensis*,
Saccharomyces oviformis, or *Yarrowia lipolytica* cell.

The fungal host cell may be a filamentous fungal cell. "Filamentous fungi" include all
10 filamentous forms of the subdivision *Eumycota* and *Oomycota* (as defined by Hawksworth et
al., 1995, *supra*). The filamentous fungi are generally characterized by a mycelial wall
composed of chitin, cellulose, glucan, chitosan, mannan, and other complex polysaccharides.
Vegetative growth is by hyphal elongation and carbon catabolism is obligately aerobic. In
contrast, vegetative growth by yeasts such as *Saccharomyces cerevisiae* is by budding of a
15 unicellular thallus and carbon catabolism may be fermentative.

The filamentous fungal host cell may be an *Acremonium*, *Aspergillus*, *Aureobasidium*,
Bjerkandera, *Ceriporiopsis*, *Chrysosporium*, *Coprinus*, *Coriolus*, *Cryptococcus*, *Filibasidium*,
Fusarium, *Humicola*, *Magnaporthe*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*,
Paecilomyces, *Penicillium*, *Phanerochaete*, *Phlebia*, *Piromyces*, *Pleurotus*, *Schizophyllum*,
20 *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trametes*, or *Trichoderma* cell.

For example, the filamentous fungal host cell may be an *Aspergillus awamori*,
Aspergillus foetidus, *Aspergillus fumigatus*, *Aspergillus japonicus*, *Aspergillus nidulans*,
Aspergillus niger, *Aspergillus oryzae*, *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis*
caregiea, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*,
25 *Ceriporiopsis subrufa*, *Ceriporiopsis subvermispora*, *Chrysosporium inops*, *Chrysosporium*
keratinophilum, *Chrysosporium lucknowense*, *Chrysosporium merdarium*, *Chrysosporium*
pannicola, *Chrysosporium queenslandicum*, *Chrysosporium tropicum*, *Chrysosporium zonatum*,
Coprinus cinereus, *Coriolus hirsutus*, *Fusarium bactridioides*, *Fusarium cerealis*, *Fusarium*
crookwellense, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*, *Fusarium*
30 *heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium*
roseum, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium*
sulphureum, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola*
insolens, *Humicola lanuginosa*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*,
Penicillium purpurogenum, *Phanerochaete chrysosporium*, *Phlebia radiata*, *Pleurotus eryngii*,
35 *Thielavia terrestris*, *Trametes villosa*, *Trametes versicolor*, *Trichoderma harzianum*,
Trichoderma koningii, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride*
cell.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* and *Trichoderma* host cells are described in EP 238023, Yelton *et al.*, 1984, *Proc. Natl. Acad. Sci. USA* 81: 1470-1474, and Christensen *et al.*, 1988, *Bio/Technology* 6: 1419-1422. Suitable methods for transforming *Fusarium* species are described by Malardier *et al.*, 1989, *Gene* 78: 147-156, and WO 96/00787. Yeast may be transformed using the procedures described by Becker and Guarente, In Abelson, J.N. and Simon, M.I., editors, *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *J. Bacteriol.* 153: 163; and Hinnen *et al.*, 1978, *Proc. Natl. Acad. Sci. USA* 75: 1920.

Methods of Production

The present invention also relates to methods of producing a polypeptide of the present invention, comprising (a) cultivating a cell, which in its wild-type form produces the polypeptide, under conditions conducive for production of the polypeptide; and optionally, (b) recovering the polypeptide. In one aspect, the cell is a *Myceliophthora* cell. In another aspect, the cell is a *Myceliophthora sepeodonium* cell. In one aspect, the cell is a *Chaetomium* cell. In another aspect, the cell is a *Chaetomium virescens* cell.

The present invention also relates to methods of producing a polypeptide of the present invention, comprising (a) cultivating a recombinant host cell of the present invention under conditions conducive for production of the polypeptide; and optionally, (b) recovering the polypeptide.

The host cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods known in the art. For example, the cells may be cultivated by shake flask cultivation, or small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the polypeptide is not secreted, it can be recovered from cell lysates.

The polypeptide may be detected using methods known in the art that are specific for the polypeptides. These detection methods include, but are not limited to, use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme substrate. For example, an enzyme assay may be used to determine the activity of the polypeptide.

The polypeptide may be recovered using methods known in the art. For example, the polypeptide may be recovered from the nutrient medium by conventional procedures including, but not limited to, collection, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation. In one aspect, a fermentation broth comprising the polypeptide is recovered.

5 The polypeptide may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (e.g., preparative isoelectric focusing), differential solubility (e.g., ammonium sulfate precipitation), SDS-PAGE, or extraction (see, e.g., *Protein Purification*, Janson and Ryden, editors, VCH Publishers, New York, 1989) to obtain
10 substantially pure polypeptides.

In an alternative aspect, the polypeptide is not recovered, but rather a host cell of the present invention expressing the polypeptide is used as a source of the polypeptide.

Production in plants

The present invention also relates to isolated plants, e.g., a transgenic plant, plant part,
15 or plant cell, comprising a polynucleotide of the present invention so as to express and produce a polypeptide or domain in recoverable quantities. The polypeptide or domain may be recovered from the plant or plant part. Alternatively, the plant or plant part containing the polypeptide or domain may be used as such for improving the quality of a food or feed, e.g., improving nutritional value, palatability, and rheological properties, or to destroy an antinutritive
20 factor.

The transgenic plant can be dicotyledonous (a dicot) or monocotyledonous (a monocot). Examples of monocot plants are grasses, such as meadow grass (blue grass, *Poa*), forage grass such as *Festuca*, *Lolium*, temperate grass, such as *Agrostis*, and cereals, e.g., wheat, oats, rye, barley, rice, sorghum, and maize (corn).

25 Examples of dicot plants are tobacco, legumes, such as lupins, potato, sugar beet, pea, bean and soybean, and cruciferous plants (family *Brassicaceae*), such as cauliflower, rape seed, and the closely related model organism *Arabidopsis thaliana*.

Examples of plant parts are stem, callus, leaves, root, fruits, seeds, and tubers as well as the individual tissues comprising these parts, e.g., epidermis, mesophyll, parenchyme,
30 vascular tissues, meristems.

Plant cells and specific plant cell compartments, such as chloroplasts, apoplasts, mitochondria, vacuoles, peroxisomes and cytoplasm are also considered to be a plant part.

Also included within the scope of the present invention are the progeny of such plants, plant parts, and plant cells.

35 The transgenic plant or plant cell expressing the polypeptide or domain may be constructed in accordance with methods known in the art.

The present invention also relates to methods of producing a polypeptide or domain of the present invention comprising (a) cultivating a transgenic plant or a plant cell comprising a polynucleotide encoding the polypeptide or domain under conditions conducive for production of the polypeptide or domain; and (b) recovering the polypeptide or domain.

5 Fermentation Broth Formulations or Cell Compositions

The present invention also relates to a fermentation broth formulation or a cell composition comprising a polypeptide of the present invention. The fermentation broth product further comprises additional ingredients used in the fermentation process, such as, for example, cells (including, the host cells containing the gene encoding the polypeptide of the present
10 invention which are used to produce the polypeptide of interest), cell debris, biomass, fermentation media and/or fermentation products. In some embodiments, the composition is a cell-killed whole broth containing organic acid(s), killed cells and/or cell debris, and culture medium.

The term "fermentation broth" as used herein refers to a preparation produced by
15 cellular fermentation that undergoes no or minimal recovery and/or purification. For example, fermentation broths are produced when microbial cultures are grown to saturation, incubated under carbon-limiting conditions to allow protein synthesis (e.g., expression of enzymes by host cells) and secretion into cell culture medium. The fermentation broth can contain unfractionated or fractionated contents of the fermentation materials derived at the end of the fermentation.
20 Typically, the fermentation broth is unfractionated and comprises the spent culture medium and cell debris present after the microbial cells (e.g., filamentous fungal cells) are removed, e.g., by centrifugation. In some embodiments, the fermentation broth contains spent cell culture medium, extracellular enzymes, and viable and/or nonviable microbial cells.

In an embodiment, the fermentation broth formulation and cell compositions comprise a
25 first organic acid component comprising at least one 1-5 carbon organic acid and/or a salt thereof and a second organic acid component comprising at least one 6 or more carbon organic acid and/or a salt thereof. In a specific embodiment, the first organic acid component is acetic acid, formic acid, propionic acid, a salt thereof, or a mixture of two or more of the foregoing and the second organic acid component is benzoic acid, cyclohexanecarboxylic acid,
30 4-methylvaleric acid, phenylacetic acid, a salt thereof, or a mixture of two or more of the foregoing.

In one aspect, the composition contains an organic acid(s), and optionally further contains killed cells and/or cell debris. In one embodiment, the killed cells and/or cell debris are removed from a cell-killed whole broth to provide a composition that is free of these
35 components.

The fermentation broth formulations or cell compositions may further comprise a preservative and/or anti-microbial (e.g., bacteriostatic) agent, including, but not limited to, sorbitol, sodium chloride, potassium sorbate, and others known in the art.

The cell-killed whole broth or composition may contain the unfractionated contents of the fermentation materials derived at the end of the fermentation. Typically, the cell-killed whole broth or composition contains the spent culture medium and cell debris present after the microbial cells (e.g., filamentous fungal cells) are grown to saturation, incubated under carbon-limiting conditions to allow protein synthesis. In some embodiments, the cell-killed whole broth or composition contains the spent cell culture medium, extracellular enzymes, and killed filamentous fungal cells. In some embodiments, the microbial cells present in the cell-killed whole broth or composition can be permeabilized and/or lysed using methods known in the art.

A whole broth or cell composition as described herein is typically a liquid, but may contain insoluble components, such as killed cells, cell debris, culture media components, and/or insoluble enzyme(s). In some embodiments, insoluble components may be removed to provide a clarified liquid composition.

The whole broth formulations and cell compositions of the present invention may be produced by a method described in WO 90/15861 or WO 2010/096673.

Enzyme Compositions

The present invention also relates to compositions comprising a polypeptide having trehalase activity of the present invention. Preferably, the compositions are enriched in such a polypeptide. The term "enriched" indicates that the trehalase activity of the composition has been increased, e.g., with an enrichment factor of at least 1.1.

The compositions may comprise a polypeptide of the present invention as the major enzymatic component, e.g., a mono-component composition. Alternatively, the compositions may comprise multiple enzymatic activities, such as one or more (e.g., several) enzymes selected from the group consisting of hydrolase, isomerase, ligase, lyase, oxidoreductase, or transferase, e.g., an alpha-galactosidase, alpha-glucosidase, aminopeptidase, amylase, beta-galactosidase, beta-glucosidase, beta-xylosidase, carbohydrase, carboxypeptidase, catalase, cellobiohydrolase, cellulase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, endoglucanase, esterase, glucoamylase, invertase, laccase, lipase, mannosidase, mutanase, oxidase, pectinolytic enzyme, peroxidase, phytase, polyphenoloxidase, proteolytic enzyme, ribonuclease, transglutaminase, or xylanase.

In an embodiment the composition comprises a trehalase of the invention and a glucoamylase. In an embodiment the composition comprises a trehalase of the invention and a glucoamylase derived from *Talaromyces emersonii* (e.g., SEQ ID NO: 13). In an embodiment the composition comprises a trehalase of the invention and a glucoamylase derived from

Gloeophyllum, such as *G. serpiarium* (e.g., SEQ ID NO: 19) or *G. trabeum* (e.g., SEQ ID NO: 20). In an embodiment the composition comprises a trehalase of the invention, a glucoamylase and an alpha-amylase. In an embodiment the composition comprises a trehalase of the invention, a glucoamylase and an alpha-amylase derived from *Rhizomucor*, preferably a strain the *Rhizomucor pusillus*, such as a *Rhizomucor pusillus* alpha-amylase hybrid having an linker (e.g., from *Aspergillus niger*) and starch-bonding domain (e.g., from *Aspergillus niger*). In an embodiment the composition comprises a trehalase of the invention, a glucoamylase, an alpha-amylase and a cellulolytic enzyme composition. In an embodiment the composition comprises a trehalase of the invention, a glucoamylase, an alpha-amylase and a cellulolytic enzyme composition, wherein the cellulolytic composition is derived from *Trichoderma reesei*. In an embodiment the composition comprises a trehalase of the invention, a glucoamylase, an alpha-amylase and a protease. In an embodiment the composition comprises a trehalase of the invention, a glucoamylase, an alpha-amylase and a protease. The protease may be derived from *Thermoascus aurantiacus*. In an embodiment the composition comprises a trehalase of the invention, a glucoamylase, an alpha-amylase, a cellulolytic enzyme composition and a protease. In an embodiment the composition comprises a trehalase of the invention, a glucoamylase, e.g., derived from *Talaromyces emersonii*, *Gloeophyllum serpiarium* or *Gloeophyllum trabeum*, an alpha-amylase, e.g., derived from *Rhizomucor pusillus*, in particular one having a linker and starch-binding domain, in particular derived from *Aspergillus niger*, in particular one having the following substitutions: G128D+D143N (using SEQ ID NO: 15 for numbering); a cellulolytic enzyme composition derived from *Trichoderma reesei*, and a protease, e.g., derived from *Thermoascus aurantiacus*.

Examples of specifically contemplated secondary enzymes, e.g., a glucoamylase from *Talaromyces emersonii* shown in SEQ ID NO: 13 herein or a glucoamylase having, e.g., at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% identity to SEQ ID NO: 13 herein can be found in the "Enzymes" section below.

The compositions may be prepared in accordance with methods known in the art and may be in the form of a liquid or a dry composition. The compositions may be stabilized in accordance with methods known in the art.

Examples are given below of preferred uses of the compositions of the present invention. The dosage of the composition and other conditions under which the composition is used may be determined on the basis of methods known in the art.

Processes Of The Invention

Producing a fermentation product from Gelatinized Starch Material Using a Trehalase of the Invention

In this aspect the present invention relates to producing a fermentation product, in particular ethanol, from gelatinized and/or ungelatinized starch-containing material or cellulosic material. Fermentable sugars generated during saccharification/hydrolysis are converted to the desired fermentation in question, in particular ethanol, during fermentation by a fermenting organism, in particular yeast.

In an embodiment the invention relates to processes of producing a fermentation product, in particular ethanol, comprising

(a) liquefying a starch-containing material with an alpha-amylase;
optionally pre-saccharifying the liquefied material before step (b);

(b) saccharifying the liquefied material;

(c) fermenting using a fermentation organism;

wherein

i) a glucoamylase;

ii) a trehalase of the invention;

iii) optionally a cellulolytic enzyme composition and/or a protease;

are present and/or added during

- saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

Liquefaction step (a)

According to processes of the invention, liquefaction in step (a) is carried out by subjecting starch-containing material at a temperature above the initial gelatinization temperature, in particular at a temperature between 80-90°C, to an alpha-amylase and optionally a protease and other enzymes, such as a glucoamylase, a pullulanase and/or a phytase.

The term "initial gelatinization temperature" means the lowest temperature at which gelatinization of the starch-containing material commences. In general, starch heated in water begins to gelatinize between about 50°C and 75°C; the exact temperature of gelatinization depends on the specific starch and can readily be determined by the skilled artisan. Thus, the initial gelatinization temperature may vary according to the plant species, to the particular variety of the plant species as well as with the growth conditions. In the context of this invention the initial gelatinization temperature of a given starch-containing material may be determined as the temperature at which birefringence is lost in 5% of the starch granules using the method described by Gorinstein and Lii, 1992, *Starch/Stärke* 44(12): 461-466.

According to the invention liquefaction in step (a) is typically carried out at a temperature in the range from 70-100°C. In an embodiment the temperature in liquefaction is between 75-95°C, such as between 75-90°C, preferably between 80-90°C, such as 82-88°C, such as around 85°C.

5 The pH in liquefaction may be in the range between 3 and 7, preferably from 4 to 6, or more preferably from 4.5 to 5.5.

According to the invention a jet-cooking step may be carried out prior to liquefaction in step (a). The jet-cooking may be carried out at a temperature between 110-145°C, preferably 120-140°C, such as 125-135°C, preferably around 130°C for about 1-15 minutes, preferably for
10 about 3-10 minutes, especially around about 5 minutes.

In an embodiment, the process of the invention further comprises, prior to the liquefaction step (a), the steps of:

- x) reducing the particle size of the starch-containing material, preferably by dry milling;
- z) forming a slurry comprising the starch-containing material and water.

15 According to the invention the dry solid content (DS) in liquefaction lies in the range from 20-55 wt.-%, preferably 25-45 wt.-%, more preferably 30-40 wt.-% or 30-45 wt.-%.

The starch-containing starting material, such as whole grains, may be reduced in particle size, e.g., by milling, in order to open up the structure, to increase surface area, and allowing for further processing. Generally there are two types of processes: wet and dry milling.
20 In dry milling whole kernels are milled and used. Wet milling gives a good separation of germ and meal (starch granules and protein). Wet milling is often applied at locations where the starch hydrolysate is used in production of, e.g., syrups. Both dry milling and wet milling are well known in the art of starch processing. According to the present invention dry milling is preferred.

25 In an embodiment the particle size is reduced to between 0.05 to 3.0 mm, preferably 0.1-0.5 mm, or so that at least 30%, preferably at least 50%, more preferably at least 70%, even more preferably at least 90% of the starch-containing material fit through a sieve with a 0.05 to 3.0 mm screen, preferably 0.1-0.5 mm screen. In another embodiment at least 50%, preferably at least 70%, more preferably at least 80%, especially at least 90% of the starch-
30 containing material fit through a sieve with # 6 screen.

Liquefaction in step (a) may be carried out for 0.5-5 hours, such as 1-3 hours, such as typically around 2 hours.

The alpha-amylase and other optional enzymes, such as protease, may initially be added to the aqueous slurry to initiate liquefaction (thinning). In an embodiment only a portion
35 of the enzymes (e.g., about 1/3) is added to the aqueous slurry, while the rest of the enzymes (e.g., about 2/3) are added in liquefaction step (a).

A non-exhaustive list of examples of alpha-amylases can be found below in the "Alpha-Amylase Present and/or Added In Liquefaction"-section. In a preferred embodiment the alpha-amylase is a bacterial alpha-amylase. Bacterial alpha-amylases are typically thermostable. In a preferred embodiment the alpha-amylase is derived from the genus *Bacillus*, such as a strain of *Bacillus stearothermophilus*, in particular a variant of a *Bacillus stearothermophilus* alpha-amylase, such as the one shown in SEQ ID NO: 3 in WO 99/019467 or SEQ ID NO: 18 herein.

In an embodiment the alpha-amylase used in liquefaction step (a) is a variant of the *Bacillus stearothermophilus* alpha-amylase shown in SEQ ID NO: 18 herein, in particular with the double deletions in I181*+G182*, and optionally with a N193F substitution, and truncated to be around 491 amino acids long, e.g., from 480-495 amino acids long.

Examples of suitable *Bacillus stearothermophilus* alpha-amylase variants can be found below in the "Thermostable Alpha-Amylase"-section and include one from the following group of *Bacillus stearothermophilus* alpha-amylase variants with double deletions I181*+G182*, and optionally substitution N193F, and additionally the following substitutions:

- E129V+K177L+R179E;
- V59A+Q89R+E129V+K177L+R179E+H208Y+K220P+N224L+Q254S;
- V59A+ Q89R+ E129V+ K177L+ R179E+ Q254S+ M284V;
- V59A+ E129V+ K177L+ R179E+ Q254S+ M284V; and
- E129V+K177L+R179E+K220P+N224L+S242Q+Q254S (using SEQ ID NO: 18 for numbering).

According to processes of the invention, liquefaction in step (a) may be carried out using a combination of alpha-amylase (e.g., *Bacillus stearothermophilus* alpha-amylase shown in SEQ ID NO: 18) and protease (e.g., *Pyrococcus furiosus* (pfu protease) shown in SEQ ID NO: 22). A glucoamylase may also be present, such as the one derived from *Penicillium oxalicum* shown in SEQ ID NO: 23 herein (see the "Glucoamylase Present and/or Added In Liquefaction Step (a)"-section below).

Saccharification and Fermentation

A trehalase of the invention, a glucoamylase and optionally a protease and/or a cellulolytic enzyme composition may be present and/or added in saccharification step (b); fermentation step (c); simultaneous saccharification and fermentation (SSF); optionally a presaccharification step before step (b).

In a preferred embodiment the glucoamylase is added together with a fungal alpha-amylase, in particular acid fungal alpha-amylase. Examples of glucoamylases can be found in the "Glucoamylases Present and/or Added In Saccharification and/or Fermentation"-section below.

When doing sequential saccharification and fermentation, saccharification step (b) may be carried out at conditions well-known in the art, i.e., suitable for enzyme saccharification. For instance, the saccharification step (b) may last up to from about 24 to about 72 hours.

In an embodiment pre-saccharification is done before saccharification in step (b). Pre-saccharification is typically done for 40-90 minutes at a temperature between 30-65°C, typically about 60°C. Pre-saccharification is in an embodiment followed by saccharification during fermentation in simultaneous saccharification and fermentation (SSF). Saccharification is typically carried out at temperatures from 20-75°C, preferably from 40-70°C, typically around 60°C, and at a pH between 4 and 5, normally at about pH 4.5.

Simultaneous saccharification and fermentation ("SSF") is widely used in industrial scale fermentation product production processes, especially ethanol production processes. When doing SSF the saccharification step (b) and the fermentation step (c) are carried out simultaneously. There is no holding stage for the saccharification, meaning that a fermenting organism, in particular yeast, and enzymes, may be added together. However, it is also contemplated to add the fermenting organism and enzymes separately. SSF is according to the invention typically carried out at a temperature from 25°C to 40°C, such as from 28°C to 35°C, such as from 30°C to 34°C, preferably around about 32°C. In an embodiment fermentation is ongoing for 6 to 120 hours, in particular 24 to 96 hours. In an embodiment the pH is between 4-5.

In an embodiment of the invention a cellulolytic composition is present and/or added in saccharification step (b), fermentation step (c) or simultaneous saccharification and fermentation (SSF) or pre-saccharification before step (b). Examples of such cellulolytic compositions can be found in the "Cellulolytic Enzyme Composition present and/or added during Saccharification and/or Fermentation"-section below. The optional cellulolytic enzyme composition may be present and/or added together with the glucoamylase and trehalase of the invention. Examples of proteases can be found in the "Proteases Present and/or Added In Saccharification and/or Fermentation"-section below.

In a preferred embodiment the trehalase is present and/or added in an amount between 0.01-20 ug EP trehalase/g DS, such as between 0.05-15 ug EP terhalase/g DS, such as between 0.5 and 10 ug EP trehalase/g DS.

Starch-Containing Materials

According to the invention any suitable starch-containing starting material may be used. The starting material is generally selected based on the desired fermentation product, in particular ethanol. Examples of starch-containing starting materials, suitable for use in processes of the present invention, include cereal, tubers or grains. Specifically the starch-containing material may be corn, wheat, barley, rye, milo, sago, cassava, tapioca, sorghum,

oat, rice, peas, beans, or sweet potatoes, or mixtures thereof. Contemplated are also waxy and non-waxy types of corn and barley.

In a preferred embodiment the starch-containing starting material is corn.

In a preferred embodiment the starch-containing starting material is wheat.

5 In a preferred embodiment the starch-containing starting material is barley.

In a preferred embodiment the starch-containing starting material is rye.

In a preferred embodiment the starch-containing starting material is milo.

In a preferred embodiment the starch-containing starting material is sago.

In a preferred embodiment the starch-containing starting material is cassava.

10 In a preferred embodiment the starch-containing starting material is tapioca.

In a preferred embodiment the starch-containing starting material is sorghum.

In a preferred embodiment the starch-containing starting material is rice,

In a preferred embodiment the starch-containing starting material is peas.

In a preferred embodiment the starch-containing starting material is beans.

15 In a preferred embodiment the starch-containing starting material is sweet potatoes.

In a preferred embodiment the starch-containing starting material is oats.

Producing a Fermentation Product from Ungelatinized Starch Material using a Trehalase of the Invention

20 A trehalase of the invention may suitably be used in a raw starch hydrolysis (RSH) process for producing desired fermentation products, in particular ethanol. In RSH processes the starch does not gelatinize as the process is carried out at temperatures below the initial gelatinization temperature of the starch in question (defined above).

25 The desired fermentation product may in an embodiment be ethanol produced from ungelatinized (i.e., uncooked), preferably milled, grains, such as corn, or small grains such as wheat, oats, barley, rye, rice, or cereals such as sorghum. Examples of suitable starch-containing starting materials are listed in the section "Starch-Containing Materials"-section above.

30 Accordingly, in this aspect the invention relates to processes of producing fermentation products from starch-containing material comprising:

(a) saccharifying a starch-containing material at a temperature below the initial gelatinization temperature; and

(b) fermenting using a fermentation organism; and

(c) optionally recovering the fermentation product;

35 wherein saccharification and/or fermentation is done in the presence of the following enzymes: glucoamylase, alpha-amylase, trehalase of the invention, and optionally a cellulolytic enzyme composition and/or a protease.

Before step (a) an aqueous slurry of starch-containing material, such as granular starch, having 10-55 wt.-% dry solids (DS), preferably 25-45 wt.-% dry solids, more preferably 30-40% dry solids of starch-containing material may be prepared. The slurry may include water and/or process waters, such as stillage (backset), scrubber water, evaporator condensate or distillate, side-stripper water from distillation, or process water from other fermentation product plants. Because a raw starch hydrolysis process of the invention is carried out below the initial gelatinization temperature, and thus no significant viscosity increase takes place, high levels of stillage may be used, if desired. In an embodiment the aqueous slurry contains from about 1 to about 70 vol.-%, preferably 15-60% vol.-%, especially from about 30 to 50 vol.-% water and/or process waters, such as stillage (backset), scrubber water, evaporator condensate or distillate, side-stripper water from distillation, or process water from other fermentation product plants, or combinations thereof, or the like.

In an embodiment backset, or another recycled stream, is added to the slurry before step (a), or to the saccharification (step (a)), or to the simultaneous saccharification and fermentation steps (combined step (a) and step (b)).

A RSH process of the invention is conducted at a temperature below the initial gelatinization temperature, which means that the temperature at which a separate step (a) is carried out typically lies in the range between 25-75°C, such as between 30-70°C, or between 45-60°C.

In a preferred embodiment the temperature during fermentation in step (b) or simultaneous saccharification and fermentation in steps (a) and (b) is between 25°C and 40°C, preferably between 28°C and 36°C, such as between 28°C and 35°C, such as between 28°C and 34°C, such as around 32°C.

In an embodiment of the invention fermentation is carried out for 30 to 150 hours, preferably 48 to 96 hours.

In an embodiment fermentation is carried out so that the sugar level, such as glucose level, is kept at a low level, such as below 6 wt.-%, such as below about 3 wt.-%, such as below about 2 wt.-%, such as below about 1 wt.-%, such as below about 0.5%, or below 0.25% wt.-%, such as below about 0.1 wt.-%. Such low levels of sugar can be accomplished by simply employing adjusted quantities of enzymes and fermenting organism. A skilled person in the art can easily determine which doses/quantities of enzyme and fermenting organism to use. The employed quantities of enzyme and fermenting organism may also be selected to maintain low concentrations of maltose in the fermentation broth. For instance, the maltose level may be kept below about 0.5 wt.-%, such as below about 0.2 wt.-%.

The process of the invention may be carried out at a pH from 3 and 7, preferably from 3 to 6, or more preferably from 3.5 to 5.0.

The term "granular starch" means raw uncooked starch, i.e., starch in its natural form found in, e.g., cereal, tubers or grains. Starch is formed within plant cells as tiny granules insoluble in water. When put in cold water, the starch granules may absorb a small amount of the liquid and swell. At temperatures up to around 50°C to 75°C the swelling may be reversible. However, at higher temperatures an irreversible swelling called "gelatinization" begins. The granular starch may be a highly refined starch, preferably at least 90%, at least 95%, at least 97% or at least 99.5% pure, or it may be a more crude starch-containing materials comprising (e.g., milled) whole grains including non-starch fractions such as germ residues and fibers.

The raw material, such as whole grains, may be reduced in particle size, e.g., by milling, in order to open up the structure and allowing for further processing. Examples of suitable particle sizes are disclosed in US4514496 and WO2004/081193 (both references are incorporated by reference). Two processes are preferred according to the invention: wet and dry milling. In dry milling whole kernels are milled and used. Wet milling gives a good separation of germ and meal (starch granules and protein) and is often applied at locations where the starch hydrolysate is used in production of, e.g., syrups. Both dry and wet milling is well known in the art of starch processing.

In an embodiment the particle size is reduced to between 0.05 to 3.0 mm, preferably 0.1-0.5 mm, or so that at least 30%, preferably at least 50%, more preferably at least 70%, even more preferably at least 90% of the starch-containing material fit through a sieve with a 0.05 to 3.0 mm screen, preferably 0.1-0.5 mm screen. In a preferred embodiment starch-containing material is prepared by reducing the particle size of the starch-containing material, preferably by milling, such that at least 50% of the starch-containing material has a particle size of 0.1-0.5 mm.

In a preferred embodiment the trehalase is present and/or added in an amount between 0.01-20 ug EP trehalase/g DS, such as between 0.05-15 ug EP terhalase/g DS, such as between 0.5 and 10 ug EP trehalase/g DS.

According to the invention the enzymes are added so that the glucoamylase is present in an amount of 0.001 to 10 AGU/g DS, preferably from 0.01 to 5 AGU/g DS, especially 0.1 to 0.5 AGU/g DS.

According to the invention the enzymes are added so that the alpha-amylase is present or added in an amount of 0.001 to 10 AFAU/g DS, preferably from 0.01 to 5 AFAU/g DS, especially 0.3 to 2 AFAU/g DS or 0.001 to 1 FAU-F/g DS, preferably 0.01 to 1 FAU-F/g DS.

According to the invention the enzymes are added so that the cellulolytic enzyme composition is present or added in an amount 1-10,000 micro grams EP/g DS, such as 2-5,000, such as 3 and 1,000, such as 4 and 500 micro grams EP/g DS.

According to the invention the enzymes are added so that the cellulolytic enzyme composition is present or added in an amount in the range from 0.1-100 FPU per gram total solids (TS), preferably 0.5-50 FPU per gram TS, especially 1-20 FPU per gram TS.

In an embodiment of the invention the enzymes are added so that the protease is present in an amount of 0.0001-1 mg enzyme protein per g DS, preferably 0.001 to 0.1 mg enzyme protein per g DS. Alternatively, the protease is present and/or added in an amount of 0.0001 to 1 LAPU/g DS, preferably 0.001 to 0.1 LAPU/g DS and/or 0.0001 to 1 mAU-RH/g DS, preferably 0.001 to 0.1 mAU-RH/g DS.

In an embodiment of the invention the enzymes are added so that the protease is present or added in an amount in the range 1-1,000 µg EP/g DS, such as 2-500 µg EP/g DS, such as 3-250 µg EP/g DS.

In a preferred embodiment ratio between glucoamylase and alpha-amylase is between 99:1 and 1:2, such as between 98:2 and 1:1, such as between 97:3 and 2:1, such as between 96:4 and 3:1, such as 97:3, 96:4, 95:5, 94:6, 93:7, 90:10, 85:15, 83:17 or 65:35 (mg EP glucoamylase: mg EP alpha-amylase).

In a preferred embodiment the total dose of glucoamylase and alpha-amylase is according to the invention from 10-1,000 µg/g DS, such as from 50-500 µg/g DS, such as 75-250 µg/g DS.

In a preferred embodiment the total dose of cellulolytic enzyme composition added is from 10-500 µg/g DS, such as from 20-400 µg/g DS, such as 20-300 µg/g DS.

In an embodiment the glucoamylase, such as one derived from *Trametes cingulata*, used in fermentation or SSF exhibits at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to the mature part of SEQ ID NO: 14.

In an embodiment the glucoamylase, such as one derived from *Pycnoporus sanguineus*, used in fermentation or SSF exhibits at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to the mature part of SEQ ID NO: 28.

In an embodiment the alpha-amylase used in fermentation or SSF exhibits at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to the mature part of SEQ ID NO: 15.

In a preferred embodiment the invention relates to processes of producing fermentation products from starch-containing material comprising:

- (a) saccharifying a starch-containing material at a temperature below the initial gelatinization temperature; and
- (b) fermenting using a fermentation organism;

wherein saccharification and/or fermentation is done in the presence of the following enzymes:

- i) glucoamylase;
- ii) alpha-amylase;
- 5 iii) trehalase of the invention;
- iii) optionally a cellulolytic enzyme composition and/or a protease.

In a preferred embodiment the enzymes may be added as an enzyme composition of the invention. In a preferred embodiment steps (a) and (b) are carried out simultaneously (i.e., one-step fermentation). However, step (a) and (b) may also be carried out sequentially.

Fermentation

Fermentation is carried out in a fermentation medium. The fermentation medium includes the fermentation substrate, that is, the carbohydrate source that is metabolized by the fermenting organism. According to the invention the fermentation medium may comprise 15 nutrients and growth stimulator(s) for the fermenting organism(s). Nutrient and growth stimulators are widely used in the art of fermentation and include nitrogen sources, such as ammonia; urea, vitamins and minerals, or combinations thereof.

Fermenting Organisms for Starch Based Fermentation

20 The term "fermenting organism" refers to any organism, including bacterial and fungal organisms, especially yeast, suitable for use in a fermentation process and capable of producing the desired fermentation product. Especially suitable fermenting organisms are able to ferment, i.e., convert, sugars, such as glucose or maltose, directly or indirectly into the desired fermentation product, such as in particular ethanol. Examples of fermenting organisms 25 include fungal organisms, such as in particular yeast. Preferred yeast includes strains of *Saccharomyces* spp., in particular, *Saccharomyces cerevisiae*.

Suitable concentrations of the viable fermenting organism during fermentation, such as SSF, are well known in the art or can easily be determined by the skilled person in the art. In one embodiment the fermenting organism, such as ethanol fermenting yeast, (e.g., 30 *Saccharomyces cerevisiae*) is added to the fermentation medium so that the viable fermenting organism, such as yeast, count per mL of fermentation medium is in the range from 10^5 to 10^{12} , preferably from 10^7 to 10^{10} , especially about 5×10^7 .

Examples of commercially available yeast includes, e.g., RED STAR™ and ETHANOL RED™ yeast (available from Fermentis/Lesaffre, USA), FALI (available from Fleischmann's 35 Yeast, USA), SUPERSTART and THERMOSACC™ fresh yeast (available from Ethanol Technology, WI, USA), BIOFERM AFT and XR (available from NABC - North American

Bioproducts Corporation, GA, USA), GERT STRAND (available from Gert Strand AB, Sweden), and FERMIOL (available from DSM Specialties).

Recovery

Subsequent to fermentation, e.g., SSF, the fermentation product, in particular ethanol may be separated from the fermentation medium. The slurry may be distilled to recover/extract the desired fermentation product (i.e., ethanol). Alternatively the desired fermentation product (i.e., ethanol) may be extracted from the fermentation medium by micro or membrane filtration techniques. The fermentation product (i.e., ethanol) may also be recovered by stripping or other method well known in the art.

Alpha-Amylase Present and/or Added In Liquefaction

According to the invention an alpha-amylase is present and/or added in liquefaction optionally together with other enzymes such as a protease, a glucoamylase, phytase and/or pullulanase.

The alpha-amylase added in liquefaction step (a) may be any alpha-amylase. Preferred are bacterial alpha-amylases, which typically are stable at temperatures used in liquefaction.

Bacterial Alpha-Amylase

The term "bacterial alpha-amylase" means any bacterial alpha-amylase classified under EC 3.2.1.1. A bacterial alpha-amylase used according to the invention may, e.g., be derived from a strain of the genus *Bacillus*, which is sometimes also referred to as the genus *Geobacillus*. In an embodiment the *Bacillus* alpha-amylase is derived from a strain of *Bacillus amyloliquefaciens*, *Bacillus licheniformis*, *Bacillus stearothermophilus*, or *Bacillus subtilis*, but may also be derived from other *Bacillus* sp.

Specific examples of bacterial alpha-amylases include the *Bacillus stearothermophilus* alpha-amylase of SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 18 herein, the *Bacillus amyloliquefaciens* alpha-amylase of SEQ ID NO: 5 in WO 99/19467, and the *Bacillus licheniformis* alpha-amylase of SEQ ID NO: 4 in WO 99/19467 (all sequences are hereby incorporated by reference). In an embodiment the alpha-amylase may be an enzyme having a degree of identity of at least 60%, e.g., at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% to any of the sequences shown in SEQ ID NOS: 3, 4 or 5, respectively, in WO 99/19467.

In an embodiment the alpha-amylase may be an enzyme having a degree of identity of at least 60%, e.g., at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99% to any of the sequences shown in SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 18 herein.

In a preferred embodiment the alpha-amylase is derived from *Bacillus stearothermophilus*. The *Bacillus stearothermophilus* alpha-amylase may be a mature wild-type or a mature variant thereof. The mature *Bacillus stearothermophilus* alpha-amylases may naturally be truncated during recombinant production. For instance, the *Bacillus stearothermophilus* alpha-amylase may be a truncated so it has around 491 amino acids, e.g., so that it is between 480-495 amino acids long, so it lacks a functional starch binding domain (compared to SEQ ID NO: 3 in WO 99/19467) or SEQ ID NO: 18 herein.

The *Bacillus* alpha-amylase may also be a variant and/or hybrid. Examples of such a variant can be found in any of WO 96/23873, WO 96/23874, WO 97/41213, WO 99/19467, WO 00/60059, and WO 02/10355 (all documents are hereby incorporated by reference). Specific alpha-amylase variants are disclosed in U.S. Patent Nos. 6,093,562, 6,187,576, 6,297,038, and 7,713,723 (hereby incorporated by reference) and include *Bacillus stearothermophilus* alpha-amylase (often referred to as BSG alpha-amylase) variants having a deletion of one or two amino acids at positions R179, G180, I181 and/or G182, preferably a double deletion disclosed in WO 96/23873 – see, e.g., page 20, lines 1-10 (hereby incorporated by reference), preferably corresponding to deletion of positions I181 and G182 compared to the amino acid sequence of *Bacillus stearothermophilus* alpha-amylase set forth in SEQ ID NO: 3 disclosed in WO 99/19467 or SEQ ID NO: 1 herein or the deletion of amino acids R179 and G180 using SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 1 herein for numbering (which reference is hereby incorporated by reference). Even more preferred are *Bacillus* alpha-amylases, especially *Bacillus stearothermophilus* alpha-amylases, which have a double deletion corresponding to a deletion of positions 181 and 182 and further comprise a N193F substitution (also denoted I181* + G182* + N193F) compared to the wild-type BSG alpha-amylase amino acid sequence set forth in SEQ ID NO: 3 disclosed in WO 99/19467 or SEQ ID NO: 18 herein. The bacterial alpha-amylase may also have a substitution in a position corresponding to S239 in the *Bacillus licheniformis* alpha-amylase shown in SEQ ID NO: 4 in WO 99/19467, or a S242 and/or E188P variant of the *Bacillus stearothermophilus* alpha-amylase of SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 18 herein.

In an embodiment the variant is a S242A, E or Q variant, preferably a S242Q variant, of the *Bacillus stearothermophilus* alpha-amylase (using SEQ ID NO: 18 herein for numbering).

In an embodiment the variant is a position E188 variant, preferably E188P variant of the *Bacillus stearothermophilus* alpha-amylase (using SEQ ID NO: 18 herein for numbering).

The bacterial alpha-amylase may in an embodiment be a truncated *Bacillus* alpha-amylase. Especially the truncation is so that, e.g., the *Bacillus stearothermophilus* alpha-amylase shown in SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 18 herein, is around 491 amino acids long, such as from 480 to 495 amino acids long, or so it lack a functional starch binding domain.

Bacterial Hybrid Alpha-Amylases

The bacterial alpha-amylase may also be a hybrid bacterial alpha-amylase, e.g., an alpha-amylase comprising 445 C-terminal amino acid residues of the *Bacillus licheniformis* alpha-amylase (shown in SEQ ID NO: 4 of WO 99/19467) and the 37 N-terminal amino acid residues of the alpha-amylase derived from *Bacillus amyloliquefaciens* (shown in SEQ ID NO: 5 of WO 99/19467). In a preferred embodiment this hybrid has one or more, especially all, of the following substitutions:

G48A+T49I+G107A+H156Y+A181T+N190F+I201F+A209V+Q264S (using the *Bacillus licheniformis* numbering in SEQ ID NO: 4 of WO 99/19467). Also preferred are variants having one or more of the following mutations (or corresponding mutations in other *Bacillus* alpha-amylases): H154Y, A181T, N190F, A209V and Q264S and/or the deletion of two residues between positions 176 and 179, preferably the deletion of E178 and G179 (using SEQ ID NO: 5 of WO 99/19467 for position numbering).

In an embodiment the bacterial alpha-amylase is the mature part of the chimeric alpha-amylase disclosed in Richardson et al. (2002), The Journal of Biological Chemistry, Vol. 277, No 29, Issue 19 July, pp. 267501-26507, referred to as BD5088 or a variant thereof. This alpha-amylase is the same as the one shown in SEQ ID NO: 2 in WO 2007134207. The mature enzyme sequence starts after the initial "Met" amino acid in position 1.

Thermostable Alpha-Amylase

The alpha-amylase may be a thermostable alpha-amylase, such as a thermostable bacterial alpha-amylase, preferably from *Bacillus stearothermophilus*.

In an embodiment of the invention the alpha-amylase is an bacterial alpha-amylase, preferably derived from the genus *Bacillus*, especially a strain of *Bacillus stearothermophilus*, in particular the *Bacillus stearothermophilus* as disclosed in WO 99/019467 as SEQ ID NO: 3 (SEQ ID NO: 18 herein) with one or two amino acids deleted at positions R179, G180, I181 and/or G182, in particular with R179 and G180 deleted, or with I181 and G182 deleted, with mutations in below list of mutations.

In preferred embodiments the *Bacillus stearothermophilus* alpha-amylases have double deletion I181 + G182, and optionally substitution N193F, further comprising mutations selected from below list:

-V59A+Q89R+G112D+E129V+K177L+R179E+K220P+N224L+Q254S;
- V59A+Q89R+E129V+K177L+R179E+H208Y+K220P+N224L+Q254S;
- V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+D269E+D281N;
- V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+I270L;
- V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+H274K;

- V59A+Q89R+E129V+K177L+R179E+K220P+N224L+Q254S+Y276F;
- V59A+E129V+R157Y+K177L+R179E+K220P+N224L+S242Q+Q254S;
- V59A+E129V+K177L+R179E+H208Y+K220P+N224L+S242Q+Q254S;
- 59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S;
- V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+H274K;
- V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+Y276F;
- V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+D281N;
- V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+M284T;
- V59A+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+G416V;
- V59A+E129V+K177L+R179E+K220P+N224L+Q254S;
- V59A+E129V+K177L+R179E+K220P+N224L+Q254S+M284T;
- A91L+M96I+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S;
- E129V+K177L+R179E;
- E129V+K177L+R179E+K220P+N224L+S242Q+Q254S;
- E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+Y276F+L427M;
- E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+M284T;
- E129V+K177L+R179E+K220P+N224L+S242Q+Q254S+N376*+I377*;
- E129V+K177L+R179E+K220P+N224L+Q254S;
- E129V+K177L+R179E+K220P+N224L+Q254S+M284T;
- E129V+K177L+R179E+S242Q;
- E129V+K177L+R179V+K220P+N224L+S242Q+Q254S;
- K220P+N224L+S242Q+Q254S;
- M284V;
- V59A+Q89R+ E129V+ K177L+ R179E+ Q254S+ M284V.
- V59A+E129V+K177L+R179E+Q254S+ M284V;

Specific information about the thermostability of above alpha-amylases variants can be found in WO12/088303 (Novozymes) which is hereby incorporated by reference.

In a preferred embodiment the alpha-amylase is selected from the group of *Bacillus* *stearothermophilus* alpha-amylase variants having a double deletion in I181+G182, and optionally a substitution in N193F, and substitutions from the following list

- E129V+K177L+R179E;
- V59A+Q89R+E129V+K177L+R179E+H208Y+K220P+N224L+Q254S;
- V59A+Q89R+ E129V+ K177L+ R179E+ Q254S+ M284V;
- V59A+E129V+K177L+R179E+Q254S+ M284V; and
- E129V+K177L+R179E+K220P+N224L+S242Q+Q254S (using SEQ ID NO: 18 herein for numbering).

It should be understood that when referring to *Bacillus stearothermophilus* alpha-amylase and variants thereof they are normally produced in truncated form. In particular, the

truncation may be so that the *Bacillus stearothermophilus* alpha-amylase shown in SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 18 herein, or variants thereof, are truncated in the C-terminal and are typically around 491 amino acids long, such as from 480-495 amino acids long, or so that it lacks a functional starch binding domain.

5 In a preferred embodiment the alpha-amylase variant may be an enzyme having a degree of identity of at least 60%, e.g., at least 70%, at least 80%, at least 90%, at least 95%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98% or at least 99%, but less than 100% to the sequence shown in SEQ ID NO: 3 in WO 99/19467 or SEQ ID NO: 18 herein.

10 In an embodiment the bacterial alpha-amylase, e.g., *Bacillus* alpha-amylase, such as especially *Bacillus stearothermophilus* alpha-amylase, or variant thereof, is dosed to liquefaction in a concentration between 0.01-10 KNU-A/g DS, e.g., between 0.02 and 5 KNU-A/g DS, such as 0.03 and 3 KNU-A, preferably 0.04 and 2 KNU-A/g DS, such as especially 0.01 and 2 KNU-A/g DS. In an embodiment the bacterial alpha-amylase, e.g., *Bacillus* alpha-
15 amylase, such as especially *Bacillus stearothermophilus* alpha-amylases, or variant thereof, is dosed to liquefaction in a concentration of between 0.0001-1 mg EP(Enzyme Protein)/g DS, e.g., 0.0005-0.5 mg EP/g DS, such as 0.001-0.1 mg EP/g DS.

Protease Present and/or Added In Liquefaction

20 According to the invention a protease may optionally be present and/or added in liquefaction together with the alpha-amylase, and an optional glucoamylase, phytase and/or pullulanase.

Proteases are classified on the basis of their catalytic mechanism into the following groups: Serine proteases (S), Cysteine proteases (C), Aspartic proteases (A), Metallo
25 proteases (M), and Unknown, or as yet unclassified, proteases (U), see Handbook of Proteolytic Enzymes, A.J.Barrett, N.D.Rawlings, J.F.Woessner (eds), Academic Press (1998), in particular the general introduction part.

In a preferred embodiment the thermostable protease used according to the invention is a "metallo protease" defined as a protease belonging to EC 3.4.24 (metalloendopeptidases);
30 preferably EC 3.4.24.39 (acid metallo proteinases).

To determine whether a given protease is a metallo protease or not, reference is made to the above "Handbook of Proteolytic Enzymes" and the principles indicated therein. Such determination can be carried out for all types of proteases, be it naturally occurring or wild-type proteases; or genetically engineered or synthetic proteases.

35 Protease activity can be measured using any suitable assay, in which a substrate is employed, that includes peptide bonds relevant for the specificity of the protease in question. Assay-pH and assay-temperature are likewise to be adapted to the protease in question.

Examples of assay-pH-values are pH 6, 7, 8, 9, 10, or 11. Examples of assay-temperatures are 30, 35, 37, 40, 45, 50, 55, 60, 65, 70 or 80°C.

Examples of protease substrates are casein, such as Azurine-Crosslinked Casein (AZCL-casein). Two protease assays are described below in the "Materials & Methods"-section, of which the so-called "AZCL-Casein Assay" is the preferred assay.

There are no limitations on the origin of the protease used in a process of the invention as long as it fulfills the thermostability properties defined below.

In one embodiment the protease is of fungal origin.

The protease may be a variant of, e.g., a wild-type protease as long as the protease is thermostable. In a preferred embodiment the thermostable protease is a variant of a metallo protease as defined above. In an embodiment the thermostable protease used in a process of the invention is of fungal origin, such as a fungal metallo protease, such as a fungal metallo protease derived from a strain of the genus *Thermoascus*, preferably a strain of *Thermoascus aurantiacus*, especially *Thermoascus aurantiacus* CGMCC No. 0670 (classified as EC 3.4.24.39).

In an embodiment the thermostable protease is a variant of the mature part of the metallo protease shown in SEQ ID NO: 2 disclosed in WO 2003/048353 or the mature part of SEQ ID NO: 1 in WO 2010/008841 and shown as SEQ ID NO: 17 herein, further with mutations selected from below list:

- S5*+D79L+S87P+A112P+D142L;
- D79L+S87P+A112P+T124V+D142L;
- S5*+N26R+D79L+S87P+A112P+D142L;
- N26R+T46R+D79L+S87P+A112P+D142L;
- T46R+D79L+S87P+T116V+D142L;
- D79L+P81R+S87P+A112P+D142L;
- A27K+D79L+S87P+A112P+T124V+D142L;
- D79L+Y82F+S87P+A112P+T124V+D142L;
- D79L+Y82F+S87P+A112P+T124V+D142L;
- D79L+S87P+A112P+T124V+A126V+D142L;
- D79L+S87P+A112P+D142L;
- D79L+Y82F+S87P+A112P+D142L;
- S38T+D79L+S87P+A112P+A126V+D142L;
- D79L+Y82F+S87P+A112P+A126V+D142L;
- A27K+D79L+S87P+A112P+A126V+D142L;
- D79L+S87P+N98C+A112P+G135C+D142L;
- D79L+S87P+A112P+D142L+T141C+M161C;
- S36P+D79L+S87P+A112P+D142L;

- A37P+D79L+S87P+A112P+D142L;
- S49P+D79L+S87P+A112P+D142L;
- S50P+D79L+S87P+A112P+D142L;
- D79L+S87P+D104P+A112P+D142L;
- 5 - D79L+Y82F+S87G+A112P+D142L;
- S70V+D79L+Y82F+S87G+Y97W+A112P+D142L;
- D79L+Y82F+S87G+Y97W+D104P+A112P+D142L;
- S70V+D79L+Y82F+S87G+A112P+D142L;
- D79L+Y82F+S87G+D104P+A112P+D142L;
- 10 - D79L+Y82F+S87G+A112P+A126V+D142L;
- Y82F+S87G+S70V+D79L+D104P+A112P+D142L;
- Y82F+S87G+D79L+D104P+A112P+A126V+D142L;
- A27K+D79L+Y82F+S87G+D104P+A112P+A126V+D142L;
- A27K+Y82F+S87G+D104P+A112P+A126V+D142L;
- 15 - A27K+D79L+Y82F+ D104P+A112P+A126V+D142L;
- A27K+Y82F+D104P+A112P+A126V+D142L;
- A27K+D79L+S87P+A112P+D142L;
- D79L+S87P+D142L.

Specific information about the thermostability of above protease variants can be found
 20 in WO12/088303 (Novozymes), which is hereby incorporated by reference.

In an preferred embodiment the thermostable protease is a variant of the metallo
 protease disclosed as the mature part of SEQ ID NO: 2 disclosed in WO 2003/048353 or the
 mature part of SEQ ID NO: 1 in WO 2010/008841 or SEQ ID NO: 17 herein with the following
 mutations:

- 25 D79L+S87P+A112P+D142L;
- D79L+S87P+D142L; or
- A27K+ D79L+Y82F+S87G+D104P+A112P+A126V+D142L.

In an embodiment the protease variant has at least 75% identity preferably at least 80%,
 more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more
 30 preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and
 even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at
 least 99%, but less than 100% identity to the mature part of the polypeptide of SEQ ID NO: 2
 disclosed in WO 2003/048353 or the mature part of SEQ ID NO: 1 in WO 2010/008841 or SEQ
 ID NO: 17 herein.

35 The thermostable protease may also be derived from any bacterium as long as the
 protease has the thermostability properties defined according to the invention.

In an embodiment the thermostable protease is derived from a strain of the bacterium *Pyrococcus*, such as a strain of *Pyrococcus furiosus* (pfu protease).

In an embodiment the protease is one shown as SEQ ID NO: 1 in US patent No. 6,358,726-B1 (Takara Shuzo Company), or SEQ ID NO: 22 herein.

In another embodiment the thermostable protease is one disclosed in SEQ ID NO: 22 herein or a protease having at least 80% identity, such as at least 85%, such as at least 90%, such as at least 95%, such as at least 96%, such as at least 97%, such as at least 98%, such as at least 99% identity to SEQ ID NO: 1 in US patent no. 6,358,726-B1 or SEQ ID NO: 22 herein. *Pyrococcus furiosus* protease can be purchased from Takara Bio, Japan.

Glucoamylase Present and/or Added In Liquefaction

According to the invention a glucoamylase may optionally be present and/or added in liquefaction step (a). In a preferred embodiment the glucoamylase is added together with or separately from the alpha-amylase and optional protease, phytase and/or pullulanase.

In a specific and preferred embodiment the glucoamylase, preferably of fungal origin, preferably a filamentous fungi, is from a strain of the genus *Penicillium*, especially a strain of *Penicillium oxalicum*, in particular the *Penicillium oxalicum* glucoamylase disclosed as SEQ ID NO: 2 in WO 2011/127802 (which is hereby incorporated by reference) and shown in SEQ ID NO: 23 herein.

In an embodiment the glucoamylase has at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at least 99% or 100% identity to the mature polypeptide shown in SEQ ID NO: 2 in WO 2011/127802 or SEQ ID NO: 23 herein.

In a preferred embodiment the glucoamylase is a variant of the *Penicillium oxalicum* glucoamylase disclosed as SEQ ID NO: 2 in WO 2011/127802 and shown in SEQ ID NO: 23 herein, having a K79V substitution (using the mature sequence shown in SEQ ID NO: 23 herein for numbering). The K79V glucoamylase variant has reduced sensitivity to protease degradation relative to the parent as disclosed in WO 2013/036526 (which is hereby incorporated by reference).

In an embodiment the glucoamylase is derived from *Penicillium oxalicum*.

In an embodiment the glucoamylase is a variant of the *Penicillium oxalicum* glucoamylase disclosed as SEQ ID NO: 2 in WO 2011/127802 and shown in SEQ ID NO: 23 herein. In a preferred embodiment the *Penicillium oxalicum* glucoamylase is the one disclosed as SEQ ID NO: 2 in WO 2011/127802 and shown in SEQ ID NO: 23 herein having Val (V) in position 79 (using SEQ ID NO: 23 herein for numbering).

Contemplated *Penicillium oxalicum* glucoamylase variants are disclosed in WO 2013/053801 which is hereby incorporated by reference.

In an embodiment these variants have reduced sensitivity to protease degradation. degradation.

5 In an embodiment these variant have improved thermostability compared to the parent.

More specifically, in an embodiment the glucoamylase has a K79V substitution (using SEQ ID NO: 23 herein for numbering), (PE001 variant), and further comprises at least one of the following substitutions or combination of substitutions:

T65A; or

10 Q327F; or

E501V; or

Y504T; or

Y504*; or

T65A + Q327F; or

15 T65A + E501V; or

T65A + Y504T; or

T65A + Y504*; or

Q327F + E501V; or

Q327F + Y504T; or

20 Q327F + Y504*; or

E501V + Y504T; or

E501V + Y504*; or

T65A + Q327F + E501V; or

T65A + Q327F + Y504T; or

25 T65A + E501V + Y504T; or

Q327F + E501V + Y504T; or

T65A + Q327F + Y504*; or

T65A + E501V + Y504*; or

Q327F + E501V + Y504*; or

30 T65A + Q327F + E501V + Y504T; or

T65A + Q327F + E501V + Y504*;

E501V + Y504T; or

T65A + K161S; or

T65A + Q405T; or

35 T65A + Q327W; or

T65A + Q327F; or

T65A + Q327Y; or

- P11F + T65A + Q327F; or
 R1K + D3W + K5Q + G7V + N8S + T10K + P11S + T65A + Q327F; or
 P2N + P4S + P11F + T65A + Q327F; or
 P11F + D26C + K33C + T65A + Q327F; or
 5 P2N + P4S + P11F + T65A + Q327W + E501V + Y504T; or
 R1E + D3N + P4G + G6R + G7A + N8A + T10D + P11D + T65A + Q327F; or
 P11F + T65A + Q327W; or
 P2N + P4S + P11F + T65A + Q327F + E501V + Y504T; or
 P11F + T65A + Q327W + E501V + Y504T; or
 10 T65A + Q327F + E501V + Y504T; or
 T65A + S105P + Q327W; or
 T65A + S105P + Q327F; or
 T65A + Q327W + S364P; or
 T65A + Q327F + S364P; or
 15 T65A + S103N + Q327F; or
 P2N + P4S + P11F + K34Y + T65A + Q327F; or
 P2N + P4S + P11F + T65A + Q327F + D445N + V447S; or
 P2N + P4S + P11F + T65A + I172V + Q327F; or
 P2N + P4S + P11F + T65A + Q327F + N502*; or
 20 P2N + P4S + P11F + T65A + Q327F + N502T + P563S + K571E; or
 P2N + P4S + P11F + R31S + K33V + T65A + Q327F + N564D + K571S; or
 P2N + P4S + P11F + T65A + Q327F + S377T; or
 P2N + P4S + P11F + T65A + V325T + Q327W; or
 P2N + P4S + P11F + T65A + Q327F + D445N + V447S + E501V + Y504T; or
 25 P2N + P4S + P11F + T65A + I172V + Q327F + E501V + Y504T; or
 P2N + P4S + P11F + T65A + Q327F + S377T + E501V + Y504T; or
 P2N + P4S + P11F + D26N + K34Y + T65A + Q327F; or
 P2N + P4S + P11F + T65A + Q327F + I375A + E501V + Y504T; or
 P2N + P4S + P11F + T65A + K218A + K221D + Q327F + E501V + Y504T; or
 30 P2N + P4S + P11F + T65A + S103N + Q327F + E501V + Y504T; or
 P2N + P4S + T10D + T65A + Q327F + E501V + Y504T; or
 P2N + P4S + F12Y + T65A + Q327F + E501V + Y504T; or
 K5A + P11F + T65A + Q327F + E501V + Y504T; or
 P2N + P4S + T10E + E18N + T65A + Q327F + E501V + Y504T; or
 35 P2N + T10E + E18N + T65A + Q327F + E501V + Y504T; or
 P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + T568N; or
 P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + K524T + G526A; or

P2N + P4S + P11F + K34Y + T65A + Q327F + D445N + V447S + E501V + Y504T; or
 P2N + P4S + P11F + R31S + K33V + T65A + Q327F + D445N + V447S + E501V +
 Y504T; or

P2N + P4S + P11F + D26N + K34Y + T65A + Q327F + E501V + Y504T; or

5 P2N + P4S + P11F + T65A + F80* + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + K112S + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + T516P + K524T + G526A; or

P2N + P4S + P11F + T65A + Q327F + E501V + N502T + Y504*; or

P2N + P4S + P11F + T65A + Q327F + E501V + Y504T; or

10 P2N + P4S + P11F + T65A + S103N + Q327F + E501V + Y504T; or

K5A + P11F + T65A + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + E501V + Y504T + T516P + K524T + G526A; or

P2N + P4S + P11F + T65A + V79A + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + V79G + Q327F + E501V + Y504T; or

15 P2N + P4S + P11F + T65A + V79I + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + V79L + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + V79S + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + L72V + Q327F + E501V + Y504T; or

S255N + Q327F + E501V + Y504T; or

20 P2N + P4S + P11F + T65A + E74N + V79K + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + G220N + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + Y245N + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q253N + Q327F + E501V + Y504T; or

P2N + P4S + P11F + T65A + D279N + Q327F + E501V + Y504T; or

25 P2N + P4S + P11F + T65A + Q327F + S359N + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + D370N + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + V460S + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + V460T + P468T + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + T463N + E501V + Y504T; or

30 P2N + P4S + P11F + T65A + Q327F + S465N + E501V + Y504T; or

P2N + P4S + P11F + T65A + Q327F + T477N + E501V + Y504T.

In a preferred embodiment the *Penicillium oxalicum* glucoamylase variant has a K79V substitution (using SEQ ID NO: 23 herein for numbering), corresponding to the PE001 variant, and further comprises one of the following mutations:

35 P11F + T65A + Q327F; or

P2N + P4S + P11F + T65A + Q327F; or

P11F + D26C + K33C + T65A + Q327F; or

P2N + P4S + P11F + T65A + Q327W + E501V + Y504T; or
 P2N + P4S + P11F + T65A + Q327F + E501V + Y504T; or
 P11F + T65A + Q327W + E501V + Y504T.

The glucoamylase may be added in amounts from 0.1- 100 micrograms EP/g, such as
 5 0.5-50 micrograms EP/g, such as 1-25 micrograms EP/g, such as 2-12 micrograms EP/g DS.

Trehalase Present And/Or Added in Saccharification and/or Fermentation

According to the process of the invention a trehalase of the invention is present and/or
 added during the

- 10 - saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

In a preferred embodiment the mature trehalase disclosed in SEQ ID NO: 30 or SEQ ID
 15 NO: 2. In a preferred embodiment the mature trehalase disclosed in SEQ ID NO: 4. In a
 preferred embodiment the trehalase is present and/or added in an amount between 0.01-20 ug
 EP (Enzyme Protein) trehalase/g DS, such as between 0.05-15 ug EP terhalase/g DS, such as
 between 0.5 and 10 ug EP trehalase/g DS.

20 Glucoamylase Present And/Or Added in Saccharification and/or Fermentation

The glucoamylase present and/or added_during saccharification step (b); fermentation
 step (c); simultaneous saccharification and fermentation; or presaccharification before step (b),
 may be derived from any suitable source, e.g., derived from a microorganism or a plant.
 Preferred glucoamylases are of fungal or bacterial origin, selected from the group consisting of
 25 *Aspergillus* glucoamylases, in particular *Aspergillus niger* G1 or G2 glucoamylase (Boel et al.
 (1984), EMBO J. 3 (5), p. 1097-1102), or variants thereof, such as those disclosed in WO
 92/00381, WO 00/04136 and WO 01/04273 (from Novozymes, Denmark); the *A. awamori*
 glucoamylase disclosed in WO 84/02921, *Aspergillus oryzae* glucoamylase (Agric. Biol. Chem.
 (1991), 55 (4), p. 941-949), or variants or fragments thereof. Other *Aspergillus* glucoamylase
 30 variants include variants with enhanced thermal stability: G137A and G139A (Chen et al.
 (1996), Prot. Eng. 9, 499-505); D257E and D293E/Q (Chen et al. (1995), Prot. Eng. 8, 575-
 582); N182 (Chen et al. (1994), Biochem. J. 301, 275-281); disulphide bonds, A246C (Fierobe
 et al. (1996), Biochemistry, 35, 8698-8704; and introduction of Pro residues in position A435
 and S436 (Li et al. (1997), Protein Eng. 10, 1199-1204).

35 Other glucoamylases include *Athelia rolfsii* (previously denoted *Corticium rolfsii*)
 glucoamylase (see US patent no. 4,727,026 and (Nagasaka et al. (1998) "Purification and
 properties of the raw-starch-degrading glucoamylases from *Corticium rolfsii*, Appl Microbiol

Biotechnol 50:323-330), *Talaromyces* glucoamylases, in particular derived from *Talaromyces emersonii* (WO 99/28448), *Talaromyces leycettanus* (US patent no. Re. 32,153), *Talaromyces duponti*, *Talaromyces thermophilus* (US patent no. 4,587,215). In a preferred embodiment the glucoamylase used during saccharification and/or fermentation is the *Talaromyces emersonii* glucoamylase disclosed in WO 99/28448.

Bacterial glucoamylases contemplated include glucoamylases from the genus *Clostridium*, in particular *C. thermoamylolyticum* (EP 135,138), and *C. thermohydrosulfuricum* (WO 86/01831).

Contemplated fungal glucoamylases include *Trametes cingulate* (SEQ ID NO: 20), *Pachykytospora papyracea*; and *Leucopaxillus giganteus* all disclosed in WO 2006/069289; or *Peniophora rufomarginata* disclosed in WO2007/124285; or a mixture thereof. Also hybrid glucoamylase are contemplated according to the invention. Examples include the hybrid glucoamylases disclosed in WO 2005/045018. Specific examples include the hybrid glucoamylase disclosed in Table 1 and 4 of Example 1 (which hybrids are hereby incorporated by reference).

In an embodiment the glucoamylase is derived from a strain of the genus *Pycnoporus*, in particular a strain of *Pycnoporus sanguineus* as described in WO 2011/066576 (SEQ ID NOs 2, 4 or 6), in particular the one shown a SEQ ID NO: 28 herein (corresponding to SEQ ID NO: 4 in WO 2011/066576) or from a strain of the genus *Gloeophyllum*, such as a strain of *Gloeophyllum sepiarium* or *Gloeophyllum trabeum*, in particular a strain of *Gloeophyllum* as described in WO 2011/068803 (SEQ ID NO: 2, 4, 6, 8, 10, 12, 14 or 16). In a preferred embodiment the glucoamylase is SEQ ID NO: 2 in WO 2011/068803 or SEQ ID NO: 19 herein (i.e. *Gloeophyllum sepiarium* glucoamylase). In a preferred embodiment the glucoamylase is SEQ ID NO: 20 herein (i.e., *Gloeophyllum trabeum* glucoamylase discloses as SEQ ID NO: 3 in WO2014/177546) (all references hereby incorporated by reference).

Contemplated are also glucoamylases which exhibit a high identity to any of the above mentioned glucoamylases, i.e., at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to any one of the mature parts of the enzyme sequences mentioned above, such as any of SEQ ID NOs: 13, 14, 19, 20 or 28 herein, respectively.

In an embodiment the glucoamylase used in fermentation or SSF exhibits at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to the mature part of SEQ ID NO: 20 herein.

In an embodiment the glucoamylase used in fermentation or SSF exhibits at least 60%, such as at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at

least 96%, at least 97%, at least 98%, at least 99% or even 100% identity to the mature part of SEQ ID NO: 28 herein.

5 Glucoamylases may in an embodiment be added to the saccharification and/or fermentation in an amount of 0.0001-20 AGU/g DS, preferably 0.001-10 AGU/g DS, especially between 0.01-5 AGU/g DS, such as 0.1-2 AGU/g DS.

Glucoamylases may in an embodiment be added to the saccharification and/or fermentation in an amount of 1-1,000 µg EP/g DS, preferably 10-500 µg/gDS, especially between 25-250 µg/g DS.

10 In an embodiment the glucoamylase is added as a blend further comprising an alpha-amylase. In a preferred embodiment the alpha-amylase is a fungal alpha-amylase, especially an acid fungal alpha-amylase. The alpha-amylase is typically a side activity.

In an embodiment the glucoamylase is a blend comprising *Talaromyces emersonii* glucoamylase disclosed in WO 99/28448 as SEQ ID NO: 34 or SEQ ID NO: 13 herein and *Trametes cingulata* glucoamylase disclosed as SEQ ID NO: 2 in WO 06/069289 and SEQ ID NO: 14 herein.

15 In an embodiment the glucoamylase is a blend comprising *Talaromyces emersonii* glucoamylase disclosed in SEQ ID NO: 13 herein, *Trametes cingulata* glucoamylase disclosed as SEQ ID NO: 14 herein, and *Rhizomucor pusillus* alpha-amylase with *Aspergillus niger* glucoamylase linker and SBD disclosed as V039 in Table 5 in WO 2006/069290 or SEQ ID NO: 15 herein.

20 In an embodiment the glucoamylase is a blend comprising *Gloeophyllum sepiarium* glucoamylase shown as SEQ ID NO: 19 herein and *Rhizomucor pusillus* with an *Aspergillus niger* glucoamylase linker and starch-binding domain (SBD), disclosed SEQ ID NO: 15 herein with the following substitutions: G128D+D143N.

25 In an embodiment the alpha-amylase may be derived from a strain of the genus *Rhizomucor*, preferably a strain the *Rhizomucor pusillus*, such as the one shown in SEQ ID NO: 3 in WO2013/006756, or the genus *Meripilus*, preferably a strain of *Meripilus giganteus*. In a preferred embodiment the alpha-amylase is derived from a *Rhizomucor pusillus* with an *Aspergillus niger* glucoamylase linker and starch-binding domain (SBD), disclosed as V039 in Table 5 in WO 2006/069290 or SEQ ID NO: 15 herein.

30 In an embodiment the *Rhizomucor pusillus* alpha-amylase or the *Rhizomucor pusillus* alpha-amylase with a linker and starch-binding domain (SBD), preferably *Aspergillus niger* glucoamylase linker and SBD, has at least one of the following substitutions or combinations of substitutions: D165M; Y141W; Y141R; K136F; K192R; P224A; P224R; S123H+Y141W; G20S + Y141W; A76G + Y141W; G128D + Y141W; G128D + D143N; P219C + Y141W; N142D + D143N; Y141W + K192R; Y141W + D143N; Y141W + N383R; Y141W + P219C + A265C; Y141W + N142D + D143N; Y141W + K192R V410A; G128D + Y141W + D143N; Y141W +

D143N + P219C; Y141W + D143N + K192R; G128D + D143N + K192R; Y141W + D143N + K192R + P219C; G128D + Y141W + D143N + K192R; or G128D + Y141W + D143N + K192R + P219C (using SEQ ID NO: 3 in WO 2013/006756 for numbering or SEQ ID NO: 15 herein).

In a preferred embodiment the glucoamylase blend comprises *Gloeophyllum sepiarium* glucoamylase (e.g., SEQ ID NO: 2 in WO 2011/068803 or SEQ ID NO: 19 herein) and *Rhizomucor pusillus* alpha-amylase.

In a preferred embodiment the glucoamylase blend comprises *Gloeophyllum sepiarium* glucoamylase shown as SEQ ID NO: 2 in WO 2011/068803 or SEQ ID NO: 19 herein and *Rhizomucor pusillus* with a linker and starch-binding domain (SBD), preferably *Aspergillus niger* glucoamylase linker and SBD, disclosed SEQ ID NO: 3 in WO 2013/006756 and SEQ ID NO: 15 herein with the following substitutions: G128D+D143N.

Commercially available compositions comprising glucoamylase include AMG 200L; AMG 300 L; SAN™ SUPER, SAN™ EXTRA L, SPIRIZYME™ PLUS, SPIRIZYME™ FUEL, SPIRIZYME™ B4U, SPIRIZYME™ ULTRA, SPIRIZYME™ EXCEL, SPIRIZYME ACHIEVE™, and AMG™ E (from Novozymes A/S); OPTIDEX™ 300, GC480, GC417 (from DuPont-Danisco); AMIGASE™ and AMIGASE™ PLUS (from DSM); G-ZYME™ G900, G-ZYME™ and G990 ZR (from DuPont-Danisco).

Cellulolytic Enzyme Composition Present And/Or Added In Saccharification And/Or Fermentation

According to the invention a cellulolytic enzyme composition may be present in saccharification, fermentation or simultaneous saccharification and fermentation (SSF).

The cellulolytic enzyme composition comprises a beta-glucosidase, a cellobiohydrolase and an endoglucanase.

Examples of suitable cellulolytic composition can be found in WO 2008/151079 and WO 2013/028928 which are incorporated by reference.

In preferred embodiments the cellulolytic enzyme composition is derived from a strain of *Trichoderma*, *Humicola*, or *Chrysosporium*.

In an embodiment the cellulolytic enzyme composition is derived from a strain of *Trichoderma reesei*, *Humicola insolens* and/or *Chrysosporium lucknowense*.

In an embodiment the cellulolytic enzyme composition comprises a beta-glucosidase, preferably one derived from a strain of the genus *Aspergillus*, such as *Aspergillus oryzae*, such as the one disclosed in WO 2002/095014 or the fusion protein having beta-glucosidase activity disclosed in WO 2008/057637 (in particular the *Aspergillus oryzae* beta-glucosidase variant fusion protein shown in SEQ ID NOs: 73 and 74, respectively, in WO 2008/057637 or the *Aspergillus oryzae* beta-glucosidase fusion protein shown in SEQ ID NOs: 75 and 76, respectively, in WO 2008/057637 – both hereby incorporated by reference), or *Aspergillus*

fumigatus, such as one disclosed in WO 2005/047499 or SEQ ID NO: 10 herein or an *Aspergillus fumigatus* beta-glucosidase variant disclosed in WO 2012/044915 (Novozymes), such as one with one or more, such as all, of the following substitutions: F100D, S283G, N456E, F512Y; or a strain of the genus a strain *Penicillium*, such as a strain of the *Penicillium* *brasiliense* disclosed in WO 2007/019442, or a strain of the genus *Trichoderma*, such as a strain of *Trichoderma reesei*.

In an embodiment the cellulolytic enzyme composition comprises a GH61 polypeptide having cellulolytic enhancing activity such as one derived from the genus *Thermoascus*, such as a strain of *Thermoascus aurantiacus*, such as the one described in WO 2005/074656 as SEQ ID NO: 2 or SEQ ID NO: 21 herein; or one derived from the genus *Thielavia*, such as a strain of *Thielavia terrestris*, such as the one described in WO 2005/074647 as SEQ ID NO: 7 and SEQ ID NO: 8; or one derived from a strain of *Aspergillus*, such as a strain of *Aspergillus fumigatus*, such as the one described in WO 2010/138754 as SEQ ID NO: 1 and SEQ ID NO: 2; or one derived from a strain derived from *Penicillium*, such as a strain of *Penicillium emersonii*, such as the one disclosed in WO 2011/041397 as SEQ ID NO: 2 or SEQ ID NO: 12 herein.

In an embodiment the cellulolytic composition comprises a cellobiohydrolase I (CBH I), such as one derived from a strain of the genus *Aspergillus*, such as a strain of *Aspergillus fumigatus*, such as the Cel7a CBH I disclosed in SEQ ID NO: 2 in WO 2011/057140 or SEQ ID NO: 6 herein, or a strain of the genus *Trichoderma*, such as a strain of *Trichoderma reesei*.

In an embodiment the cellulolytic composition comprises a cellobiohydrolase II (CBH II), such as one derived from a strain of the genus *Aspergillus*, such as a strain of *Aspergillus fumigatus* disclosed as SEQ ID NO: 8 herein; or a strain of the genus *Trichoderma*, such as *Trichoderma reesei*, or a strain of the genus *Thielavia*, such as a strain of *Thielavia terrestris*, such as cellobiohydrolase II CEL6A from *Thielavia terrestris*.

In an embodiment the cellulolytic enzyme composition comprises a GH61 polypeptide having cellulolytic enhancing activity and a beta-glucosidase.

In an embodiment the cellulolytic composition comprises a GH61 polypeptide having cellulolytic enhancing activity, a beta-glucosidase, and a CBH I.

In an embodiment the cellulolytic composition comprises a GH61 polypeptide having cellulolytic enhancing activity, a beta-glucosidase, a CBH I, and a CBH II.

In an embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* GH61A polypeptide having cellulolytic enhancing activity (SEQ ID NO: 2 in WO 2005/074656 or SEQ ID NO: 21 herein), and *Aspergillus oryzae* beta-glucosidase fusion protein (WO 2008/057637).

In an embodiment the cellulolytic composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* GH61A polypeptide having

cellulolytic enhancing activity (SEQ ID NO: 2 in WO 2005/074656 or SEQ ID NO: 21 herein) and *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 of WO 2005/047499 or SEQ ID NO: 10 herein).

In an embodiment the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition further comprising *Penicillium emersonii* GH61A polypeptide having cellulolytic enhancing activity disclosed as SEQ ID NO: 2 in WO 2011/041397 or SEQ ID NO: 12 herein and *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 of WO 2005/047499 or SEQ ID NO: 10 herein) or a variant thereof with the following substitutions F100D, S283G, N456E, F512Y (using SEQ ID NO: 10 herein for numbering).

In a preferred embodiment the cellulolytic enzyme composition comprising one or more of the following components:

- (i) an *Aspergillus fumigatus* cellobiohydrolase I;
- (ii) an *Aspergillus fumigatus* cellobiohydrolase II;
- (iii) an *Aspergillus fumigatus* beta-glucosidase or variant thereof; and
- (iv) a *Penicillium* sp. GH61 polypeptide having cellulolytic enhancing activity; or homologs thereof.

In an preferred embodiment the cellulolytic enzyme composition is derived from *Trichoderma reesei* comprising GH61A polypeptide having cellulolytic enhancing activity derived from a strain of *Penicillium emersonii* (SEQ ID NO: 2 in WO 2011/041397 or SEQ ID NO: 12 herein), *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 in WO 2005/047499 or SEQ ID NO: 10 herein) variant with the following substitutions: F100D, S283G, N456E, F512Y (disclosed in WO 2012/044915); *Aspergillus fumigatus* Cel7A CBH I disclosed as SEQ ID NO: 6 in WO2011/057140 or SEQ ID NO: 6 herein and *Aspergillus fumigatus* CBH II disclosed as SEQ ID NO: 18 in WO 2011/057140 or SEQ ID NO: 8 herein.

In an embodiment the cellulolytic composition is dosed from 0.0001-3 mg EP/g DS, preferably, 0.0005-2 mg EP/g DS, preferably 0.001-1 mg/g DS, more preferably 0.005-0.5 mg EP/g DS, and even more preferably 0.01-0.1 mg EP/g DS.

Proteases Present and/or Added in Saccharification and/or Fermentation

Any suitable protease may be added in saccharification and/or fermentation, such as SSF.

In a preferred embodiment the protease is a metallo protease or a serine protease.

In an embodiment the enzyme composition comprises a metallo protease, preferably derived from a strain of the genus *Thermoascus*, preferably a strain of *Thermoascus aurantiacus*, especially *Thermoascus aurantiacus* CGMCC No. 0670, such as the metallo protease disclosed as the mature part of SEQ ID NO: 2 disclosed in WO 2003/048353 or the mature polypeptide of SEQ ID NO: 17 herein.

In an embodiment the protease has at least 60%, such as at least 70%, such as at least 75% identity preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the mature part of the polypeptide of SEQ ID NO: 17 herein.

In an embodiment the protease is derived from a strain of *Pyrococcus*, such as a strain of *Pyrococcus furiosus*, such as the protease shown in SEQ ID NO: 1 in US 6,358,726 or SEQ ID NO: 22 herein.

In an embodiment the protease is the mature sequence from *Meripilus giganteus* protease 3 (peptidase family S53 protease) concerned in Example 2 in WO 2014/037438 (hereby incorporated by reference). In an embodiment the protease is the mature protease 3 sequence from a strain of *Meripilus*, in particular *Meripilus giganteus* shown as SEQ ID NO: 5 in WO 2014/037438 (hereby incorporated by reference) and SEQ ID NO: 32 herein.

In an embodiment the protease has at least 60%, such as at least 70%, such as at least 75% identity preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the mature part of the polypeptide of SEQ ID NO: 32 herein shown as amino acids 1-547.

Alpha-amylase Present and/or Added in Saccharification and/or Fermentation

Any suitable alpha-amylase, such as fungal acid alpha-amylase, may be present and/or added in saccharification and/or fermentation.

In a preferably embodiment the alpha-amylase is a fungal alpha-amylase, in particular one that has at least 60%, such as at least 70%, such as at least 75% identity preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at least 99%, or 100% identity to the mature part of the polypeptide of SEQ ID NO: 15. In a preferred embodiment the alpha-amylase has one or more of the following substitutions: G128D, D143N, in particular G128D+D143N.

Processes of Producing a Fermentation Product From Cellulosic Materials Using a Trehalase of the Invention

In an embodiment the invention relates to processes of producing a fermentation product from pretreated cellulosic material, comprising:

(a) hydrolyzing said pretreated cellulosic material with a cellulolytic enzyme composition;

(b) fermenting using a fermenting organism; and

(c) optionally recovering the fermentation product,

5 wherein a trehalase of the invention is added and/or present in hydrolysis step (a) and/or fermentation step (b).

According to the process of the invention hydrolysis and fermentation may be carried out separate or simultaneous. In an embodiment the process of the invention is carried out as separate hydrolysis and fermentation (SHF); simultaneous saccharification and fermentation
10 (SSF); simultaneous saccharification and co-fermentation (SSCF); hybrid hydrolysis and fermentation (HHF); separate hydrolysis and co-fermentation (SHCF); hybrid hydrolysis and co-fermentation (HHCF); or direct microbial conversion (DMC), also sometimes called consolidated bioprocessing (CBP). SHF uses separate process steps to first enzymatically hydrolyze the cellulosic material to fermentable sugars, e.g., glucose, cellobiose, and pentose monomers,
15 and then ferment the fermentable sugars to ethanol. In SSF, the enzymatic hydrolysis of the cellulosic material and the fermentation of sugars to ethanol are combined in one step. SSCF involves the co-fermentation of multiple sugars (Sheehan, J., and Himmel, M., 1999, *Enzymes, energy and the environment: A strategic perspective on the U.S. Department of Energy's research and development activities for bioethanol, Biotechnol. Prog.* 15: 817-827). HHF
20 involves a separate hydrolysis step, and in addition a simultaneous saccharification and hydrolysis step, which can be carried out in the same reactor. The steps in a HHF process can be carried out at different temperatures, i.e., high temperature enzymatic hydrolysis followed by SSF at a lower temperature that the fermentation strain can tolerate. DMC combines all three processes (enzyme production, hydrolysis, and fermentation) in one or more (e.g., several)
25 steps where the same organism is used to produce the enzymes for conversion of the cellulosic material to fermentable sugars and to convert the fermentable sugars into a final product.

According to the invention the cellulosic material is plant material chips, plant stem segments and/or whole plant stems. In an embodiment cellulosic material is selected from the group comprising arundo, bagasse, bamboo, corn cob, corn fiber, corn stover, miscanthus,
30 orange peel, rice straw, switchgrass, wheat straw. In a preferred embodiment the source of the cellulosic material is corn stover, corn cobs, and/or wheat straw.

According to the invention any pretreatment may be used. In a preferred embodiment chemical pretreatment, physical pretreatment, or chemical pretreatment and a physical pretreatment is used. In a preferred embodiment the cellulosic material is pretreated with an
35 acid, such as dilute acid pretreatment. In an embodiment the cellulosic material is prepared by pretreating cellulosic material at high temperature, high pressure with an acid.

In an embodiment hydrolysis is carried out at a temperature between 20-70°C, such as 30-60°C, preferably 45-55°C at a pH in the range 4-6, such as 4.5-5.5.

In an embodiment the cellulosic material is present at 1-20 (w/w) % of TS, such as 2-10 (w/w)% TS, such as around 5 (w/w) % TS during hydrolysis.

5 In an embodiment the hydrolysis is carried out for 1-20 days, preferably between from 5-15 days.

In an embodiment the cellulolytic enzyme composition is derived from *Trichoderma reesei*, *Humicola insolens* or *Chrysosporium lucknowense*.

Cellulolytic enzyme composition: The term “cellulolytic enzyme composition” means
10 one or more (e.g., several) enzymes that hydrolyze a cellulosic material. Such enzymes include endoglucanase(s), cellobiohydrolase(s), beta-glucosidase(s), or combinations thereof. The two basic approaches for measuring cellulolytic activity include: (1) measuring the total cellulolytic activity, and (2) measuring the individual cellulolytic activities (endoglucanases, cellobiohydrolases, and beta-glucosidases) as reviewed in Zhang *et al.*, Outlook for cellulase
15 improvement: Screening and selection strategies, 2006, *Biotechnology Advances* 24: 452-481. Total cellulolytic activity is usually measured using insoluble substrates, including Whatman №1 filter paper, microcrystalline cellulose, bacterial cellulose, algal cellulose, cotton, pretreated lignocellulose, *etc.* The most common total cellulolytic activity assay is the filter paper assay using Whatman №1 filter paper as the substrate. The assay was established by the
20 International Union of Pure and Applied Chemistry (IUPAC) (Ghose, 1987, Measurement of cellulase activities, *Pure Appl. Chem.* 59: 257-68).

For purposes of the present invention, cellulolytic enzyme activity is determined by measuring the increase in hydrolysis of a cellulosic material by cellulolytic enzyme(s) under the following conditions: 1-50 mg of cellulolytic enzyme protein/g of cellulose in PCS (or other
25 pretreated cellulosic material) for 3-7 days at a suitable temperature, e.g., 50°C, 55°C, or 60°C, compared to a control hydrolysis without addition of cellulolytic enzyme protein. Typical conditions are 1 ml reactions, washed or unwashed PCS, 5% insoluble solids, 50 mM sodium acetate pH 5, 1 mM MnSO₄, 50°C, 55°C, or 60°C, 72 hours, sugar analysis by AMINEX® HPX-87H column (Bio-Rad Laboratories, Inc., Hercules, CA, USA).

30 Examples of cellulolytic compositions can be found in the “Cellulolytic Enzyme Composition present and/or added during Saccharification and/or Fermentation”-section above.

Cellulosic material: The term “cellulosic material” means any material containing cellulose. The predominant polysaccharide in the primary cell wall of cellulosic material is cellulose, the second most abundant is hemicellulose, and the third is pectin. The secondary
35 cell wall, produced after the cell has stopped growing, also contains polysaccharides and is strengthened by polymeric lignin covalently cross-linked to hemicellulose. Cellulose is a homopolymer of anhydrocellobiose and thus a linear beta-(1-4)-D-glucan, while hemicelluloses

include a variety of compounds, such as xylans, xyloglucans, arabinoxylans, and mannans in complex branched structures with a spectrum of substituents. Although generally polymorphous, cellulose is found in plant tissue primarily as an insoluble crystalline matrix of parallel glucan chains. Hemicelluloses usually hydrogen bond to cellulose, as well as to other hemicelluloses, which help stabilize the cell wall matrix.

Cellulose is generally found, for example, in the stems, leaves, hulls, husks, and cobs of plants or leaves, branches, and wood of trees. The cellulosic material can be, but is not limited to, agricultural residue, herbaceous material (including energy crops), municipal solid waste, pulp and paper mill residue, waste paper, and wood (including forestry residue) (see, for example, Wiseloge *et al.*, 1995, in Handbook on Bioethanol (Charles E. Wyman, editor), pp. 105-118, Taylor & Francis, Washington D.C.; Wyman, 1994, *Bioresource Technology* 50: 3-16; Lynd, 1990, *Applied Biochemistry and Biotechnology* 24/25: 695-719; Mosier *et al.*, 1999, Recent Progress in Bioconversion of Lignocellulosics, in *Advances in Biochemical Engineering/Biotechnology*, T. Scheper, managing editor, Volume 65, pp.23-40, Springer-Verlag, New York). It is understood herein that the cellulose may be in the form of lignocellulose, a plant cell wall material containing lignin, cellulose, and hemicellulose in a mixed matrix. In a preferred aspect, the cellulosic material is any biomass material. In another preferred aspect, the cellulosic material is lignocellulose, which comprises cellulose, hemicelluloses, and lignin.

In one aspect, the cellulosic material is agricultural residue. In another aspect, the cellulosic material is herbaceous material (including energy crops). In another aspect, the cellulosic material is municipal solid waste. In another aspect, the cellulosic material is pulp and paper mill residue. In another aspect, the cellulosic material is waste paper. In another aspect, the cellulosic material is wood (including forestry residue).

In another aspect, the cellulosic material is arundo. In another aspect, the cellulosic material is bagasse. In another aspect, the cellulosic material is bamboo. In another aspect, the cellulosic material is corn cob. In another aspect, the cellulosic material is corn fiber. In another aspect, the cellulosic material is corn stover. In another aspect, the cellulosic material is miscanthus. In another aspect, the cellulosic material is orange peel. In another aspect, the cellulosic material is rice straw. In another aspect, the cellulosic material is switchgrass. In another aspect, the cellulosic material is wheat straw.

In another aspect, the cellulosic material is aspen. In another aspect, the cellulosic material is eucalyptus. In another aspect, the cellulosic material is fir. In another aspect, the cellulosic material is pine. In another aspect, the cellulosic material is poplar. In another aspect, the cellulosic material is spruce. In another aspect, the cellulosic material is willow.

In another aspect, the cellulosic material is algal cellulose. In another aspect, the cellulosic material is bacterial cellulose. In another aspect, the cellulosic material is cotton linter.

In another aspect, the cellulosic material is filter paper. In another aspect, the cellulosic material is microcrystalline cellulose. In another aspect, the cellulosic material is phosphoric-acid treated cellulose.

In another aspect, the cellulosic material is an aquatic biomass. As used herein the term "aquatic biomass" means biomass produced in an aquatic environment by a photosynthesis process. The aquatic biomass can be algae, emergent plants, floating-leaf plants, or submerged plants.

The cellulosic material may be used as is or may be subjected to pretreatment, using conventional methods known in the art, as described herein. In a preferred aspect, the cellulosic material is pretreated.

Beta-glucosidase: The term "beta-glucosidase" means a beta-D-glucoside glucohydrolase (E.C. 3.2.1.21) that catalyzes the hydrolysis of terminal non-reducing beta-D-glucose residues with the release of beta-D-glucose. For purposes of the present invention, beta-glucosidase activity is determined using *p*-nitrophenyl-beta-D-glucopyranoside as substrate according to the procedure of Venturi *et al.*, 2002, Extracellular beta-D-glucosidase from *Chaetomium thermophilum* var. *coprophilum*: production, purification and some biochemical properties, *J. Basic Microbiol.* 42: 55-66. One unit of beta-glucosidase is defined as 1.0 μ mole of *p*-nitrophenolate anion produced per minute at 25°C, pH 4.8 from 1 mM *p*-nitrophenyl-beta-D-glucopyranoside as substrate in 50 mM sodium citrate containing 0.01% TWEEN® 20 (polyoxyethylene sorbitan monolaurate).

Beta-xylosidase: The term "beta-xylosidase" means a beta-D-xyloside xylohydrolase (E.C. 3.2.1.37) that catalyzes the exo-hydrolysis of short beta (1→4)-xylooligosaccharides to remove successive D-xylose residues from non-reducing termini. For purposes of the present invention, one unit of beta-xylosidase is defined as 1.0 μ mole of *p*-nitrophenolate anion produced per minute at 40°C, pH 5 from 1 mM *p*-nitrophenyl-beta-D-xyloside as substrate in 100 mM sodium citrate containing 0.01% TWEEN® 20.

Cellobiohydrolase: The term "cellobiohydrolase" means a 1,4-beta-D-glucan cellobiohydrolase (E.C. 3.2.1.91) that catalyzes the hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, celooligosaccharides, or any beta-1,4-linked glucose containing polymer, releasing cellobiose from the reducing or non-reducing ends of the chain (Teeri, 1997, Crystalline cellulose degradation: New insight into the function of cellobiohydrolases, *Trends in Biotechnology* 15: 160-167; Teeri *et al.*, 1998, *Trichoderma reesei* cellobiohydrolases: why so efficient on crystalline cellulose?, *Biochem. Soc. Trans.* 26: 173-178). Cellobiohydrolase activity is determined according to the procedures described by Lever *et al.*, 1972, *Anal. Biochem.* 47: 273-279; van Tilbeurgh *et al.*, 1982, *FEBS Letters* 149: 152-156; van Tilbeurgh and Claeyssens, 1985, *FEBS Letters* 187: 283-288; and Tomme *et al.*, 1988, *Eur. J. Biochem.* 170:

575-581. In the present invention, the Tomme *et al.* method can be used to determine cellobiohydrolase activity.

Endoglucanase: The term “endoglucanase” means an endo-1,4-(1,3;1,4)-beta-D-glucan 4-glucanohydrolase (E.C. 3.2.1.4) that catalyzes endohydrolysis of 1,4-beta-D-glycosidic linkages in cellulose, cellulose derivatives (such as carboxymethyl cellulose and hydroxyethyl cellulose), lichenin, beta-1,4 bonds in mixed beta-1,3 glucans such as cereal beta-D-glucans or xyloglucans, and other plant material containing cellulosic components. Endoglucanase activity can be determined by measuring reduction in substrate viscosity or increase in reducing ends determined by a reducing sugar assay (Zhang *et al.*, 2006, *Biotechnology Advances* 24: 452-481). For purposes of the present invention, endoglucanase activity is determined using carboxymethyl cellulose (CMC) as substrate according to the procedure of Ghose, 1987, *Pure and Appl. Chem.* 59: 257-268, at pH 5, 40°C.

Family 61 glycoside hydrolase: The term “Family 61 glycoside hydrolase” or “Family GH61” or “GH61” means a polypeptide falling into the glycoside hydrolase Family 61 according to Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, *Biochem. J.* 280: 309-316, and Henrissat and Bairoch, 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696. The enzymes in this family were originally classified as a glycoside hydrolase family based on measurement of very weak endo-1,4-beta-D-glucanase activity in one family member. The structure and mode of action of these enzymes are non-canonical and they cannot be considered as bona fide glycosidases. However, they are kept in the CAZy classification on the basis of their capacity to enhance the breakdown of cellulose when used in conjunction with a cellulase or a mixture of cellulases.

Hemicellulolytic enzyme or hemicellulase: The term “hemicellulolytic enzyme” or “hemicellulase” means one or more (e.g., several) enzymes that hydrolyze a hemicellulosic material. See, for example, Shallom and Shoham, 2003, Microbial hemicellulases. *Current Opinion In Microbiology* 6(3): 219-228). Hemicellulases are key components in the degradation of plant biomass. Examples of hemicellulases include, but are not limited to, an acetylmannan esterase, an acetylxylan esterase, an arabinanase, an arabinofuranosidase, a coumaric acid esterase, a feruloyl esterase, a galactosidase, a glucuronidase, a glucuronoyl esterase, a mannanase, a mannosidase, a xylanase, and a xylosidase. The substrates of these enzymes, the hemicelluloses, are a heterogeneous group of branched and linear polysaccharides that are bound via hydrogen bonds to the cellulose microfibrils in the plant cell wall, crosslinking them into a robust network. Hemicelluloses are also covalently attached to lignin, forming together with cellulose a highly complex structure. The variable structure and organization of hemicelluloses require the concerted action of many enzymes for its complete degradation. The catalytic modules of hemicellulases are either glycoside hydrolases (GHs) that hydrolyze

glycosidic bonds, or carbohydrate esterases (CEs), which hydrolyze ester linkages of acetate or ferulic acid side groups. These catalytic modules, based on homology of their primary sequence, can be assigned into GH and CE families. Some families, with an overall similar fold, can be further grouped into clans, marked alphabetically (e.g., GH-A). A most informative and updated classification of these and other carbohydrate active enzymes is available in the Carbohydrate-Active Enzymes (CAZy) database. Hemicellulolytic enzyme activities can be measured according to Ghose and Bisaria, 1987, *Pure & Appl. Chem.* 59: 1739-1752, at a suitable temperature, e.g., 50°C, 55°C, or 60°C, and pH, e.g., 5.0 or 5.5.

Polypeptide having cellulolytic enhancing activity: The term “polypeptide having cellulolytic enhancing activity” means a GH61 polypeptide that catalyzes the enhancement of the hydrolysis of a cellulosic material by enzyme having cellulolytic activity. For purposes of the present invention, cellulolytic enhancing activity is determined by measuring the increase in reducing sugars or the increase of the total of cellobiose and glucose from the hydrolysis of a cellulosic material by cellulolytic enzyme under the following conditions: 1-50 mg of total protein/g of cellulose in PCS, wherein total protein is comprised of 50-99.5% w/w cellulolytic enzyme protein and 0.5-50% w/w protein of a GH61 polypeptide having cellulolytic enhancing activity for 1-7 days at a suitable temperature, e.g., 50°C, 55°C, or 60°C, and pH, e.g., 5.0 or 5.5, compared to a control hydrolysis with equal total protein loading without cellulolytic enhancing activity (1-50 mg of cellulolytic protein/g of cellulose in PCS). In an aspect, a mixture of CELLUCLAST® 1.5L (Novozymes A/S, Bagsværd, Denmark) in the presence of 2-3% of total protein weight *Aspergillus oryzae* beta-glucosidase (recombinantly produced in *Aspergillus oryzae* according to WO 02/095014) or 2-3% of total protein weight *Aspergillus fumigatus* beta-glucosidase (recombinantly produced in *Aspergillus oryzae* as described in WO 2002/095014) of cellulase protein loading is used as the source of the cellulolytic activity.

The GH61 polypeptides having cellulolytic enhancing activity enhance the hydrolysis of a cellulosic material catalyzed by enzyme having cellulolytic activity by reducing the amount of cellulolytic enzyme required to reach the same degree of hydrolysis preferably at least 1.01-fold, e.g., at least 1.05-fold, at least 1.10-fold, at least 1.25-fold, at least 1.5-fold, at least 2-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 10-fold, or at least 20-fold.

Xylanase: The term “xylanase” means a 1,4-beta-D-xylan-xylohydrolase (E.C. 3.2.1.8) that catalyzes the endohydrolysis of 1,4-beta-D-xylosidic linkages in xylans. For purposes of the present invention, xylanase activity is determined with 0.2% AZCL-arabinoxylan as substrate in 0.01% TRITON® X-100 and 200 mM sodium phosphate buffer pH 6 at 37°C. One unit of xylanase activity is defined as 1.0 µmole of azurine produced per minute at 37°C, pH 6 from 0.2% AZCL-arabinoxylan as substrate in 200 mM sodium phosphate pH 6 buffer.

Fermenting Organism For Cellulosic Based Fermentation

The term “fermenting organism” or “fermenting microorganism” refers to any organism, including bacterial and fungal organisms, suitable for use in a desired fermentation process to produce a fermentation product. The fermenting organism may be hexose and/or pentose fermenting organisms, or a combination thereof. Both hexose and pentose fermenting organisms are well known in the art. Suitable fermenting microorganisms are able to ferment, *i.e.*, convert, sugars, such as glucose, xylose, xylulose, arabinose, maltose, mannose, galactose, and/or oligosaccharides, directly or indirectly into the desired fermentation product. Examples of bacterial and fungal fermenting organisms producing ethanol are described by Lin *et al.*, 2006, *Appl. Microbiol. Biotechnol.* 69: 627-642.

Examples of fermenting organisms that can ferment hexose sugars include bacterial and fungal organisms, such as yeast. Preferred yeast includes strains of *Candida*, *Kluyveromyces*, and *Saccharomyces*, *e.g.*, *Candida sonorensis*, *Kluyveromyces marxianus*, and *Saccharomyces cerevisiae*.

Examples of fermenting organisms that can ferment pentose sugars in their native state include bacterial and fungal organisms, such as some yeast. Preferred xylose fermenting yeast include strains of *Candida*, preferably *C. sheatae* or *C. sonorensis*; and strains of *Pichia*, preferably *P. stipitis*, such as *P. stipitis* CBS 5773. Preferred pentose fermenting yeast include strains of *Pachysolen*, preferably *P. tannophilus*. Organisms not capable of fermenting pentose sugars, such as xylose and arabinose, may be genetically modified to do so by methods known in the art.

Examples of bacteria that can efficiently ferment hexose and pentose to ethanol include, for example, *Bacillus coagulans*, *Clostridium acetobutylicum*, *Clostridium thermocellum*, *Clostridium phytofermentans*, *Geobacillus* sp., *Thermoanaerobacter saccharolyticum*, and *Zymomonas mobilis* (Philippidis, 1996, *supra*).

Other fermenting organisms include strains of *Bacillus*, such as *Bacillus coagulans*; *Candida*, such as *C. sonorensis*, *C. methanosorbosa*, *C. diddensiae*, *C. parapsilosis*, *C. naedodendra*, *C. blankii*, *C. entomophila*, *C. brassicae*, *C. pseudotropicalis*, *C. boidinii*, *C. utilis*, and *C. scheidtiae*; *Clostridium*, such as *C. acetobutylicum*, *C. thermocellum*, and *C. phytofermentans*; *E. coli*, especially *E. coli* strains that have been genetically modified to improve the yield of ethanol; *Geobacillus* sp.; *Hansenula*, such as *Hansenula anomala*; *Klebsiella*, such as *K. oxytoca*; *Kluyveromyces*, such as *K. marxianus*, *K. lactis*, *K. thermotolerans*, and *K. fragilis*; *Schizosaccharomyces*, such as *S. pombe*; *Thermoanaerobacter*, such as *Thermoanaerobacter saccharolyticum*; and *Zymomonas*, such as *Zymomonas mobilis*.

In a preferred aspect, the yeast is a *Bretannomyces*. In a more preferred aspect, the yeast is *Bretannomyces clausenii*. In another preferred aspect, the yeast is a *Candida*. In another more preferred aspect, the yeast is *Candida sonorensis*. In another more preferred aspect, the yeast is *Candida boidinii*. In another more preferred aspect, the yeast is *Candida*

blankii. In another more preferred aspect, the yeast is *Candida brassicae*. In another more preferred aspect, the yeast is *Candida diddensii*. In another more preferred aspect, the yeast is *Candida entomophiliia*. In another more preferred aspect, the yeast is *Candida pseudotropicalis*. In another more preferred aspect, the yeast is *Candida scehatae*. In another more preferred aspect, the yeast is *Candida utilis*. In another preferred aspect, the yeast is a *Clavispora*. In another more preferred aspect, the yeast is *Clavispora lusitaniae*. In another more preferred aspect, the yeast is *Clavispora opuntiae*. In another preferred aspect, the yeast is a *Kluyveromyces*. In another more preferred aspect, the yeast is *Kluyveromyces fragilis*. In another more preferred aspect, the yeast is *Kluyveromyces marxianus*. In another more preferred aspect, the yeast is *Kluyveromyces thermotolerans*. In another preferred aspect, the yeast is a *Pachysolen*. In another more preferred aspect, the yeast is *Pachysolen tannophilus*. In another preferred aspect, the yeast is a *Pichia*. In another more preferred aspect, the yeast is a *Pichia stipitis*. In another preferred aspect, the yeast is a *Saccharomyces* spp. In a more preferred aspect, the yeast is *Saccharomyces cerevisiae*. In another more preferred aspect, the yeast is *Saccharomyces distaticus*. In another more preferred aspect, the yeast is *Saccharomyces uvarum*.

In a preferred aspect, the bacterium is a *Bacillus*. In a more preferred aspect, the bacterium is *Bacillus coagulans*. In another preferred aspect, the bacterium is a *Clostridium*. In another more preferred aspect, the bacterium is *Clostridium acetobutylicum*. In another more preferred aspect, the bacterium is *Clostridium phytofermentans*. In another more preferred aspect, the bacterium is *Clostridium thermocellum*. In another more preferred aspect, the bacterium is *Geobacillus* sp. In another more preferred aspect, the bacterium is a *Thermoanaerobacter*. In another more preferred aspect, the bacterium is *Thermoanaerobacter saccharolyticum*. In another preferred aspect, the bacterium is a *Zymomonas*. In another more preferred aspect, the bacterium is *Zymomonas mobilis*.

Commercially available yeast suitable for ethanol production include, e.g., BIOFERM™ AFT and XR (NABC - North American Bioproducts Corporation, GA, USA), ETHANOL RED™ yeast (Fermentis/Lesaffre, USA), FALI™ (Fleischmann's Yeast, USA), FERMIOL™ (DSM Specialties), GERT STRAND™ (Gert Strand AB, Sweden), and SUPERSTART™ and THERMOSACC™ fresh yeast (Ethanol Technology, WI, USA).

In a preferred aspect, the fermenting microorganism has been genetically modified to provide the ability to ferment pentose sugars, such as xylose utilizing, arabinose utilizing, and xylose and arabinose co-utilizing microorganisms.

The cloning of heterologous genes into various fermenting microorganisms has led to the construction of organisms capable of converting hexoses and pentoses to ethanol (co-fermentation) (Chen and Ho, 1993, Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, *Appl. Biochem. Biotechnol.* 39-40: 135-

147; Ho *et al.*, 1998, Genetically engineered *Saccharomyces* yeast capable of effectively cofermenting glucose and xylose, *Appl. Environ. Microbiol.* 64: 1852-1859; Kotter and Ciriacy, 1993, Xylose fermentation by *Saccharomyces cerevisiae*, *Appl. Microbiol. Biotechnol.* 38: 776-783; Walfridsson *et al.*, 1995, Xylose-metabolizing *Saccharomyces cerevisiae* strains overexpressing the TKL1 and TAL1 genes encoding the pentose phosphate pathway enzymes transketolase and transaldolase, *Appl. Environ. Microbiol.* 61: 4184-4190; Kuyper *et al.*, 2004, Minimal metabolic engineering of *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: a proof of principle, *FEMS Yeast Research* 4: 655-664; Beall *et al.*, 1991, Parametric studies of ethanol production from xylose and other sugars by recombinant *Escherichia coli*, *Biotech. Bioeng.* 38: 296-303; Ingram *et al.*, 1998, Metabolic engineering of bacteria for ethanol production, *Biotechnol. Bioeng.* 58: 204-214; Zhang *et al.*, 1995, Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*, *Science* 267: 240-243; Deanda *et al.*, 1996, Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering, *Appl. Environ. Microbiol.* 62: 4465-4470; WO 2003/062430, xylose isomerase).

In a preferred aspect, the genetically modified fermenting microorganism is *Candida sonorensis*. In another preferred aspect, the genetically modified fermenting microorganism is *Escherichia coli*. In another preferred aspect, the genetically modified fermenting microorganism is *Klebsiella oxytoca*. In another preferred aspect, the genetically modified fermenting microorganism is *Kluyveromyces marxianus*. In another preferred aspect, the genetically modified fermenting microorganism is *Saccharomyces cerevisiae*. In another preferred aspect, the genetically modified fermenting microorganism is *Zymomonas mobilis*.

It is well known in the art that the organisms described above can also be used to produce other substances, as described herein.

The fermenting microorganism is typically added to the degraded cellulosic material or hydrolysate and the fermentation is performed for about 8 to about 96 hours, e.g., about 24 to about 60 hours. The temperature is typically between about 26°C to about 60°C, e.g., about 32°C or 50°C, and about pH 3 to about pH 8, e.g., pH 4-5, 6, or 7.

In one aspect, the yeast and/or another microorganism are applied to the degraded cellulosic material and the fermentation is performed for about 12 to about 96 hours, such as typically 24-60 hours. In another aspect, the temperature is preferably between about 20°C to about 60°C, e.g., about 25°C to about 50°C, about 32°C to about 50°C, or about 32°C to about 50°C, and the pH is generally from about pH 3 to about pH 7, e.g., about pH 4 to about pH 7. However, some fermenting organisms, e.g., bacteria, have higher fermentation temperature optima. Yeast or another microorganism is preferably applied in amounts of approximately 10^5 to 10^{12} , preferably from approximately 10^7 to 10^{10} , especially approximately 2×10^8 viable cell count per ml of fermentation broth. Further guidance in respect of using yeast for fermentation

can be found in, e.g., "The Alcohol Textbook" (Editors K. Jacques, T.P. Lyons and D.R. Kelsall, Nottingham University Press, United Kingdom 1999), which is hereby incorporated by reference.

For ethanol production, following the fermentation the fermented slurry is distilled to extract the ethanol. The ethanol obtained according to the processes of the invention can be used as, e.g., fuel ethanol, drinking ethanol, i.e., potable neutral spirits, or industrial ethanol.

A fermentation stimulator can be used in combination with any of the processes described herein to further improve the fermentation process, and in particular, the performance of the fermenting microorganism, such as, rate enhancement and ethanol yield. A "fermentation stimulator" refers to stimulators for growth of the fermenting microorganisms, in particular, yeast. Preferred fermentation stimulators for growth include vitamins and minerals. Examples of vitamins include multivitamins, biotin, pantothenate, nicotinic acid, meso-inositol, thiamine, pyridoxine, para-aminobenzoic acid, folic acid, riboflavin, and Vitamins A, B, C, D, and E. See, for example, Alfenore *et al.*, Improving ethanol production and viability of *Saccharomyces cerevisiae* by a vitamin feeding strategy during fed-batch process, Springer-Verlag (2002), which is hereby incorporated by reference. Examples of minerals include minerals and mineral salts that can supply nutrients comprising P, K, Mg, S, Ca, Fe, Zn, Mn, and Cu.

Fermentation products

According to the invention the term "fermentation product" can be any substance derived from fermentation. The fermentation product can be, without limitation, an alcohol (e.g., arabinitol, n-butanol, isobutanol, ethanol, glycerol, methanol, ethylene glycol, 1,3-propanediol [propylene glycol], butanediol, glycerin, sorbitol, and xylitol); an alkane (e.g., pentane, hexane, heptane, octane, nonane, decane, undecane, and dodecane), a cycloalkane (e.g., cyclopentane, cyclohexane, cycloheptane, and cyclooctane), an alkene (e.g. pentene, hexene, heptene, and octene); an amino acid (e.g., aspartic acid, glutamic acid, glycine, lysine, serine, and threonine); a gas (e.g., methane, hydrogen (H₂), carbon dioxide (CO₂), and carbon monoxide (CO)); isoprene; a ketone (e.g., acetone); an organic acid (e.g., acetic acid, acetic acid, adipic acid, ascorbic acid, citric acid, 2,5-diketo-D-gluconic acid, formic acid, fumaric acid, glucaric acid, gluconic acid, glucuronic acid, glutaric acid, 3-hydroxypropionic acid, itaconic acid, lactic acid, malic acid, malonic acid, oxalic acid, oxaloacetic acid, propionic acid, succinic acid, and xylonic acid); and polyketide. The fermentation product can also be protein as a high value product.

In a preferred aspect, the fermentation product is an alcohol. It will be understood that the term "alcohol" encompasses a substance that contains one or more (e.g., several) hydroxyl moieties. In a more preferred aspect, the alcohol is n-butanol. In another more preferred aspect, the alcohol is isobutanol. In another more preferred aspect, the alcohol is ethanol. In

another more preferred aspect, the alcohol is methanol. In another more preferred aspect, the alcohol is arabinitol. In another more preferred aspect, the alcohol is butanediol. In another more preferred aspect, the alcohol is ethylene glycol. In another more preferred aspect, the alcohol is glycerin. In another more preferred aspect, the alcohol is glycerol. In another more preferred aspect, the alcohol is 1,3-propanediol. In another more preferred aspect, the alcohol is sorbitol. In another more preferred aspect, the alcohol is xylitol. See, for example, Gong, C. S., Cao, N. J., Du, J., and Tsao, G. T., 1999, Ethanol production from renewable resources, in *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207-241; Silveira, M. M., and Jonas, R., 2002, The biotechnological production of sorbitol, *Appl. Microbiol. Biotechnol.* 59: 400-408; Nigam and Singh, 1995, Processes for fermentative production of xylitol – a sugar substitute, *Process Biochemistry* 30(2): 117-124; Ezeji *et al.*, 2003, Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and *in situ* recovery by gas stripping, *World Journal of Microbiology and Biotechnology* 19(6): 595-603.

In another preferred aspect, the fermentation product is an alkane. The alkane can be an unbranched or a branched alkane. In another more preferred aspect, the alkane is pentane. In another more preferred aspect, the alkane is hexane. In another more preferred aspect, the alkane is heptane. In another more preferred aspect, the alkane is octane. In another more preferred aspect, the alkane is nonane. In another more preferred aspect, the alkane is decane. In another more preferred aspect, the alkane is undecane. In another more preferred aspect, the alkane is dodecane.

In another preferred aspect, the fermentation product is a cycloalkane. In another more preferred aspect, the cycloalkane is cyclopentane. In another more preferred aspect, the cycloalkane is cyclohexane. In another more preferred aspect, the cycloalkane is cycloheptane. In another more preferred aspect, the cycloalkane is cyclooctane.

In another preferred aspect, the fermentation product is an alkene. The alkene can be an unbranched or a branched alkene. In another more preferred aspect, the alkene is pentene. In another more preferred aspect, the alkene is hexene. In another more preferred aspect, the alkene is heptene. In another more preferred aspect, the alkene is octene.

In another preferred aspect, the fermentation product is an amino acid. In another more preferred aspect, the organic acid is aspartic acid. In another more preferred aspect, the amino acid is glutamic acid. In another more preferred aspect, the amino acid is glycine. In another more preferred aspect, the amino acid is lysine. In another more preferred aspect, the amino acid is serine. In another more preferred aspect, the amino acid is threonine. See, for example, Richard and Margaritis, 2004, Empirical modeling of batch fermentation kinetics for poly(glutamic acid) production and other microbial biopolymers, *Biotechnology and Bioengineering* 87(4): 501-515.

In another preferred aspect, the fermentation product is a gas. In another more preferred aspect, the gas is methane. In another more preferred aspect, the gas is H₂. In another more preferred aspect, the gas is CO₂. In another more preferred aspect, the gas is CO. See, for example, Kataoka, Miya, and Kiriya, 1997, Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria, *Water Science and Technology* 36(6-7): 41-47; and Gunaseelan, 1997, Anaerobic digestion of biomass for methane production: A review, *Biomass and Bioenergy*, 13(1-2): 83-114.

In another preferred aspect, the fermentation product is isoprene.

In another preferred aspect, the fermentation product is a ketone. It will be understood that the term "ketone" encompasses a substance that contains one or more (e.g., several) ketone moieties. In another more preferred aspect, the ketone is acetone. See, for example, Qureshi and Blaschek, 2003, *supra*.

In another preferred aspect, the fermentation product is an organic acid. In another more preferred aspect, the organic acid is acetic acid. In another more preferred aspect, the organic acid is acetic acid. In another more preferred aspect, the organic acid is adipic acid. In another more preferred aspect, the organic acid is ascorbic acid. In another more preferred aspect, the organic acid is citric acid. In another more preferred aspect, the organic acid is 2,5-diketo-D-gluconic acid. In another more preferred aspect, the organic acid is formic acid. In another more preferred aspect, the organic acid is fumaric acid. In another more preferred aspect, the organic acid is glucaric acid. In another more preferred aspect, the organic acid is gluconic acid. In another more preferred aspect, the organic acid is glucuronic acid. In another more preferred aspect, the organic acid is glutaric acid. In another preferred aspect, the organic acid is 3-hydroxypropionic acid. In another more preferred aspect, the organic acid is itaconic acid. In another more preferred aspect, the organic acid is lactic acid. In another more preferred aspect, the organic acid is malic acid. In another more preferred aspect, the organic acid is malonic acid. In another more preferred aspect, the organic acid is oxalic acid. In another more preferred aspect, the organic acid is propionic acid. In another more preferred aspect, the organic acid is succinic acid. In another more preferred aspect, the organic acid is xylonic acid. See, for example, Chen and Lee, 1997, Membrane-mediated extractive fermentation for lactic acid production from cellulosic biomass, *Appl. Biochem. Biotechnol.* 63-65: 435-448.

Recovery

The fermentation product(s) are optionally recovered after fermentation using any method known in the art including, but not limited to, chromatography, electrophoretic procedures, differential solubility, distillation, or extraction. For example, alcohol, such as ethanol, is separated from the fermented material and purified by conventional methods of

distillation. Ethanol with a purity of up to about 96 vol. % can be obtained, which can be used as, for example, fuel ethanol, drinking ethanol, *i.e.*, potable neutral spirits, or industrial ethanol.

The present invention is further described by the following examples that should not be construed as limiting the scope of the invention.

Examples

The invention described and claimed herein is not to be limited in scope by the specific aspects herein disclosed, since these aspects are intended as illustrations of several aspects of the invention. Any equivalent aspects are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In the case of conflict, the present disclosure including definitions will control.

MATERIALS & METHODS

Enzymes:

Trehalase (Ms37 tr or Ms trehalase) (P33WJF) from *Myceliophthora sepeodonium* is shown as amino acids 21-697 in SEQ ID NO: 30 herein.

Trehalase (Cv37 tr or Cv trehalase) (P33W9X) from *Chaetomium virescens* is shown as amino acids 21-690 in SEQ ID NO: 4 herein.

Trehalase (Tr37 tr or Tr37 trehalase) (P337ZG) from *Trichoderma reesei* is shown as SEQ ID NO: 12 in WO 2013/148993 and SEQ ID NO: 16 herein.

Trehalase (Tr65 tr or Tr65 trehalase) (P24TTB) from *Trichoderma reesei* is shown as SEQ ID NO: 31 herein.

Cellulase VD: Cellulolytic composition derived from *Trichoderma reesei* comprising GH61A polypeptide having cellulolytic enhancing activity derived from a strain of *Penicillium emersonii* (SEQ ID NO: 2 in WO 2011/041397 or SEQ ID NO: 8 herein), *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 in WO 2005/047499 or SEQ ID NO: 6 herein) variant with the following substitutions: F100D, S283G, N456E, F512Y, disclosed in WO 2012/044915; *Aspergillus fumigatus* Cel7A CBH I disclosed as SEQ ID NO: 6 in WO2011/057140 and SEQ ID NO: 6 herein and *Aspergillus fumigatus* CBH II disclosed as SEQ ID NO: 18 in WO 2011/057140 and as SEQ ID NO: 8 herein.

Alpha-Amylase A (AAA): *Bacillus stearothermophilus* alpha-amylase with the mutations I181*+G182*+N193F truncated to 491 amino acids (SEQ ID NO: 18 herein).

Alpha-Amylase 369 (AA369): *Bacillus stearothermophilus* alpha-amylase with the mutations: I181*+G182*+N193F+V59A+Q89R+E129V+K177L+R179E+Q254S+M284V truncated to 491 amino acids (SEQ ID NO: 18).

Alpha-Amylase Blend B (AABB): Blend of Alpha-Amylase 369, Glucoamylase PoAMG498 and Pfu protease in a ratio of approximately 55: 120: 1 on µg Enzyme Protein basis.

Penicillium oxalicum glucoamylase variant PE498 ("PoAMG498"): *Penicillium oxalicum* glucoamylase variant having the following mutations: K79V+ P2N+ P4S+ P11F+ T65A+ Q327F (using SEQ ID NO: 23 herein for numbering):

Protease Pfu ("PFU"): Protease derived from *Pyrococcus furiosus* shown in SEQ ID NO: 22 herein.

Glucoamylase E: comprises a blend comprising *Talaromyces emersonii* glucoamylase disclosed in WO99/28448 (SEQ ID NO: 13 herein), *Trametes cingulata* glucoamylase disclosed as SEQ ID NO: 2 in WO 06/69289 and SEQ ID NO: 14 herein, and *Rhizomucor pusillus* alpha-amylase with *Aspergillus niger* glucoamylase linker and SBD disclosed as SEQ ID NO: 15 herein with the following substitutions: G128D+D143N (activity ratio AGU:AGU:FAU(F): approx. 30:7:1).

Glucoamylase U: Blend comprising *Talaromyces emersonii* glucoamylase disclosed as SEQ ID NO: 34 in WO99/28448, *Trametes cingulata* glucoamylase disclosed as SEQ ID NO: 2 in WO 06/69289 and *Rhizomucor pusillus* alpha-amylase with *Aspergillus niger* glucoamylase linker and starch binding domain (SBD) disclosed in SEQ ID NO: 15 herein with the following substitutions: G128D+D143N (activity ratio in AGU: AGU:FAU-F is about 65:15:1).

Glucoamylase A: Blend of Glucoamylase E and Cellulase VD in a ratio of approximately 10:3 on µg Enzyme Protein basis.

Yeast:

ETHANOL RED™ from Fermentis, USA

Sequence Identity

For purposes of the present invention, the sequence identity between two amino acid sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *Trends Genet.* 16: 276-277), preferably version 5.0.0 or later. The parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix. The output of Needle labeled "longest identity" (obtained using the -nobrief option) is used as the percent identity and is calculated as follows:

(Identical Residues x 100)/(Length of Alignment – Total Number of Gaps in Alignment)

Trehalase Assay:

PRINCIPLE:

5 Trehalose + H₂O $\xrightarrow{\text{Trehalase}}$ 2 Glucose

T = 37°C, pH = 5.7, A_{340nm}, Light path = 1 cm

Spectrophotometric Stop Rate Determination

10 Unit definition:

One unit will convert 1.0 mmole of trehalose to 2.0 mmoles of glucose per minute at pH 5.7 at 37°C (liberated glucose determined at pH 7.5).

(See Dahlqvist, A. (1968) Analytical Biochemistry 22, 99-107)

15 **Protease assays**

AZCL-casein assay

A solution of 0.2% of the blue substrate AZCL-casein is suspended in Borax/NaH₂PO₄ buffer pH9 while stirring. The solution is distributed while stirring to microtiter plate (100 microL to each well), 30 microL enzyme sample is added and the plates are incubated in an Eppendorf
20 Thermomixer for 30 minutes at 45°C and 600 rpm. Denatured enzyme sample (100°C boiling for 20min) is used as a blank. After incubation the reaction is stopped by transferring the microtiter plate onto ice and the colored solution is separated from the solid by centrifugation at 3000rpm for 5 minutes at 4°C. 60 microL of supernatant is transferred to a microtiter plate and the absorbance at 595nm is measured using a BioRad Microplate Reader.

25

pNA-assay

50 microL protease-containing sample is added to a microtiter plate and the assay is started by adding 100 microL 1mM pNA substrate (5 mg dissolved in 100 microL DMSO and further diluted to 10 mL with Borax/NaH₂PO₄ buffer pH 9.0). The increase in OD₄₀₅ at room
30 temperature is monitored as a measure of the protease activity.

Glucoamylase activity (AGU)

Glucoamylase activity may be measured in Glucoamylase Units (AGU).

35 The Novo Glucoamylase Unit (AGU) is defined as the amount of enzyme, which hydrolyzes 1 micromole maltose per minute under the standard conditions 37°C, pH 4.3, substrate: maltose 23.2 mM, buffer: acetate 0.1 M, reaction time 5 minutes.

An autoanalyzer system may be used. Mutarotase is added to the glucose

dehydrogenase reagent so that any alpha-D-glucose present is turned into beta-D-glucose. Glucose dehydrogenase reacts specifically with beta-D-glucose in the reaction mentioned above, forming NADH which is determined using a photometer at 340 nm as a measure of the original glucose concentration.

<u>AMG incubation:</u>	
Substrate:	maltose 23.2 mM
Buffer:	acetate 0.1 M
pH:	4.30 ± 0.05
Incubation temperature:	37°C ± 1
Reaction time:	5 minutes
Enzyme working range:	0.5-4.0 AGU/mL

5

<u>Color reaction:</u>	
GlucDH:	430 U/L
Mutarotase:	9 U/L
NAD:	0.21 mM
Buffer:	phosphate 0.12 M; 0.15 M NaCl
pH:	7.60 ± 0.05
Incubation temperature:	37°C ± 1
Reaction time:	5 minutes
Wavelength:	340 nm

A folder ([EB-SM-0131.02/01](#)) describing this analytical method in more detail is available on request from Novozymes A/S, Denmark, which folder is hereby included by reference.

10

Acid Alpha-Amylase Activity

When used according to the present invention the activity of an acid alpha-amylase may be measured in AFAU (Acid Fungal Alpha-amylase Units) or FAU-F.

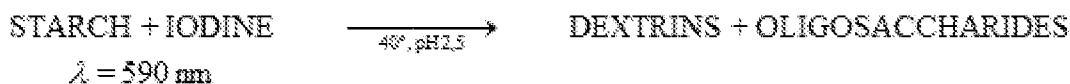
15 Acid alpha-amylase activity (AFAU)

Acid alpha-amylase activity may be measured in AFAU (Acid Fungal Alpha-amylase Units), which are determined relative to an enzyme standard. 1 AFAU is defined as the amount

of enzyme which degrades 5.260 mg starch dry matter per hour under the below mentioned standard conditions.

Acid alpha-amylase, an endo-alpha-amylase (1,4-alpha-D-glucan-glucanohydrolase, E.C. 3.2.1.1) hydrolyzes alpha-1,4-glucosidic bonds in the inner regions of the starch molecule to form dextrans and oligosaccharides with different chain lengths. The intensity of color formed with iodine is directly proportional to the concentration of starch. Amylase activity is determined using reverse colorimetry as a reduction in the concentration of starch under the specified analytical conditions.

ALPHA - AMYLASE



10 blue/violet t = 23 sec. decoloration

Standard conditions/reaction conditions:

Substrate:	Soluble starch, approx. 0.17 g/L
Buffer:	Citrate, approx. 0.03 M
Iodine (I ₂):	0.03 g/L
CaCl ₂ :	1.85 mM
pH:	2.50 ± 0.05
Incubation	40°C

temperature:

Reaction time:	23 seconds
----------------	------------

Wavelength:	590 nm
-------------	--------

Enzyme	0.025 AFAU/mL
--------	---------------

concentration:

Enzyme working	0.01-0.04 AFAU/mL
----------------	-------------------

range:

A folder [EB-SM-0259.02/01](#) describing this analytical method in more detail is available upon request to Novozymes A/S, Denmark, which folder is hereby included by reference.

Determination of FAU-F

15 FAU-F Fungal Alpha-Amylase Units (Fungamyl) is measured relative to an enzyme standard of a declared strength.

Reaction conditions	
Temperature	37°C
pH	7.15
Wavelength	405 nm
Reaction time	5 min
Measuring time	2 min

A folder (EB-SM-0216.02) describing this standard method in more detail is available on request from Novozymes A/S, Denmark, which folder is hereby included by reference.

5 Alpha-amylase activity (KNU)

The alpha-amylase activity may be determined using potato starch as substrate. This method is based on the break-down of modified potato starch by the enzyme, and the reaction is followed by mixing samples of the starch/enzyme solution with an iodine solution. Initially, a blackish-blue color is formed, but during the break-down of the starch the blue color gets weaker and gradually turns into a reddish-brown, which is compared to a colored glass standard.

One Kilo Novo alpha amylase Unit (KNU) is defined as the amount of enzyme which, under standard conditions (i.e., at 37°C +/- 0.05; 0.0003 M Ca²⁺; and pH 5.6) dextrinizes 5260 mg starch dry substance Merck Amylum soluble.

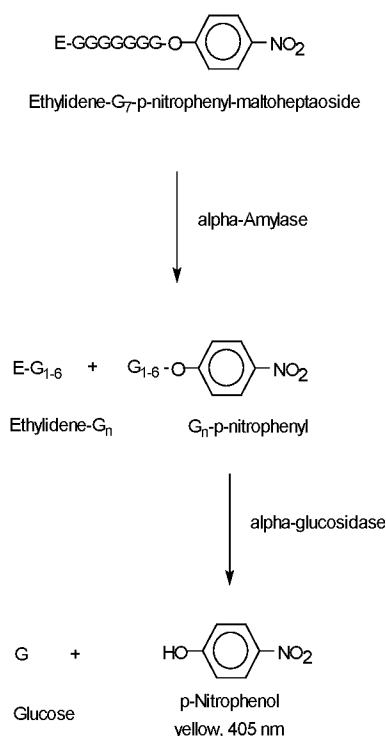
A folder EB-SM-0009.02/01 describing this analytical method in more detail is available upon request to Novozymes A/S, Denmark, which folder is hereby included by reference.

Alpha-amylase Activity (KNU-A)

Alpha amylase activity is measured in KNU(A) Kilo Novozymes Units (A), relative to an enzyme standard of a declared strength.

Alpha amylase in samples and α -glucosidase in the reagent kit hydrolyze the substrate (4,6-ethylidene(G₇)-*p*-nitrophenyl(G₁)- α ,D-maltoheptaoside (ethylidene-G₇PNP) to glucose and the yellow-colored *p*-nitrophenol.

The rate of formation of *p*-nitrophenol can be observed by Konelab 30. This is an expression of the reaction rate and thereby the enzyme activity.



The enzyme is an alpha-amylase with the enzyme classification number EC 3.2.1.1.

Parameter	Reaction conditions
Temperature	37°C
pH	7.00 (at 37°C)
Substrate conc.	Ethylidene-G ₇ PNP, R2: 1.86 mM
Enzyme conc. (conc. of high/low standard in reaction mixture)	1.35 -4.07 KNU(A)/L
Reaction time	2 min
Interval kinetic measuring time	7 / 18 sec.
Wave length	405nm
Conc. of reagents/chemicals critical for the analysis	α -glucosidase, R1: ≥ 3.39 kU/L

- 5 A folder EB-SM-5091.02-D on determining KNU-A activity is available upon request to Novozymes A/S, Denmark, which folder is hereby included by reference.

pNP-G7 assay

- 10 The alpha-amylase activity may be determined by a method employing the G7-pNP substrate. G7-pNP which is an abbreviation for 4,6-ethylidene(G₇)-p-nitrophenyl(G₁)- α ,D-maltoheptaoside, a blocked oligosaccharide which can be cleaved by an endo-amylase, such as an alpha-amylase. Following the cleavage, the alpha-Glucosidase included in the kit digest the

hydrolysed substrate further to liberate a free PNP molecule which has a yellow color and thus can be measured by visible spectrophotometry at $\lambda=405\text{nm}$ (400-420 nm.). Kits containing G7-pNP substrate and alpha-Glucosidase is manufactured by Roche/Hitachi (cat. No.11876473).

5 REAGENTS:

The G7-pNP substrate from this kit contains 22 mM 4,6-ethylidene- G7-pNP and 52.4 mM HEPES (2-[4-(2-hydroxyethyl)-1-piperazinyl]-ethanesulfonic acid), pH 7.0) .

10 The alpha-Glucosidase reagent contains 52.4 mM HEPES, 87 mM NaCl, 12.6 mM MgCl_2 , 0.075 mM CaCl_2 , ≥ 4 kU/L alpha-glucosidase).

The substrate working solution is made by mixing 1 mL of the alpha-Glucosidase reagent with 0.2 mL of the G7-pNP substrate. This substrate working solution is made immediately before use.

15 Dilution buffer: 50 mM MOPS, 0.05% (w/v) Triton X100 (polyethylene glycol p-(1,1,3,3-tetramethylbutyl)-phenyl ether ($\text{C}_{14}\text{H}_{22}\text{O}(\text{C}_2\text{H}_4\text{O})_n$ ($n = 9-10$))), 1mM CaCl_2 , pH8.0.

PROCEDURE:

20 The amylase sample to be analyzed is diluted in dilution buffer to ensure the pH in the diluted sample is 7. The assay is performed by transferring 20 μl diluted enzyme samples to 96 well microtiter plate and adding 80 μl substrate working solution. The solution is mixed and pre-incubated 1 minute at room temperature and absorption is measured every 20 sec. over 5 minutes at OD 405 nm.

25 The slope (absorbance per minute) of the time dependent absorption-curve is directly proportional to the specific activity (activity per mg enzyme) of the alpha-amylase in question under the given set of conditions. The amylase sample should be diluted to a level where the slope is below 0.4 absorbance units per minute.

30 **Phadebas activity assay**

The alpha-amylase activity may also be determined by a method using the Phadebas substrate (from for example Magle Life Sciences, Lund, Sweden). A Phadebas tablet includes interlinked starch polymers that are in the form of globular microspheres that are insoluble in water. A blue dye is covantly bound to these microspheres. The interlinked starch polymers in the microsphere are degraded at a speed that is proportional to the alpha-amylase activity. When the alpha-amylse degrades the starch polymers, the released blue dye is water soluble

35

and concentration of dye can be determined by measuring absorbance at 620nm. The concentration of blue is proportional to the alpha-amylase activity in the sample.

The amylase sample to be analysed is diluted in activity buffer with the desired pH. One substrate tablet is suspended in 5mL activity buffer and mixed on magnetic stirrer. During
5 mixing of substrate transfer 150µl to microtiter plate (MTP) or PCR-MTP. Add 30µl diluted amylase sample to 150µl substrate and mix. Incubate for 15 minutes at 37°C. The reaction is stopped by adding 30µl 1M NaOH and mix. Centrifuge MTP for 5 minutes at 4000xg. Transfer 100µl to new MTP and measure absorbance at 620nm.

The amylase sample should be diluted so that the absorbance at 620nm is between 0
10 and 2.2, and is within the linear range of the activity assay.

Reducing sugar activity assay:

The alpha-amylase activity may also be determined by reducing sugar assay with for example corn starch substrate. The number of reducing ends formed by the alpha-amylase
15 hydrolysing the alpha-1,4-glycosidic linkages in starch is determined by reaction with p-Hydroxybenzoic acid hydrazide (PHBAH). After reaction with PHBAH the number of reducing ends can be measured by absorbance at 405nm and the concentration of reducing ends is proportional to the alpha-amylase activity in the sample.

The corns starch substrate (3mg/ml) is solubilised by cooking for 5 minutes in milliQ
20 water and cooled down before assay. For the stop solution prepare a Ka-Na-tartrate/NaOH solution (K-Na-tartrate (Merck 8087) 50g/l, NaOH 20g/l) and prepare freshly the stop solution by adding p-Hydroxybenzoic acid hydrazide (PHBAH, Sigma H9882) to Ka-Na-tartrate/NaOH solution to 15mg/ml.

In PCR-MTP 50µl activity buffer is mixed with 50µl substrate. Add 50µl diluted enzyme
25 and mix. Incubate at the desired temperature in PCR machine for 5 minutes. Reaction is stopped by adding 75µl stop solution (Ka-Na-tartrate/NaOH/PHBAH). Incubate in PCR machine for 10 minutes at 95°C. Transfer 150µl to new MTP and measure absorbance at 405nm.

The amylase sample should be diluted so that the absorbance at 405nm is between 0
30 and 2.2, and is within the linear range of the activity assay.

EnzChek® assay:

For the determination of residual amylase activity an EnzChek® Ultra Amylase Assay Kit (E33651, Invitrogen, La Jolla, CA, USA) may be used.

The substrate is a corn starch derivative, DQ™ starch, which is corn starch labeled with
35 BODIPY® FL dye to such a degree that fluorescence is quenched. One vial containing approx. 1 mg lyophilized substrate is dissolved in 100 microliters of 50 mM sodium acetate (pH 4.0). The vial is vortexed for 20 seconds and left at room temperature, in the dark, with occasional

mixing until dissolved. Then 900 microliters of 100 mM acetate, 0.01% (w/v) TRITON® X100, 0.125 mM CaCl₂, pH 5.5 is added, vortexed thoroughly and stored at room temperature, in the dark until ready to use. The stock substrate working solution is prepared by diluting 10-fold in residual activity buffer (100 mM acetate, 0.01% (w/v) TRITON® X100, 0.125 mM CaCl₂, pH 5.5). Immediately after incubation the enzyme is diluted to a concentration of 10-20 ng enzyme protein/ml in 100 mM acetate, 0.01% (W/v) TRITON® X100, 0.125 mM CaCl₂, pH 5.5.

For the assay, 25 microliters of the substrate working solution is mixed for 10 second with 25 microliters of the diluted enzyme in a black 384 well microtiter plate. The fluorescence intensity is measured (excitation: 485 nm, emission: 555 nm) once every minute for 15 minutes in each well at 25°C and the V_{max} is calculated as the slope of the plot of fluorescence intensity against time. The plot should be linear and the residual activity assay has been adjusted so that the diluted reference enzyme solution is within the linear range of the activity assay.

Determination of Pullulanase Activity (NPUN)

Endo-pullulanase activity in NPUN is measured relative to a Novozymes pullulanase standard. One pullulanase unit (NPUN) is defined as the amount of enzyme that releases 1 micro mol glucose per minute under the standard conditions (0.7% red pullulan (Megazyme), pH 5, 40°C, 20 minutes). The activity is measured in NPUN/ml using red pullulan.

1 mL diluted sample or standard is incubated at 40°C for 2 minutes. 0.5 mL 2% red pullulan, 0.5 M KCl, 50 mM citric acid, pH 5 are added and mixed. The tubes are incubated at 40°C for 20 minutes and stopped by adding 2.5 ml 80% ethanol. The tubes are left standing at room temperature for 10-60 minutes followed by centrifugation 10 minutes at 4000 rpm. OD of the supernatants is then measured at 510 nm and the activity calculated using a standard curve.

Media and Solutions

YP + 2% glucose medium was composed of 1% yeast extract, 2% peptone and 2% glucose.

YP + 2% maltodextrin medium was composed of 1% yeast extract, 2% peptone and 2% maltodextrin.

PDA agar plates were composed of potato infusion (potato infusion was made by boiling 300 g of sliced (washed but unpeeled) potatoes in water for 30 minutes and then decanting or straining the broth through cheesecloth. Distilled water was then added until the total volume of the suspension was one liter, followed by 20 g of dextrose and 20 g of agar powder. The

medium was sterilized by autoclaving at 15 psi for 15 minutes (Bacteriological Analytical Manual, 8th Edition, Revision A, 1998).

LB plates were composed of 10 g of Bacto-Tryptone, 5 g of yeast extract, 10 g of sodium chloride, 15 g of Bacto-agar, and deionized water to 1 liter.

LB medium was composed of 10 g of Bacto-Tryptone, 5 g of yeast extract, 10 g of sodium chloride, and deionized water to 1 liter.

COVE sucrose plates were composed of 342 g of sucrose, 20 g of agar powder, 20 ml of COVE salts solution, and deionized water to 1 liter. The medium was sterilized by autoclaving at 15 psi for 15 minutes (Bacteriological Analytical Manual, 8th Edition, Revision A, 1998). The medium was cooled to 60°C and 10 mM acetamide, 15 mM CsCl, TRITON® X-100 (50 µl/500 ml) were added.

COVE salts solution was composed of 26 g of MgSO₄•7H₂O, 26 g of KCL, 26 g of KH₂PO₄, 50 ml of COVE trace metals solution, and deionized water to 1 liter.

COVE trace metals solution was composed of 0.04 g of Na₂B₄O₇•10H₂O, 0.4 g of CuSO₄•5H₂O, 1.2 g of FeSO₄•7H₂O, 0.7 g of MnSO₄•H₂O, 0.8 g of Na₂MoO₄•2H₂O, 10 g of ZnSO₄•7H₂O, and deionized water to 1 liter.

Example 1

Cloning of *Chaetomium virescens* trehalase (Cv37 tr) polypeptide (P33W9X) coding sequence from *Chaetomium virescens* CBS547.75.

The *Chaetomium virescens* trehalase (Cv Trehalase) polypeptide coding sequence was cloned from *Chaetomium virescens* CBS547.75 DNA by PCR.

Chaetomium virescens CBS547.75 was purchased from the Centraalbureau voor Schimmelcultures (Utrecht, the Netherlands). The fungal strain was cultivated in 100 ml of YP + 2% glucose medium in 1000 ml Erlenmeyer shake flasks for 5 days at 20°C. Mycelia were harvested from the flasks by filtration of the medium through a Buchner vacuum funnel lined with MIRACLOTH® (EMD Millipore, Billerica, MA, USA). Mycelia were frozen in liquid nitrogen and stored at -80°C until further use. Genomic DNA was isolated using a DNEASY® Plant Maxi Kit (QIAGEN GMBH, Hilden Germany) according to the manufacturer's instructions.

Genomic sequence information was generated by Illumina MySeq (Illumina Inc., San Diego, CA). 5 µgs of the isolated *Chaetomium virescens* genomic DNA was used for library preparation and analysis according to the manufacturer's instructions. Genes were called using

GeneMark.hmm ES version 2.3c and identification of the catalytic domain was made using "Trehalase PF01204" Hidden Markov Model provided by Pfam. The polypeptide coding sequence for the entire coding region was cloned from *Chaetomium virescens* CBS547.75 genomic DNA by PCR using the primers (SEQ ID NO: 24 and SEQ ID NO: 25) described below.

KKSC0334-F 5'- ACACAAC**TGGGGATCC**ACCATGACGCTCCGACACCTCGG -3'
(SEQ ID NO: 24)

KKSC0334-R 5'- CTAGATCTCGAGA**AAGCTT**TCACGACCTCCTCCCTACCC -3' (SEQ ID NO: 25)

Bold letters represent *Chaetomium virescens* enzyme coding sequence. Restriction sites are underlined. The sequence to the left of the restriction sites is homologous to the insertion sites of pDau109 (WO 2005/042735).

The amplification reaction (50 µls) was performed according to the manufacturer's instructions (Phusion HiFi DNA polymerase cat # M0530S, New England Biolabs Inc.) with the following final concentrations:

PCR mix

5x HF Phusion Buffer	10 ul
dNTP 10 mM	1ul
DMSO	1.5 ul
Primer F	1 ul
Prime R	1 ul
Phusion polymerase	0,5 uls
H2O	35 uls
<u>Genomic DNA</u>	<u>1 uls</u>
Total volume	50 uls

The PCR reaction was incubated in a DYAD® Dual-Block Thermal Cycler (BioRad, USA) programmed for 1 cycle at 98°C for 30 seconds; 35 cycles each at 98°C for 10 seconds, 72°C for 2 minutes followed by 1 cycle at 72°C for 7 minutes. Samples were cooled to 10°C before removal and further processing.

Four µl of the PCR reaction was analyzed by 1% agarose gel electrophoresis using 40 mM Tris base, 20 mM sodium acetate, 1 mM disodium EDTA (TAE) buffer. A major band of about 2241 bp was observed. The remaining PCR reaction was purified directly with an ILLUSTRA™ GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions.

Two µg of plasmid pDau109 was digested with Bam HI and Hind III and the digested plasmid was run on a 1% agarose gel using 50 mM Tris base-50 mM boric acid-1 mM disodium EDTA (TBE) buffer in order to remove the stuffer fragment from the restricted plasmid. The bands were visualized by the addition of SYBR® Safe DNA gel stain (Life Technologies Corporation, Grand Island, NY, USA) and use of a 470 nm wavelength transilluminator. The band corresponding to the restricted plasmid was excised and purified using an ILLUSTRA™ GFX™ PCR DNA and Gel Band Purification Kit (Cat.no28-9034-70 from GE Healthcare). The plasmid was eluted into 10 mM Tris pH 8.0 and its concentration adjusted to 20 ng per µl. An IN-FUSION® HD EcoDry-Down PCR Cloning Kit (BD Biosciences, Palo Alto, CA, USA), was used to clone the 2241 bp PCR fragment into pDau109 digested with Bam HI and Hind III (20 ng). The IN-FUSION® total reaction volume was 10 µl. The IN-FUSION® total reaction volume was 10 µl. The IN-FUSION® reaction was transformed into FUSION-BLUE™ E. coli cells (BD Biosciences, Palo Alto, CA, USA) according to the manufacturer's protocol and plated onto LB agar plates supplemented with 50 µg of ampicillin per ml. After incubation overnight at 37°C, transformant colonies were observed growing under selection on the LB plates supplemented with 50 µg of ampicillin per ml.

One plasmid with the correct *C. virescens* trehalase coding sequence (SEQ ID NO: 3) was chosen. The plasmid was designated pKKSC0334-1. Cloning of the *Chaetomium virescens* trehalase gene into Bam HI-HindIII digested pDau109 resulted in transcription of the *C. virescens* trehalase gene under the control of a NA2-tpi double promoter. NA2-tpi is a modified promoter from an *Aspergillus niger* neutral alpha-amylase gene in which the untranslated leader has been replaced by an untranslated leader from an *Aspergillus nidulans* triose phosphate isomerase gene.

The expression plasmid pKKSC0334-1 was transformed into protoplasts of *Aspergillus oryzae* MT3568 according to the method of European Patent EP 0238023, pages 14-15. *Aspergillus oryzae* MT3568 is an amdS (acetamidase) disrupted derivative of *A. oryzae* JaL355 (WO 2002/40694) in which pyrG auxotrophy was restored in the process of inactivating the *A. oryzae* amdS gene.

E. coli 3701 containing pKKSC0334-1 was grown overnight according to the manufacturer's instructions (Genomed) and plasmid DNA of pKKSC0334-1 was isolated using a Plasmid Midi Kit (Genomed JETquick kit, cat.nr. 400250, GENOMED GmbH, Germany) according to the manufacturer's instructions. The purified plasmid DNA was transformed into *Aspergillus oryzae* MT3568. *A. oryzae* MT3568 protoplasts were prepared according to the method of Christensen et al., 1988, Bio/Technology 6: 1419-1422. The selection plates consisted of COVE sucrose with +10 mM acetamide +15 mM CsCl + TRITON® X-100 (50µl/500ml). The plates were incubated at 37°C. Briefly, 8µl of plasmid DNA representing 3µg of DNA was added to 100µl MT3568 protoplasts. 250 µl of 60% PEG solution was added

and the tubes were gently mixed and incubate at 37°C for 30 minutes. The mix was added to 10 ml of pre melted Cove top agarose (The top agarose melted and then the temperature equilibrated to 40°C in a warm water bath before being added to the protoplast mixture). The combined mixture was then plated on two Cove-sucrose selection petri plates with 10mM Acetamide. The plates are incubated at 37°C for 4 days. Single *Aspergillus* transformed colonies were identified by growth on the selection Acetamide as a carbon source. Each of the four *A. oryzae* transformants were inoculated into 750 µl of YP medium supplemented with 2% glucose and also 750 µl of 2% maltodextrin and also DAP4C in 96 well deep plates and incubated at 37°C stationary for 4 days. At same time the four transformants were restreaked on COVE-2 sucrose agar medium.

Culture broth from the *Aspergillus oryzae* transformants were then analyzed for production of the P33W9X trehalase polypeptide by SDS-PAGE using NUPAGE® 10% Bis-Tris SDS gels (Invitrogen, Carlsbad, CA, USA) according to the manufacturer. A band at approximately 75 kDa was observed for each of the *Aspergillus oryzae* transformants. One *A. oryzae* transformant producing the P33W9X polypeptide was designated *A. oryzae* EXP09256.

For larger scale production, *A. oryzae* EXP09256 spores were spread onto a PDA plate and incubated for five days at 37°C. The confluent spore plate was washed twice with 5 ml of 0.01% TWEEN® 20 to maximize the number of spores collected. The spore suspension was then used to inoculate nine 500 ml flasks containing 100 ml of Dap-4C medium. The cultures were incubated at 30°C with constant shaking at 150 rpm. At day four post-inoculation, the culture broth was collected by filtration through a bottle top MF75™ SUPOR® MachV 0.2 µm PES filter (Thermo Fisher Scientific, Roskilde, Denmark). Fresh culture broth from this transformant produced a band of GH37 protein of approximately 75 kDa (designated EXP09256). The identity of this band as the *C. virescens* trehalase polypeptide was verified by peptide sequencing.

Characterization of the EXP09256 trehalase polypeptide coding sequence from *Chaetomium virescens* CBS547.75

The genomic DNA sequence and deduced amino acid sequence of the *Chaetomium virescens* CBS547.75.

The *Chaetomium virescens* trehalase polypeptide (P33W9X) genomic coding sequence is shown in SEQ ID NO: 3 and SEQ ID NO: 4, respectively.. The encoded predicted protein is 690 amino acids. Using the SignalP program (Nielsen et al., 1997, supra), a signal peptide of 20 residues was predicted. The predicted mature protein contains 670 amino acids with a predicted molecular mass of 75 kDa and a predicted isoelectric point of 5.5.

Example 2

Expression of trehalase (P33WJF) from *Myceliophthora sepedonium* (Ms Tr37)

This strain was purchased from UAMH and received on the 12th of Sept. 1988 as UAMH5004. UAMH is the University of Alberta Microfungus Collection and Herbarium.

Myceliophthora sepedonium strain UAHM5004 was obtained from the University of Alberta Microfungus Collection and Herbarium UAMH, Edmonton, Alberta, Canada T6G 2R3.

5 The earlier genus and species designation *Cornyascus sepedonium* is still employed.

The trehalase polypeptide (P33WJF) coding sequence was cloned from *Myceliophthora sepedonium* strain UAHM5004 DNA by PCR.

The fungal strain was cultivated, DNA isolated genome sequenced and the trehalase enzyme candidate identified as in example 1.

10 The polypeptide coding sequence for the entire coding region was cloned from *Myceliophthora sepedonium* strain UAHM5004 genomic DNA by PCR using the primers (SEQ ID NO: 26 and SEQ ID NO: 27) described below.

KKSC0335-F 5'- ACACAACTGGGGATCCACCATGGCGCTACGACACATCGC-3'
(SEQ ID NO: 26)

15 KKSC0335-R 5'- CTAGATCTCGAGAAGCTTTTACGAGACGGAGACACTAAACA
(SEQ ID NO: 27)

Bold letters represent *Myceliophthora sepedonium* enzyme coding sequence. Restriction sites are underlined. The sequence to the left of the restriction sites is homologous to the insertion sites of pDau109 (WO 2005/042735).

20 The amplification reaction (50 µls) was performed according to the protocol in example 1 according the manufacturer's instructions (Phusion HiFi DNA polymerase cat # M0530S, New England Biolabs Inc

Four µl of the PCR reaction was analyzed by 1% agarose gel electrophoresis using 40 mM Tris base, 20 mM sodium acetate, 1 mM disodium EDTA (TAE) buffer. A major band of about 2354 bp was observed. The remaining PCR reaction was purified directly with an ILLUSTRATM GFX™ PCR DNA and Gel Band Purification Kit (GE Healthcare, Piscataway, NJ, USA) and cloned into pDau109 as in example 1.

One plasmid with the correct *M. sepedonium* trehalase coding sequence (SEQ ID NO: x) was chosen. The plasmid was designated pKKSC0335-1. The expression plasmid pKKSC0335-1 was transformed into protoplasts of *Aspergillus oryzae* MT3568 and a transformants chosen and propagated also according example 1.

Production of the trehalase polypeptide (P33WJF) by SDS-PAGE using NUPAGE® 10% Bis-Tris SDS gels (Invitrogen, Carlsbad, CA, USA) according to the manufacturer. A band at approximately 76 kDa was observed for each of the *Aspergillus oryzae* transformants. One *A. oryzae* transformant producing the P33W9X polypeptide was designated *A. oryzae* EXP09258.

Larger scale production, *A. oryzae* EXP09258 was performed also according to Example 1.

Characterization of the EXP09258 trehalase polypeptide coding sequence from *Myceliophthora sepidonium* strain UAHM5004 The genomic DNA sequence and deduced amino acid sequence of the *Myceliophthora sepidonium* strain UAHM5004

Trehalase polypeptide (P33WJF) genomic coding sequence are shown in SEQ ID NO: 1 and SEQ ID NO: 2, respectively. The encoded predicted protein is 697 amino acids. Using the SignalP program (Nielsen et al., 1997, supra), a signal peptide of 20 residues was predicted. The predicted mature protein contains 677 amino acids with a predicted molecular mass of 76 kDa and a predicted isoelectric point of 5.1.

Example 3

Glucose production with trehalase addition using supernatant of ferm drop

Ferm drop sample at the end of fermentation from industrial corn ethanol plant was first centrifuged, and supernatant was collected for the testing. The trehalose level in the supernatant was analyzed to be around 2 g/L by an Ion Chromatography assay. The pH of the supernatant was measured as 4.6, and LACTROL® was supplemented as antibacterial agent at the level of 3ppm. Next, 2 ml of supernatant was aliquoted into 15 ml polypropylene tube. Each tube was dosed with trehalase listed in Table 1. The enzyme dosage was 0, 0.05, 0.5 and 5 ug/ml, respectively. Each treatment ran three replicate. Afterwards, all tubes were incubated in 32°C water bath for 5 hours. Samples were taken at 5 hour of incubation for HPLC analysis. The HPLC preparation consisted of stopping the reaction by addition of 20 micro liters of 40% H₂SO₄, and filtering through a 0.45 micrometer filter. Agilent™ 1100 HPLC system coupled with RI detector was used to determine glucose concentration. The separation column was aminex HPX-87H ion exclusion column (300mm x 7.8mm) from BioRad™.

Table 1. List of trehalase tested

Purified ID	Sequence and ID Code	Donor source
U3EAN	SEQ ID NO: 30 - P33WJG	<i>Myceliophthora sepidonium</i> (Ms37 tr)
U3EAJ	SEQ ID NO: 4 - P33W9X	<i>Chaetomium virescens</i> (Cv37 tr)
U3EAH	SEQ ID NO: 16 - P337ZG	<i>Trichoderma reesei</i> (Tr37 tr)

Results

The glucose results from HPLC analysis is summarized in Fig. 1. Compared to control without any trehalase addition, all treatments with trehalase showed glucose increase. The amount of glucose increase was depended on the dosage of trehalase and its source. Results

showed *Myceliophthora sepdonium* trehalase (Ms trehalase) performed the best. At dosage of 0.5 ug/ml, all the trehalose was converted into glucose with 5 hour incubation. *Chaetomium virescens* trehalase was a little less efficient than *Myceliophthora sepdonium* trehalase, but still much better than *Trichoderma reesei* trehalase (Tr tr).

5

Example 4

Application of Ms trehalase in Conventional SSF process for ethanol production

All treatments were evaluated via 5 g small assay. Each treatment ran five replicate. Three corn mashes liquefied by Alpha-Amylase Blend B (AABB) (Mash A and B) and Alpha-Amylase A (AAA) (Mash C) from industrial corn ethanol plants were used for the testing. 3 ppm penicillin and 1000 ppm urea were supplemented into each mash. The pH of the slurries was adjusted to 5.0 with 40% H₂SO₄ or 50% NaOH. Approximately 5 g of the slurry was added to 15 ml polypropylene tube. The tubes were prepared by drilling a 1/32 inch hole and the empty tubes were then weighed before corn slurry was added. The tubes were weighed again after mash was added to determine the exact weight of mash in each tube. Each tube was dosed with actual enzyme dosage based on the exact weight of corn slurry in each tube. The enzyme dosage for each treatment was listed in Table 2. Trehalase from *Myceliophthora sepdonium* (Ms37 trehalase) (SEQ ID NO: 30) was used in this study. Afterwards, the tubes were dosed with 50 ul of yeast propagate to around 5 g corn mash, and then were incubated in 32°C water batch for SSF. Samples were taken at 53 hour of fermentation for HPLC analysis. The HPLC preparation consisted of stopping the reaction by addition of 50 micro liters of 40% H₂SO₄, centrifuging, and filtering through a 0.45 micrometer filter. Agilent™ 1100 HPLC system coupled with RI detector was used to determine sugars, acids and ethanol concentration. The separation column was aminex HPX-87H ion exclusion column (300mm x 7.8mm) from BioRad™.

Table 2. AMG and trehalase dosage for each treatment in SSF

	Mash	AMG	AMG, AGU/g-DS	Trehalase, ug EP/g-DS
1	Mash A	Glucoamylase U	0.56	0
2				5
3		Glucoamylase E	0.60	0
4				5
5		Glucoamylase A	0.60	0
6				5
7	Mash B	Glucoamylase U	0.56	0
8				5

9	Mash C	Glucoamylase E	0.60	0
10				5
11		Glucoamylase A	0.60	0
12				5
13		GlucoamylaseU	0.56	0
14				5
15		Glucoamylase E	0.60	0
16				5
17		Glucoamylase A	0.60	0
18				5

Results

The results from HPLC analysis were summarized in Table 3. Compared to the control, treatment with trehalase addition achieved much higher ethanol titer, ranging from between 0.5 to 1.2% yield increase.

Table 3. Ethanol yield increase with trehalase addition compared to the control in SSF

Treatment	Mash A	Mash B	Mash C
Glucoamylase U (control)	-	-	-
Glucoamylase U + Ms37 trehalase	0.70%	0.64%	0.90%
Glucoamylase E (control)	-	-	-
Glucoamylase E + Ms37 trehalase	1.13%	0.91%	1.16%
Glucoamylase A (control)	-	-	-
Glucoamylase A + Ms37 trehalase	0.54%	0.64%	1.01%

Example 5

Purification of GH37 trehalases

Filtered culture broth from fermentation of *A. oryzae* harboring the GH37 trehalase gene was added solid ammonium sulfate to a final concentration of 2 M and pH adjusted to 7. The solution, containing the trehalase, was applied to a hydrophobic interaction column (butyl Toyopearl, approximately 50 ml in a XK26 column, equilibrated with buffer A), using as buffer A 50 mM Hepes + 2 M ammonium sulfate pH 7.0, and as buffer B 50 mM Hepes pH 7.0. Unbound material was washed off the column with equilibration buffer and the trehalase was

eluted with a linear gradient (100% to 0% A) over 5 column volumes and 10 ml fractions were collected. Based on the chromatogram trehalase containing fractions were pooled and dialyzed against a large volume of 20 mM Hepes pH 7.0.

5 The purified GH37 trehalases are listed in the table below:

Purified Sample ID	Donor Scientific Name	Protein Sequence Reference ID
U3EAN	<i>Myceliophthora sepedonium</i>	SEQ ID NO: 30 - P33WJF
U3EAJ	<i>Chaetomium virescens</i>	SEQ ID NO: 4 - P33W9X
U3EAH	<i>Trichoderma reesei</i>	SEQ ID NO:16 - P337ZG

Molecular weight of the purified trehalases

The molecular weight, as estimated from SDS-PAGE, was approximately 100 kDa for all three trehalases and the purity was in all cases > 95%.

10

Example 6

Purification of a GH65 trehalase from *Trichoderma reesei* (SEQ ID NO: 31)

Filtered culture broth from fermentation of *A. oryzae* harboring the GH65 trehalase gene from *Trichoderma reesei* (P24TTB) was loaded on a gel filtration column (about 780 ml Sephadex G-25 medium) equilibrated with 50 mM Na-acetate pH 5.0 and fractions collected. Based on the chromatogram fractions containing protein (A280>A260) that ran straight through the column were pooled. The trehalase containing pool was concentrated using a Viva cell with a 30 kDa cut-off membrane.

15

20 Molecular weight of the purified trehalases

The molecular weight, as estimated from SDS-PAGE, was approximately 120 kDa and the purity was > 95%.

Example 7

25 Determination of trehalase activity (GOD)

GOD-perid

50 microliter trehalase-containing enzyme solution (appropriately diluted with 20 mM MES buffer pH 5.0) is dispensed in a microtiter plate well (e. g. NUNC 269620) and 50 microliter substrate (10 mg/ml trehalose in 20 mM MES buffer pH 5.0) is added. The plate is sealed and incubated 15 min., shaken with 750 rpm at 32 °C. After the incubation 20 microliter of reaction solution is transferred to an empty microtiter plate and 180 microliter GOD-perid

30

[600 ppm glucose oxidase (Sigma G1625), 20 ppm peroxidase (Sigma P8125), 0.1 % ATBS (Roche 102946) in 0.1 M K-phosphate buffer pH 7.0] is added. The plate is left for 30 minutes at 20-30 °C and following this the absorbance at 420 nm is measured in a microtiter plate spectrophotometer.

5

Example 8

pH profile

The pH profile was determined at 32°C in the pH range of 2.0 to 7.5 (in 0.5 pH-unit steps) as described above in the section "Determination of trehalase activity, GOD-perid", except that a buffer cocktail (110 mM acetic acid, 110 mM citric acid and 110 mM MES) was used instead of the 20 mM MES buffer pH 5.0 buffer. The results are summarized in table 5 below. The values given for each pH in the range of 2.0 - 7.5 are the relative activity in % normalized to the value at optimum.

15

Relative activity												
pH	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8
Tr37 tr SEQ ID NO: 16 U3EAH	36	100	98	68	35	18	16	14	15	13	10	7
Cv37 tr SEQ ID NO: 4 U3EAJ	27	73	81	100	86	77	70	76	90	77	53	25
Ms37 tr SEQ ID NO: 30 U3EAN	26	85	88	100	59	52	52	46	2	32	24	18

Example 9

Thermostability using DSC data

20

25

The thermostability of trehalases were determined by Differential Scanning Calorimetry (DSC) using a VP-capillary DSC instrument (MicroCal Inc., Piscataway, NJ, USA) equipped with an auto sampler. Aliquots of the trehalase, purified as described in Example 1, were buffer-changed (see buffer in table below) using prepacked NAP-5 columns. The samples were diluted with the corresponding buffer to approximately 0.5 mg/ml and the buffer was used as reference solution. Sample and reference solutions (approx. 0.5 ml) were thermally pre-equilibrated for 10 minutes at 20°C and the DSC scan was performed from 20 to 100°C at a

scan rate of 200 K/hour. Data-handling was performed using the MicroCal Origin software (version 7.0383). The thermal denaturation temperature, T_d (°C), was taken as the top of denaturation peak (major endothermic peak) in thermograms (C_p vs. T) obtained after heating the trehalase solution in the buffer at a constant programmed heating rate. Denaturation

The results of the DSC measurements are summarized in the table below.

Purified Sample ID	Donor Scientific Name	Protein Sequence SEQ ID and Reference ID	Buffer	T _d (°C)
U3EAN	<i>Myceliophthora sepedonium</i>	SEQ ID NO: 30 P33WJF	50 mM MES pH 5.0	68.9
U3EAJ	<i>Chaetomium virescens</i>	SEQ ID NO: 4 P33W9X	50 mM MES pH 5.0	65.1
U3EAH	<i>Trichoderma reesei</i>	SEQ ID NO: 16 P337ZG	50 mM MES pH 5.0	67.0
U6B9P	<i>Trichoderma reesei</i>	SEQ ID NO: 31 P24TTB	50 mM MES pH 5.0	64.7

Example 10

Glucose production by Ms37 trehalase (SEQ ID NO: 30) and Tr65 trehalase (SEQ ID NO: 31)

Ferm drop samples at the end of fermentation from two industrial corn ethanol plants were collected and pH was measured as 4.68 (A) and 4.49 (B) respectively. Ferm drop samples were centrifuged at 3500 rpm for 15 minutes, and then supernatants were collected and used for the trehalase application test. First, penicillium was supplemented as antibacterial agent at the level of 3 ppm. Next, 2 ml of supernatant was aliquoted into 15 ml polypropylene tube. Each tube was dosed with trehalase listed in the table below. The enzyme doses tested were 0, 0.05, 0.15, 0.5 and 5 ug per ml of supernatant. Each treatment ran three replicate. Afterwards, all tubes were incubated in 32°C water bath for 5 hours. Samples were taken at 5 hour of incubation for HPLC analysis. The HPLC preparation consisted of stopping the reaction by addition of 20 micro liters of 40% H₂SO₄, and filtering through a 0.45 micrometer filter. Agilent™ 1100 HPLC system coupled with RI detector was used to determine glucose concentration. The separation column was aminex HPX-87H ion exclusion column (300mm x 7.8mm) from BioRad™.

Trehalases tested

Purified ID	GH family	Donor source and SEQ ID
-------------	-----------	-------------------------

U3EAN	GH37	<i>Myceliophthora sepdonium</i> (SEQ ID NO: 30)
U6BP9	GH65	<i>Trichoderma reesei</i> (SEQ ID NO: 31)

Results

The glucose results from HPLC analysis were summarized by JMP in Figure 2. Compared to control without any trehalase addition, treatments with trehalase showed glucose increase, except the lowest dose of 0.05 ug/ml in Ferm Drop B. The amount of glucose increase was dependent on the dosage of trehalase and its source. Results showed that *Myceliophthora sepdonium* GH37 trehalase (SEQ ID NO: 30) performed better than *Trichoderma reesei* GH65 trehalase (SEQ ID NO: 31). At high dose of 5 ug/ml trehalase addition, the trehalose in ferm drop samples was completely converted to glucose by the trehalases.

The present invention is presented in the following numbered paragraphs:

1. A polypeptide having trehalase activity, selected from the group consisting of:

(a) a polypeptide having at least 90% sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 or at least 80% sequence identity to the mature polypeptide of SEQ ID NO: 4;

(b) a polypeptide encoded by a polynucleotide that hybridizes under high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 29, SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii);

(c) a polypeptide encoded by a polynucleotide having at least 60% sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 29 or SEQ ID NO: 1 or SEQ ID NO: 3 or the cDNA sequence thereof;

(d) a variant of the mature polypeptide of SEQ ID NO: 30, SEQ ID NO: 2 or SEQ ID NO: 4 comprising a substitution, deletion, and/or insertion at one or more (e.g., several) positions; and

(e) a fragment of the polypeptide of (a), (b), (c), or (d) that has trehalase activity.

2. The polypeptide of paragraph 1, having at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2.

3. The polypeptide of paragraph 1 or 2, which is encoded by a polynucleotide that hybridizes under very high stringency conditions with (i) the mature polypeptide coding sequence of SEQ

ID NO: 29 or SEQ ID NO: 1, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii).

4. The polypeptide of any of paragraphs 1-3, which is encoded by a polynucleotide having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 29 or SEQ ID NO: 1 or the cDNA sequence thereof.

5. The polypeptide of any of paragraphs 1-4, comprising or consisting of SEQ ID NO: 30 or SEQ ID NO: 2 or the mature polypeptide of SEQ ID NO: 30 shown as amino acids 21-694 or SEQ ID NO: 2 shown as amino acids 21-697 of SEQ ID NO: 2.

6. The polypeptide of any of paragraphs 1-6, which is a variant of the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 comprising a substitution, deletion, and/or insertion at one or more positions.

7. The polypeptide of any of paragraphs 1-6, which is a fragment of SEQ ID NO: 30 or SEQ ID NO: 2, wherein the fragment has trehalase activity.

8. The polypeptide of any of paragraphs 1-7, having at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to the mature polypeptide of SEQ ID NO: 4.

9. The polypeptide of any of paragraphs 1-8, which is encoded by a polynucleotide that hybridizes under very high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 3, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii).

10. The polypeptide of any of paragraphs 1-9, which is encoded by a polynucleotide having at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or 100% sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 3 or the cDNA sequence thereof.

11. The polypeptide of any of paragraphs 1-10, comprising or consisting of SEQ ID NO: 4 or the mature polypeptide of SEQ ID NO: 4 shown as amino acids 21-690 of SEQ ID NO: 4.

12. The polypeptide of any of paragraphs 1-11, which is a variant of the mature polypeptide of SEQ ID NO: 4 comprising a substitution, deletion, and/or insertion at one or more positions.

5 13. The polypeptide of any of paragraphs 1-12, which is a fragment of SEQ ID NO: 4 wherein the fragment has trehalase activity.

14. A composition comprising the polypeptide of any of paragraphs 1-13.

10 15. A whole broth formulation or cell culture composition comprising the polypeptide of any of paragraphs 1-14.

16. A polynucleotide encoding the polypeptide of any of paragraphs 1-14.

15 17. A nucleic acid construct or expression vector comprising the polynucleotide of paragraph 16 operably linked to one or more control sequences that direct the production of the polypeptide in an expression host.

20 18. A recombinant host cell comprising the polynucleotide of paragraph 16 operably linked to one or more control sequences that direct the production of the polypeptide.

25 19. A method of producing the polypeptide of any of paragraphs 1-14, comprising cultivating a cell, which in its wild-type form produces the polypeptide, under conditions conducive for production of the polypeptide.

20. The method of paragraph 19, further comprising recovering the polypeptide.

30 21. A method of producing a polypeptide having trehalase activity, comprising cultivating the host cell of paragraph 18 under conditions conducive for production of the polypeptide.

22. The method of paragraph 21, further comprising recovering the polypeptide.

23. A transgenic plant, plant part or plant cell transformed with a polynucleotide encoding the polypeptide of any of paragraphs 1-14.

35

24. A method of producing a polypeptide having trehalase activity, comprising cultivating the transgenic plant or plant cell of paragraph 21 under conditions conducive for production of the polypeptide.

5 25. The method of paragraph 24, further comprising recovering the polypeptide.

26. A process of producing a fermentation product, comprising

(a) liquefying a starch-containing material with an alpha-amylase;
optionally pre-saccharifying the liquefied material before step (b);

10 (b) saccharifying the liquefied material;

(c) fermenting using a fermentation organism;
wherein

i) a glucoamylase;

ii) a trehalase of any of paragraphs 1-13;

15 iii) optionally a cellulolytic enzyme composition and/or a protease;

are present and/or added during

- saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- 20 - optionally presaccharification step before step (b).

27. The process of paragraph 26, wherein the alpha-amylase is a bacterial alpha-amylase, in particular of the genus *Bacillus*, such as a strain of *Bacillus stearothermophilus*, in particular a variant of a *Bacillus stearothermophilus* alpha-amylase, such as the one shown in SEQ ID NO:
25 3 in WO 99/019467 or SEQ ID NO: 18 herein.

28. The process of paragraph 27, wherein the *Bacillus stearothermophilus* alpha-amylase or variant thereof is truncated, preferably to be from 485-495 amino acids long, such as around 491 amino acids long.

29. The process of any of paragraphs 26 or 28, wherein the *Bacillus stearothermophilus* alpha-amylase has a double deletion at positions I181 + G182, and optionally a N193F substitution, or deletion of R179 + G180 (using SEQ ID NO: 18 for numbering).

30. The process of any of paragraphs 26-29, wherein the *Bacillus stearothermophilus* alpha-amylase has a substitution in position S242, preferably a S242A, E or Q substitution (using SEQ ID NO: 18 for numbering).

31. The process of any of paragraphs 26-30, wherein the *Bacillus stearrowthermophilus* alpha-amylase has a substitution in position E188, preferably E188P substitution (using SEQ ID NO: 18 for numbering).

5

32. The process of any of paragraphs 26-31, wherein the alpha-amylase in liquefaction step (a) is selected from the following group of *Bacillus stearrowthermophilus* alpha-amylase variants:

- I181*+G182*+N193F+E129V+K177L+R179E;

- I181*+G182*+N193F+V59A+Q89R+E129V+K177L+R179E+H208Y+K220P+N224L+Q254S

10 - I181*+G182*+N193F+V59A+Q89R+E129V+K177L+R179E+Q254S+M284V;

- I181*+G182*+N193F+V59A+E129V+K177L+R179E+Q254S+M284V and

- I181*+G182*+N193F+E129V+K177L+R179E+K220P+N224L+S242Q+Q254S (using SEQ ID NO: 18 herein for numbering).

15 33. The process of any of paragraphs 26-32, wherein the glucoamylase is of fungal origin, preferably from a strain of *Aspergillus*, preferably *A. niger*, *A. awamori*, or *A. oryzae*; or a strain of *Trichoderma*, preferably *T. reesei*; or a strain of *Talaromyces*, preferably *T. emersonii*, or a strain of *Gloeophyllum*, such as *G. serpiarium* or *G. trabeum*.

20 34. The process of any of paragraphs 26-33, wherein the glucoamylase is derived from *Talaromyces emersonii*, such as the one shown in SEQ ID NO: 13 herein.

35. The process of any of paragraphs 26-34, wherein the glucoamylase is selected from the group consisting of:

25 (i) a glucoamylase comprising the mature polypeptide of SEQ ID NO: 13 herein;

(ii) a glucoamylase comprising an amino acid sequence having at least 60%, at least 70%, e.g., at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the mature polypeptide of SEQ ID NO: 13 herein.

30

36. The process of any of paragraphs 26-35, wherein the glucoamylase is derived from *Gloeophyllum serpiarium*, such as the one shown in SEQ ID NO: 19 herein.

37. The process of any of paragraphs 26-36, wherein the glucoamylase is selected from the group consisting of:

35

(i) a glucoamylase comprising the mature polypeptide of SEQ ID NO: 19 herein;

(ii) a glucoamylase comprising an amino acid sequence having at least 60%, at least 70%, e.g., at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the mature polypeptide of SEQ ID NO: 19 herein.

5

38. The process of any of paragraphs 26-37, wherein the glucoamylase is derived from *Gloeophyllum trabeum* such as the one shown in SEQ ID NO: 20 herein.

39. The process of any of paragraphs 26-38, wherein the glucoamylase present and/or added in saccharification is selected from the group consisting of:

10

- (i) a glucoamylase comprising the mature polypeptide of SEQ ID NO: 20 herein;
- (ii) a glucoamylase comprising an amino acid sequence having at least 60%, at least 70%, e.g., at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the mature polypeptide of SEQ ID NO: 20 herein.

15

40. The process of any of paragraphs 26-39, further wherein an alpha-amylase is present and/or added during saccharification step (b); fermentation step (c); simultaneous saccharification and fermentation; or the optional presaccharification step before step (b).

20

41. The process of paragraph 40, wherein the alpha-amylase is of fungal or bacterial origin.

42. The process of paragraph 40 or 41, wherein the alpha-amylase is derived from a strain of the genus *Rhizomucor*, preferably a strain the *Rhizomucor pusillus*, such as the one shown in SEQ ID NO: 3 in WO 2013/006756, such as a *Rhizomucor pusillus* alpha-amylase hybrid having an *Aspergillus niger* linker and starch-bonding domain, such as the one shown in SEQ ID NO: 15 herein.

25

43. The process of any of paragraphs 40-42, wherein the alpha-amylase present and/or added in saccharification and/or fermentation is selected from the group consisting of:

30

- (i) an alpha-amylase comprising the mature polypeptide of SEQ ID NO: 15 herein;
- (ii) an alpha-amylase comprising an amino acid sequence having at least 60%, at least 70%, e.g., at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the mature polypeptide of SEQ ID NO: 15 herein.

35

44. The process of any of paragraphs 40-43, wherein the alpha-amylase is a variant of the alpha-amylase shown in SEQ ID NO: 15 having at least one of the following substitutions or combinations of substitutions: D165M; Y141W; Y141R; K136F; K192R; P224A; P224R; S123H + Y141W; G20S + Y141W; A76G + Y141W; G128D + Y141W; G128D + D143N; P219C + Y141W; N142D + D143N; Y141W + K192R; Y141W + D143N; Y141W + N383R; Y141W + P219C + A265C; Y141W + N142D + D143N; Y141W + K192R V410A; G128D + Y141W + D143N; Y141W + D143N + P219C; Y141W + D143N + K192R; G128D + D143N + K192R; Y141W + D143N + K192R + P219C; G128D + Y141W + D143N + K192R; or G128D + Y141W + D143N + K192R + P219C (using SEQ ID NO: 15 for numbering).

45. The process of any of paragraphs 36-44, wherein the alpha-amylase is derived from a *Rhizomucor pusillus*, in particular with an *Aspergillus niger* glucoamylase linker and starch-binding domain (SBD), preferably the one disclosed as SEQ ID NO: 15 herein, preferably having one or more of the following substitutions: G128D, D143N, preferably G128D+D143N (using SEQ ID NO: 15 for numbering).

46. The process of any of paragraphs 44-45, wherein the alpha-amylase variant has at least 60% identity, such as at least 70%, preferably at least 75% identity, preferably at least 80%, more preferably at least 85%, more preferably at least 90%, more preferably at least 91%, more preferably at least 92%, even more preferably at least 93%, most preferably at least 94%, and even most preferably at least 95%, such as even at least 96%, at least 97%, at least 98%, at least 99%, but less than 100% identity to the mature part of the polypeptide of SEQ ID NO: 15 herein.

47. The process of any of paragraphs 26-46, wherein the cellulolytic enzyme composition is derived from *Trichoderma reesei*, *Humicola insolens* or *Chrysosporium lucknowense*.

49. The process of any of paragraphs 26-48, wherein the cellulolytic enzyme composition comprising a beta-glucosidase, a cellobiohydrolase, an endoglucanase and optionally a GH61 polypeptide.

50. The process of any of paragraph 26-49, wherein the cellulolytic enzyme composition comprises a beta-glucosidase, preferably one derived from a strain of the genus *Aspergillus*, such as *Aspergillus oryzae*, such as the one disclosed in WO 2002/095014 or the fusion protein having beta-glucosidase activity disclosed in WO 2008/057637, or *Aspergillus fumigatus*, such as one disclosed in SEQ ID NO: 2 in WO 2005/047499 or SEQ ID NO: 10 herein or an *Aspergillus fumigatus* beta-glucosidase variant disclosed in WO 2012/044915; in particular an

Aspergillus fumigatus beta-glucosidase variant with one or more, such as all, of the following substitutions: F100D, S283G, N456E, F512Y; or a strain of the genus a strain *Penicillium*, such as a strain of the *Penicillium brasilianum* disclosed in WO 2007/019442, or a strain of the genus *Trichoderma*, such as a strain of *Trichoderma reesei*.

5

51. The process of any of paragraphs 26-50, wherein the cellulolytic enzyme composition comprises a cellobiohydrolase I (CBH I), such as one derived from a strain of the genus *Aspergillus*, such as a strain of *Aspergillus fumigatus*, such as the Cel7a CBH I disclosed in SEQ ID NO: 6 in WO 2011/057140 or SEQ ID NO: 6 herein, or a strain of the genus

10

Trichoderma, such as a strain of *Trichoderma reesei*.

52. The process of any of paragraphs 26-50, wherein the cellulolytic enzyme composition comprises a cellobiohydrolase II (CBH II), such as one derived from a strain of the genus *Aspergillus*, such as a strain of *Aspergillus fumigatus*; such as the one disclosed as SEQ ID

15

NO: 8 herein or a strain of the genus *Trichoderma*, such as *Trichoderma reesei*, or a strain of the genus *Thielavia*, such as a strain of *Thielavia terrestris*, such as cellobiohydrolase II CEL6A from *Thielavia terrestris*.

53. The process of any of paragraphs 26-52, wherein the cellulolytic enzyme composition further comprises a GH61 polypeptide having cellulolytic enhancing activity such as one derived from the genus *Thermoascus*, such as a strain of *Thermoascus aurantiacus*, such as the one described in WO 2005/074656 as SEQ ID NO: 2 or SEQ ID NO: 21 herein; or one derived from the genus *Thielavia*, such as a strain of *Thielavia terrestris*, such as the one described in WO 2005/074647 as SEQ ID NO: 7 and SEQ ID NO: 8; or one derived from a

20

25

strain of *Aspergillus*, such as a strain of *Aspergillus fumigatus*, such as the one described in WO 2010/138754 as SEQ ID NO: 1 and SEQ ID NO: 2; or one derived from a strain derived from *Penicillium*, such as a strain of *Penicillium emersonii*, such as the one disclosed in WO 2011/041397 as SEQ ID NO: 2 or SEQ ID NO: 12 herein.

54. The process of any of paragraphs 26-54, wherein the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* GH61A polypeptide having cellulolytic enhancing activity (SEQ ID NO: 2 in WO 2005/074656 or SEQ ID NO: 21 herein) and *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 of WO 2005/047499) or SEQ ID NO: 6 herein.

30

35

55. The process of any of paragraphs 26-54, wherein the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition further comprising *Penicillium emersonii*

GH61A polypeptide having cellulolytic enhancing activity disclosed in WO 2011/041397 as SEQ ID NO: 2 or SEQ ID NO: 12 herein; and *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 of WO 2005/047499 or SEQ ID NO: 10 herein) or a variant thereof with one or more, such as all, of the following substitutions: F100D, S283G, N456E, F512Y.

5

56. The process of any of paragraphs 26-55, wherein the cellulolytic enzyme composition is dosed from 0.0001-3 mg EP/g DS, preferably 0.0005-2 mg EP/g DS, preferably 0.001-1 mg/g DS, more preferred from 0.005-0.5 mg EP/g DS, even more preferred 0.01-0.1 mg EP/g DS.

10

57. The process of any of paragraphs 26-56, wherein the presaccharification is carried out at a temperature from 40-75°C, such as 50-70°C, preferably 60°C; a pH between 4-6, preferably 5; for a period of 30-360 minutes, such as from 60-420 minutes, such as around between 150-180 minutes.

15

58. A process of any of paragraphs 26-57, comprising the steps of :

- (a) liquefying a starch-containing material with an alpha-amylase; optionally pre-saccharifying the liquefied material before step (b);
- (b) saccharifying the liquefied material;
- (c) fermenting using a fermentation organism;

20

wherein

- i) a glucoamylase;
- ii) a trehalase of any of paragraphs 1-13;

are present and/or added during

25

- saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

59. A process of any of paragraphs 26-58, comprising the steps of :

30

- (a) liquefying a starch-containing material with an alpha-amylase; optionally pre-saccharifying the liquefied material before step (b);
- (b) saccharifying the liquefied material;
- (c) fermenting using a fermentation organism;

wherein

35

- i) a glucoamylase from *Talaromyces emersonii* or *Gloeophyllum serpiarium*;

ii) a trehalase shown in any of paragraphs 1-13;
are present and/or added during

- saccharification step (b);
- fermentation step (c);
- 5 - simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

60. A process of any of paragraphs 26-58, comprising the steps of :

- (a) liquefying a starch-containing material with an alpha-amylase;
- 10 optionally pre-saccharifying the liquefied material before step (b);
- (b) saccharifying the liquefied material;
- (c) fermenting using a fermentation organism;

wherein

- i) a glucoamylase from *Talaromyces emersonii* or *Gloeophyllum*
- 15 *serpiarium*;
- ii) a trehalase shown any of paragraphs 1-13;
- iii) a cellulolytic enzyme composition derived from *Trichoderma reesei*;

are present and/or added during

- saccharification step (b);
- 20 - fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

61. The process of any of paragraphs 26-60, wherein saccharification step (a) and
25 fermentation step (b) are done separately or simultaneously.

62. The process of any of paragraphs 26-61, wherein the fermentation product is recovered
after fermentation.

30 63. The process of any of paragraphs 26-62, wherein the starch-containing material is plant
material selected from the corn (maize), cobs, wheat, barley, rye, milo, sago, cassava, tapioca,
sorghum, rice, peas, beans, sweet potatoes, or a mixture thereof, preferably corn.

64. The process of any of paragraphs 26-63, wherein the temperature in liquefaction is
35 above the initial gelatinization temperature, in particular in the range from 70-100°C, such as
between 75-95°C, such as between 75-90°C, preferably between 80-90°C, such as 82-88°C,
such as around 85°C.

65. The process of any of paragraphs 26-64, wherein liquefaction step (a) is carried out at a pH in the range between 3 and 7, preferably from 4 to 6, or more preferably from 4.5 to 5.5.

66. The process of any of paragraphs 26-65, wherein the dry solid content (DS) in liquefaction lies in the range from 20-55 wt.-%, preferably 25-45 wt.-%, more preferably 30-40 wt.-% or 30-45 wt.-%.

66. The process of any of paragraphs 26-65, further comprises, prior to the liquefaction step (a), the steps of:

- x) reducing the particle size of the starch-containing material, preferably by dry milling;
- y) forming a slurry comprising the starch-containing material and water.

67. The process of any of paragraphs 26-66, wherein a jet-cooking step is carried out prior to liquefaction in step (a).

68. The process of any of paragraphs 26-67, wherein the starch-containing material is reduced in particle size, such by dry milling or wet milling or using particle size emulsion technology.

69. The process of any of paragraphs 26-68, wherein the fermentation is carried out for 30 to 150 hours, preferably 48 to 96 hours.

70. The process of any of paragraphs 26-69, wherein the temperature during fermentation in step (b) or simultaneous saccharification and fermentation in steps (a) and (b) is between 25°C and 40°C, preferably between 28°C and 36°C, such as between 28°C and 35°C, such as between 28°C and 34°C, such as around 32°C.

71. The process of any of paragraphs 26-70, wherein further a protease is present during saccharification and/or fermentation.

72. The process of any of paragraphs 26-71, wherein glucoamylase is present and/or added in an amount of 0.001 to 10 AGU/g DS, preferably from 0.01 to 5 AGU/g DS, especially 0.1 to 0.5 AGU/g DS.

73. The process of any of paragraphs 26-72, wherein the fermentation product is an alcohol, preferably ethanol, especially fuel ethanol, potable ethanol and/or industrial ethanol.

74. The process of any of paragraphs 26-73, further wherein a protease is present and/or added during

- saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

75. The process of any of paragraphs 26-74, wherein the protease is derived from *Thermoascus*, in particular *Thermoascus aurantiacus*, especially *Thermoascus aurantiacus* CGMCC No. 0670 (classified as EC 3.4.24.39) shown in SEQ ID NO: 17 herein.

76. The process of paragraph 75, wherein the protease is the one shown in SEQ ID NO: 17 herein or a protease being at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 96%, such as at least 97%, as as at least 98%, such as at least 99% identical to SEQ ID NO: 17 herein.

77. The process of any of paragraphs 26-74, wherein the protease is derived from a strain of *Meripilus*, in particular *Meripilus giganteus*, in particular the one shown as SEQ ID NO: 32 herein.

78. The process of paragraph 77, wherein the protease is the one shown in SEQ ID NO: 32 herein or a protease being at least 60%, such as at least 70%, such as at least 80%, such as at least 90%, such as at least 95%, such as at least 96%, such as at least 97%, as as at least 98%, such as at least 99% identical to SEQ ID NO: 32 herein.

79. The process of any of paragraphs 26-78, wherein the fermenting organism is derived from a strain of *Saccharomyces*, such as *Saccharomyces cerevisiae*.

80. A process of producing fermentation products from starch-containing material comprising:
(i) saccharifying a starch-containing material at a temperature below the initial gelatinization temperature; and
(ii) fermenting using a fermentation organism;
wherein saccharification and/or fermentation is done in the presence of the following enzymes: glucoamylase, alpha-amylase, trehalase of any of paragraphs 1-13, and optionally a protease and/or a cellulytic enzyme composition.

81. A process of producing a fermentation product from pretreated cellulosic material, comprising:

(a) hydrolyzing said pretreated cellulosic material with a cellulolytic enzyme composition;

(b) fermenting using a fermenting organism; and

(c) optionally recovering the fermentation product,

wherein a trehalase of any of paragraphs 1-13 is added and/or present in hydrolysis step (a) and/or fermentation step (b).

82. The process of any of paragraphs 26-81, wherein the trehalase is added in an amount between 0.01-20 ug EP trehalase/g DS, such as between 0.05-15 ug EP terhalase/g DS, such as between 0.5 and 10 ug EP trehalase/g DS.

83. The process of any of paragraphs 81-82, wherein the cellulolytic enzyme composition is derived from *Trichoderma reesei*, *Humicola insolens* or *Chrysosporium lucknowense*.

84. The process of any of paragraphs 81-83, wherein the cellulolytic enzyme composition comprising a beta-glucosidase, a cellobiohydrolase, an endoglucanase and optionally a GH61 polypeptide.

85. The process of any of paragraph 81-84, wherein the cellulolytic enzyme composition comprises a beta-glucosidase, preferably one derived from a strain of the genus *Aspergillus*, such as *Aspergillus oryzae*, such as the one disclosed in WO 2002/095014 or the fusion protein having beta-glucosidase activity disclosed in WO 2008/057637, or *Aspergillus fumigatus*, such as one disclosed in SEQ ID NO: 2 in WO 2005/047499 or SEQ ID NO: 10 herein or an *Aspergillus fumigatus* beta-glucosidase variant disclosed in WO 2012/044915; in particular an *Aspergillus fumigatus* beta-glucosidase variant with one or more, such as all, of the following substitutions: F100D, S283G, N456E, F512Y; or a strain of the genus a strain *Penicillium*, such as a strain of the *Penicillium brasilianum* disclosed in WO 2007/019442, or a strain of the genus *Trichoderma*, such as a strain of *Trichoderma reesei*.

86. The process of any of paragraphs 81-85, wherein the cellulolytic enzyme composition comprises a cellobiohydrolase I (CBH I), such as one derived from a strain of the genus *Aspergillus*, such as a strain of *Aspergillus fumigatus*, such as the Cel7a CBH I disclosed in SEQ ID NO: 6 in WO 2011/057140 or SEQ ID NO: 6 herein, or a strain of the genus *Trichoderma*, such as a strain of *Trichoderma reesei*.

87. The process of any of paragraphs 81-86, wherein the cellulolytic enzyme composition comprises a cellobiohydrolase II (CBH II, such as one derived from a strain of the genus *Aspergillus*, such as a strain of *Aspergillus fumigatus*; such as the one disclosed as SEQ ID NO: 8 herein or a strain of the genus *Trichoderma*, such as *Trichoderma reesei*, or a strain of the genus *Thielavia*, such as a strain of *Thielavia terrestris*, such as cellobiohydrolase II CEL6A from *Thielavia terrestris*.

88. The process of any of paragraphs 81-87, wherein the cellulolytic enzyme composition further comprises a GH61 polypeptide having cellulolytic enhancing activity such as one derived from the genus *Thermoascus*, such as a strain of *Thermoascus aurantiacus*, such as the one described in WO 2005/074656 as SEQ ID NO: 2 or SEQ ID NO: 21 herein; or one derived from the genus *Thielavia*, such as a strain of *Thielavia terrestris*, such as the one described in WO 2005/074647 as SEQ ID NO: 7 and SEQ ID NO: 8; or one derived from a strain of *Aspergillus*, such as a strain of *Aspergillus fumigatus*, such as the one described in WO 2010/138754 as SEQ ID NO: 1 and SEQ ID NO: 2; or one derived from a strain derived from *Penicillium*, such as a strain of *Penicillium emersonii*, such as the one disclosed in WO 2011/041397 as SEQ ID NO: 2 or SEQ ID NO: 12 herein.

89. The process of any of paragraphs 81-88, wherein the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition, further comprising *Thermoascus aurantiacus* GH61A polypeptide having cellulolytic enhancing activity (SEQ ID NO: 2 in WO 2005/074656 or SEQ ID NO: 21 herein) and *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 of WO 2005/047499) or SEQ ID NO: 6 herein.

90. The process of any of paragraphs 81-89, wherein the cellulolytic enzyme composition is a *Trichoderma reesei* cellulolytic enzyme composition further comprising *Penicillium emersonii* GH61A polypeptide having cellulolytic enhancing activity disclosed in WO 2011/041397 as SEQ ID NO: 2 or SEQ ID NO: 12 herein; and *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 2 of WO 2005/047499 or SEQ ID NO: 10 herein) or a variant thereof with one or more, such as all, of the following substitutions: F100D, S283G, N456E, F512Y.

91. The process of any of paragraphs 81-90, wherein the cellulolytic enzyme composition is dosed from 0.0001-3 mg EP/g DS, preferably 0.0005-2 mg EP/g DS, preferably 0.001-1 mg/g DS, more preferred from 0.005-0.5 mg EP/g DS, even more preferred 0.01-0.1 mg EP/g DS.

Claims

1. A polypeptide having trehalase activity, selected from the group consisting of:

5 (a) a polypeptide having at least 90% sequence identity to the mature polypeptide of SEQ ID NO: 30 or SEQ ID NO: 2 or at least 80% sequence identity to the mature polypeptide of SEQ ID NO: 4;

(b) a polypeptide encoded by a polynucleotide that hybridizes under high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 29, SEQ ID NO: 1 or
10 SEQ ID NO: 3, (ii) the cDNA sequence thereof, or (iii) the full-length complement of (i) or (ii);

(c) a polypeptide encoded by a polynucleotide having at least 60% sequence identity to the mature polypeptide coding sequence of SEQ ID NO: 29, SEQ ID NO: 1 or SEQ ID NO: 3 or the cDNA sequence thereof;

(d) a variant of the mature polypeptide of SEQ ID NO: 30, SEQ ID NO: 2 or SEQ ID
15 NO: 4 comprising a substitution, deletion, and/or insertion at one or more (e.g., several) positions; and

(e) a fragment of the polypeptide of (a), (b), (c), or (d) that has trehalase activity.

2. A composition comprising the polypeptide of claim 1.

3. A whole broth formulation or cell culture composition comprising the polypeptide of claim 1.

4. A polynucleotide encoding the polypeptide of claim 1.

25 5. A nucleic acid construct or expression vector comprising the polynucleotide of claim 4 operably linked to one or more control sequences that direct the production of the polypeptide in an expression host.

30 6. A recombinant host cell comprising the polynucleotide of claim 4 operably linked to one or more control sequences that direct the production of the polypeptide.

7. The recombinant host cell of claim 6, wherein the recombinant host cell is a yeast, such as a strain of *Saccharomyces*, in particular *Saccharomyces cerevisiae*.

35 8. A method of producing the polypeptide of claim 1, comprising cultivating a cell, which in its wild-type form produces the polypeptide, under conditions conducive for production of the polypeptide.

9. A method of producing a polypeptide having trehalase activity, comprising cultivating the host cell of claims 6 or 7 under conditions conducive for production of the polypeptide.

10. A transgenic plant, plant part or plant cell transformed with a polynucleotide encoding the polypeptide of claim 1.

11. A method of producing a polypeptide having trehalase activity, comprising cultivating the transgenic plant or plant cell of claim 10 under conditions conducive for production of the polypeptide.

12. A process of producing a fermentation product, comprising

(a) liquefying a starch-containing material with an alpha-amylase;
optionally pre-saccharifying the liquefied material before step (b);

(b) saccharifying the liquefied material;

(c) fermenting using a fermentation organism;

wherein

i) a glucoamylase;

ii) a trehalase of claim 1;

iii) optionally a cellulolytic enzyme composition and/or a protease;

are present and/or added during

- saccharification step (b);
- fermentation step (c);
- simultaneous saccharification and fermentation;
- optionally presaccharification step before step (b).

13. A process of producing fermentation products from starch-containing material comprising:

(i) saccharifying a starch-containing material at a temperature below the initial gelatinization temperature; and

(ii) fermenting using a fermentation organism;

wherein saccharification and/or fermentation is done in the presence of the following enzymes: glucoamylase, alpha-amylase, trehalase of claim 1, and optionally a protease and/or a cellulolytic enzyme composition.

14. A process of producing a fermentation product from pretreated cellulosic material, comprising:

(a) hydrolyzing said pretreated cellulosic material with a cellulolytic enzyme composition;

(b) fermenting using a fermenting organism; and

(c) optionally recovering the fermentation product,

5 wherein a trehalase of claim 1 is added and/or present in hydrolysis step (a) and/or fermentation step (b).

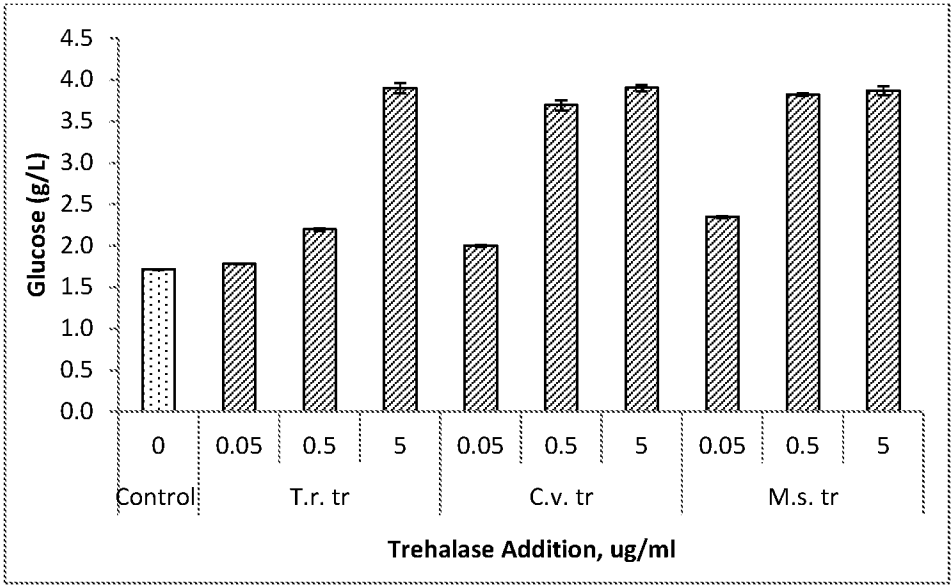


Fig. 1

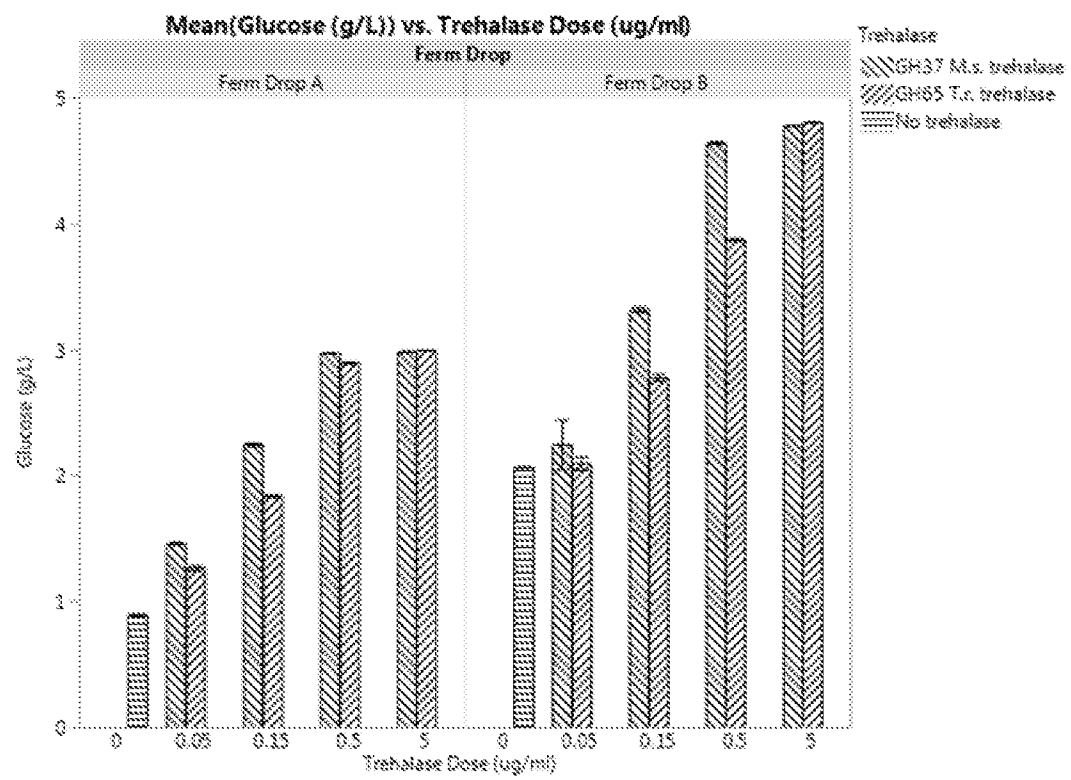


Fig. 2

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/037224

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
- a. ☒ forming part of the international application as filed:
- ☒ in the form of an Annex C/ST.25 text file.
- ☐ on paper or in the form of an image file.
- b. ☐ furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
- c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
- ☐ in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
- ☐ on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/037224

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-14(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/037224A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/24
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)
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, Sequence Search, BIOSIS, 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	DATABASE UniProt [Online] 16 November 2011 (2011-11-16), "RecName: Full=Trehalase {ECO:0000256 RuleBase:RU361180}; EC=3.2.1.28 {ECO:0000256 RuleBase:RU361180}; AltName: Full=Alpha-trehalose glucohydrolase {ECO:0000256 RuleBase:RU361180};", XP002760542, retrieved from EBI accession no. UNIPROT:G2PZS2 Database accession no. G2PZS2 sequence -/--	1-14



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

4 August 2016

Date of mailing of the international search report

23/11/2016

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

Chavanne, Franz

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2016/037224

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	& RANDY M BERKA ET AL: "Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris", NATURE BIOTECHNOLOGY, vol. 29, no. 10, 1 January 2011 (2011-01-01), pages 922-927, XP055034838, ISSN: 1087-0156, DOI: 10.1038/nbt.1976 the whole document ----- DATABASE UniProt [Online] 16 November 2011 (2011-11-16), "RecName: Full=Trehalase {ECO:0000256 RuleBase:RU361180}; EC=3.2.1.28 {ECO:0000256 RuleBase:RU361180}; AltName: Full=Alpha-trehalose glucohydrolase {ECO:0000256 RuleBase:RU361180};", XP002760543, retrieved from EBI accession no. UNIPROT:G2RDT9 Database accession no. G2RDT9 sequence & Randy M Berka ET AL: "Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris", Nature Biotechnology, vol. 29, no. 10, 1 January 2011 (2011-01-01), pages 922-927, XP055034838, ISSN: 1087-0156, DOI: 10.1038/nbt.1976 -----	1-14
X	WO 2012/027374 A2 (DYADIC INTERNAT USA INC [US]; VISSER JOHANNES [NL]; HINZ SANDRA [NL];) 1 March 2012 (2012-03-01) abstract paragraphs [0019], [0020], [0030], [0052], [0053], [0174] claims 1, 7, 8, 65, 66 -----	1-14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2016/037224

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012027374	A2	01-03-2012	NONE

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-14(partially)

Polypeptide having at least 90% sequence identity with SEQ ID No.2 or 30; composition comprising said polypeptide, polynucleotide encoding said polypeptide, recombinant host cell comprising said polynucleotide, methods for the production of said polypeptide, transgenic plant, plant part or plant cell transformed with said polynucleotide, processes for producing fermentation products using said polypeptide.

2. claims: 1-14(partially)

Polypeptide having at least 80% sequence identity with SEQ ID No.4; composition comprising said polypeptide, polynucleotide encoding said polypeptide, recombinant host cell comprising said polynucleotide, methods for the production of said polypeptide, transgenic plant, plant part or plant cell transformed with said polynucleotide, processes for producing fermentation products using said polypeptide.
