

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 August 2010 (12.08.2010)

PCT

(10) International Publication Number
WO 2010/091221 A1

- (51) **International Patent Classification:**
C12N 9/26 (2006.01)
- (21) **International Application Number:**
PCT/US20 10/023272
- (22) **International Filing Date:**
5 February 2010 (05.02.2010)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
61/150,528 6 February 2009 (06.02.2009) US
- (71) **Applicants** (for all designated States except US):
NOVOZYMES A/S [DK/DK]; Krogshøjvej 36,
DK-2880 Bagsvaerd (DK). **NOVOZYMES NORTH
AMERICA, INC.** [US/US]; 77 Perry Chapel Church
Road, Franklinton, North Carolina 27525 (US).
- (72) **Inventors; and**
- (75) **Inventors/Applicants** (for US only): **LIU, Jim** [US/US];
2100 Wide River Drive, Raleigh, North Carolina 27614
(US). **LI, Ming** [CN/CN]; 1# Qinghuayuan, Haidian Dis-
trict, Beijing 100080 (CN). **LAN, Tang** [CN/CN];
ZhiChun Road 57-1-707, Haidian District, Beijing
100080 (CN). **YE, Liu** [CN/CN]; #14 Zizhuyuan Lu,
Haidian District, Beijing 100044 (CN). **ZHENG, Liu**
[CN/CN]; 28 Xi Hua Men Da Jie, Xi Cheng District, Bei-
jing 100017 (CN).
- (74) **Agent:** LAMBIRIS, **Elias**; Novozymes North America,
Inc., 500 Fifth Avenue, Suite 1600, New York, New York
101 10 (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**, 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**, **IS**, **JP**, KE, KG, **KM**, KN, **KP**,
KR, KZ, LA, LC, **LK**, **LR**, LS, **LT**, LU, LY, MA, MD,
ME, MG, **MK**, MN, MW, MX, MY, MZ, NA, NG, **NI**,
NO, NZ, OM, PE, PG, **PH**, **PL**, **PT**, RO, RS, RU, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ,
TM), European (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, SE, SI, SK, SM,
TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
ML, MR, NE, SN, TD, TG).
- Published:**
- with international search report (Art. 21(3))
 - with sequence listing part of description (Rule 5.2(a))



WO 2010/091221 A1

(54) **Title:** POLYPEPTIDES HAVING ALPHA-AMYLASE ACTIVITY AND POLYNUCLEOTIDES ENCODING SAME

(57) **Abstract:** The present invention relates to isolated polypeptides having alpha-amylase activity and isolated 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 and using the polypeptides.

**POLYPEPTIDES HAVING ALPHA-AMYLASE ACTIVITY
AND POLYNUCLEOTIDES ENCODING SAME**

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 isolated polypeptides having alpha-amylase activity and isolated 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 and using the polypeptides.

15 **Description of the Related Art**

 Alpha-amylases (alpha-1,4-glucan-4-glucanohydrolases, EC. 3.2.1 .1) constitute a group of enzymes which catalyze hydrolysis of starch and other linear and branched 1,4-glucosidic oligo- and polysaccharides.

20 For a number of years alpha-amylase enzymes have been used for a variety of different purposes, the most important of which are starch liquefaction, textile desizing, textile washing, starch modification in the paper and pulp industry, and for brewing, ethanol production and baking.

25 The object of the present invention is to provide alpha-amylases for conversion of starch into maltodextrins, mono- and disaccharides and/or useful in processes involving starch liquefaction, textile washing, textile desizing, starch modification in the paper and pulp industry, and for brewing, ethanol production and baking.

 An polypeptide from *Aspergillus fumigatus* having alpha-amylase activity is disclosed in WO 2003/012071 (GeneseqP:ABB80178).

30 **Summary of the Invention**

 It is an object of the present invention to provide polypeptides having alpha-amylase activity and polynucleotides encoding the polypeptides. The polypeptides of the invention are derived from *Aspergillus terreus*. The polypeptides of the invention have improved properties such as increased stability at low pH, activity at over a broader pH spectrum, high specific
35 activity.

The present invention relates to an isolated polypeptide having alpha-amylase activity, selected from the group consisting of:

- (a) a polypeptide comprising an amino acid sequence having preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 85%, most preferably at least 90%, and even most preferably 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: 2 and/or the catalytic domain of SEQ ID NO: 2;
- (b) a polypeptide comprising an amino acid sequence having preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 85%, most preferably at least 90%, and even most preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the catalytic domain of SEQ ID NO: 2;
- (c) a polypeptide encoded by a polynucleotide that hybridizes under at least high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 and/or the catalytic domain coding sequence of SEQ ID NO: 1, (ii) a full-length complementary strand of (i);
- (d) a polypeptide encoded by a polynucleotide comprising a nucleotide sequence having preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 85%, most preferably at least 90%, and even most preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 and/or the catalytic domain coding sequence of SEQ ID NO: 1; and
- (e) a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2.

The present invention relates in a second aspect to an isolated polynucleotide comprising a nucleotide sequence that encodes the polypeptide of the first aspect.

In further aspects the present invention relates to a nucleic acid construct, a recombinant expression vector and a recombinant host cell comprising the isolated polynucleotide of the second aspect.

In still further aspects the present invention relates to a method of producing the polypeptide of the first aspect as well as a transgenic plant, plant part or plant cell transformed with a polynucleotide encoding the polypeptide of the first aspect.

In yet further aspects the present invention relates to a composition comprising the polypeptide of the first aspect as well as to uses of such a composition, including use for production of ethanol in a process comprising hydrolyzing an ungelatinized starch.

Definitions

Alpha-amylase activity: The term "alpha-amylase activity" is defined herein as an activity that catalyzes the endohydrolysis of $\rightarrow(4)$ -alpha-D-glucosidic linkages in polysaccharides containing three or more (1 \rightarrow 4)-alpha-linked D-glucose units. The term "alpha-amylase activity" corresponds to the enzymes grouped in E.C. 3.2.1.1. For purposes of the present invention, alpha-amylase activity is determined according to the procedure described in the "Materials and Methods": Acid alpha-amylase activity.

The polypeptides of the present invention have at least 20%, preferably at least 40%, more preferably at least 50%, more preferably at least 60%, more preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 100% of the alpha-amylase activity of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2.

Isolated polypeptide: The term "isolated polypeptide" as used herein refers to a polypeptide that is isolated from a source. In a preferred aspect, the polypeptide is at least 1% pure, preferably at least 5% pure, more preferably at least 10% pure, more preferably at least 20% pure, more preferably at least 40% pure, more preferably at least 60% pure, even more preferably at least 80% pure, and most preferably at least 90% pure, as determined by SDS-PAGE.

Substantially pure polypeptide: The term "substantially pure polypeptide" denotes herein a polypeptide preparation that contains at most 10%, preferably at most 8%, more preferably at most 6%, more preferably at most 5%, more preferably at most 4%, more preferably at most 3%, even more preferably at most 2%, most preferably at most 1%, and even most preferably at most 0.5% by weight of other polypeptide material with which it is natively or recombinantly associated. It is, therefore, preferred that the substantially pure polypeptide is at least 92% pure, preferably at least 94% pure, more preferably at least 95% pure, more preferably at least 96% pure, more preferably at least 96% pure, more preferably at least 97% pure, more preferably at least 98% pure, even more preferably at least 99%, most preferably at least 99.5% pure, and even most preferably 100% pure by weight of the total polypeptide material present in the preparation. The polypeptides of the present invention are preferably in a substantially pure form, *i.e.*, that the polypeptide preparation is essentially free of other polypeptide material with which it is natively or recombinantly associated. This can be accomplished, for example, by preparing the polypeptide by well-known recombinant methods or by classical purification methods.

Mature polypeptide: The term "mature polypeptide" is defined herein as a polypeptide having alpha-amylase activity that is in its final form following translation and any post-translational modifications, such as N-terminal processing, C-terminal truncation, glycosylation,

phosphorylation, etc. In SEQ ID NO: 2 the mature polypeptide is amino acids 1 to 587 and/or amino acids -20 to -1 of SEQ ID NO: 2 are a signal peptide.

Catalytic domain: The term "catalytic domain" is defined herein as an amino acid sequence comprising a GH13 domain. In SEQ ID NO: 2 amino acids 1 to 478 are the catalytic domain.

Carbohydrate binding domain: The term "carbohydrate binding domain" or "CBM" is defined herein as an amino acid sequence comprising a CBM of family 20, also known as a starch binding domain. In SEQ ID NO: 2, amino acids 480 to 587 are the CBM.

Polypeptide coding sequence: The term "coding sequence" is defined herein as a nucleotide sequence that encodes a polypeptide. In the present context the coding sequence includes the signal peptide coding sequence and the mature polypeptide coding sequence, i.e. SEQ ID NO: 1 excluding any introns.

Mature polypeptide coding sequence: The term "mature polypeptide coding sequence" is defined herein as a nucleotide sequence that encodes a mature polypeptide having alpha-amylase activity.

In SEQ ID NO: 1, the polypeptide coding sequence is nucleotides 24 to 188, 238 to 276, 326 to 441, 484 to 592, 644 to 872, 923 to 1085, 1135 to 1281, 1329 to 1569 and 1615 to 2226. The signal peptide coding sequence starts at nucleotide 24 and ends at nucleotide 83. The mature polypeptide coding sequence starts at nucleotide 84 and ends at nucleotide 2226.

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

For purposes of the present invention, the degree of 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 in Genetics* 16: 276-277), preferably version 3.0.0 or later. The optional 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:

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

For purposes of the present invention, the degree of identity between two deoxyribonucleotide sequences is determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *supra*) as implemented in the Needle program of the EMBOSS package (EMBOSS: The European Molecular Biology Open Software Suite, Rice *et al.*, 2000, *supra*), preferably version 3.0.0 or later. The optional parameters used are gap open penalty of 10, gap extension penalty of 0.5, and the EDNAFULL (EMBOSS version of NCBI NUC4.4)

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:

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

5 **Homologous sequence:** The term "homologous sequence" is defined herein as a predicted protein that gives an E value (or expectancy score) of less than 0.001 in a tfasty search (Pearson, W.R., 1999, in *Bioinformatics Methods and Protocols*, S. Misener and S. A. Krawetz, ed., pp. 185-219) with a specified sequence, e.g. the amino acid sequence of SEQ ID NO: 2.

10 Alternatively, the term "homologous sequence" is defined as a sequence having a degree of identity to a specified sequence, e.g. the nucleotide sequences in SEQ ID NO: 1, or the amino acid sequence of SEQ ID NO: 2, of preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 96%, at least 97%, at least 98%, or at least 99%.

15 **Homologous polypeptide:** The term "homologous polypeptide" is defined herein as a "homologous sequence" of amino acids, which have alpha-amylase activity.

Polypeptide fragment: The term "polypeptide fragment" is defined herein as a polypeptide having one or more (several) amino acids deleted from the amino and/or carboxyl terminus of the mature polypeptide of SEQ ID NO: 2; or a homologous sequence thereof; 20 wherein the fragment has alpha-amylase activity. In a preferred aspect, a fragment contains at least 400 amino acid residues, more preferably at least 425 amino acid residues, and most preferably at least 450 amino acid residues of the mature polypeptide of SEQ ID NO: 2 or a homologous sequence thereof.

Subsequence: The term "subsequence" is defined herein as a nucleotide sequence 25 having one or more (several) nucleotides deleted from the 5' and/or 3' end of the mature polypeptide coding sequence of SEQ ID NO: 1; or a homologous sequence thereof; wherein the subsequence encodes a polypeptide fragment having alpha-amylase activity. In a preferred aspect, a subsequence contains at least 1200 nucleotides, more preferably at least 1275 nucleotides, and most preferably at least 1350 nucleotides of the mature polypeptide coding 30 sequence of SEQ ID NO: 1 or a homologous sequence thereof.

Allelic variant: The term "allelic variant" denotes herein 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 35 amino acid sequences. An allelic variant of a polypeptide is a polypeptide encoded by an allelic variant of a gene.

Isolated polynucleotide: The term "isolated polynucleotide" as used herein refers to a polynucleotide that is isolated from a source. In a preferred aspect, the polynucleotide is at least

1% pure, preferably at least 5% pure, more preferably at least 10% pure, more preferably at least 20% pure, more preferably at least 40% pure, more preferably at least 60% pure, even more preferably at least 80% pure, and most preferably at least 90% pure, as determined by agarose electrophoresis.

5 **Substantially pure polynucleotide:** The term "substantially pure polynucleotide" as used herein refers to a polynucleotide preparation free of other extraneous or unwanted nucleotides and in a form suitable for use within genetically engineered protein production systems. Thus, a substantially pure polynucleotide contains at most 10%, preferably at most 8%, more preferably at most 6%, more preferably at most 5%, more preferably at most 4%,
10 more preferably at most 3%, even more preferably at most 2%, most preferably at most 1%, and even most preferably at most 0.5% by weight of other polynucleotide material with which it is natively or recombinantly associated. A substantially pure polynucleotide may, however, include naturally occurring 5' and 3' untranslated regions, such as promoters and terminators. It is preferred that the substantially pure polynucleotide is at least 90% pure, preferably at least
15 92% pure, more preferably at least 94% pure, more preferably at least 95% pure, more preferably at least 96% pure, more preferably at least 97% pure, even more preferably at least 98% pure, most preferably at least 99%, and even most preferably at least 99.5% pure by weight. The polynucleotides of the present invention are preferably in a substantially pure form, *i.e.*, that the polynucleotide preparation is essentially free of other polynucleotide material with
20 which it is natively or recombinantly associated. The polynucleotides may be of genomic, cDNA, RNA, semisynthetic, synthetic origin, or any combinations thereof.

Coding sequence: When used herein the term "coding sequence" means a nucleotide sequence, which directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by an open reading frame, which
25 usually begins with the ATG start codon or alternative start codons such as GTG and TTG and ends with a stop codon such as TAA, TAG, and TGA. The coding sequence may be a DNA, cDNA, synthetic, or recombinant nucleotide sequence.

Nucleic acid construct: The term "nucleic acid construct" as used herein refers to a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally
30 occurring gene or which is modified to contain segments of nucleic acids in a manner that would not otherwise exist in nature or which is synthetic. The term nucleic acid construct is synonymous with the term "expression cassette" when the nucleic acid construct contains the control sequences required for expression of a coding sequence of the present invention.

Control sequences: The term "control sequences" is defined herein to include all
35 components necessary for the expression of a polynucleotide encoding a polypeptide of the present invention. Each control sequence may be native or foreign to the nucleotide sequence encoding the polypeptide or native or foreign to each other. Such control sequences include, but are not limited to, a leader, polyadenylation 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 nucleotide sequence encoding a polypeptide.

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

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

Expression vector: The term "expression vector" is defined herein as a linear or circular DNA molecule that comprises a polynucleotide encoding a polypeptide of the present invention and is operably linked to additional nucleotides that provide for its expression.

Host cell: The term "host cell", as used herein, includes any cell type that is susceptible to transformation, transfection, transduction, and the like with a nucleic acid construct or expression vector comprising a polynucleotide of the present invention.

Modification: The term "modification" means herein any chemical modification of the polypeptide consisting of the mature polypeptide of SEQ ID NO: 2; or a homologous sequence thereof; as well as genetic manipulation of the DNA encoding such a polypeptide. The modification can be a substitution, a deletion and/or an insertion of one or more (several) amino acids as well as replacements of one or more (several) amino acid side chains.

Artificial variant: When used herein, the term "artificial variant" means a polypeptide having alpha-amylase activity produced by an organism expressing a modified polynucleotide sequence of the mature polypeptide coding sequence of SEQ ID NO: 1; or a homologous sequence thereof. The modified nucleotide sequence is obtained through human intervention by modification of the polynucleotide sequence disclosed in SEQ ID NO: 1; or a homologous sequence thereof.

Detailed Description of the Invention

Polypeptides having alpha-amylase activity

In a first aspect, the present invention relates to isolated polypeptides comprising an amino acid sequence having a degree of identity to the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2 of preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and

even most preferably at least 96%, at least 97%, at least 98%, or at least 99%, which have alpha-amylase activity (hereinafter "homologous polypeptides"). In a preferred embodiment of the first aspect, the homologous polypeptides have an amino acid sequence that differs by ten amino acids, preferably by five amino acids, more preferably by four amino acids, even more preferably by three amino acids, most preferably by two amino acids, and even most preferably by one amino acid from the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2.

A polypeptide of the present invention preferably comprises the amino acid sequence of SEQ ID NO: 2 or an allelic variant thereof; or a fragment thereof having alpha-amylase activity. In a preferred embodiment of the first aspect, the polypeptide comprises the amino acid sequence of SEQ ID NO: 2. In another preferred embodiment of the first aspect, the polypeptide comprises the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2. In another preferred embodiment of the first aspect, the polypeptide consists of the amino acid sequence of SEQ ID NO: 2 or an allelic variant thereof; or a fragment thereof having alpha-amylase activity. In another preferred embodiment of the first aspect, the polypeptide consists of the amino acid sequence of SEQ ID NO: 2. In another preferred embodiment of the first aspect, the polypeptide consists of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2. In another preferred aspect, the polypeptide consists of amino acids 1 to 587 of SEQ ID NO: 2 or an allelic variant thereof; or a fragment thereof having alpha-amylase activity. In another preferred embodiment of the first aspect, the polypeptide consists of amino acids 1 to 478 of SEQ ID NO: 2.

In a preferred embodiment of the first aspect, the present invention relates to isolated polypeptides having alpha-amylase activity that are encoded by polynucleotides that hybridize under preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with (i) the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, (ii) the genomic DNA sequence comprising] the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, (iii) a subsequence of (i) or (ii), or (iv) a full-length complementary strand of (i), (ii), or (iii) (J. Sambrook, E.F. Fritsch, and T. Maniatis, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York). A subsequence of the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1 contains at least 100 contiguous nucleotides or preferably at least 200 contiguous nucleotides. Moreover, the subsequence may encode a polypeptide fragment having alpha-amylase activity. In a preferred embodiment of the first aspect, the complementary strand is the full-length complementary strand of the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1.

The nucleotide sequence of SEQ ID NO: 1; or a subsequence thereof; as well as the amino acid sequence of SEQ ID NO: 2; or a fragment thereof; may be used to design nucleic acid probes to identify and clone DNA encoding polypeptides having alpha-amylase 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 or cDNA of the genus or species 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 14, preferably at least 25, more preferably at least 35, and most preferably at least 70 nucleotides in length. It is, however, preferred that the nucleic acid probe is at least 100 nucleotides in length. For example, the nucleic acid probe may be at least 200 nucleotides, preferably at least 300 nucleotides, more preferably at least 400 nucleotides, or most preferably at least 500 nucleotides in length. Even longer probes may be used, e.g., nucleic acid probes that are preferably at least 600 nucleotides, more preferably at least 700 nucleotides, even more preferably at least 800 nucleotides, or most preferably 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 prepared from such other strains may, therefore, be screened for DNA that hybridizes with the probes described above and encodes a polypeptide having alpha-amylase 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 is homologous with SEQ ID NO: 1; or a subsequence thereof; the carrier material is preferably used in a Southern blot.

For purposes of the present invention, hybridization indicates that the nucleotide sequence hybridizes to a labeled nucleic acid probe corresponding to the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1; the genomic DNA sequence comprising] the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1; its full-length complementary strand; or a subsequence thereof; under very low 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.

In a preferred embodiment, the nucleic acid probe is the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1. In another preferred embodiment, the nucleic acid probe is a polynucleotide sequence that encodes the polypeptide of SEQ ID NO: 2, or a subsequence thereof. In another preferred embodiment, the nucleic acid probe is SEQ ID NO: 1.

For long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 µg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures for 12 to 24 hours optimally.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS preferably at 45°C (very low stringency), more preferably at 50°C (low stringency), more preferably at 55°C (medium stringency), more preferably at 60°C (medium-high stringency), even more preferably at 65°C (high stringency), and most preferably at 70°C (very high stringency).

For short probes that are about 15 nucleotides to about 70 nucleotides in length, stringency conditions are defined as prehybridization, hybridization, and washing post-hybridization at about 5°C to about 10°C below the calculated T_m using the calculation according to Bolton and McCarthy (1962, *Proceedings of the National Academy of Sciences USA* 48:1390) in 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1X Denhardt's solution, 1 mM sodium pyrophosphate, 1 mM sodium monobasic phosphate, 0.1 mM ATP, and 0.2 mg of yeast RNA per ml following standard Southern blotting procedures for 12 to 24 hours optimally.

For short probes that are about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SSC plus 0.1% SDS for 15 minutes and twice each for 15 minutes using 6X SSC at 5°C to 10°C below the calculated T_m .

In an embodiment, the present invention relates to isolated polypeptides having alpha-amylase activity encoded by polynucleotides comprising or consisting of nucleotide sequences that have a degree of identity to the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1 of preferably at least 60%, more preferably at least 65%, more preferably at least 70%, more preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and even most preferably 96%, 97%, 98%, or 99%, which encode an active polypeptide. See polynucleotide section herein.

In an embodiment, the present invention relates to artificial variants comprising a substitution, deletion, and/or insertion of one or more (or several) amino acids of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2; or a homologous sequence thereof. Preferably, amino acid changes are 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 one to about 30 amino acids; small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue; a small linker peptide of up to about 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.

Examples of conservative substitutions are within the group 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 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. The most commonly occurring exchanges 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/Ala, Ala/Glu, and Asp/Gly.

5 In addition to the 20 standard amino acids, non-standard amino acids (such as 4-hydroxyproline, 6-*N*-methyl lysine, 2-aminoisobutyric acid, isovaline, and alpha-methyl serine) may be substituted for amino acid residues of a wild-type polypeptide. A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, and unnatural amino acids may be substituted for amino acid residues. "Unnatural amino acids" have been
10 modified after protein synthesis, and/or have a chemical structure in their side chain(s) different from that of the standard amino acids. Unnatural amino acids can be chemically synthesized, and preferably, are commercially available, and include pipercolic acid, thiazolidine carboxylic acid, dehydroproline, 3- and 4-methylproline, and 3,3-dimethylproline.

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 the parent polypeptide can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning
20 mutagenesis (Cunningham 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 mutant molecules are tested for biological activity (*i.e.*, alpha-amylase 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
25 determined by physical analysis of structure, as 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 identities of essential amino acids can also be inferred from
30 analysis of identities with polypeptides that are related to a polypeptide according to the invention.

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,
35 *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, phage display (e.g., Lowman *et al.*, 1991, *Biochem.* 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 of interest, and can be applied to polypeptides of unknown structure.

The total number of amino acid substitutions, deletions and/or insertions of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2, such as amino acids 1 to 587 of SEQ ID NO: 2, is 10, preferably 9, more preferably 8, more preferably 7, more preferably at most 6, more preferably 5, more preferably 4, even more preferably 3, most preferably 2, and even most preferably 1.

Sources of polypeptides having alpha-amylase activity

A polypeptide 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 nucleotide sequence is produced by the source or by a strain in which the nucleotide sequence from the source has been inserted. In a preferred an embodiment, the polypeptide obtained from a given source is secreted extracellularly.

A polypeptide having alpha-amylase activity of the present invention may be a fungal polypeptide, and more preferably a yeast polypeptide such as a *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* polypeptide having alpha-amylase activity; or more preferably a filamentous fungal polypeptide such as an *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Aureobasidium*, *Botryosphaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptosphaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania*, *Pseudotriconympha*, *Rhizomucor*, *Schizophyllum*, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trichoderma*, *Trichophaea*, *Verticillium*, *Volvariella*, or *Xylaria* polypeptide having alpha-amylase activity.

In a preferred an embodiment, the polypeptide is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*,

Saccharomyces kluyveri, *Saccharomyces norbensis*, or *Saccharomyces oviformis* polypeptide having alpha-amylase activity.

In another preferred an embodiment, the polypeptide is an *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*,
5 *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Aspergillus terreus* *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium tropicum*, *Chrysosporium merdarium*, *Chrysosporium inops*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium zonatum*, *Fusarium bacthdioides*, *Fusarium cerealis*, *Fusarium crookwellense*, *Fusarium culmorum*, *Fusarium graminearum*, *Fusarium graminum*,
10 *Fusarium heterosporum*, *Fusarium negundi*, *Fusarium oxysporum*, *Fusarium reticulatum*, *Fusarium roseum*, *Fusarium sambucinum*, *Fusarium sarcochroum*, *Fusarium sporotrichioides*, *Fusarium sulphureum*, *Fusarium torulosum*, *Fusarium trichothecioides*, *Fusarium venenatum*, *Humicola grisea*, *Humicola insolens*, *Humicola lanuginosa*, *Irpex lacteus*, *Mucor miehei*, *Myceliophthora thermophila*, *Neurospora crassa*, *Penicillium funiculosum*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Thielavia achromatica*, *Thielavia albomyces*,
15 *Thielavia albopilosa*, *Thielavia australeinsis*, *Thielavia fimeti*, *Thielavia microspora*, *Thielavia ovispora*, *Thielavia peruviana*, *Thielavia spededonium*, *Thielavia setosa*, *Thielavia subthermophila*, *Thielavia terrestris*, *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride* polypeptide having
20 having alpha-amylase activity.

In a more preferred an embodiment, the polypeptide is an *Aspergillus terreus* polypeptide having alpha-amylase activity, e.g., the polypeptide comprising the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2.

It will be understood that for the aforementioned species the invention encompasses
25 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
30 Mikroorganismen und Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

Furthermore, such polypeptides may be identified and obtained from other sources including microorganisms isolated from nature (e.g., soil, composts, water, etc.) using the
35 above-mentioned probes. Techniques for isolating microorganisms from natural habitats are well known in the art. The polynucleotide may then be obtained by similarly screening a genomic or cDNA library of such a microorganism. Once a polynucleotide sequence encoding a polypeptide has been detected with the probe(s), the polynucleotide can be isolated or cloned

by utilizing techniques that are well known to those of ordinary skill in the art (see, e.g., Sambrook *et al.*, 1989, *supra*).

Polypeptides of the present invention also include fused polypeptides or cleavable fusion polypeptides in which another polypeptide is fused at the N-terminus or the C-terminus of the polypeptide or fragment thereof. A fused polypeptide is produced by fusing a nucleotide sequence (or a portion thereof) encoding another polypeptide to a nucleotide sequence (or a portion thereof) 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 fused polypeptide is under control of the same promoter(s) and terminator.

A fusion polypeptide can further comprise a cleavage site. Upon secretion of the fusion protein, the site is cleaved releasing the polypeptide having alpha-amylase activity from the fusion protein. Examples of cleavage sites include, but are not limited to, a Kex2 site that encodes the dipeptide Lys-Arg (Martin *et al.*, 2003, *J. Ind. Microbiol. Biotechnol.* 3: 568-76; 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), an Ne-(Glu or Asp)-Gly-Arg (SEQ ID NO: 3) site, which is cleaved by a Factor Xa protease after the arginine residue (Eaton *et al.*, 1986, *Biochem.* 25: 505-512); an Asp-Asp-Asp-Asp-Lys (SEQ ID NO: 4) site, which is cleaved by an enterokinase after the lysine (Collins-Racie *et al.*, 1995, *Biotechnology* 13: 982-987); a His-Tyr-Glu or His-Tyr-Asp site, which is cleaved by Genenase I (Carter *et al.*, 1989, *Proteins: Structure, Function, and Genetics* 6: 240-248); a Leu-Val-Pro-Arg-Gly-Ser (SEQ ID NO: 5) site, which is cleaved by thrombin after the Arg (Stevens, 2003, *Drug Discovery World* 4: 35-48.); a Glu-Asn-Leu-Tyr-Phe-Gln-Gly (SEQ ID NO: 6) site, which is cleaved by TEV protease after the Gln (Stevens, 2003, *supra*); and a Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro (SEQ ID NO: 7) site, which is cleaved by a genetically engineered form of human rhinovirus 3C protease after the Gln (Stevens, 2003, *supra*).

Polynucleotides

The present invention also relates to isolated polynucleotides comprising or consisting of nucleotide sequences that encode polypeptides having alpha-amylase activity of the present invention.

In a preferred embodiment, the nucleotide sequence comprises or consists of SEQ ID NO: 1. In another preferred embodiment, the nucleotide sequence comprises or consists of the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1. In another preferred embodiment, the nucleotide sequence comprises or consists of nucleotides 61 to 1821 of SEQ ID NO: 1. The present invention also encompasses nucleotide sequences that encode polypeptides comprising or consisting of the amino acid sequence of SEQ ID NO: 2 or

the mature polypeptide and/or the catalytic domain thereof, which differ from SEQ ID NO: 1 or the mature polypeptide and/or the catalytic domain coding sequence thereof by virtue of the degeneracy of the genetic code. The present invention also relates to subsequences of SEQ ID NO: 1 that encode fragments of SEQ ID NO: 2 that have alpha-amylase activity.

5 The present invention also relates to mutant polynucleotides comprising or consisting of at least one mutation in the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, in which the mutant nucleotide sequence encodes the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2.

10 The techniques used to isolate or clone a polynucleotide encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the polynucleotides of the present invention from such 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*,
15 Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleotide sequence-based amplification (NASBA) may be used. The polynucleotides may be cloned from a strain of *Aspergillus*, or another or related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleotide sequence.

20 The present invention also relates to isolated polynucleotides comprising or consisting of nucleotide sequences that have a degree of identity to the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1 of at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 96%, at least 97%, at least 98%, or at least 99%
25 identity, which encode an active polypeptide.

30 Modification of a nucleotide sequence encoding a polypeptide of the present invention may be necessary for the synthesis of polypeptides substantially similar to the polypeptide. The term "substantially similar" to the polypeptide refers to non-naturally occurring forms of the polypeptide. These polypeptides may differ in some engineered way from the polypeptide isolated from its native source, e.g., artificial variants that differ in specific activity, thermostability, pH optimum, or the like. The variant sequence may be constructed on the basis of the nucleotide sequence presented as the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, e.g., a subsequence thereof, and/or by introduction of
35 nucleotide substitutions that do not give rise to another amino acid sequence of the polypeptide encoded by the nucleotide sequence, but which correspond to the codon usage of the host organism intended for production of the enzyme, or by introduction of nucleotide substitutions that may give rise to a different amino acid sequence. For a general description of nucleotide substitution, see, e.g., Ford *et al.*, 1991, *Protein Expression and Purification 2*: 95-107.

It will be apparent to those skilled in the art that such substitutions can be made outside the regions critical to the function of the molecule and still result in an active polypeptide. Amino acid residues essential to the activity of the polypeptide encoded by an isolated polynucleotide of the invention, and therefore preferably not subject to substitution, may be identified according
5 to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (see, e.g., Cunningham and Wells, 1989, *supra*). In the latter technique, mutations are introduced at every positively charged residue in the molecule, and the resultant mutant molecules are tested for alpha-amylase activity to identify amino acid residues that are critical to the activity of the molecule. Sites of substrate-enzyme interaction can also be determined by
10 analysis of the three-dimensional structure as determined by such techniques as nuclear magnetic resonance analysis, crystallography or photoaffinity labeling (see, e.g., de Vos *et al.*, 1992, *supra*; Smith *et al.*, 1992, *supra*; Wlodaver *et al.*, 1992, *supra*).

The present invention also relates to isolated polynucleotides encoding polypeptides of the present invention, which hybridize under very low stringency conditions, preferably low
15 stringency conditions, more preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with (i) the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, (ii) the genomic DNA sequence comprising the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, or (iii) a full-
20 length complementary strand of (i) or (ii); or allelic variants and subsequences thereof (Sambrook *et al.*, 1989, *supra*), as defined herein. In a preferred embodiment, the complementary strand is the full-length complementary strand of the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1.

The present invention also relates to isolated polynucleotides obtained by (a) hybridizing
25 a population of DNA under very low, low, medium, medium-high, high, or very high stringency conditions with (i) the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, (ii) the genomic DNA sequence comprising the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, or (iii) a full-length complementary strand of (i) or (ii); and (b) isolating the hybridizing polynucleotide, which encodes a polypeptide having
30 alpha-amylase activity. In a preferred embodiment, the complementary strand is the full-length complementary strand of the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1.

Nucleic Acid Constructs

35 The present invention also relates to nucleic acid constructs comprising an isolated polynucleotide of the present invention operably linked to one or more (several) control sequences that direct the expression of the coding sequence in a suitable host cell under conditions compatible with the control sequences.

An isolated polynucleotide encoding a polypeptide of the present invention may be manipulated in a variety of ways to provide for expression of the polypeptide. Manipulation of the polynucleotide's sequence prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotide sequences utilizing recombinant DNA methods are well known in the art.

The control sequence may be an appropriate promoter sequence, a nucleotide sequence that is recognized by a host cell for expression of a polynucleotide encoding a polypeptide of the present invention. The promoter sequence contains transcriptional control sequences that mediate the expression of the polypeptide. The promoter may be any nucleotide sequence that shows transcriptional activity in the host cell of choice including mutant, 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 the transcription of the nucleic acid constructs of the present invention in a filamentous fungal host cell are promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, *Fusarium venenatum* amyloglucosidase (WO 00/56900), *Fusarium venenatum* Daria (WO 00/56900), *Fusarium venenatum* Quinn (WO 00/56900), *Fusarium oxysporum* trypsin-like protease (WO 96/00787), *Trichoderma reesei* beta-glucosidase, *Trichoderma reesei* cellobiohydrolase I, *Trichoderma reesei* cellobiohydrolase II, *Trichoderma reesei* endoglucanase I, *Trichoderma reesei* endoglucanase II, *Trichoderma reesei* endoglucanase III, *Trichoderma reesei* endoglucanase IV, *Trichoderma reesei* endoglucanase V, *Trichoderma reesei* xylanase I, *Trichoderma reesei* xylanase II, *Trichoderma reesei* beta-xylosidase, as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase); and mutant, truncated, and hybrid promoters thereof.

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

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

The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA that is important for translation by the host cell. The leader sequence is operably

linked to the 5' terminus of the nucleotide sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used in the present invention.

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

5 The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleotide sequence 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 of choice may be used in the present invention.

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

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

15 The control sequence may also be a signal peptide coding sequence that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleotide sequence may inherently contain a signal peptide coding sequence naturally linked in translation reading frame with the segment of the coding sequence that encodes the
20 secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding sequence that is foreign to the coding sequence. The foreign signal peptide coding sequence may be required where the coding sequence does not naturally contain a signal peptide coding sequence. Alternatively, the foreign signal peptide coding sequence may simply replace the natural signal peptide coding sequence in order to enhance secretion of the
25 polypeptide. However, any signal peptide coding sequence that directs the expressed polypeptide into the secretory pathway of a host cell of choice, *i.e.*, secreted into a culture medium, may be used in the present invention.

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

In a preferred embodiment, the signal peptide comprises or consists of amino acids 1 to
35 20 of SEQ ID NO: 2. In another preferred embodiment, the signal peptide coding sequence comprises or consists of nucleotides 24 to 83 of SEQ ID NO: 1.

It may also be desirable to add regulatory sequences that allow the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those that cause the 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 systems in prokaryotic systems include the *lac*, *tac*, *xyl* and *trp* operator systems. In yeast, the ADH2 system or GAL1 system may be used. In filamentous fungi, the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter may be used as regulatory sequences. 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 nucleotide sequence encoding the polypeptide would be operably linked with the regulatory sequence.

10

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 nucleic acids and control sequences described herein may be joined together to produce a recombinant expression vector that may include one or more (several) convenient restriction sites to allow for insertion or substitution of the nucleotide sequence encoding the polypeptide at such sites. Alternatively, a polynucleotide sequence of the present invention may be expressed by inserting the nucleotide sequence or a nucleic acid construct comprising the sequence 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 nucleotide sequence. 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 vectors may be linear or closed circular plasmids.

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 vectors of the present invention preferably contain one or more (several) 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.

Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *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 the *amdS* and *pyrG* genes of *Aspergillus nidulans* or *Aspergillus oryzae* and the *bar* gene of *Streptomyces hygrosopicus*.

The vectors of the present invention preferably contain 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 nonhomologous recombination. Alternatively, the vector may contain additional nucleotide sequences 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 preferably contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which have a high degree of 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 nucleotide sequences. 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" is defined herein as a nucleotide sequence that enables a plasmid or vector to replicate *in vivo*.

Examples of origins of replication useful in a filamentous fungal cell are AMA1 and ANSI (Gems *et al.*, 1991, *Gene* 98: 61-67; Cullen *et al.*, 1987, *Nucleic Acids Research* 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 WO 00/24883.

More than one copy of a polynucleotide of the present invention may be inserted into a host cell to increase production of the gene product. 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
5 (see, e.g., Sambrook *et al.*, 1989, *supra*).

Host Cells

The present invention also relates to recombinant host cells, comprising an isolated polynucleotide of the present invention, which are advantageously used in the recombinant
10 production of the polypeptides. A vector comprising a polynucleotide of the present invention is introduced into a host cell so that the 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
15 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 bacterium or a Gram negative bacterium. Gram positive bacteria include, but not limited to, *Bacillus*, *Streptococcus*,
20 *Streptomyces*, *Staphylococcus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus*, and *Oceanobacillus*. Gram negative bacteria include, but not limited to, *E. coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*, *Flavobacterium*, *Fusobacterium*, *Ilyobacter*, *Neisseria*, and *Ureaplasma*.

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

In a preferred embodiment, the host cell is a fungal cell. "Fungi" as used herein includes the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota (as defined by Hawksworth *et al.*, *In, Ainsworth and Bisby's Dictionary of The Fungi*, 8th edition, 1995, CAB International, University Press, Cambridge, UK) as well as the Oomycota (as cited in
30 Hawksworth *et al.*, 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, *supra*).

In a preferred embodiment, the fungal host cell is 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
35 in the future, for the purposes of this invention, yeast shall be defined as described in *Biology and Activities of Yeast* (Skinner, F.A., Passmore, S.M., and Davenport, R.R., eds, *Soc. App. Bacteriol. Symposium Series No. 9*, 1980).

In a preferred embodiment, the yeast host cell is a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* cell.

In a preferred embodiment, the yeast host cell is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*,
5 *Saccharomyces kluyveri*, *Saccharomyces norbensis*, or *Saccharomyces oviformis* cell. In another most preferred embodiment, the yeast host cell is a *Kluyveromyces lactis* cell. In another most preferred embodiment, the yeast host cell is a *Yarrowia lipolytica* cell.

In another more preferred embodiment, the fungal host cell is a filamentous fungal cell. "Filamentous fungi" include all filamentous forms of the subdivision Eumycota and Oomycota
10 (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 unicellular thallus and carbon catabolism may be
15 fermentative.

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

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

harzianum, *Thchoderma koningii*, *Thchoderma longibrachiatum*, *Thchoderma 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*.
5 Suitable procedures for transformation of *Aspergillus* and *Trichoderma* host cells are described in EP 238 023 and Yelton *et al.*, 1984, *Proceedings of the National Academy of Sciences USA* 81: 1470-1474. 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,
10 *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, Volume 194, pp 182-187, Academic Press, Inc., New York; Ito *et al.*, 1983, *Journal of Bacteriology* 153: 163; and Hinnen *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 1920.

Methods of Production

15 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 (b) recovering the polypeptide. In a preferred embodiment, the cell is of the genus *Aspergillus*. In a more preferred embodiment, the cell is *Aspergillus terreus*.

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

The present invention also relates to methods of producing a polypeptide of the present invention, comprising: (a) cultivating a recombinant host cell under conditions conducive for
25 production of the polypeptide, wherein the host cell comprises a mutant nucleotide sequence having at least one mutation in the mature polypeptide and/or the catalytic domain coding sequence of SEQ ID NO: 1, wherein the mutant nucleotide sequence encodes a polypeptide that comprises or consists of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2, and (b) recovering the polypeptide.

30 In the production methods of the present invention, the cells are cultivated in a nutrient medium suitable for production of the polypeptide using methods well known in the art. For example, the cell may be cultivated by shake flask cultivation, and small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the
35 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 into the medium, it can be recovered from cell lysates.

5 The polypeptides may be detected using methods known in the art that are specific for the polypeptides. These detection methods may include 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 as described herein.

10 The resulting 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, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation.

15 The polypeptides of the present invention 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*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989) to obtain substantially pure polypeptides.

Plants

20 The present invention also relates to plants, e.g., a transgenic plant, plant part, or plant cell, comprising an isolated polynucleotide encoding a polypeptide having alpha-amylase activity of the present invention so as to express and produce the polypeptide in recoverable quantities. The polypeptide may be recovered from the plant or plant part. Alternatively, the plant or plant part containing the recombinant polypeptide may be used as such for improving
25 the quality of a food or feed, e.g., improving nutritional value, palatability, and rheological properties, or to destroy an antinutritive 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,
30 oats, rye, barley, rice, sorghum, and maize (corn).

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*.

35 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, vascular tissues, meristems. Specific plant cell compartments, such as chloroplasts, apoplasts, mitochondria, vacuoles, peroxisomes and cytoplasm are also considered to be a plant part. Furthermore, any plant cell, whatever the tissue origin, is considered to be a plant part.

Likewise, plant parts such as specific tissues and cells isolated to facilitate the utilisation of the invention are also considered plant parts, e.g., embryos, endosperms, aleurone and seeds coats.

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

The transgenic plant or plant cell expressing a polypeptide of the present invention may be constructed in accordance with methods known in the art. In short, the plant or plant cell is constructed by incorporating one or more (several) expression constructs encoding a polypeptide of the present invention into the plant host genome or chloroplast genome and
10 propagating the resulting modified plant or plant cell into a transgenic plant or plant cell.

The expression construct is conveniently a nucleic acid construct that comprises a polynucleotide encoding a polypeptide of the present invention operably linked with appropriate regulatory sequences required for expression of the nucleotide sequence in the plant or plant part of choice. Furthermore, the expression construct may comprise a selectable marker useful
15 for identifying host cells into which the expression construct has been integrated and DNA sequences necessary for introduction of the construct into the plant in question (the latter depends on the DNA introduction method to be used).

The choice of regulatory sequences, such as promoter and terminator sequences and optionally signal or transit sequences are determined, for example, on the basis of when,
20 where, and how the polypeptide is desired to be expressed. For instance, the expression of the gene encoding a polypeptide of the present invention may be constitutive or inducible, or may be developmental, stage or tissue specific, and the gene product may be targeted to a specific tissue or plant part such as seeds or leaves. Regulatory sequences are, for example, described by Tague *et al.*, 1988, *Plant Physiology* 86: 506.

For constitutive expression, the 35S-CaMV, the maize ubiquitin 1, and the rice actin 1
25 promoter may be used (Franck *et al.*, 1980, *Cell* 21: 285-294, Christensen *et al.*, 1992, *Plant Mo. Biol.* 18: 675-689; Zhang *et al.*, 1991, *Plant Cell* 3: 1155-1165). organ-specific promoters may be, for example, a promoter from storage sink tissues such as seeds, potato tubers, and fruits (Edwards & Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), or from metabolic sink tissues
30 such as meristems (Ito *et al.*, 1994, *Plant Mol. Biol.* 24: 863-878), a seed specific promoter such as the glutelin, prolamin, globulin, or albumin promoter from rice (Wu *et al.*, 1998, *Plant and Cell Physiology* 39: 885-889), a *Vicia faba* promoter from the legumin B4 and the unknown seed protein gene from *Vicia faba* (Conrad *et al.*, 1998, *Journal of Plant Physiology* 152: 708-711), a promoter from a seed oil body protein (Chen *et al.*, 1998, *Plant and Cell Physiology* 39:
35 935-941), the storage protein *napA* promoter from *Brassica napus*, or any other seed specific promoter known in the art, e.g., as described in WO 91/14772. Furthermore, the promoter may be a leaf specific promoter such as the *rbcS* promoter from rice or tomato (Kyoizuka *et al.*, 1993, *Plant Physiology* 102: 991-1000, the chlorella virus adenine methyltransferase gene promoter

(Mitra and Higgins, 1994, *Plant Molecular Biology* 26: 85-93), or the *aldP* gene promoter from rice (Kagaya *et al.*, 1995, *Molecular and General Genetics* 248: 668-674), or a wound inducible promoter such as the potato *pin2* promoter (Xu *et al.*, 1993, *Plant Molecular Biology* 22: 573-588). Likewise, the promoter may be inducible by abiotic treatments such as temperature, drought, or alterations in salinity or induced by exogenously applied substances that activate the promoter, e.g., ethanol, oestrogens, plant hormones such as ethylene, abscisic acid, and gibberellic acid, and heavy metals.

A promoter enhancer element may also be used to achieve higher expression of a polypeptide of the present invention in the plant. For instance, the promoter enhancer element may be an intron that is placed between the promoter and the nucleotide sequence encoding a polypeptide of the present invention. For instance, Xu *et al.*, 1993, *supra*, disclose the use of the first intron of the rice actin 1 gene to enhance expression.

The selectable marker gene and any other parts of the expression construct may be chosen from those available in the art.

The nucleic acid construct is incorporated into the plant genome according to conventional techniques known in the art, including *Agrobacterium-mediated* transformation, virus-mediated transformation, microinjection, particle bombardment, biolistic transformation, and electroporation (Gasser *et al.*, 1990, *Science* 244: 1293; Potrykus, 1990, *Bio/Technology* 8: 535; Shimamoto *et al.*, 1989, *Nature* 338: 274).

Presently, *Agrobacterium tumefaciens-mediated* gene transfer is the method of choice for generating transgenic dicots (for a review, see Hooykas and Schilperoort, 1992, *Plant Molecular Biology* 19: 15-38) and can also be used for transforming monocots, although other transformation methods are often used for these plants. Presently, the method of choice for generating transgenic monocots is particle bombardment (microscopic gold or tungsten particles coated with the transforming DNA) of embryonic calli or developing embryos (Christou, 1992, *Plant Journal* 2: 275-281 ; Shimamoto, 1994, *Current Opinion Biotechnology* 5: 158-162; Vasil *et al.*, 1992, *Bio/Technology* 10: 667-674). An alternative method for transformation of monocots is based on protoplast transformation as described by Omirulleh *et al.*, 1993, *Plant Molecular Biology* 21: 415-428.

Following transformation, the transformants having incorporated the expression construct are selected and regenerated into whole plants according to methods well-known in the art. Often the transformation procedure is designed for the selective elimination of selection genes either during regeneration or in the following generations by using, for example, co-transformation with two separate T-DNA constructs or site specific excision of the selection gene by a specific recombinase.

The present invention also relates to methods of producing a polypeptide of the present invention comprising: (a) cultivating a transgenic plant or a plant cell comprising a polynucleotide encoding the polypeptide having alpha-amylase activity of the present invention

under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

Compositions

5 The present invention also relates to compositions comprising a polypeptide of the present invention.

The composition may further comprise an enzyme selected from the group comprising of; a fungal alpha-amylase (EC 3.2.1.1), a beta-amylase (E.C. 3.2.1.2), a glucoamylase (E.C.3.2.1.3), a pullulanases (E.C. 3.2.1.41), a phytase (E.C.3.1.2.28) and a protease (E.C. 10 3.4.). The glucoamylase may preferably be derived from a strain of *Aspergillus* sp., such as *Aspergillus niger*, or from a strain of *Talaromyces* sp. and in particular derived from *Talaromyces leycettanus* such as the glucoamylase disclosed in U.S. patent no. Re. 32,153, *Talaromyces duponti* and/or *Talaromyces thermopiles* such as the glucoamylases disclosed in U.S. Patent No. 4,587,215 and more preferably derived from *Talaromyces emersonii*. Most 15 preferably the glucoamylase is derived from *Talaromyces emersonii* strain CBS 793.97 and/or having the sequence disclosed as SEQ ID NO: 7 in WO 99/28448. Further preferred is a glucoamylase which has an amino acid sequence having at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or even at least 99% identity to the aforementioned amino acid sequence. A commercial *Talaromyces* 20 glucoamylase preparation is supplied by Novozymes A/S as SPIRIZYME FUEL.

Also preferred for a composition comprising the polypeptide of the present invention and a glucoamylase are polypeptides having glucoamylase activity which are derived from a strain of the genus *Trametes*, preferably *Trametes cingulata*. Further preferred is polypeptide having glucoamylase activity and having at least 50%, at least 60%, at least 70%, at least 80%, at 25 least 90% or even at least 95% identity with amino acids for mature polypeptide of SEQ ID NO: 2 in WO 2006/069289.

Also preferred for a composition comprising the polypeptide of the present invention and a glucoamylase are polypeptides having glucoamylase activity which are derived from a strain of the genus *Pachykytospora*, preferably *Pachykytospora papyracea* or the *E. coli* strain 30 deposited at DSMZ and given the no. DSM 17105. Further preferred are polypeptides having glucoamylase activity and having at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or even at least 95% identity with amino acids for mature polypeptide of SEQ ID NO: 5 in WO 2006/069289.

The composition described above may preferably comprise acid alpha-amylase present in 35 an amount of 0.01 to 10 AFAU/g DS, preferably 0.1 to 5 AFAU/g DS, more preferably 0.5 to 3 AFAU/AGU, and most preferably 0.3 to 2 AFAU/g DS. The dosage of the polypeptide composition of the invention and other conditions under which the composition is used may be determined on the basis of methods known in the art.

The polypeptide 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. For instance, the polypeptide composition may be in the form of granulate or a microgranulate. The polypeptide to be included in the composition may be stabilized in accordance with methods known in the art.

5

Uses

The present invention is also directed to methods for using the polypeptides having alpha-amylase activity, or compositions thereof.

The polypeptide or the composition of the present invention may be used in starch conversion, starch to sugar conversion and ethonal production etc, e.g., in liquefying and/or saccharifying a gelatinized starch or a granular starch, as well as a partly gelatinized starch. A partly gelatinized starch is a starch which to some extent is gelatinized, *i.e.*, wherein part of the starch has irreversibly swelled and gelatinized and part of the starch is still present in a granular state. It can be used in a process for liquefying starch, wherein a gelatinized or granular starch substrate is treated in aqueous medium with the enzyme. The polypeptide or the composition of the present invention may also be used in a process for saccharification of a liquefied starch substrate. A preferred use is in a fermentation process wherein a starch substrate is liquefied and/or saccharified in the presence of the polypeptide or the composition of the present invention to produce glucose and/or maltose suitable for conversion into a fermentation product by a fermenting organism, preferably a yeast. Such fermentation processes include a process for producing ethanol for fuel or drinking ethanol (portable alcohol), a process for producing a beverage, a process for producing desired organic compounds, such as citric acid, itaconic acid, lactic acid, gluconic acid, sodium gluconate, calcium gluconate, potassium gluconate, glucono delta lactone, or sodium erythorbate; ketones; amino acids, such as glutamic acid (sodium monoglutamate), but also more complex compounds such as antibiotics, such as penicillin, tetracyclin; enzymes; vitamins, such as riboflavin, B12, beta-carotene; hormones, which are difficult to produce synthetically.

Furthermore, due to the superior hydrolysis activity of the polypeptide of the first aspect the amount of glucoamylase during the saccharification step can be reduced. The glucoamylase may preferably be derived from a strain within *Aspergillus* sp., *Talaromyces* sp., *Pachykytospora* sp. or *Trametes* sp., more preferably from *Aspergillus niger*, *Talaromyces emersonii*, *Trametes cingulata* or *Pachykytospora papyracea*.

In a preferred embodiment, the polypeptide of the present invention is used in a process comprising fermentation to produce a fermentation product, e.g., ethanol, from a gelatinized starch. Such a process for producing ethanol from gelatinized starch by fermentation comprises: (i) liquefying the gelatinized starch with a polypeptide with alpha-amylase activity of the present invention; (ii) saccharifying the liquefied mash obtained; (iii) fermenting the material obtained in step (ii) in the presence of a fermenting organism. Optionally the process further

35

comprises recovery of the ethanol. The saccharification and fermentation may be carried out as a simultaneous saccharification and fermentation process (SSF process).

In another preferred embodiment, the polypeptide of the present invention is used in a process comprising fermentation to produce a fermentation product, e.g., ethanol, from an ungelatinized ("raw") starch. Such a process for producing ethanol from ungelatinized starch-containing material by fermentation comprises: (i) contacting the ungelatinized starch with a polypeptide with alpha-amylase activity of the present invention to degrade the ungelatinized starch; (ii) saccharifying the mash obtained; (iii) fermenting the material obtained in step (ii) in the presence of a fermenting organism. Optionally the process further comprises recovery of the ethanol. The saccharification and fermentation may be carried out as a simultaneous saccharification and fermentation process (SSF process).

In further embodiments, the polypeptide of the present invention may also be useful in textile, fabric or garment desizing or washing, in baking, detergent and pulp and paper production.

Methods and Materials

Examples

Chemicals used as buffers and substrates were commercial products of at least reagent grade.

Example 1 - Determination of alpha-amylase activity

The activity of any acid alpha-amylase 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, *i.e.*, acid stable alpha-amylase, an endo-alpha-amylase (1,4-alpha-D-glucan-glucano-hydrolase, E.C. 3.2.1.1) hydrolyzes alpha-1,4-glycosidic 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.

Reaction condition: 10 microliters standard or enzyme sample, 70 microliters H₂O, and 80 microliters starch working solution (The final concentration was starch 0.35 g/L, Acetate buffer 50 mM pH 5.0, NaCl 0.1 M, CaCl₂ 3 mM) mixed and react for 2 minutes with shaking at 37°C. Add 40 microliters Iodine working solution (the final iodine concentration was 0.04 g/L) and react at 37°C for 1 minute. Reading OD590 (Before reading, shaking 10 seconds).

FUNGAMYL™ (available from Novozymes A/S) is used as standard.

The activity of the mature polypeptide in SEQ ID NO: 2 was determined to be 810.2 AFAU/g.

Example 2 - AZCL-HE-amylose Assay

5 10 microliters enzyme sample and 120 microliters 0.1% AZCL-HE-amylose (Megazyme International Ireland Ltd.) at pH 7 were mixed in a Microtiter plate and placed on ice before reaction. The Microtiter plate was incubated in an Eppendorf thermomixer for 30 min at 50°C. Then 70 microL supernatant was transferred to a new microtiter plate and OD595 was read as a measure of amylase activity. All reaction was done with duplicate and a buffer blind was
10 included in the assay (instead of enzyme).

Example 3 - pH Profile

The pH profile of the polypeptide of the invention was determined. 5 microliters enzyme solution and 20 microliters 1 2% AZCL-HE-amylose in 100 microliters B&R buffer (Britton-
15 Robinson buffer: 0.1 M boric acid, 0.1 M acetic acid, and 0.1 M phosphoric acid) adjusted to pH-values 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0 and 11.0 with HCl or NaOH were mixed in a Microtiter plate and placed on ice before reaction. The Microtiter plate was incubated in an Eppendorf thermomixer for 30 min at 50°C. Then 80 microliters supernatant was transferred to a new microtiter plate and OD595 was read as a measure of amylase activity. All reaction was
20 done with duplicate.

The polypeptide of the invention showed activity from pH 3 to pH 10 with optimal pH between pH 4 and 5.

Example 4 - pH Stability

25 The pH stability of the polypeptide of the invention was determined by adding 14 microliters enzyme solution into 126 microL buffer (100 mM Na-acetate) at pH4, incubating at 40°C for 0, 5, 10, 30, 60 and 120 min. 20 microL taken for reaction at each time point. The enzyme was added into 120 microliters buffer at pH 5.0 containing 0.2% AZCL-HE-amylose and incubated at 50°C for 30 min before 60 microL supernatant was taken for OD595. The
30 polypeptide of the invention retained 39% residual activity after incubation 2 hours at pH 4 and 40°C.

Example 5 - Temperature Profile

The temperature profile of the polypeptide of the invention was determined by adding 20
35 microliters enzyme solution into 150 microliters buffer (100 mM Tris-HCl) at pH 7 containing 0.1% AZCL-HE-amylose, and incubating for 5 min at 20, 30, 40, 50, 60, 70, 80, and 90°C. Then the sample was placed on ice for 10 min before 80 microL supernatant was taken for OD595.

The polypeptide of the invention showed activity at temperatures from 20 to 70°C with an optimum temperature around 50°C.

Example 6 - Temperature stability

5 The temperature stability of the polypeptide of the invention was determined by adding 14 microliters enzyme solution into 126 microliters H₂O was incubated at 70°C for 0, 5, 10, 30, 60 and 120 min, 20 microliters taken and put on ice at each time point. Then 120 microliters pH 5 0.3% AZCL-H E-amylose was added at 50°C for 4 min and 60 microL taken for OD595. The polypeptide of the invention retained 58% residual activity after 2 hours.

10 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
15 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.

Claims

1. An isolated polypeptide having alpha-amylase activity, selected from the group consisting of:
- 5 (a) a polypeptide comprising an amino acid sequence having preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 85%, most preferably at least 90%, and even most preferably 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: 2 and/or the catalytic domain of SEQ ID NO: 2;
- 10 (b) a polypeptide comprising an amino acid sequence having preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 85%, most preferably at least 90%, and even most preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the catalytic domain of SEQ ID NO: 2;
- (c) a polypeptide encoded by a polynucleotide that hybridizes under at least high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 and/or the catalytic domain coding sequence of SEQ ID NO: 1, (ii) a full-length complementary strand of (i);
- 15 (d) a polypeptide encoded by a polynucleotide comprising a nucleotide sequence having preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 85%, most preferably at least 90%, and even most preferably at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 and/or the catalytic domain coding sequence of SEQ ID NO: 1; and
- 20 (e) a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide and/or the catalytic domain of SEQ ID NO: 2.
- 25
2. The polypeptide according to claim 1, comprising or consisting of the amino acid sequence of SEQ ID NO: 2; or a fragment thereof having alpha-amylase activity.
- 30
3. The polypeptide according to claim 1 or 2, which is encoded by a polynucleotide comprising or consisting of the nucleotide sequence of SEQ ID NO: 1; or a subsequence thereof encoding a fragment having alpha-amylase activity.
- 35
4. An isolated polynucleotide comprising a nucleotide sequence that encodes the polypeptide of any of claims 1-3.

5. A nucleic acid construct comprising the polynucleotide of claim 4 operably linked to one or more (several) control sequences that direct the production of the polypeptide in an expression host.
- 5 6. A recombinant expression vector comprising the nucleic acid construct of claim 5.
7. A recombinant host cell comprising the nucleic acid construct of claim 5.
8. A method of producing the polypeptide of any of claims 1-3, comprising:
- 10 (a) cultivating a cell, which in its wild-type form produces the polypeptide, under conditions conducive for production of the polypeptide; and
- (b) recovering the polypeptide.
9. A method of producing the polypeptide of any of claims 1-3, comprising:
- 15 (a) cultivating a host cell comprising a nucleic acid construct comprising a nucleotide sequence encoding the polypeptide under conditions conducive for production of the polypeptide; and
- (b) recovering the polypeptide.
- 20 10. A method of producing the polypeptide of any of claims 1-3, comprising:
- (a) cultivating a transgenic plant or a plant cell comprising a polynucleotide encoding the polypeptide under conditions conducive for production of the polypeptide; and
- (b) recovering the polypeptide.
- 25 11. A transgenic plant, plant part or plant cell transformed with a polynucleotide encoding the polypeptide of any of claims 1-3.
12. A composition comprising the polypeptide of any of claims 1-3 and an enzyme selected from the group consisting of; a fungal alpha-amylase (EC 3.2.1.1), a beta-amylase (E.C. 3.2.1.2), a glucoamylase (E.C.3.2.1.3), a pullulanases (E.C. 3.2.1.41), a phytase (E.C.3.1.2.28) and a protease (E.C. 3.4.).
- 30 13. Use of the polypeptide of any of claims 1-3 or the composition of claim 12.
- 35 14. Use according to claim 13 for starch modification in the food industry, starch modification in the paper and pulp industry, starch liquefaction, textile washing, textile desizing, brewing, ethanol production and/or baking.

15. Use according to claim 13 or 14 for production of ethanol in a process comprising hydrolyzing an ungelatinized starch.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/023272

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N9/26
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 practical, search terms used)

EPO-Internal, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
X	<p>DATABASE UniProt [Online] 17 October 2006 (2006-10-17), "SubName: FuI l=Al pha-amyl ase ;" XP002576027 retrieved from EBI accession no. UNIPROT:Q0C881 Database accession no. Q0C881 abstract; sequence 1 -----</p>	1-7
X	<p>WO 2008/080093 A2 (VERENIUM CORP [US]; SYNGENTA PARTICIPATIONS AG [CH]; SLUPSKA MALGORZAT) 3 July 2008 (2008-07-03) abstract sequences 51,52 page 3, line 12 - page 6, line 14 page 9, line 6 - page 10, line 2 page 11, line 12 - page 13, line 35 page 15, line 4 - page 16, line 21 ----- -/-</p>	1-15

Further documents are listed in the continuation of Box C

See patent family annex

* Special categories of cited documents

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document 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

31 March 2010

Date of mailing of the international search report

28/04/2010

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

Montrone, Marco

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/023272

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 03/012071 A2 (ELITRA PHARMACEUTICALS INC [US]; JIANG BO [CA]; STORMS REGINALD [CA];) 13 February 2003 (2003-02-13) abstract sequences 43-45 claims 1-27 page 28, lines 3-14 page 48, line 15 - page 51, line 26 -----	1-15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2010/023272

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008080093	A2	03-07-2008	
		AU 2007336732 A1	03-07-2008
		CA 2673525 A1	03-07-2008
		EP 2069490 A2	17-06-2009
WO 03012071	A2	13-02-2003	
		AU 2002319764 A1	17-02-2003