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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 July 2009 (09.07.2009)

PCT

(10) International Publication Number
WO 2009/085935 A2

(51) International Patent Classification:
C12N 9/24 (2006.01) *C12N 1/21* (2006.01)
C12N 9/42 (2006.01) *C12N 5/10* (2006.01)
C12N 15/56 (2006.01) *C12S 3/02* (2006.01)
C12N 1/15 (2006.01) *C12P 19/14* (2006.01)

(21) International Application Number:
PCT/US2008/087402

(22) International Filing Date:
18 December 2008 (18.12.2008)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/014,900 19 December 2007 (19.12.2007) US
61/014,980 19 December 2007 (19.12.2007) US

(71) Applicant *(for all designated States except US)*: **NOVOZYMES A/S** [DK/DK]; Krogshoejvej 36, DK-2880 Bagsvaerd (DK).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HARRIS, Paul** [US/US]; 11927 342nd Avenue, N.E, Carnation, WA 98014 (US). **MAIYURAN, Suchindra** [AU/US]; 11717 New Albion Drive, Gold River, California 95670 (US). **BROWN, Kimberly** [US/US]; 8322 Windswept Court, Elk Grove, CA 95758 (US).

(74) Agent: **STARNES, Robert**; Novozymes, Inc., 1445 Drew Avenue, Davis, CA 95618 (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, 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, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, 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, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

WO 2009/085935 A2

(54) **Title:** POLYPEPTIDES HAVING CELLULOLYTIC ENHANCING ACTIVITY AND POLYNUCLEOTIDES ENCODING SAME

(57) **Abstract:** The present invention relates to isolated polypeptides having cellulolytic enhancing 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 CELLULOLYTIC ENHANCING ACTIVITY AND POLYNUCLEOTIDES ENCODING SAME

Reference to a Sequence Listing

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

Reference to Deposit of Biological Material

This application contains a reference to a deposit of biological material, which deposit is incorporated herein by reference.

Background of the Invention

Field of the Invention

The present invention relates to isolated polypeptides having cellulolytic enhancing 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.

Description of the Related Art

Cellulose is a polymer of the simple sugar glucose linked by beta-1,4-bonds. Many microorganisms produce enzymes that hydrolyze beta-linked glucans. These enzymes include endoglucanases, cellobiohydrolases, and beta-glucosidases. Endoglucanases digest the cellulose polymer at random locations, opening it to attack by cellobiohydrolases. Cellobiohydrolases sequentially release molecules of cellobiose from the ends of the cellulose polymer. Cellobiose is a water-soluble beta-1,4-linked dimer of glucose. Beta-glucosidases hydrolyze cellobiose to glucose.

The conversion of lignocellulosic feedstocks into ethanol has the advantages of the ready availability of large amounts of feedstock, the desirability of avoiding burning or land filling the materials, and the cleanliness of the ethanol fuel. Wood, agricultural residues, herbaceous crops, and municipal solid wastes have been considered as feedstocks for ethanol production. These materials primarily consist of cellulose, hemicellulose, and lignin. Once the cellulose is converted to glucose, the glucose is easily fermented by yeast into ethanol.

It would be advantageous in the art to improve the ability to convert cellulosic feedstocks.

WO 2005/074647 discloses isolated polypeptides having cellulolytic enhancing activity and polynucleotides thereof from *Thielavia terrestris*. WO 2005/074656 discloses an isolated polypeptide having cellulolytic enhancing activity and the polynucleotide thereof from *Thermoascus aurantiacus*. WO 2007/089290 discloses an isolated polypeptide having cellulolytic enhancing activity and the polynucleotide thereof from *Trichoderma reesei*.

The present invention relates to polypeptides having cellulolytic enhancing activity and polynucleotides encoding the polypeptides.

Summary of the Invention

The present invention relates to isolated polypeptides having cellulolytic enhancing activity selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence having at least 60% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4;

(b) a polypeptide encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii);

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

(d) a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

The present invention also relates to isolated polynucleotides encoding polypeptides having cellulolytic enhancing activity, selected from the group consisting of:

(a) a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least 60% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4;

(b) a polynucleotide that hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii);

(c) a polynucleotide comprising a nucleotide sequence having at least 60%

identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3; and

(d) a polynucleotide encoding a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

The present invention also relates to nucleic acid constructs, recombinant expression vectors, recombinant host cells comprising the polynucleotides, and methods of producing a polypeptide having cellulolytic enhancing activity.

The present invention also relates to methods of inhibiting the expression of a polypeptide having cellulolytic enhancing activity in a cell, comprising administering to the cell or expressing in the cell a double-stranded RNA (dsRNA) molecule, wherein the dsRNA comprises a subsequence of a polynucleotide of the present invention. The present also relates to such a double-stranded inhibitory RNA (dsRNA) molecule, wherein optionally the dsRNA is a siRNA or a miRNA molecule.

The present invention also relates to methods for degrading or converting a cellulosic material, comprising: treating the cellulosic material with a cellulolytic enzyme composition in the presence of such a polypeptide having cellulolytic enhancing activity, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity.

The present invention also relates to methods of producing a fermentation product, comprising: (a) saccharifying a cellulosic material with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity; (b) fermenting the saccharified cellulosic material of step (a) with one or more fermenting microorganisms to produce the fermentation product; and (c) recovering the fermentation product from the fermentation.

The present invention also relates to methods of fermenting a cellulosic material, comprising: fermenting the cellulosic material with one or more fermenting microorganisms, wherein the cellulosic material is hydrolyzed with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity of the present invention and the presence of the polypeptide having cellulolytic enhancing activity increases the hydrolysis of the cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity.

The present invention also relates to plants comprising an isolated polynucleotide encoding a polypeptide having cellulolytic enhancing activity.

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

The present invention further relates to nucleic acid constructs comprising a gene encoding a protein, wherein the gene is operably linked to a nucleotide sequence encoding a signal peptide comprising or consisting of amino acids 1 to 17 of SEQ ID NO: 2 or amino acids 1 to 15 of SEQ ID NO: 4, wherein the gene is foreign to the nucleotide sequence.

Brief Description of the Figures

Figure 1 shows the genomic DNA sequence and the deduced amino acid sequence of a *Myceliophthora thermophila* CBS 202.75 GH61A polypeptide having cellulolytic enhancing activity (SEQ ID NOs: 1 and 2, respectively).

Figure 2 shows a restriction map of pSMai190.

Figure 3 shows the genomic DNA sequence and the deduced amino acid sequence of a *Myceliophthora thermophila* CBS 202.75 GH61F polypeptide having cellulolytic enhancing activity (SEQ ID NOs: 3 and 4, respectively).

Figure 4 shows a restriction map of pSMai192.

Figure 5 shows a restriction map of pSMai185.

Figure 6 shows a restriction map of pSMai198.

Figure 7 shows the effect of *Myceliophthora thermophila* GH61A and GH61F polypeptides having cellulolytic enhancing activity on enzymatic hydrolysis of pretreated corn stover.

Definitions

Cellulolytic enhancing activity: The term "cellulolytic enhancing activity" is defined herein as a biological activity that enhances the hydrolysis of a cellulosic material by proteins having cellulolytic activity. For purposes of the present invention, cellulolytic enhancing activity is determined by measuring the increase in reducing sugars or in the increase of the total of cellobiose and glucose from the hydrolysis of a cellulosic material by cellulase protein under the following conditions: 1-50 mg of total protein/g of cellulose in PCS, wherein total protein is comprised of 80-99.5% w/w cellulase protein/g of cellulose in PCS and 0.5-20% w/w protein of cellulolytic enhancing

activity for 1-7 days at 50°C compared to a control hydrolysis with equal total protein loading without cellulolytic enhancing activity (1-50 mg of cellulolytic protein/g of cellulose in PCS). In a preferred aspect, a mixture of CELLUCLAST® 1.5L (Novozymes A/S, Bagsvaerd, Denmark) in the presence of 3% of total protein weight *Aspergillus oryzae* beta-glucosidase (recombinantly produced in *Aspergillus oryzae* according to WO 02/095014) or 3% of total protein weight *Aspergillus fumigatus* beta-glucosidase (recombinantly produced in *Aspergillus oryzae* according to Example 22 of WO 02/095014) of cellulase protein loading is used as the source of the cellulolytic activity.

The polypeptides having cellulolytic enhancing activity 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 cellulolytic enhancing activity of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

The polypeptides having cellulolytic enhancing activity enhance the hydrolysis of a cellulosic material catalyzed by proteins having cellulolytic activity by reducing the amount of cellulolytic enzyme required to reach the same degree of hydrolysis preferably at least 0.1-fold, more at least 0.2-fold, more preferably at least 0.3-fold, more preferably at least 0.4-fold, more preferably at least 0.5-fold, more preferably at least 1-fold, more preferably at least 3-fold, more preferably at least 4-fold, more preferably at least 5-fold, more preferably at least 10-fold, more preferably at least 20-fold, even more preferably at least 30-fold, most preferably at least 50-fold, and even most preferably at least 100-fold.

Cellulolytic activity: The term "cellulolytic activity" is defined herein as a biological activity which hydrolyzes a cellulosic material. Cellulolytic protein may hydrolyze or hydrolyzes carboxymethyl cellulose (CMC), thereby decreasing the viscosity of the incubation mixture. The resulting reduction in viscosity may be determined by a vibration viscosimeter (*e.g.*, MIVI 3000 from Sofraser, France). Determination of cellulase activity, measured in terms of Cellulase Viscosity Unit (CEVU), quantifies the amount of catalytic activity present in a sample by measuring the ability of the sample to reduce the viscosity of a solution of carboxymethyl cellulose (CMC). The assay is performed at the temperature and pH suitable for the cellulolytic protein and substrate. For CELLUCLAST™ (Novozymes A/S, Bagsvaerd, Denmark) the assay is carried out at 40°C in 0.1 M phosphate pH 9.0 buffer for 30 minutes with CMC as substrate (33.3 g/L carboxymethyl cellulose Hercules 7 LFD) and an enzyme concentration of approximately 3.3-4.2 CEVU/ml. The CEVU activity is calculated

relative to a declared enzyme standard, such as CELLUZYME™ Standard 17-1194 (obtained from Novozymes A/S, Bagsvaerd, Denmark).

For purposes of the present invention, cellulolytic activity is determined by measuring the increase in hydrolysis of a cellulosic material by a cellulolytic mixture under the following conditions: 1-10 mg of cellulolytic protein/g of cellulose in PCS for 5-7 day at 50°C compared to a control hydrolysis without addition of cellulolytic protein.

Endoglucanase: The term "endoglucanase" is defined herein as an endo-1,4-(1,3;1,4)-beta-D-glucan 4-glucanohydrolase (E.C. No. 3.2.1.4), which catalyses endohydrolysis of 1,4-beta-D-glycosidic linkages in cellulose, cellulose derivatives (such as carboxymethyl cellulose and hydroxyethyl cellulose), lichenin, beta-1,4 bonds in mixed beta-1,3 glucans such as cereal beta-D-glucans or xyloglucans, and other plant material containing cellulosic components. For purposes of the present invention, endoglucanase activity is determined using carboxymethyl cellulose (CMC) hydrolysis according to the procedure of Ghose, 1987, *Pure and Appl. Chem.* 59: 257-268.

Cellobiohydrolase: The term "cellobiohydrolase" is defined herein as a 1,4-beta-D-glucan cellobiohydrolase (E.C. 3.2.1.91), which catalyzes the hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, cellooligosaccharides, or any beta-1,4-linked glucose containing polymer, releasing cellobiose from the reducing or non-reducing ends of the chain. For purposes of the present invention, cellobiohydrolase activity is determined according to the procedures described by Lever *et al.*, 1972, *Anal. Biochem.* 47: 273-279 and by van Tilbeurgh *et al.*, 1982, *FEBS Letters* 149: 152-156; van Tilbeurgh and Claeyssens, 1985, *FEBS Letters* 187: 283-288. In the present invention, the Lever *et al.* method was employed to assess hydrolysis of cellulose in corn stover, while the method of van Tilbeurgh *et al.* was used to determine the cellobiohydrolase activity on a fluorescent disaccharide derivative.

Beta-glucosidase: The term "beta-glucosidase" is defined herein as a beta-D-glucoside glucohydrolase (E.C. 3.2.1.21), which catalyzes the hydrolysis of terminal non-reducing beta-D-glucose residues with the release of beta-D-glucose. For purposes of the present invention, beta-glucosidase activity is determined according to the basic procedure described by Venturi *et al.*, 2002, *J. Basic Microbiol.* 42: 55-66, except different conditions were employed as described herein. One unit of beta-glucosidase activity is defined as 1.0 µmole of p-nitrophenol produced per minute at 50°C, pH 5 from 4 mM p-nitrophenyl-beta-D-glucopyranoside as substrate in 100 mM sodium citrate, 0.01% TWEEN® 20.

Family 61 glycoside hydrolase: The term "Family 61 glycoside hydrolase" or "Family GH61" is defined herein as a polypeptide falling into the glycoside hydrolase Family 61 according to Henrissat B., 1991, A classification of glycosyl hydrolases based

on amino-acid sequence similarities, *Biochem. J.* 280: 309-316, and Henrissat B., and Bairoch A., 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696. Presently, Henrissat lists the GH61 Family as unclassified indicating that properties such as mechanism, catalytic nucleophile/base, catalytic proton donors, and 3-D structure are not known for polypeptides belonging to this family.

Cellulosic material: The cellulosic material can be any material containing cellulose. The predominant polysaccharide in the primary cell wall of biomass is cellulose, the second most abundant is hemi-cellulose, and the third is pectin. The secondary cell wall, produced after the cell has stopped growing, also contains polysaccharides and is strengthened by polymeric lignin covalently cross-linked to hemicellulose. Cellulose is a homopolymer of anhydrocellobiose and thus a linear beta-(1-4)-D-glucan, while hemicelluloses include a variety of compounds, such as xylans, xyloglucans, arabinoxylans, and mannans in complex branched structures with a spectrum of substituents. Although generally polymorphous, cellulose is found in plant tissue primarily as an insoluble crystalline matrix of parallel glucan chains. Hemicelluloses usually hydrogen bond to cellulose, as well as to other hemicelluloses, which help stabilize the cell wall matrix.

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

In one aspect, the cellulosic material is herbaceous material. In another aspect, the cellulosic material is agricultural residue. In another aspect, the cellulosic material is forestry residue. In another aspect, the cellulosic material is municipal solid waste. In another aspect, the cellulosic material is waste paper. In another aspect, the cellulosic

material is pulp and paper mill residue.

In another aspect, the cellulosic material is corn stover. In another preferred aspect, the cellulosic material is corn fiber. In another aspect, the cellulosic material is corn cob. In another aspect, the cellulosic material is orange peel. In another aspect, the cellulosic material is rice straw. In another aspect, the cellulosic material is wheat straw. In another aspect, the cellulosic material is switch grass. In another aspect, the cellulosic material is miscanthus. In another aspect, the cellulosic material is bagasse.

In another aspect, the cellulosic material is microcrystalline cellulose. In another aspect, the cellulosic material is bacterial cellulose.

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

Pre-treated corn stover: The term "PCS" or "Pre-treated Corn Stover" is defined herein as a cellulosic material derived from corn stover by treatment with heat and dilute acid.

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 97% pure, more preferably at least 98% pure, even more preferably at least 99% pure, 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 in its final form following translation and any post-translational modifications, such as N-terminal processing, C-terminal truncation, glycosylation, phosphorylation, etc. In a preferred aspect, the mature polypeptide is amino acids 18 to 232 of SEQ ID NO: 2 based on the SignalP program (Nielsen *et al.*, 1997, *Protein Engineering* 10: 1-6) that predicts amino acids 1 to 17 of SEQ ID NO: 2 are a signal peptide. In another preferred aspect, the mature polypeptide is amino acids 16 to 235 of SEQ ID NO: 4 based on the SignalP program that predicts amino acids 1 to 15 of SEQ ID NO: 4 are a signal peptide.

Mature polypeptide coding sequence: The term "mature polypeptide coding sequence" is defined herein as a nucleotide sequence that encodes a mature polypeptide having cellulolytic enhancing activity. In a preferred aspect, the mature polypeptide coding sequence is nucleotides 52 to 921 of SEQ ID NO: 1 based on the SignalP program that predicts nucleotides 1 to 51 of SEQ ID NO: 1 encode a signal peptide. In another preferred aspect, the mature polypeptide coding sequence is nucleotides 46 to 851 of SEQ ID NO: 3 based on the SignalP program that predicts nucleotides 1 to 45 of SEQ ID NO: 3 encode a signal peptide.

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:

(Identical Residues x 100)/(Length of Alignment – 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:

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

Homologous sequence: The term "homologous sequence" is defined herein as a predicted protein having 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 the *Myceliophthora thermophila* polypeptide having cellulolytic enhancing activity of SEQ ID NO: 2 or SEQ ID NO: 4, or the mature polypeptide thereof.

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 SEQ ID NO: 4; or a homologous sequence thereof; wherein the fragment has cellulolytic enhancing activity. In a preferred aspect, a fragment contains at least 185 amino acid residues, more preferably at least 195 amino acid residues, and most preferably at least 205 amino acid residues of the mature polypeptide of SEQ ID NO: 2 or a homologous sequence thereof. In another preferred aspect, a fragment contains at least 190 amino acid residues, more preferably at least 200 amino acid residues, and most preferably at least 210 amino acid residues of the mature polypeptide of SEQ ID NO: 4 or a homologous sequence thereof.

Subsequence: The term "subsequence" is defined herein as a nucleotide sequence 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 SEQ ID NO: 3; or a homologous sequence thereof; wherein the subsequence encodes a polypeptide fragment having cellulolytic enhancing activity. In a preferred aspect, a subsequence contains at least 555 nucleotides, more preferably at least 585 nucleotides, and most preferably at least 615 nucleotides of the mature polypeptide coding sequence of SEQ ID NO: 1 or a homologous sequence thereof. In another preferred aspect, a subsequence contains at least 570 nucleotides, more preferably at least 600 nucleotides, and most preferably at least 630 nucleotides of the mature polypeptide coding sequence of SEQ ID NO: 3 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 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.

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%, 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 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% pure, 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 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 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.

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

mRNA. These steps include the removal of intron sequences by a process called splicing. cDNA derived from mRNA lacks, therefore, any intron sequences.

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 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 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 SEQ ID NO: 4; 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 cellulolytic enhancing activity produced by an organism expressing a modified polynucleotide sequence of the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3; 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 SEQ ID NO: 3; or a homologous sequence thereof.

Detailed Description of the Invention

Polypeptides Having Cellulolytic Enhancing 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 of SEQ ID NO: 2 or SEQ ID NO: 4 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 at least 96%, at least 97%, at least 98%, or at least 99%, which have cellulolytic enhancing activity (hereinafter "homologous polypeptides"). In a preferred 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 of SEQ ID NO: 2 or SEQ ID NO: 4.

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 cellulolytic enhancing activity. In a preferred aspect, the polypeptide comprises the amino acid sequence of SEQ ID NO: 2. In another preferred aspect, the polypeptide comprises the mature polypeptide of SEQ ID NO: 2. In another preferred aspect, the polypeptide comprises amino acids 18 to 232 of SEQ ID NO: 2, or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In another preferred aspect, the polypeptide comprises amino acids 18 to 232 of SEQ ID NO: 2. In another preferred aspect, the polypeptide consists of the amino acid sequence of SEQ ID NO: 2 or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In another preferred aspect, the polypeptide consists of the amino acid sequence of SEQ ID NO: 2. In another preferred aspect, the polypeptide consists of

the mature polypeptide of SEQ ID NO: 2. In another preferred aspect, the polypeptide consists of amino acids 18 to 232 of SEQ ID NO: 2 or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In another preferred aspect, the polypeptide consists of amino acids 18 to 232 of SEQ ID NO: 2.

A polypeptide of the present invention preferably comprises the amino acid sequence of SEQ ID NO: 4 or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In a preferred aspect, the polypeptide comprises the amino acid sequence of SEQ ID NO: 4. In another preferred aspect, the polypeptide comprises the mature polypeptide of SEQ ID NO: 4. In another preferred aspect, the polypeptide comprises amino acids 16 to 235 of SEQ ID NO: 4, or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In another preferred aspect, the polypeptide comprises amino acids 16 to 235 of SEQ ID NO: 4. In another preferred aspect, the polypeptide consists of the amino acid sequence of SEQ ID NO: 4 or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In another preferred aspect, the polypeptide consists of the amino acid sequence of SEQ ID NO: 4. In another preferred aspect, the polypeptide consists of the mature polypeptide of SEQ ID NO: 4. In another preferred aspect, the polypeptide consists of amino acids 16 to 235 of SEQ ID NO: 4 or an allelic variant thereof; or a fragment thereof having cellulolytic enhancing activity. In another preferred aspect, the polypeptide consists of amino acids 16 to 235 of SEQ ID NO: 4.

In a second aspect, the present invention relates to isolated polypeptides having cellulolytic enhancing activity that are encoded by polynucleotides that hybridize under preferably very low stringency conditions, more preferably low 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 coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (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 coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3 contains at least 100 contiguous nucleotides or preferably at least 200 contiguous nucleotides. Moreover, the subsequence may encode a polypeptide fragment having cellulolytic enhancing activity. In a preferred aspect, the complementary strand is the full-length complementary strand of the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

The nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 3; or a subsequence

thereof; as well as the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4; or a fragment thereof; may be used to design nucleic acid probes to identify and clone DNA encoding polypeptides having cellulolytic enhancing 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, or most preferably at least 800 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, ^{3}H, ^{35}S, biotin, or avidin). Such probes are encompassed by the present invention.

A genomic DNA or cDNA library prepared from such other strains may, therefore, be screened for DNA that hybridizes with the probes described above and encodes a polypeptide having cellulolytic enhancing 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 SEQ ID NO: 3; 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 coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3; the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3; 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 aspect, the nucleic acid probe is the mature polypeptide coding sequence of SEQ ID NO: 1. In another preferred aspect, the nucleic acid probe is nucleotides 52 to 921 of SEQ ID NO: 1. In another preferred aspect, the nucleic acid probe is a polynucleotide sequence that encodes the polypeptide of SEQ ID NO: 2, or a

subsequence thereof. In another preferred aspect, the nucleic acid probe is SEQ ID NO: 1. In another preferred aspect, the nucleic acid probe is the polynucleotide sequence contained in plasmid pSMai190 which is contained in *E. coli* NRRL B-50083, wherein the polynucleotide sequence thereof encodes a polypeptide having cellulolytic enhancing activity. In another preferred aspect, the nucleic acid probe is the mature polypeptide coding region contained in plasmid pSMai190 which is contained in *E. coli* NRRL B-50083.

In another preferred aspect, the nucleic acid probe is the mature polypeptide coding sequence of SEQ ID NO: 3. In another preferred aspect, the nucleic acid probe is nucleotides 46 to 851 of SEQ ID NO: 3. In another preferred aspect, the nucleic acid probe is a polynucleotide sequence that encodes the polypeptide of SEQ ID NO: 4, or a subsequence thereof. In another preferred aspect, the nucleic acid probe is SEQ ID NO: 3. In another preferred aspect, the nucleic acid probe is the polynucleotide sequence contained in plasmid pSMai192 which is contained in *E. coli* NRRL B-50085, wherein the polynucleotide sequence thereof encodes a polypeptide having cellulolytic enhancing activity. In another preferred aspect, the nucleic acid probe is the mature polypeptide coding region contained in plasmid pSMai192 which is contained in *E. coli* NRRL B-50085.

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 of 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 of about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SCC 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 a third aspect, the present invention relates to isolated polypeptides having cellulolytic enhancing activity encoded by polynucleotides comprising or consisting of nucleotide sequences that have a degree of identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3 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 at least 96%, at least 97%, at least 98%, or at least 99%, which encode a polypeptide having cellulolytic enhancing activity. See polynucleotide section herein.

In a fourth aspect, 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 of SEQ ID NO: 2 or SEQ ID NO: 4; 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/Val, Ala/Glu, and Asp/Gly.

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 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 pipecolic 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 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 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.*, cellulolytic enhancing activity) to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton *et al.*, 1996, *J. Biol. Chem.* 271: 4699-4708. The active site of the enzyme or other biological interaction can also be determined by physical analysis of structure, as 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 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, *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 of SEQ ID NO: 2 or SEQ ID NO: 4, 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 Cellulolytic Enhancing 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 aspect, the polypeptide obtained from a given source is secreted extracellularly.

A polypeptide having cellulolytic enhancing activity of the present invention may be a bacterial polypeptide. For example, the polypeptide may be a gram positive bacterial polypeptide such as a *Bacillus*, *Streptococcus*, *Streptomyces*, *Staphylococcus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus*, or *Oceanobacillus* polypeptide having cellulolytic enhancing activity, or a Gram negative bacterial polypeptide such as an *E. coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*, *Flavobacterium*, *Fusobacterium*, *Ilyobacter*, *Neisseria*, or *Ureaplasma* polypeptide having cellulolytic enhancing activity.

In a preferred aspect, the polypeptide is a *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis* polypeptide having cellulolytic enhancing activity.

In another preferred aspect, the polypeptide is a *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, or *Streptococcus equi* subsp. *Zooepidemicus* polypeptide having cellulolytic enhancing activity.

In another preferred aspect, the polypeptide is a *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, or *Streptomyces lividans* polypeptide having cellulolytic enhancing activity.

A polypeptide having cellulolytic enhancing activity of the present invention may also be a fungal polypeptide, and more preferably a yeast polypeptide such as a

Candida, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* polypeptide having cellulolytic enhancing activity; or more preferably a filamentous fungal polypeptide such as an *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Aureobasidium*, *Botryospaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptospaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania*, *Pseudotrichonympha*, *Rhizomucor*, *Schizophyllum*, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trichoderma*, *Trichophaea*, *Verticillium*, *Volvariella*, or *Xylaria* polypeptide having cellulolytic enhancing activity.

In a preferred aspect, the polypeptide is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, or *Saccharomyces oviformis* polypeptide having cellulolytic enhancing activity.

In another preferred aspect, the polypeptide is an *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium tropicum*, *Chrysosporium merdarium*, *Chrysosporium inops*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium zonatum*, *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*, *Fusarium venenatum*, *Humicola grisea*, *Humicola insolens*, *Humicola lanuginosa*, *Irpex lacteus*, *Mucor miehei*, *Neurospora crassa*, *Penicillium funiculosum*, *Penicillium purpurogenum*, *Phanerochaete chrysosporium*, *Thielavia achromatica*, *Thielavia albomyces*, *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 cellulolytic enhancing activity.

In another preferred aspect, the polypeptide is a *Myceliophthora hinnulea*, *Myceliophthora lutea*, *Myceliophthora thermophila*, or *Myceliophthora vellerea*

polypeptide having cellulolytic enhancing activity.

In a more preferred aspect, the polypeptide is a *Myceliophthora thermophila* polypeptide having cellulolytic enhancing activity. In a most preferred aspect, the polypeptide is a *Myceliophthora thermophila* CBS 202.75 polypeptide having cellulolytic enhancing activity, *e.g.*, the polypeptide comprising the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

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

Strains of these species are readily accessible to the public in a number of culture collections, such as the American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSM), 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 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 cellulolytic enhancing 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 Ile-(Glu or Asp)-Gly-Arg site, which is cleaved by a Factor Xa protease after the arginine residue (Eaton *et al.*, 1986, *Biochem.* 25: 505-512); a Asp-Asp-Asp-Asp-Lys site, which is cleaved by an enterokinase after the lysine (Collins-Racie *et al.*, 1995, *Biotechnology* 13: 982-987); a His-Tyr-Glu site 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 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 site, which is cleaved by TEV protease after the Gln (Stevens, 2003, *supra*); and a Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro 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 cellulolytic enhancing activity of the present invention.

In a preferred aspect, the nucleotide sequence comprises or consists of SEQ ID NO: 1. In another more preferred aspect, the nucleotide sequence comprises or consists of the sequence contained in plasmid pSMai190 which is contained in *E. coli* NRRL B-50083. In another preferred aspect, the nucleotide sequence comprises or consists of the mature polypeptide coding sequence of SEQ ID NO: 1. In another preferred aspect, the nucleotide sequence comprises or consists of nucleotides 52 to 921 of SEQ ID NO: 1. In another more preferred aspect, the nucleotide sequence comprises or consists of the mature polypeptide coding sequence contained in plasmid pSMai190 which is contained in *E. coli* NRRL B-50083.

In a preferred aspect, the nucleotide sequence comprises or consists of SEQ ID NO: 3. In another more preferred aspect, the nucleotide sequence comprises or consists of the sequence contained in plasmid pSMai192 which is contained in *E. coli* NRRL B-50085. In another preferred aspect, the nucleotide sequence comprises or consists of the mature polypeptide coding sequence of SEQ ID NO: 3. In another preferred aspect, the nucleotide sequence comprises or consists of nucleotides 46 to 851 of SEQ ID NO: 3. In another more preferred aspect, the nucleotide sequence comprises or consists of the mature polypeptide coding sequence contained in plasmid pSMai192 which is contained in *E. coli* NRRL B-50085.

The present invention also encompasses nucleotide sequences that encode polypeptides comprising or consisting of the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4 or the mature polypeptide thereof, which differ from SEQ ID NO: 1 or

SEQ ID NO: 3, or the mature polypeptide coding sequence thereof, respectively, by virtue of the degeneracy of the genetic code. The present invention also relates to subsequences of SEQ ID NO: 1 or SEQ ID NO: 3 that encode fragments of SEQ ID NO: 2 or SEQ ID NO: 4 that have cellulolytic enhancing activity, respectively.

The present invention also relates to mutant polynucleotides comprising or consisting of at least one mutation in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, in which the mutant nucleotide sequence encodes the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4, respectively.

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*, 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 *Myceliophthora*, 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.

The present invention also relates to isolated polynucleotides comprising or consisting of nucleotide sequences that have a degree of identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3 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 at least 96%, at least 97%, at least 98%, or at least 99% identity, which encode a polypeptide having cellulolytic enhancing activity.

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 coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, *e.g.*, a subsequence thereof, and/or by introduction of 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 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 cellulolytic enhancing 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 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 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 coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii); or allelic variants and subsequences thereof (Sambrook *et al.*, 1989, *supra*), as defined herein. In a preferred aspect, the complementary strand is the full-length complementary strand of the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

The present invention also relates to isolated polynucleotides obtained by (a) hybridizing a population of DNA under very low, low, medium, medium-high, high, or very high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii); and (b) isolating the hybridizing polynucleotide, which encodes a polypeptide having cellulolytic enhancing activity. In a preferred aspect, the complementary strand is the full-length complementary strand of the mature polypeptide

coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

Nucleic Acid Constructs

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, especially in a bacterial host cell, are the promoters obtained from the *E. coli lac* operon, *Streptomyces coelicolor* agarase gene (*dagA*), *Bacillus subtilis* levansucrase gene (*sacB*), *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* penicillinase gene (*penP*), *Bacillus subtilis xylA* and *xylB* genes, and prokaryotic beta-lactamase gene (Villa-Kamaroff *et al.*, 1978, *Proceedings of the National Academy of Sciences USA* 75: 3727-3731), as well as the *tac* promoter (DeBoer *et al.*, 1983, *Proceedings of the National Academy of Sciences USA* 80: 21-25). Further promoters are described in "Useful proteins from recombinant bacteria" in *Scientific American*, 1980, 242: 74-94; and in Sambrook *et al.*, 1989, *supra*.

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.

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

The control sequence may also be a 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.

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

The control sequence may also be 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.

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

The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the 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.

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.

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 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 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 bacterial host cells are the signal peptide coding sequences obtained from the genes for *Bacillus* NCIB 11837 maltogenic

amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (*nprT*, *nprS*, *nprM*), and *Bacillus subtilis prsA*. Further signal peptides are described by Simonen and Palva, 1993, *Microbiological Reviews* 57: 109-137.

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

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

In a preferred aspect, the signal peptide comprises or consists of amino acids 1 to 17 of SEQ ID NO: 2. In another preferred aspect, the signal peptide coding sequence comprises or consists of nucleotides 1 to 51 of SEQ ID NO: 1.

In another preferred aspect, the signal peptide comprises or consists of amino acids 1 to 15 of SEQ ID NO: 4. In another preferred aspect, the signal peptide coding sequence comprises or consists of nucleotides 1 to 45 of SEQ ID NO: 3.

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

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

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*,

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.

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.

Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers that confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol, or tetracycline resistance. Suitable markers for yeast host cells are ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. 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 adenyltransferase), 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 hygroscopicus*.

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 bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAMß1 permitting replication in *Bacillus*.

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

Examples of origins of replication useful in a filamentous fungal cell are AMA1 and ANS1 (Gems *et al.*, 1991, *Gene* 98: 61-67; Cullen *et al.*, 1987, *Nucleic Acids 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 (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 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 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, *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 bacterial host cell may be any *Bacillus* cell. *Bacillus* cells useful in the practice of the present invention include, but are not limited to, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, and *Bacillus thuringiensis* cells.

In a preferred aspect, the bacterial host cell is a *Bacillus amyloliquefaciens*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus stearothermophilus* or *Bacillus subtilis* cell. In a more preferred aspect, the bacterial host cell is a *Bacillus amyloliquefaciens* cell. In another more preferred aspect, the bacterial host cell is a *Bacillus clausii* cell. In another more preferred aspect, the bacterial host cell is a *Bacillus licheniformis* cell. In another more preferred aspect, the bacterial host cell is a *Bacillus subtilis* cell.

The bacterial host cell may also be any *Streptococcus* cell. *Streptococcus* cells useful in the practice of the present invention include, but are not limited to, *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, and *Streptococcus equi* subsp. *Zooepidemicus* cells.

In a preferred aspect, the bacterial host cell is a *Streptococcus equisimilis* cell. In another preferred aspect, the bacterial host cell is a *Streptococcus pyogenes* cell. In another preferred aspect, the bacterial host cell is a *Streptococcus uberis* cell. In another preferred aspect, the bacterial host cell is a *Streptococcus equi* subsp. *Zooepidemicus* cell.

The bacterial host cell may also be any *Streptomyces* cell. *Streptomyces* cells useful in the practice of the present invention include, but are not limited to, *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, and *Streptomyces lividans* cells.

In a preferred aspect, the bacterial host cell is a *Streptomyces achromogenes* cell. In another preferred aspect, the bacterial host cell is a *Streptomyces avermitilis* cell. In another preferred aspect, the bacterial host cell is a *Streptomyces coelicolor* cell. In another preferred aspect, the bacterial host cell is a *Streptomyces griseus* cell. In another preferred aspect, the bacterial host cell is a *Streptomyces lividans* cell.

The introduction of DNA into a *Bacillus* cell may, for instance, be effected by protoplast transformation (see, *e.g.*, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), by using competent cells (see, *e.g.*, Young and Spizizen, 1961,

Journal of Bacteriology 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), by electroporation (see, *e.g.*, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or by conjugation (see, *e.g.*, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5271-5278). The introduction of DNA into an *E coli* cell may, for instance, be effected by protoplast transformation (see, *e.g.*, Hanahan, 1983, *J. Mol. Biol.* 166: 557-580) or electroporation (see, *e.g.*, Dower *et al.*, 1988, *Nucleic Acids Res.* 16: 6127-6145). The introduction of DNA into a *Streptomyces* cell may, for instance, be effected by protoplast transformation and electroporation (see, *e.g.*, Gong *et al.*, 2004, *Folia Microbiol. (Praha)* 49: 399-405), by conjugation (see, *e.g.*, Mazodier *et al.*, 1989, *J. Bacteriol.* 171: 3583-3585), or by transduction (see, *e.g.*, Burke *et al.*, 2001, *Proc. Natl. Acad. Sci. USA* 98: 6289-6294). The introduction of DNA into a *Pseudomonas* cell may, for instance, be effected by electroporation (see, *e.g.*, Choi *et al.*, 2006, *J. Microbiol. Methods* 64: 391-397) or by conjugation (see, *e.g.*, Pinedo and Smets, 2005, *Appl. Environ. Microbiol.* 71: 51-57). The introduction of DNA into a *Streptococcus* cell may, for instance, be effected by natural competence (see, *e.g.*, Perry and Kuramitsu, 1981, *Infect. Immun.* 32: 1295-1297), by protoplast transformation (see, *e.g.*, Catt and Jollick, 1991, *Microbios.* 68: 189-2070, by electroporation (see, *e.g.*, Buckley *et al.*, 1999, *Appl. Environ. Microbiol.* 65: 3800-3804) or by conjugation (see, *e.g.*, Clewell, 1981, *Microbiol. Rev.* 45: 409-436). However, any method known in the art for introducing DNA into a host cell can be used.

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

In a preferred aspect, 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 Hawksworth *et al.*, 1995, *supra*, page 171) and all mitosporic fungi (Hawksworth *et al.*, 1995, *supra*).

In a more preferred aspect, 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 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 an even more preferred aspect, the yeast host cell is a *Candida*, *Hansenula*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* cell.

In a most preferred aspect, the yeast host cell is a *Saccharomyces*

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

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

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

In a most preferred aspect, 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 aspect, 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 another most preferred aspect, the filamentous fungal host cell is a *Bjerkandera adusta*, *Ceriporiopsis aneirina*, *Ceriporiopsis aneirina*, *Ceriporiopsis caregiea*, *Ceriporiopsis gilvescens*, *Ceriporiopsis pannocinta*, *Ceriporiopsis rivulosa*, *Ceriporiopsis subrufa*, *Ceriporiopsis subvermispora*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium tropicum*, *Chrysosporium merdarium*, *Chrysosporium inops*, *Chrysosporium pannicola*, *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*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride* cell.

Fungal cells may be transformed by a process involving protoplast formation, transformation of the protoplasts, and regeneration of the cell wall in a manner known *per se*. Suitable procedures for transformation of *Aspergillus* and *Trichoderma* host cells are described in EP 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, *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

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 aspect, the cell is of the genus *Myceliophthora*. In a more preferred aspect, the cell is *Myceliophthora thermophila*. In a most preferred aspect, the cell is *Myceliophthora thermophila* CBS 202.75. In another most preferred aspect, the cell is *Myceliophthora thermophila* CBS 117.65.

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 production of the polypeptide, wherein the host cell comprises a mutant nucleotide sequence having at least one mutation in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, wherein the mutant nucleotide sequence encodes a polypeptide that comprises or consists of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4; and (b) recovering the polypeptide.

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 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.

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.

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.

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

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 cellulolytic enhancing 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 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, 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*.

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, 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 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 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, is determined, for example, on the basis of when, 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 promoter may be used (Franck *et al.*, 1980, *Cell* 21: 285-294, Christensen *et al.*, 1992, *Plant Mol. 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 and Coruzzi, 1990, *Ann. Rev. Genet.* 24: 275-303), or from metabolic sink tissues 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: 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 (Kyozuka *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 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 cellulolytic enhancing activity of the present invention under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

Removal or Reduction of Cellulolytic Enhancing Activity

The present invention also relates to methods of producing a mutant of a parent cell, which comprises disrupting or deleting a polynucleotide sequence, or a portion thereof, encoding a polypeptide of the present invention, which results in the mutant cell producing less of the polypeptide than the parent cell when cultivated under the same conditions.

The mutant cell may be constructed by reducing or eliminating expression of a nucleotide sequence encoding a polypeptide of the present invention using methods well known in the art, for example, insertions, disruptions, replacements, or deletions. In a preferred aspect, the nucleotide sequence is inactivated. The nucleotide sequence to be modified or inactivated may be, for example, the coding region or a part thereof essential for activity, or a regulatory element required for the expression of the coding region. An example of such a regulatory or control sequence may be a promoter sequence or a functional part thereof, *i.e.*, a part that is sufficient for affecting expression of the nucleotide sequence. Other control sequences for possible

modification include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, signal peptide sequence, transcription terminator, and transcriptional activator.

Modification or inactivation of the nucleotide sequence may be performed by subjecting the parent cell to mutagenesis and selecting for mutant cells in which expression of the nucleotide sequence has been reduced or eliminated. The mutagenesis, which may be specific or random, may be performed, for example, by use of a suitable physical or chemical mutagenizing agent, by use of a suitable oligonucleotide, or by subjecting the DNA sequence to PCR generated mutagenesis. Furthermore, the mutagenesis may be performed by use of any combination of these mutagenizing agents.

Examples of a physical or chemical mutagenizing agent suitable for the present purpose include ultraviolet (UV) irradiation, hydroxylamine, N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), O-methyl hydroxylamine, nitrous acid, ethyl methane sulphonate (EMS), sodium bisulphite, formic acid, and nucleotide analogues.

When such agents are used, the mutagenesis is typically performed by incubating the parent cell to be mutagenized in the presence of the mutagenizing agent of choice under suitable conditions, and screening and/or selecting for mutant cells exhibiting reduced or no expression of the gene.

Modification or inactivation of the nucleotide sequence may be accomplished by introduction, substitution, or removal of one or more (several) nucleotides in the gene or a regulatory element required for the transcription or translation thereof. For example, nucleotides may be inserted or removed so as to result in the introduction of a stop codon, the removal of the start codon, or a change in the open reading frame. Such modification or inactivation may be accomplished by site-directed mutagenesis or PCR generated mutagenesis in accordance with methods known in the art. Although, in principle, the modification may be performed *in vivo*, *i.e.*, directly on the cell expressing the nucleotide sequence to be modified, it is preferred that the modification be performed *in vitro* as exemplified below.

An example of a convenient way to eliminate or reduce expression of a nucleotide sequence by a cell is based on techniques of gene replacement, gene deletion, or gene disruption. For example, in the gene disruption method, a nucleic acid sequence corresponding to the endogenous nucleotide sequence is mutagenized *in vitro* to produce a defective nucleic acid sequence that is then transformed into the parent cell to produce a defective gene. By homologous recombination, the defective nucleic acid sequence replaces the endogenous nucleotide sequence. It may be desirable that the defective nucleotide sequence also encodes a marker that may be

used for selection of transformants in which the nucleotide sequence has been modified or destroyed. In a particularly preferred aspect, the nucleotide sequence is disrupted with a selectable marker such as those described herein.

Alternatively, modification or inactivation of the nucleotide sequence may be performed by established anti-sense or RNAi techniques using a sequence complementary to the nucleotide sequence. More specifically, expression of the nucleotide sequence by a cell may be reduced or eliminated by introducing a sequence complementary to the nucleotide sequence of the gene that may be transcribed in the cell and is capable of hybridizing to the mRNA produced in the cell. Under conditions allowing the complementary anti-sense nucleotide sequence to hybridize to the mRNA, the amount of protein translated is thus reduced or eliminated.

The present invention further relates to a mutant cell of a parent cell that comprises a disruption or deletion of a nucleotide sequence encoding the polypeptide or a control sequence thereof, which results in the mutant cell producing less of the polypeptide or no polypeptide compared to the parent cell.

The polypeptide-deficient mutant cells so created are particularly useful as host cells for the expression of native and/or heterologous polypeptides. Therefore, the present invention further relates to methods of producing a native or heterologous polypeptide comprising: (a) cultivating the mutant cell under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide. The term "heterologous polypeptides" is defined herein as polypeptides that are not native to the host cell, a native protein in which modifications have been made to alter the native sequence, or a native protein whose expression is quantitatively altered as a result of a manipulation of the host cell by recombinant DNA techniques.

In a further aspect, the present invention relates to a method of producing a protein product essentially free of cellulolytic enhancing activity by fermentation of a cell that produces both a polypeptide of the present invention as well as the protein product of interest by adding an effective amount of an agent capable of inhibiting cellulolytic enhancing activity to the fermentation broth before, during, or after the fermentation has been completed, recovering the product of interest from the fermentation broth, and optionally subjecting the recovered product to further purification.

In a further aspect, the present invention relates to a method of producing a protein product essentially free of cellulolytic enhancing activity by cultivating the cell under conditions permitting the expression of the product, subjecting the resultant culture broth to a combined pH and temperature treatment so as to reduce the cellulolytic enhancing activity substantially, and recovering the product from the culture broth. Alternatively, the combined pH and temperature treatment may be performed on

an enzyme preparation recovered from the culture broth. The combined pH and temperature treatment may optionally be used in combination with a treatment with an cellulolytic enhancing inhibitor.

In accordance with this aspect of the invention, it is possible to remove at least 60%, preferably at least 75%, more preferably at least 85%, still more preferably at least 95%, and most preferably at least 99% of the cellulolytic enhancing activity. Complete removal of cellulolytic enhancing activity may be obtained by use of this method.

The combined pH and temperature treatment is preferably carried out at a pH in the range of 2-4 or 9-11 and a temperature in the range of at least 60-70°C for a sufficient period of time to attain the desired effect, where typically, 30 to 60 minutes is sufficient.

The methods used for cultivation and purification of the product of interest may be performed by methods known in the art.

The methods of the present invention for producing an essentially cellulolytic enhancing-free product is of particular interest in the production of eukaryotic polypeptides, in particular fungal proteins such as enzymes. The enzyme may be selected from, *e.g.*, an amylolytic enzyme, lipolytic enzyme, proteolytic enzyme, cellulolytic enzyme, oxidoreductase, or plant cell-wall degrading enzyme. Examples of such enzymes include an aminopeptidase, amylase, amyloglucosidase, carbohydrase, carboxypeptidase, catalase, cellobiohydrolase, cellulase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, endoglucanase, esterase, galactosidase, beta-galactosidase, glucoamylase, glucose oxidase, glucosidase, haloperoxidase, hemicellulase, invertase, isomerase, laccase, ligase, lipase, lyase, mannosidase, oxidase, pectinolytic enzyme, peroxidase, phytase, phenoloxidase, polyphenoloxidase, proteolytic enzyme, ribonuclease, transferase, transglutaminase, or xylanase. The cellulolytic enhancing-deficient cells may also be used to express heterologous proteins of pharmaceutical interest such as hormones, growth factors, receptors, and the like.

It will be understood that the term "eukaryotic polypeptides" includes not only native polypeptides, but also those polypeptides, *e.g.*, enzymes, which have been modified by amino acid substitutions, deletions or additions, or other such modifications to enhance activity, thermostability, pH tolerance and the like.

In a further aspect, the present invention relates to a protein product essentially free from cellulolytic enhancing activity that is produced by a method of the present invention.

Methods of Inhibiting Expression of a Polypeptide Having Cellulolytic Enhancing Acitivity

The present invention also relates to methods of inhibiting the expression of a polypeptide having cellulolytic enhancing activity in a cell, comprising administering to the cell or expressing in the cell a double-stranded RNA (dsRNA) molecule, wherein the dsRNA comprises a subsequence of a polynucleotide of the present invention. In a preferred aspect, the dsRNA is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more duplex nucleotides in length.

The dsRNA is preferably a small interfering RNA (siRNA) or a micro RNA (miRNA). In a preferred aspect, the dsRNA is small interfering RNA (siRNAs) for inhibiting transcription. In another preferred aspect, the dsRNA is micro RNA (miRNAs) for inhibiting translation.

The present invention also relates to such double-stranded RNA (dsRNA) molecules, comprising a portion of the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3 for inhibiting expression of a polypeptide in a cell. While the present invention is not limited by any particular mechanism of action, the dsRNA can enter a cell and cause the degradation of a single-stranded RNA (ssRNA) of similar or identical sequences, including endogenous mRNAs. When a cell is exposed to dsRNA, mRNA from the homologous gene is selectively degraded by a process called RNA interference (RNAi).

The dsRNAs of the present invention can be used in gene-silencing therapeutics. In one aspect, the invention provides methods to selectively degrade RNA using the dsRNAis of the present invention. The process may be practiced *in vitro, ex vivo* or *in vivo*. In one aspect, the dsRNA molecules can be used to generate a loss-of-function mutation in a cell, an organ or an animal. Methods for making and using dsRNA molecules to selectively degrade RNA are well known in the art, see, for example, U.S. Patent No. 6,506,559; U.S. Patent No. 6,511,824; U.S. Patent No. 6,515,109; and U.S. Patent No. 6,489,127.

Compositions

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

The composition may comprise a polypeptide of the present invention as the major enzymatic component, *e.g.*, a mono-component composition. Alternatively, the composition may comprise multiple enzymatic activities, such as an aminopeptidase,

amylase, carbohydrase, carboxypeptidase, catalase, cellulase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-galactosidase, glucoamylase, alpha-glucosidase, beta-glucosidase, haloperoxidase, invertase, laccase, lipase, mannosidase, oxidase, pectinolytic enzyme, peptidoglutaminase, peroxidase, phytase, polyphenoloxidase, proteolytic enzyme, ribonuclease, transglutaminase, or xylanase. The additional enzyme(s) may be produced, for example, by a microorganism belonging to the genus *Aspergillus*, preferably *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, or *Aspergillus oryzae*; *Fusarium*, preferably *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 sulphureum*, *Fusarium toruloseum*, *Fusarium trichothecioides*, or *Fusarium venenatum*; *Humicola*, preferably *Humicola insolens* or *Humicola lanuginosa*; or *Trichoderma*, preferably *Trichoderma harzianum*, *Trichoderma koningii*, *Trichoderma longibrachiatum*, *Trichoderma reesei*, or *Trichoderma viride*.

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

Examples are given below of preferred uses of the polypeptide compositions of the invention. 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.

Processing of Cellulosic Material

The present invention also relates to methods for degrading or converting a cellulosic material, comprising: treating the cellulosic material with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity of the present invention. In a preferred aspect, the method further comprises recovering the degraded or converted cellulosic material.

The present invention also relates to methods of producing a fermentation product, comprising: (a) saccharifying a cellulosic material with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity of the present invention; (b) fermenting the saccharified cellulosic material of step (a) with one

or more fermenting microorganisms to produce the fermentation product; and (c) recovering the fermentation product from the fermentation.

The present invention also relates to methods of fermenting a cellulosic material, comprising: fermenting the cellulosic material with one or more fermenting microorganisms, wherein the cellulosic material is hydrolyzed with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity of the present invention and the presence of the polypeptide having cellulolytic enhancing activity increases the hydrolysis of the cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity. In a preferred aspect, the fermenting of the cellulosic material produces a fermentation product. In another preferred aspect, the method further comprises recovering the fermentation product from the fermentation.

The composition comprising the polypeptide having cellulolytic enhancing activity can be in the form of a crude fermentation broth with or without the cells removed or in the form of a semi-purified or purified enzyme preparation or the composition can comprise a host cell of the present invention as a source of the polypeptide having cellulolytic enhancing activity in a fermentation process with the biomass.

The methods of the present invention can be used to saccharify a cellulosic material to fermentable sugars and convert the fermentable sugars to many useful substances, *e.g.*, chemicals and fuels. The production of a desired fermentation product from cellulosic material typically involves pretreatment, enzymatic hydrolysis (saccharification), and fermentation.

The processing of cellulosic material according to the present invention can be accomplished using processes conventional in the art. Moreover, the methods of the present invention can be implemented using any conventional biomass processing apparatus configured to operate in accordance with the invention.

Hydrolysis (saccharification) and fermentation, separate or simultaneous, include, but are not limited to, separate hydrolysis and fermentation (SHF); simultaneous saccharification and fermentation (SSF); simultaneous saccharification and cofermentation (SSCF); hybrid hydrolysis and fermentation (HHF); SHCF (separate hydrolysis and co-fermentation), HHCF (hybrid hydrolysis and fermentation), and direct microbial conversion (DMC). SHF uses separate process steps to first enzymatically hydrolyze lignocellulose to fermentable sugars, *e.g.*, glucose, cellobiose, cellotriose, and pentose sugars, and then ferment the fermentable sugars to ethanol. In SSF, the enzymatic hydrolysis of lignocellulose and the fermentation of sugars to ethanol are combined in one step (Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor &

Francis, Washington, DC, 179-212). SSCF involves the cofermentation of multiple sugars (Sheehan, J., and Himmel, M., 1999, Enzymes, energy and the environment: A strategic perspective on the U.S. Department of Energy's research and development activities for bioethanol, *Biotechnol. Prog.* 15: 817-827). HHF involves a separate hydrolysis separate step, and in addition a simultaneous saccharification and hydrolysis step, which can be carried out in the same reactor. The steps in an HHF process can be carried out at different temperatures, *i.e.*, high temperature enzymatic saccharification followed by SSF at a lower temperature that the fermentation strain can tolerate. DMC combines all three processes (enzyme production, lignocellulose hydrolysis, and fermentation) in one or more steps where the same organism is used to produce the enzymes for conversion of the lignocellulose to fermentable sugars and to convert the fermentable sugars into a final product (Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S., 2002, Microbial cellulose utilization: Fundamentals and biotechnology, *Microbiol. Mol. Biol. Reviews* 66: 506-577). It is understood herein that any method known in the art comprising pretreatment, enzymatic hydrolysis (saccharification), fermentation, or a combination thereof can be used in the practicing the methods of the present invention.

A conventional apparatus can include a fed-batch stirred reactor, a batch stirred reactor, a continuous flow stirred reactor with ultrafiltration, and/or a continuous plug-flow column reactor (Fernanda de Castilhos Corazza, Flávio Faria de Moraes, Gisella Maria Zanin and Ivo Neitzel, 2003, Optimal control in fed-batch reactor for the cellobiose hydrolysis, *Acta Scientiarum. Technology* 25: 33-38; Gusakov, A. V., and Sinitsyn, A. P., 1985, Kinetics of the enzymatic hydrolysis of cellulose: 1. A mathematical model for a batch reactor process, *Enz. Microb. Technol.* 7: 346-352), an attrition reactor (Ryu, S. K., and Lee, J. M., 1983, Bioconversion of waste cellulose by using an attrition bioreactor, *Biotechnol. Bioeng.* 25: 53-65), or a reactor with intensive stirring induced by an electromagnetic field (Gusakov, A. V., Sinitsyn, A. P., Davydkin, I. Y., Davydkin, V. Y., Protas, O. V., 1996, Enhancement of enzymatic cellulose hydrolysis using a novel type of bioreactor with intensive stirring induced by electromagnetic field, *Appl. Biochem. Biotechnol.* 56: 141-153). Additional reactor types include: Fluidized bed, upflow blanket, immobilized, and extruder type reactors for hydrolysis and/or fermentation.

Pretreatment. In practicing the methods of the present invention, any pretreatment process known in the art can be used to disrupt the plant cell wall components. The cellulosic material can also be subjected to pre-soaking, wetting, or conditioning prior to pretreatment using methods known in the art. Conventional pretreatments include, but are not limited to, steam pretreatment (with or without

explosion), dilute acid pretreatment, hot water pretreatment, lime pretreatment, wet oxidation, wet explosion, ammonia fiber explosion, organosolv pretreatment, and biological pretreatment. Additional pretreatments include ultrasound, electroporation, microwave, supercritical CO_2, supercritical H_2O, and ammonia percolation pretreatments.

The cellulosic material can be pretreated before hydrolysis and/or fermentation. Pretreatment is preferably performed prior to the hydrolysis. Alternatively, the pretreatment can be carried out simultaneously with hydrolysis, such as simultaneously with treatment of the cellulosic material with one or more cellulolytic enzymes, or other enzyme activities, to release fermentable sugars, such as glucose and/or maltose. In most cases the pretreatment step itself results in some conversion of biomass to fermentable sugars (even in absence of enzymes).

Steam Pretreatment. In steam pretreatment, the cellulosic material is heated to disrupt the plant cell wall components, including lignin, hemicellulose, and cellulose to make the cellulose and other fractions, *e.g.*, hemicellulase, accessible to enzymes. The lignocellulose material is passed to or through a reaction vessel where steam is injected to increase the temperature to the required temperature and pressure and is retained therein for the desired reaction time. Steam pretreatment is preferably done at 140-230°C, more preferably 160-200°C, and most preferably 170-190°C, where the optimal temperature range depends on any addition of a chemical catalyst. Residence time for the steam pretreatment is preferably 1-15 minutes, more preferably 3-12 minutes, and most preferably 4-10 minutes, where the optimal residence time depends on temperature range and any addition of a chemical catalyst. Steam pretreatment allows for relatively high solids loadings, so that the cellulosic material is generally only moist during the pretreatment. The steam pretreatment is often combined with an explosive discharge of the material after the pretreatment, which is known as steam explosion, that is, rapid flashing to atmospheric pressure and turbulent flow of the material to increase the accessible surface area by fragmentation (Duff and Murray, 1996, *Bioresource Technology* 855: 1-33; Galbe and Zacchi, 2002, *Appl. Microbiol. Biotechnol.* 59: 618-628; U.S. Patent Application No. 20020164730). During steam pretreatment, hemicellulose acetyl groups are cleaved and the resulting acid autocatalyzes partial hydrolysis of the hemicellulose to monosaccharides and oligosaccharides. Lignin is removed to only a limited extent.

A catalyst such as H_2SO_4 or SO_2 (typically 0.3 to 3% w/w) is often added prior to steam pretreatment, which decreases the time and temperature, increases the recovery, and improves enzymatic hydrolysis (Ballesteros *et al.*, 2006, *Appl. Biochem. Biotechnol.* 129-132: 496-508; Varga *et al.*, 2004, *Appl. Biochem. Biotechnol.* 113-116:

509-523; Sassner *et al.*, 2006, *Enzyme Microb. Technol.* 39: 756-762).

Chemical Pretreatment: The term "chemical treatment" refers to any chemical pretreatment that promotes the separation and/or release of cellulose, hemicellulose, and/or lignin. Examples of suitable chemical pretreatment processes include, for example, dilute acid pretreatment, lime pretreatment, wet oxidation, ammonia fiber/freeze explosion (AFEX), ammonia percolation (APR), and organosolv pretreatments.

In dilute acid pretreatment, the cellulosic material is mixed with dilute acid, typically H_2SO_4, and water to form a slurry, heated by steam to the desired temperature, and after a residence time flashed to atmospheric pressure. The dilute acid pretreatment can be performed with a number of reactor designs, *e.g.*, plug-flow reactors, counter-current reactors, or continuous counter-current shrinking bed reactors (Duff and Murray, 1996, *supra*; Schell *et al.*, 2004, *Bioresource Technol.* 91: 179-188; Lee *et al.*, 1999, *Adv. Biochem. Eng. Biotechnol.* 65: 93-115).

Several methods of pretreatment under alkaline conditions can also be used. These alkaline pretreatments include, but are not limited to, lime pretreatment, wet oxidation, ammonia percolation (APR), and ammonia fiber/freeze explosion (AFEX).

Lime pretreatment is performed with calcium carbonate, sodium hydroxide, or ammonia at low temperatures of 85-150°C and residence times from 1 hour to several days (Wyman *et al.*, 2005, *Bioresource Technol.* 96: 1959-1966; Mosier *et al.*, 2005, *Bioresource Technol.* 96: 673-686). WO 2006/110891, WO 2006/11899, WO 2006/11900, and WO 2006/110901 disclose pretreatment methods using ammonia.

Wet oxidation is a thermal pretreatment performed typically at 180-200°C for 5-15 minutes with addition of an oxidative agent such as hydrogen peroxide or over-pressure of oxygen (Schmidt and Thomsen, 1998, *Bioresource Technol.* 64: 139-151; Palonen *et al.*, 2004, *Appl. Biochem. Biotechnol.* 117: 1-17; Varga *et al.*, 2004, *Biotechnol. Bioeng.* 88: 567-574; Martin *et al.*, 2006, *J. Chem. Technol. Biotechnol.* 81: 1669-1677). The pretreatment is performed at preferably 1-40% dry matter, more preferably 2-30% dry matter, and most preferably 5-20% dry matter, and often the initial pH is increased by the addition of alkali such as sodium carbonate.

A modification of the wet oxidation pretreatment method, known as wet explosion (combination of wet oxidation and steam explosion), can handle dry matter up to 30%. In wet explosion, the oxidizing agent is introduced during pretreatment after a certain residence time. The pretreatment is then ended by flashing to atmospheric pressure (WO 2006/032282).

Ammonia fiber explosion (AFEX) involves treating cellulosic material with liquid or gaseous ammonia at moderate temperatures such as 90-100°C and high pressure such as 17-20 bar for 5-10 minutes, where the dry matter content can be as high as 60%

(Gollapalli *et al.*, 2002, *Appl. Biochem. Biotechnol.* 98: 23-35; Chundawat *et al.*, 2007, *Biotechnol. Bioeng.* 96: 219-231; Alizadeh *et al.*, 2005, *Appl. Biochem. Biotechnol.* 121:1133-1141; Teymouri *et al.*, 2005, *Bioresource Technol.* 96: 2014-2018). AFEX pretreatment results in the depolymerization of cellulose and partial hydrolysis of hemicellulose. Lignin-carbohydrate complexes are cleaved.

Organosolv pretreatment delignifies cellulosic material by extraction using aqueous ethanol (40-60% ethanol) at 160-200°C for 30-60 minutes (Pan *et al.*, 2005, *Biotechnol. Bioeng.* 90: 473-481; Pan *et al.*, 2006, *Biotechnol. Bioeng.* 94: 851-861; Kurabi *et al.*, 2005, *Appl. Biochem. Biotechnol.* 121:219-230). Sulphuric acid is usually added as a catalyst. In organosolv pretreatment, the majority of the hemicellulose is removed.

Other examples of suitable pretreatment methods are described by Schell *et al.*, 2003, *Appl. Biochem. and Biotechnol.* Vol. 105-108, p. 69-85, and Mosier *et al.*, 2005, *Bioresource Technology* 96: 673-686, and U.S. Published Application 2002/0164730.

In one aspect, the chemical pretreatment is preferably carried out as an acid treatment, and more preferably as a continuous dilute and/or mild acid treatment. The acid is typically sulfuric acid, but other acids can also be used, such as acetic acid, citric acid, nitric acid, phosphoric acid, tartaric acid, succinic acid, hydrogen chloride or mixtures thereof. Mild acid treatment is conducted in the pH range of preferably 1-5, more preferably 1-4, and most preferably 1-3. In one aspect, the acid concentration is in the range from preferably 0.01 to 20 wt % acid, more preferably 0.05 to 10 wt % acid, even more preferably 0.1 to 5 wt % acid, and most preferably 0.2 to 2.0 wt % acid. The acid is contacted with the cellulosic material and held at a temperature in the range of preferably 160-220°C, and more preferably 165-195°C, for periods ranging from seconds to minutes to, *e.g.*, 1 second to 60 minutes.

In another aspect, pretreatment is carried out as an ammonia fiber explosion step (AFEX pretreatment step).

In another aspect, pretreatment takes place in an aqueous slurry. In preferred aspects, the cellulosic material is present during pretreatment in amounts preferably between 10-80 wt%, more preferably between 20-70 wt%, and most preferably between 30-60 wt%, such as around 50 wt%. The pretreated cellulosic material can be unwashed or washed using any method known in the art, *e.g.*, washed with water.

Mechanical Pretreatment: The term "mechanical pretreatment" refers to various types of grinding or milling (*e.g.*, dry milling, wet milling, or vibratory ball milling).

Physical Pretreatment: The term "physical pretreatment" refers to any pretreatment that promotes the separation and/or release of cellulose, hemicellulose, and/or lignin from cellulosic material. For example, physical pretreatment can involve

irradiation (*e.g.*, microwave irradiation), steaming/steam explosion, hydrothermolysis, and combinations thereof.

Physical pretreatment can involve high pressure and/or high temperature (steam explosion). In one aspect, high pressure means pressure in the range of preferably about 300 to about 600 psi, more preferably about 350 to about 550 psi, and most preferably about 400 to about 500 psi, such as around 450 psi. In another aspect, high temperature means temperatures in the range of about 100 to about 300°C, preferably about 140 to about 235°C. In a preferred aspect, mechanical pretreatment is performed in a batch-process, steam gun hydrolyzer system that uses high pressure and high temperature as defined above, *e.g.*, a Sunds Hydrolyzer available from Sunds Defibrator AB, Sweden.

Combined Physical and Chemical Pretreatment: The cellulosic material can be pretreated both physically and chemically. For instance, the pretreatment step can involve dilute or mild acid treatment and high temperature and/or pressure treatment. The physical and chemical pretreatments can be carried out sequentially or simultaneously, as desired. A mechanical pretreatment can also be included.

Accordingly, in a preferred aspect, the cellulosic material is subjected to mechanical, chemical, or physical pretreatment, or any combination thereof to promote the separation and/or release of cellulose, hemicellulose and/or lignin.

Biological Pretreatment: The term "biological pretreatment" refers to any biological pretreatment that promotes the separation and/or release of cellulose, hemicellulose, and/or lignin from the cellulosic material. Biological pretreatment techniques can involve applying lignin-solubilizing microorganisms (see, for example, Hsu, T.-A., 1996, Pretreatment of biomass, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212; Ghosh and Singh, 1993, Physicochemical and biological treatments for enzymatic/microbial conversion of cellulosic biomass, *Adv. Appl. Microbiol.* 39: 295-333; McMillan, J. D., 1994, Pretreating lignocellulosic biomass: a review, in *Enzymatic Conversion of Biomass for Fuels Production*, Himmel, M. E., Baker, J. O., and Overend, R. P., eds., ACS Symposium Series 566, American Chemical Society, Washington, DC, chapter 15; Gong, C. S., Cao, N. J., Du, J., and Tsao, G. T., 1999, Ethanol production from renewable resources, in *Advances in Biochemical Engineering/Biotechnology*, Scheper, T., ed., Springer-Verlag Berlin Heidelberg, Germany, 65: 207-241; Olsson and Hahn-Hagerdal, 1996, Fermentation of lignocellulosic hydrolysates for ethanol production, *Enz. Microb. Tech.* 18: 312-331; and Vallander and Eriksson, 1990, Production of ethanol from lignocellulosic materials: State of the art, *Adv. Biochem. Eng./Biotechnol.* 42: 63-95).

Saccharification. In the hydrolysis step, also known as saccharification, the

pretreated cellulosic material is hydrolyzed to break down cellulose and alternatively also hemicellulose to fermentable sugars, such as glucose, xylose, xylulose, arabinose, maltose, mannose, galactose, or soluble oligosaccharides. The hydrolysis is performed enzymatically by a cellulolytic enzyme composition comprising a polypeptide having cellulolytic enhancing activity of the present invention, which can further comprise one or more hemicellulolytic enzymes. The enzymes of the compositions can also be added sequentially.

Enzymatic hydrolysis is preferably carried out in a suitable aqueous environment under conditions that can be readily determined by one skilled in the art. In a preferred aspect, hydrolysis is performed under conditions suitable for the activity of the enzyme(s), *i.e.*, optimal for the enzyme(s). The hydrolysis can be carried out as a fed batch or continuous process where the pretreated cellulosic material (substrate) is fed gradually to, for example, an enzyme containing hydrolysis solution.

The saccharification is generally performed in stirred-tank reactors or fermentors under controlled pH, temperature, and mixing conditions. Suitable process time, temperature and pH conditions can readily be determined by one skilled in the art. For example, the saccharification can last up to 200 hours, but is typically performed for preferably about 12 to about 96 hours, more preferably about 16 to about 72 hours, and most preferably about 24 to about 48 hours. The temperature is in the range of preferably about 25°C to about 70°C, more preferably about 30°C to about 65°C, and more preferably about 40°C to 60°C, in particular about 50°C. The pH is in the range of preferably about 3 to about 8, more preferably about 3.5 to about 7, and most preferably about 4 to about 6, in particular about pH 5. The dry solids content is in the range of preferably about 5 to about 50 wt %, more preferably about 10 to about 40 wt %, and most preferably about 20 to about 30 wt %.

In addition to a polypeptide having cellulolytic enhancing activity of the present invention, the cellulolytic enzyme components of the composition are preferably enzymes having endoglucanase, cellobiohydrolase, and beta-glucosidase activities. In a preferred aspect, the cellulolytic enzyme composition comprises one or more (several) cellulolytic enzymes selected from the group consisting of a cellulase, endoglucanase, cellobiohydrolase, and beta-glucosidase. In another preferred aspect, the cellulolytic enzyme preparation is supplemented with one or more additional enzyme activities selected from the group consisting of hemicellulases, esterases (*e.g.*, lipases, phospholipases, and/or cutinases), proteases, laccases, peroxidases, or mixtures thereof. In the methods of the present invention, the additional enzyme(s) can be added prior to or during fermentation, including during or after propagation of the fermenting microorganism(s).

The enzymes can be derived or obtained from any suitable origin, including, bacterial, fungal, yeast, plant, or mammalian origin. The term "obtained" means herein that the enzyme may have been isolated from an organism that naturally produces the enzyme as a native enzyme. The term "obtained" also means herein that the enzyme may have been produced recombinantly in a host organism employing methods described herein, wherein the recombinantly produced enzyme is either native or foreign to the host organism or has a modified amino acid sequence, *e.g.*, having one or more amino acids that are deleted, inserted and/or substituted, *i.e.*, a recombinantly produced enzyme that is a mutant and/or a fragment of a native amino acid sequence or an enzyme produced by nucleic acid shuffling processes known in the art. Encompassed within the meaning of a native enzyme are natural variants and within the meaning of a foreign enzyme are variants obtained recombinantly, such as by site-directed mutagenesis or shuffling.

The enzymes used in the present invention can be in any form suitable for use in the methods described herein, such as a crude fermentation broth with or without cells or substantially pure polypeptides. The enzyme(s) can be a dry powder or granulate, a non-dusting granulate, a liquid, a stabilized liquid, or a protected enzyme(s). Granulates can be produced, *e.g.*, as disclosed in U.S. Patent Nos. 4,106,991 and 4,661,452, and can optionally be coated by process known in the art. Liquid enzyme preparations can, for instance, be stabilized by adding stabilizers such as a sugar, a sugar alcohol or another polyol, and/or lactic acid or another organic acid according to established process. Protected enzymes can be prepared according to the process disclosed in EP 238,216.

The optimum amounts of the enzymes and polypeptides having cellulolytic enhancing activity depend on several factors including, but not limited to, the mixture of component cellulolytic enzymes, the cellulosic substrate, the concentration of cellulosic substrate, the pretreatment(s) of the cellulosic substrate, temperature, time, pH, and inclusion of fermenting organism (*e.g.*, yeast for Simultaneous Saccharification and Fermentation).

In a preferred aspect, an effective amount of cellulolytic enzyme(s) to cellulosic material is about 0.5 to about 50 mg, preferably at about 0.5 to about 40 mg, more preferably at about 0.5 to about 25 mg, more preferably at about 0.75 to about 20 mg, more preferably at about 0.75 to about 15 mg, even more preferably at about 0.5 to about 10 mg, and most preferably at about 2.5 to about 10 mg per g of cellulosic material.

In another preferred aspect, an effective amount of a polypeptide having cellulolytic enhancing activity to cellulosic material is about 0.01 to about 50 mg,

preferably at about 0.5 to about 40 mg, more preferably at about 0.5 to about 25 mg, more preferably at about 0.75 to about 20 mg, more preferably at about 0.75 to about 15 mg, even more preferably at about 0.5 to about 10 mg, and most preferably at about 2.5 to about 10 mg per g of cellulosic material.

In another preferred aspect, an effective amount of polypeptide(s) having cellulolytic enhancing activity to cellulosic material is about 0.01 to about 50.0 mg, preferably about 0.01 to about 40 mg, more preferably about 0.01 to about 30 mg, more preferably about 0.01 to about 20 mg, more preferably about 0.01 to about 10 mg, more preferably about 0.01 to about 5 mg, more preferably at about 0.025 to about 1.5 mg, more preferably at about 0.05 to about 1.25 mg, more preferably at about 0.075 to about 1.25 mg, more preferably at about 0.1 to about 1.25 mg, even more preferably at about 0.15 to about 1.25 mg, and most preferably at about 0.25 to about 1.0 mg per g of cellulosic material.

In another preferred aspect, an effective amount of polypeptide(s) having cellulolytic enhancing activity to cellulolytic enzyme(s) is about 0.005 to about 1.0 g, preferably at about 0.01 to about 1.0 g, more preferably at about 0.15 to about 0.75 g, more preferably at about 0.15 to about 0.5 g, more preferably at about 0.1 to about 0.5 g, even more preferably at about 0.1 to about 0.5 g, and most preferably at about 0.05 to about 0.2 g per g of cellulolytic enzyme(s).

Fermentation. The fermentable sugars obtained from the pretreated and hydrolyzed cellulosic material can be fermented by one or more fermenting microorganisms capable of fermenting the sugars directly or indirectly into a desired fermentation product. "Fermentation" or "fermentation process" refers to any fermentation process or any process comprising a fermentation step. Fermentation processes also include fermentation processes used in the consumable alcohol industry (*e.g.*, beer and wine), dairy industry (*e.g.*, fermented dairy products), leather industry, and tobacco industry. The fermentation conditions depend on the desired fermentation product and fermenting organism and can easily be determined by one skilled in the art.

In the fermentation step, sugars, released from the cellulosic material as a result of the pretreatment and enzymatic hydrolysis steps, are fermented to a product, *e.g.*, ethanol, by a fermenting organism, such as yeast. Hydrolysis (saccharification) and fermentation can be separate or simultaneous. Such methods include, but are not limited to, separate hydrolysis and fermentation (SHF); simultaneous saccharification and fermentation (SSF); simultaneous saccharification and cofermentation (SSCF); hybrid hydrolysis and fermentation (HHF); SHCF (separate hydrolysis and co-fermentation), HHCF (hybrid hydrolysis and fermentation), and direct microbial conversion (DMC).

Any suitable hydrolyzed cellulosic material can be used in the fermentation step in practicing the present invention. The material is generally selected based on the desired fermentation product, *i.e.*, the substance to be obtained from the fermentation, and the process employed, as is well known in the art. Examples of substrates suitable for use in the methods of present invention, include cellulosic materials, such as wood or plant residues or low molecular sugars DP1-3 obtained from processed cellulosic material that can be metabolized by the fermenting microorganism, and which can be supplied by direct addition to the fermentation medium.

The term "fermentation medium" is understood herein to refer to a medium before the fermenting microorganism(s) is(are) added, such as, a medium resulting from a saccharification process, as well as a medium used in a simultaneous saccharification and fermentation process (SSF).

"Fermenting microorganism" refers to any microorganism, including bacterial and fungal organisms, suitable for use in a desired fermentation process to produce a fermentation product. The fermenting organism can be C_6 and/or C_5 fermenting organisms, or a combination thereof. Both C_6 and C_5 fermenting organisms are well known in the art. Suitable fermenting microorganisms are able to ferment, *i.e.*, convert, sugars, such as glucose, xylose, xylulose, arabinose, maltose, mannose, galactose, or oligosaccharides, directly or indirectly into the desired fermentation product.

Examples of bacterial and fungal fermenting organisms producing ethanol are described by Lin *et al.*, 2006, *Appl. Microbiol. Biotechnol.* 69: 627-642.

Examples of fermenting microorganisms that can ferment C6 sugars include bacterial and fungal organisms, such as yeast. Preferred yeast includes strains of the *Saccharomyces* spp., preferably *Saccharomyces cerevisiae*.

Examples of fermenting organisms that can ferment C5 sugars include bacterial and fungal organisms, such as yeast. Preferred C5 fermenting yeast include strains of *Pichia*, preferably *Pichia stipitis*, such as *Pichia stipitis* CBS 5773; strains of *Candida*, preferably *Candida boidinii*, *Candida brassicae*, *Candida sheatae*, *Candida diddensii*, *Candida pseudotropicalis*, or *Candida utilis*.

Other fermenting organisms include strains of *Zymomonas*, such as *Zymomonas mobilis*; *Hansenula*, such as *Hansenula anomala*; *Klyveromyces*, such as *K. fragilis*; *Schizosaccharomyces*, such as *S. pombe*; and *E. coli*, especially *E. coli* strains that have been genetically modified to improve the yield of ethanol.

In a preferred aspect, the yeast is a *Saccharomyces* spp. In a more preferred aspect, the yeast is *Saccharomyces cerevisiae*. In another more preferred aspect, the yeast is *Saccharomyces distaticus*. In another more preferred aspect, the yeast is *Saccharomyces uvarum*. In another preferred aspect, the yeast is a *Kluyveromyces*. In

another more preferred aspect, the yeast is *Kluyveromyces marxianus*. In another more preferred aspect, the yeast is *Kluyveromyces fragilis*. In another preferred aspect, the yeast is a *Candida*. In another more preferred aspect, the yeast is *Candida boidinii*. In another more preferred aspect, the yeast is *Candida brassicae*. In another more preferred aspect, the yeast is *Candida diddensii*. In another more preferred aspect, the yeast is *Candida pseudotropicalis*. In another more preferred aspect, the yeast is *Candida utilis*. In another preferred aspect, the yeast is a *Clavispora*. In another more preferred aspect, the yeast is *Clavispora lusitaniae*. In another more preferred aspect, the yeast is *Clavispora opuntiae*. In another preferred aspect, the yeast is a *Pachysolen*. In another more preferred aspect, the yeast is *Pachysolen tannophilus*. In another preferred aspect, the yeast is a *Pichia*. In another more preferred aspect, the yeast is a *Pichia stipitis*. In another preferred aspect, the yeast is a *Bretannomyces*. In another more preferred aspect, the yeast is *Bretannomyces clausenii* (Philippidis, G. P., 1996, Cellulose bioconversion technology, in *Handbook on Bioethanol: Production and Utilization*, Wyman, C. E., ed., Taylor & Francis, Washington, DC, 179-212).

Bacteria that can efficiently ferment hexose and pentose to ethanol include, for example, *Zymomonas mobilis* and *Clostridium thermocellum* (Philippidis, 1996, *supra*).

In a preferred aspect, the bacterium is a *Zymomonas*. In a more preferred aspect, the bacterium is *Zymomonas mobilis*. In another preferred aspect, the bacterium is a *Clostridium*. In another more preferred aspect, the bacterium is *Clostridium thermocellum*.

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

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

The cloning of heterologous genes into various fermenting microorganisms has led to the construction of organisms capable of converting hexoses and pentoses to ethanol (cofermentation) (Chen and Ho, 1993, Cloning and improving the expression of *Pichia stipitis* xylose reductase gene in *Saccharomyces cerevisiae*, *Appl. Biochem. Biotechnol.* 39-40: 135-147; Ho *et al.*, 1998, Genetically engineered *Saccharomyces* yeast capable of effectively cofermenting glucose and xylose, *Appl. Environ. Microbiol.*

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

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

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

The fermenting microorganism is typically added to the degraded lignocellulose or hydrolysate and the fermentation is performed for about 8 to about 96 hours, such as about 24 to about 60 hours. The temperature is typically between about 26°C to about 60°C, in particular about 32°C or 50°C, and at about pH 3 to about pH 8, such as around pH 4-5, 6, or 7.

In a preferred aspect, the yeast and/or another microorganism is applied to the degraded lignocellulose or hydrolysate and the fermentation is performed for about 12 to about 96 hours, such as typically 24-60 hours. In a preferred aspect, the temperature is preferably between about 20°C to about 60°C, more preferably about 25°C to about 50°C, and most preferably about 32°C to about 50°C, in particular about 32°C or 50°C, and the pH is generally from about pH 3 to about pH 7, preferably around pH 4-7. However, some bacterial fermenting organisms, for example, have higher fermentation temperature optima. Yeast or another microorganism is preferably applied in amounts of approximately 10^5 to 10^{12}, preferably from approximately 10^7 to 10^{10}, especially approximately 2×10^8 viable cell count per ml of fermentation broth. Further guidance in respect of using yeast for fermentation can be found in, *e.g.*, "The Alcohol Textbook"

(Editors K. Jacques, T.P. Lyons and D.R. Kelsall, Nottingham University Press, United Kingdom 1999), which is hereby incorporated by reference.

The most widely used process in the art is the simultaneous saccharification and fermentation (SSF) process where there is no holding stage for the saccharification, meaning that yeast and enzyme are added together.

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

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

Fermentation products: A fermentation product can be any substance derived from the fermentation. The fermentation product can be, without limitation, an alcohol (*e.g.*, arabinitol, butanol, ethanol, glycerol, methanol, 1,3-propanediol, sorbitol, and xylitol); an organic acid (*e.g.*, acetic acid, acetonic acid, adipic acid, ascorbic acid, citric acid, 2,5-diketo-D-gluconic acid, formic acid, fumaric acid, glucaric acid, gluconic acid, glucuronic acid, glutaric acid, 3-hydroxypropionic acid, itaconic acid, lactic acid, malic acid, malonic acid, oxalic acid, propionic acid, succinic acid, and xylonic acid); a ketone (*e.g.*, acetone); an amino acid (*e.g.*, aspartic acid, glutamic acid, glycine, lysine, serine, and threonine); and a gas (*e.g.*, methane, hydrogen (H_2), carbon dioxide (CO_2), and carbon monoxide (CO)). The fermentation product can also be protein as a high value product.

In a preferred aspect, the fermentation product is an alcohol. It will be understood that the term "alcohol" encompasses a substance that contains one or more hydroxyl moieties. In a more preferred aspect, the alcohol is arabinitol. In another more preferred aspect, the alcohol is butanol. In another more preferred aspect, the alcohol

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

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

In another preferred aspect, the fermentation product is a ketone. It will be understood that the term "ketone" encompasses a substance that contains one or more ketone moieties. In another more preferred aspect, the ketone is acetone. See, for

example, Qureshi and Blaschek, 2003, *supra*.

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

In another preferred aspect, the fermentation product is a gas. In another more preferred aspect, the gas is methane. In another more preferred aspect, the gas is H_2. In another more preferred aspect, the gas is CO_2. In another more preferred aspect, the gas is CO. See, for example, Kataoka, N., A. Miya, and K. Kiriyama, 1997, Studies on hydrogen production by continuous culture system of hydrogen-producing anaerobic bacteria, *Water Science and Technology* 36 (6-7): 41-47; and Gunaseelan V.N. in *Biomass and Bioenergy*, Vol. 13 (1-2), pp. 83-114, 1997, Anaerobic digestion of biomass for methane production: A review.

Recovery. The fermentation product(s) can be optionally recovered from the fermentation medium using any method known in the art including, but not limited to, chromatography, electrophoretic procedures, differential solubility, distillation, or extraction. For example, alcohol is separated from the fermented cellulosic material and purified by conventional methods of distillation. Ethanol with a purity of up to about 96 vol.% can be obtained, which can be used as, for example, fuel ethanol, drinking ethanol, *i.e.*, potable neutral spirits, or industrial ethanol.

Cellulolytic Enzyme Compositions

In the methods of the present invention, the cellulolytic enzyme composition may comprise any protein involved in the processing of a cellulose-containing material to glucose, or hemicellulose to xylose, mannose, galactose, and arabinose, their polymers, or products derived from them as described below. In one aspect, the cellulolytic enzyme composition comprises one or more enzymes selected from the group consisting of an endoglucanase, a cellobiohydrolase, and a beta-glucosidase. In another aspect, the cellulolytic enzyme composition further comprises one or more additional enzyme activities to improve the degradation of the cellulose-containing material. Preferred additional enzymes are hemicellulases, esterases (*e.g.*, lipases, phospholipases, and/or cutinases), proteases, laccases, peroxidases, or mixtures

thereof.

The cellulolytic enzyme composition may be a monocomponent preparation, *e.g.*, an endoglucanase, a multicomponent preparation, *e.g.*, endoglucanase(s), cellobiohydrolase(s), and beta-glucosidase(s), or a combination of multicomponent and monocomponent protein preparations. The cellulolytic proteins may have activity, *i.e.*, hydrolyze the cellulose-containing material, either in the acid, neutral, or alkaline pH-range.

As mentioned above, the cellulolytic proteins used in the present invention may be monocomponent preparations, *i.e.*, a component essentially free of other cellulolytic components. The single component may be a recombinant component, *i.e.*, produced by cloning of a DNA sequence encoding the single component and subsequent cell transformed with the DNA sequence and expressed in a host (see, for example, WO 91/17243 and WO 91/17244). The host cell may be a heterologous host (enzyme is foreign to host) or the host may also be a wild-type host (enzyme is native to host). Monocomponent cellulolytic proteins may also be prepared by purifying such a protein from a fermentation broth.

The enzymes used in the present invention may be in any form suitable for use in the processes described herein, such as, for example, a crude fermentation broth with or without cells, a dry powder or granulate, a non-dusting granulate, a liquid, a stabilized liquid, or a protected enzyme. Granulates may be produced, *e.g.*, as disclosed in U.S. Patent Nos. 4,106,991 and 4,661,452, and may optionally be coated by process known in the art. Liquid enzyme preparations may, for instance, be stabilized by adding stabilizers such as a sugar, a sugar alcohol or another polyol, and/or lactic acid or another organic acid according to established process. Protected enzymes may be prepared according to the process disclosed in EP 238,216.

A polypeptide having cellulolytic enzyme activity may be a bacterial polypeptide. For example, the polypeptide may be a gram positive bacterial polypeptide such as a *Bacillus*, *Streptococcus*, *Streptomyces*, *Staphylococcus*, *Enterococcus*, *Lactobacillus*, *Lactococcus*, *Clostridium*, *Geobacillus*, or *Oceanobacillus* polypeptide having cellulolytic enzyme activity, or a Gram negative bacterial polypeptide such as an *E. coli*, *Pseudomonas*, *Salmonella*, *Campylobacter*, *Helicobacter*, *Flavobacterium*, *Fusobacterium*, *Ilyobacter*, *Neisseria*, or *Ureaplasma* polypeptide having cellulolytic enzyme activity.

In a preferred aspect, the polypeptide is a *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or

Bacillus thuringiensis polypeptide having cellulolytic enzyme activity.

In another preferred aspect, the polypeptide is a *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, or *Streptococcus equi* subsp. *Zooepidemicus* polypeptide having cellulolytic enzyme activity.

In another preferred aspect, the polypeptide is a *Streptomyces achromogenes*, *Streptomyces avermitilis*, *Streptomyces coelicolor*, *Streptomyces griseus*, or *Streptomyces lividans* polypeptide having cellulolytic enzyme activity.

The polypeptide having cellulolytic enzyme activity may also be a fungal polypeptide, and more preferably a yeast polypeptide such as a *Candida*, *Kluyveromyces*, *Pichia*, *Saccharomyces*, *Schizosaccharomyces*, or *Yarrowia* polypeptide having cellulolytic enzyme activity; or more preferably a filamentous fungal polypeptide such as aan *Acremonium*, *Agaricus*, *Alternaria*, *Aspergillus*, *Aureobasidium*, *Botryospaeria*, *Ceriporiopsis*, *Chaetomidium*, *Chrysosporium*, *Claviceps*, *Cochliobolus*, *Coprinopsis*, *Coptotermes*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Diplodia*, *Exidia*, *Filibasidium*, *Fusarium*, *Gibberella*, *Holomastigotoides*, *Humicola*, *Irpex*, *Lentinula*, *Leptospaeria*, *Magnaporthe*, *Melanocarpus*, *Meripilus*, *Mucor*, *Myceliophthora*, *Neocallimastix*, *Neurospora*, *Paecilomyces*, *Penicillium*, *Phanerochaete*, *Piromyces*, *Poitrasia*, *Pseudoplectania*, *Pseudotrichonympha*, *Rhizomucor*, *Schizophyllum*, *Scytalidium*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Tolypocladium*, *Trichoderma*, *Trichophaea*, *Verticillium*, *Volvariella*, or *Xylaria* polypeptide having cellulolytic enzyme activity.

In a preferred aspect, the polypeptide is a *Saccharomyces carlsbergensis*, *Saccharomyces cerevisiae*, *Saccharomyces diastaticus*, *Saccharomyces douglasii*, *Saccharomyces kluyveri*, *Saccharomyces norbensis*, or *Saccharomyces oviformis* polypeptide having cellulolytic enzyme activity.

In another preferred aspect, the polypeptide is an *Acremonium cellulolyticus*, *Aspergillus aculeatus*, *Aspergillus awamori*, *Aspergillus fumigatus*, *Aspergillus foetidus*, *Aspergillus japonicus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Chrysosporium keratinophilum*, *Chrysosporium lucknowense*, *Chrysosporium tropicum*, *Chrysosporium merdarium*, *Chrysosporium inops*, *Chrysosporium pannicola*, *Chrysosporium queenslandicum*, *Chrysosporium zonatum*, *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*, *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*, *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*, *Trichoderma viride*, or *Trichophaea saccata* polypeptide having cellulolytic enzyme activity.

Chemically modified or protein engineered mutants of cellulolytic proteins may also be used.

One or more components of the cellulolytic enzyme composition may be a recombinant component, *i.e.*, produced by cloning of a DNA sequence encoding the single component and subsequent cell transformed with the DNA sequence and expressed in a host (see, for example, WO 91/17243 and WO 91/17244). The host is preferably a heterologous host (enzyme is foreign to host), but the host may under certain conditions also be a homologous host (enzyme is native to host). Monocomponent cellulolytic proteins may also be prepared by purifying such a protein from a fermentation broth.

Examples of commercial cellulolytic protein preparations suitable for use in the present invention include, for example, CELLUCLAST™ (available from Novozymes A/S) and NOVOZYM™ 188 (available from Novozymes A/S). Other commercially available preparations comprising cellulase that may be used include CELLUZYME™, CEREFLO™ and ULTRAFLO™ (Novozymes A/S), LAMINEX™ and SPEZYME™ CP (Genencor Int.), ROHAMENT™ 7069 W (Röhm GmbH), and FIBREZYME® LDI, FIBREZYME® LBR, or VISCOSTAR® 150L (Dyadic International, Inc., Jupiter, FL, USA). The cellulase enzymes are added in amounts effective from about 0.001% to about 5.0 % wt. of solids, more preferably from about 0.025% to about 4.0% wt. of solids, and most preferably from about 0.005% to about 2.0% wt. of solids.

Examples of bacterial endoglucanases that can be used in the methods of the present invention, include, but are not limited to, an *Acidothermus cellulolyticus* endoglucanase (WO 91/05039; WO 93/15186; U.S. Patent No. 5,275,944; WO 96/02551; U.S. Patent No. 5,536,655, WO 00/70031, WO 05/093050); *Thermobifida fusca* endoglucanase III (WO 05/093050); and *Thermobifida fusca* endoglucanase V (WO 05/093050).

Examples of fungal endoglucanases that can be used in the methods of the present invention, include, but are not limited to, a *Trichoderma reesei* endoglucanase I (Penttila *et al.*, 1986, *Gene* 45: 253-263; GENBANK™ accession no. M15665); *Trichoderma reesei* endoglucanase II (Saloheimo, *et al.*, 1988, Gene 63:11-22;

GENBANK™ accession no. M19373); *Trichoderma reesei* endoglucanase III (Okada *et al.*, 1988, *Appl. Environ. Microbiol.* 64: 555-563; GENBANK™ accession no. AB003694); *Trichoderma reesei* endoglucanase IV (Saloheimo *et al.*, 1997, *Eur. J. Biochem.* 249: 584-591; GENBANK™ accession no. Y11113); and *Trichoderma reesei* endoglucanase V (Saloheimo *et al.*, 1994, *Molecular Microbiology* 13: 219-228; GENBANK™ accession no. Z33381); *Aspergillus aculeatus* endoglucanase (Ooi *et al.*, 1990, *Nucleic Acids Research* 18: 5884); *Aspergillus kawachii* endoglucanase (Sakamoto *et al.*, 1995, *Current Genetics* 27: 435-439); *Erwinia carotovara* endoglucanase (Saarilahti *et al.*, 1990, *Gene* 90: 9-14); *Fusarium oxysporum* endoglucanase (GENBANK™ accession no. L29381); *Humicola grisea* var. *thermoidea* endoglucanase (GENBANK™ accession no. AB003107); *Melanocarpus albomyces* endoglucanase (GENBANK™ accession no. MAL515703); *Neurospora crassa* endoglucanase (GENBANK™ accession no. XM_324477); *Humicola insolens* endoglucanase V (SEQ ID NO: 25); *Myceliophthora thermophila* CBS 117.65 endoglucanase (SEQ ID NO: 27); basidiomycete CBS 495.95 endoglucanase (SEQ ID NO: 29); basidiomycete CBS 494.95 endoglucanase (SEQ ID NO: 31); *Thielavia terrestris* NRRL 8126 CEL6B endoglucanase (SEQ ID NO: 33); *Thielavia terrestris* NRRL 8126 CEL6C endoglucanase (SEQ ID NO: 35); *Thielavia terrestris* NRRL 8126 CEL7C endoglucanase (SEQ ID NO: 37); *Thielavia terrestris* NRRL 8126 CEL7E endoglucanase (SEQ ID NO: 39); *Thielavia terrestris* NRRL 8126 CEL7F endoglucanase (SEQ ID NO: 41); *Cladorrhinum foecundissimum* ATCC 62373 CEL7A endoglucanase (SEQ ID NO: 43); and *Trichoderma reesei* strain No. VTT-D-80133 endoglucanase (SEQ ID NO: 45; GENBANK™ accession no. M15665). The endoglucanases of SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35, SEQ ID NO: 37, SEQ ID NO: 39, SEQ ID NO: 41, SEQ ID NO: 43, and SEQ ID NO: 45 described above are encoded by the mature polypeptide coding sequence of SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 30, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 36, SEQ ID NO: 38, SEQ ID NO: 40, SEQ ID NO: 42, and SEQ ID NO: 44, respectively.

Examples of cellobiohydrolases useful in the methods of the present invention include, but are not limited to, *Trichoderma reesei* cellobiohydrolase I (SEQ ID NO: 47); *Trichoderma reesei* cellobiohydrolase II (SEQ ID NO: 49); *Humicola insolens* cellobiohydrolase I (SEQ ID NO: 51), *Myceliophthora thermophila* cellobiohydrolase II (SEQ ID NO: 53 and SEQ ID NO: 55), *Thielavia terrestris* cellobiohydrolase II (CEL6A) (SEQ ID NO: 57), *Chaetomium thermophilum* cellobiohydrolase I (SEQ ID NO: 59), and *Chaetomium thermophilum* cellobiohydrolase II (SEQ ID NO: 61). The cellobiohydrolases of SEQ ID NO: 47, SEQ ID NO: 49, SEQ ID NO: 51, SEQ ID NO: 53,

SEQ ID NO: 55, SEQ ID NO: 57, SEQ ID NO: 59, and SEQ ID NO: 61 described above are encoded by the mature polypeptide coding sequence of SEQ ID NO: 46, SEQ ID NO: 48, SEQ ID NO: 50, SEQ ID NO: 52, SEQ ID NO: 54, SEQ ID NO: 56, SEQ ID NO: 58, and SEQ ID NO: 60, respectively.

Examples of beta-glucosidases useful in the methods of the present invention include, but are not limited to, *Aspergillus oryzae* beta-glucosidase (SEQ ID NO: 63); *Aspergillus fumigatus* beta-glucosidase (SEQ ID NO: 65); *Penicillium brasilianum* IBT 20888 beta-glucosidase (SEQ ID NO: 67); *Aspergillus niger* beta-glucosidase (SEQ ID NO: 69); and *Aspergillus aculeatus* beta-glucosidase (SEQ ID NO: 71). The beta-glucosidases of SEQ ID NO: 63, SEQ ID NO: 65, SEQ ID NO: 67, SEQ ID NO: 69, and SEQ ID NO: 71 described above are encoded by the mature polypeptide coding sequence of SEQ ID NO: 62, SEQ ID NO: 64, SEQ ID NO: 66, SEQ ID NO: 68, and SEQ ID NO: 70, respectively.

The *Aspergillus oryzae* polypeptide having beta-glucosidase activity can be obtained according to WO 2002/095014. The *Aspergillus fumigatus* polypeptide having beta-glucosidase activity can be obtained according to WO 2005/047499. The *Penicillium brasilianum* polypeptide having beta-glucosidase activity can be obtained according to WO 2007/019442. The *Aspergillus niger* polypeptide having beta-glucosidase activity can be obtained according to Dan *et al.*, 2000, *J. Biol. Chem.* 275: 4973-4980. The *Aspergillus aculeatus* polypeptide having beta-glucosidase activity can be obtained according to Kawaguchi *et al.*, 1996, *Gene* 173: 287-288.

The beta-glucosidase may be a fusion protein. In one aspect, the beta-glucosidase is the *Aspergillus oryzae* beta-glucosidase variant BG fusion protein of SEQ ID NO: 73 or the *Aspergillus oryzae* beta-glucosidase fusion protein of SEQ ID NO: 75. In another aspect, the *Aspergillus oryzae* beta-glucosidase variant BG fusion protein is encoded by the polynucleotide of SEQ ID NO: 72 or the *Aspergillus oryzae* beta-glucosidase fusion protein is encoded by the polynucleotide of SEQ ID NO: 74.

Other endoglucanases, cellobiohydrolases, and beta-glucosidases are disclosed in numerous Glycosyl Hydrolase families using the classification according to Henrissat B., 1991, A classification of glycosyl hydrolases based on amino-acid sequence similarities, *Biochem. J.* 280: 309-316, and Henrissat B., and Bairoch A., 1996, Updating the sequence-based classification of glycosyl hydrolases, *Biochem. J.* 316: 695-696.

Other cellulolytic enzymes that may be used in the present invention are described in EP 495,257, EP 531,315, EP 531,372, WO 89/09259, WO 94/07998, WO 95/24471, WO 96/11262, WO 96/29397, WO 96/034108, WO 97/14804, WO 98/08940, WO 98/012307, WO 98/13465, WO 98/015619, WO 98/015633, WO 98/028411, WO

99/06574, WO 99/10481, WO 99/025846, WO 99/025847, WO 99/031255, WO 2000/009707, WO 2002/050245, WO 2002/0076792, WO 2002/101078, WO 2003/027306, WO 2003/052054, WO 2003/052055, WO 2003/052056, WO 2003/052057, WO 2003/052118, WO 2004/016760, WO 2004/043980, WO 2004/048592, WO 2005/001065, WO 2005/028636, WO 2005/093050, WO 2005/093073, WO 2006/074005, WO 2006/117432, WO 2007/071818, WO 2007/071820, WO 2008/008070, WO 2008/008793, U.S. Patent No. 4,435,307, U.S. Patent No. 5,457,046, U.S. Patent No. 5,648,263, U.S. Patent No. 5,686,593, U.S. Patent No. 5,691,178, U.S. Patent No. 5,763,254, and U.S. Patent No. 5,776,757.

The cellulolytic enzymes used in the methods of the present invention may be produced by fermentation of the above-noted microbial strains on a nutrient medium containing suitable carbon and nitrogen sources and inorganic salts, using procedures known in the art (see, *e.g.*, Bennett, J.W. and LaSure, L. (eds.), *More Gene Manipulations in Fungi*, Academic Press, CA, 1991). 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). Temperature ranges and other conditions suitable for growth and cellulolytic enzyme production are known in the art (see, *e.g.*, Bailey, J.E., and Ollis, D.F., *Biochemical Engineering Fundamentals*, McGraw-Hill Book Company, NY, 1986).

The fermentation can be any method of cultivation of a cell resulting in the expression or isolation of a cellulolytic enzyme. Fermentation may, therefore, be understood as comprising shake flask cultivation, or small- 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 cellulolytic enzyme to be expressed or isolated. The resulting cellulolytic enzymes produced by the methods described above may be recovered from the fermentation medium and purified by conventional procedures.

Signal Peptide

The present invention also relates to nucleic acid constructs comprising a gene encoding a protein, wherein the gene is operably linked to a nucleotide sequence encoding a signal peptide comprising or consisting of amino acids 1 to 17 of SEQ ID NO: 2 or amino acids 1 to 15 of SEQ ID NO: 4, wherein the gene is foreign to the nucleotide sequence.

In a preferred aspect, the nucleotide sequence comprises or consists of nucleotides 1 to 51 of SEQ ID NO: 1 or nucleotides 1 to 45 of SEQ ID NO: 3.

The present invention also relates to recombinant expression vectors and

recombinant host cells comprising such nucleic acid constructs.

The present invention also relates to methods of producing a protein comprising (a) cultivating such a recombinant host cell under conditions suitable for production of the protein; and (b) recovering the protein.

The protein may be native or heterologous to a host cell. The term "protein" is not meant herein to refer to a specific length of the encoded product and, therefore, encompasses peptides, oligopeptides, and proteins. The term "protein" also encompasses two or more polypeptides combined to form the encoded product. The proteins also include hybrid polypeptides that comprise a combination of partial or complete polypeptide sequences obtained from at least two different proteins wherein one or more (several) may be heterologous or native to the host cell. Proteins further include naturally occurring allelic and engineered variations of the above mentioned proteins and hybrid proteins.

Preferably, the protein is a hormone or variant thereof, enzyme, receptor or portion thereof, antibody or portion thereof, or reporter. In a more preferred aspect, the protein is an oxidoreductase, transferase, hydrolase, lyase, isomerase, or ligase. In an even more preferred aspect, the protein is an aminopeptidase, amylase, carbohydrase, carboxypeptidase, catalase, cellulase, chitinase, cutinase, cyclodextrin glycosyltransferase, deoxyribonuclease, esterase, alpha-galactosidase, beta-galactosidase, glucoamylase, alpha-glucosidase, beta-glucosidase, invertase, laccase, another lipase, mannosidase, mutanase, oxidase, pectinolytic enzyme, peroxidase, phytase, polyphenoloxidase, proteolytic enzyme, ribonuclease, transglutaminase or xylanase.

The gene may be obtained from any prokaryotic, eukaryotic, or other source.

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

Examples

Materials

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

Media

BA medium was composed per liter of 10 g of corn steep liquor dry matter, 10 g of NH_4NO_3, 10 g of KH_2PO_4, 0.75 g of $MgSO_4 \cdot 7H_2O$, 0.1 ml of pluronic, and 0.5 g of $CaCO_3$. The pH was adjusted to 6.5 before autoclaving.

YEG medium was composed per liter of 20 g of dextrose and 5 g of yeast extract.

Minimal medium plates were composed per liter of 6 g of $NaNO_3$, 0.52 g of KCl, 1.52 g of KH_2PO_4, 1 ml of COVE trace elements solution, 20 g of Noble agar, 20 ml of 50% glucose, 2.5 ml of $MgSO_4 \cdot 7H_2O$, and 20 ml of a 0.02% biotin solution.

COVE trace metals solution was composed per liter of 0.04 g of $Na_2B_4O_7 \cdot 10H_2O$, 0.4 g of $CuSO_4 \cdot 5H_2O$, 1.2 g of $FeSO_4 \cdot 7H_2O$, 0.7 g of $MnSO_4 \cdot H_2O$, 0.8 g of $Na_2MoO_2 \cdot 2H_2O$, and 10 g of $ZnSO_4 \cdot 7H_2O$.

M410 medium was composed per liter of 50 g of maltose, 50 g of glucose, 2 g of $MgSO_4 \cdot 7H_2O$, 2 g of KH_2PO_4, 4 g of anhydrous citric acid, 8 g of yeast extract, 2 g of urea, 0.5 g of $CaCl_2$, and 0.5 ml of AMG trace metals solution.

AMG trace metals was composed per liter of 14.3 g of $ZnSO_4 \cdot 7H_2O$, 2.5 g of $CuSO_4 \cdot 5H_2O$, 0.5 g of $NiCl_2 \cdot 6H_2O$, 13.8 g of $FeSO_4 \cdot 7H_2O$, 8.5 g of $MnSO_4 \cdot 7H_2O$, and 3 g of citric acid.

Example 1: Identification of Family 61 peptides

SDS-PAGE analysis. A commercial product was diluted 1:10 with water. Twenty µl was separated on a CRITERION™ 8–16% Tris-HCl SDS-PAGE gel according to the manufacturer's suggested conditions (Bio-Rad Laboratories, Inc., Hercules, CA, USA). PRECISION PLUS PROTEIN™ standards (Bio-Rad Laboratories, Inc., Hercules, CA, USA) were used as molecular weight markers. The gel was stained with BIO-SAFE™ Coomassie Stain (Bio-Rad Laboratories, Inc., Hercules, CA, USA), and visible bands were excised with a razor blade for protein identification analysis.

In-gel digestion of polypeptides for peptide sequencing. A MultiPROBE® II Liquid Handling Robot (PerkinElmer Life and Analytical Sciences, Boston, MA, USA) was used to perform the in-gel digestions. Gel bands containing protein were reduced with 50 µl of 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate pH 8.0 for 30 minutes. Following reduction, the gel piece was alkylated with 50 µl of 55 mM iodoacetamide in 100 mM ammonium bicarbonate pH 8.0 for 20 minutes. The dried gel piece was allowed to swell in 25 µl of a trypsin digestion solution (6 ng/µl sequencing grade trypsin (Promega, Madison, WI, USA) in 50 mM ammonium bicarbonate pH 8 for 30 minutes at room temperature, followed by an 8 hour digestion at 40°C. Each of the reaction steps described above was followed by numerous washes and pre-washes with

the appropriate solutions following the manufacturer's standard protocol. Fifty µl of acetonitrile was used to de-hydrate the gel piece between reactions and the gel piece was air dried between steps. Peptides were extracted twice with 1% formic acid/2% acetonitrile in HPLC grade water for 30 minutes. Peptide extraction solutions were transferred to a 96 well skirted PCR type plate (ABGene, Rochester, NY, USA) that had been cooled to 10–15°C and covered with a 96-well plate lid (PerkinElmer Life and Analytical Sciences, Boston, MA, USA) to prevent evaporation. Plates were further stored at 4°C until mass spectrometry analysis could be performed.

Protein identification. For *de novo* peptide sequencing by tandem mass spectrometry, a Q-TOF*MICRO*™ (Waters Micromass MS Technologies, Milford, MA, USA), a hybrid orthogonal quadrupole time-of-flight mass spectrometer was used for LC/MS/MS analysis. The Q-TOF *MICRO*™ is fully microprocessor controlled using MASSLYNX™ software version 4.1 (Waters Micromass MS Technologies, Milford, MA, USA). The Q-TOF *MICRO*™ was fitted with an ULTIMATE™ capillary and nano-flow HPLC system, which was coupled with a FAMOS™ micro autosampler and a SWITCHOS™ II column switching device (LCPackings/Dionex, Sunnyvale, CA, USA) for concentrating and desalting samples. Samples were loaded onto a guard column (300 µm ID X 5 cm, PEPMAP™ C18) fitted in the injection loop and washed with 0.1% formic acid in water at 40 µl per minute for 2 minutes using a Switchos II pump. Peptides were separated on a 75 µm ID x 15 cm, C18, 3 µm, 100 Å PEPMAP™ (LC Packings, San Francisco, CA, USA) nanoflow fused capillary column at a flow rate of 175 nl/minute from a split flow of 175 µl/minute using a NAN-75 calibrator (Dionex, Sunnyvale, CA, USA). A step elution gradient of 5% to 80% acetonitrile in 0.1% formic acid was applied over a 45 minute interval. The column eluent was monitored at 215 nm and introduced into the Q-TOF *MICRO*™ through an electrospray ion source fitted with the nanospray interface.

Data was acquired in survey scan mode and from a mass range of m/z 400 to 1990 with switching criteria for MS to MS/MS to include an ion intensity of greater than 10.0 counts per second and charge states of +2, +3, and +4. Analysis spectra of up to 4 co-eluting species with a scan time of 1.9 seconds and inter-scan time of 0.1 seconds could be obtained. A cone voltage of 45 volts was typically used and the collision energy was programmed to be varied according to the mass and charge state of the eluting peptide and in the range of 10–60 volts. The acquired spectra were combined, smoothed, and centered in an automated fashion and a peak list generated. This peak list was searched against selected databases using PROTEINLYNX™ Global Server 2.2.05 software (Waters Micromass MS Technologies, Milford, MA, USA) and PEAKS Studio version 4.5 (SP1) (Bioinformatic Solutions Inc., Waterloo, Ontario, Canada)

Results from the PROTEINLYNX™ and PEAKS Studio searches were evaluated and un-identified proteins were analyzed further by evaluating the MS/MS spectra of each ion of interest and *de novo* sequence was determined by identifying the y and b ion series and matching mass differences to the appropriate amino acid.

Peptide sequences were obtained from several multiply charged ions for the in-gel digested approximately 24 kDa polypeptide gel band. A doubly charged tryptic peptide ion of 871.56 m/z sequence was determined to be [Leu]-Pro-Ala-Ser-Asn-Ser-Pro-Val-Thr-Asp-Val-Thr-Ser-Asn-Ala-[Leu]-Arg (SEQ ID NO: 5). A doubly charged tryptic peptide ion of 615.84 m/z sequence was determined to be Val-Asp-Asn-Ala-Ala-Thr-Ala-Ser-Pro-Ser-Gly-[Leu]-Lys (SEQ ID NO: 6). A doubly charged tryptic peptide ion of 715.44 m/z sequence was determined to be [Leu]-Pro-Ala-Asp-[Leu]-Pro-Ser-Gly-Asp-Tyr-[Leu]-[Leu]-Arg (SEQ ID NO: 7). A doubly charged tryptic peptide ion of 988.58 m/z sequence was determined to be Gly-Pro-[Leu]-[Gln]-Val-Tyr-[Leu]-Ala-Lys (SEQ ID NO: 8). A double charged tryptic peptide ion of 1272.65 m/z sequence was determined to be Val-Ser-Val-Asn-Gly-[Gln]-Asp-[Gln]-Gly-[Gln]-[Leu]-Lys (SEQ ID NO: 9). [Leu] above may be Ile or Leu and [Gln] above may be Gln or Lys because they could not be distinguished due to equivalent masses.

Example 2: Preparation of *Myceliophthora thermophila* CBS 117.65 cDNA pool

Myceliophthora thermophila CBS 117.65 was cultivated in 200 ml of BA medium at 30°C for five days at 200 rpm. Mycelia from the shake flask culture were harvested by filtering the contents through a funnel lined with MIRACLOTH™ (CalBiochem, San Diego, CA, USA). The mycelia were then sandwiched between two MIRACLOTH™ pieces and blotted dry with absorbent paper towels. The mycelial mass was then transferred to plastic centrifuge tubes and frozen in liquid nitrogen. Frozen mycelia were stored in a -80°C freezer until use.

The extraction of total RNA was performed with guanidinium thiocyanate followed by ultracentrifugation through a 5.7 M CsCl cushion, and isolation of poly(A)+RNA was carried out by oligo(dT)-cellulose affinity chromatography, using the procedures described in WO 94/14953.

Double-stranded cDNA was synthesized from 5 µg of poly(A)+ RNA by the RNase H method (Gubler and Hoffman, 1983, *Gene* 25: 263-269, Sambrook *et al.*, 1989, Molecular cloning: A laboratory manual, Cold Spring Harbor lab., Cold Spring Harbor, NY, USA). The poly(A)+ RNA (5 µg in 5 µl of DEPC (0.1% diethylpyrocarbonate)-treated water) was heated at 70°C for 8 minutes in a pre-siliconized, RNase-free EPPENDORF® tube, quenched on ice, and combined in a final volume of 50 µl with reverse transcriptase buffer composed of 50 mM Tris-HCl, pH 8.3,

75 mM KCl, 3 mM $MgCl_2$, 10 mM dithiothreitol (DTT) (Bethesda Research Laboratories, Bethesda, MD, USA), 1 mM of dATP, dGTP and dTTP, and 0.5 mM 5-methyl-dCTP (GE Healthcare, Piscataway, NJ, USA), 40 units of human placental ribonuclease inhibitor (RNasin; Promega, Madison, WI, USA), 1.45 µg of oligo(dT)18-Not I primer (GE Healthcare, Piscataway, NJ, USA), and 1000 units of SuperScript II RNase H reverse transcriptase (Bethesda Research Laboratories, Bethesda, MD, USA). First-strand cDNA was synthesized by incubating the reaction mixture at 45°C for 1 hour. After synthesis, the mRNA:cDNA hybrid mixture was gel filtrated through a MICROSPIN™ S-400 HR spin column (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions.

After gel filtration, the hybrids were diluted in 250 µl of second strand buffer (20 mM Tris-HCl, pH 7.4, 90 mM KCl, 4.6 mM $MgCl_2$, 10 mM $(NH_4)_2SO_4$, 0.16 mM NAD) containing 200 µM of each dNTP, 60 units of *E. coli* DNA polymerase I (GE Healthcare, Piscataway, NJ, USA), 5.25 units of RNase H (Promega, Madison, WI, USA), and 15 units of *E. coli* DNA ligase (Boehringer Mannheim, Manheim, Germany). Second strand cDNA synthesis was performed by incubating the reaction tube at 16°C for 2 hours and an additional 15 minutes at 25°C. The reaction was stopped by addition of EDTA to a final concentration of 20 mM followed by phenol and chloroform extractions.

The double-stranded cDNA was precipitated at -20°C for 12 hours by addition of 2 volumes of 96% ethanol and 0.2 volume of 10 M ammonium acetate, recovered by centrifugation at 13,000 *x g*, washed in 70% ethanol, dried, and resuspended in 30 µl of Mung bean nuclease buffer (30 mM sodium acetate pH 4.6, 300 mM NaCl, 1 mM $ZnSO_4$, 0.35 mM DTT, 2% glycerol) containing 25 units of Mung bean nuclease (GE Healthcare, Piscataway, NJ, USA). The single-stranded hair-pin DNA was clipped by incubating the reaction at 30°C for 30 minutes, followed by addition of 70 µl of 10 mM Tris-HCl-1 mM EDTA pH 7.5, phenol extraction, and precipitation with 2 volumes of 96% ethanol and 0.1 volume of 3 M sodium acetate pH 5.2 on ice for 30 minutes.

The double-stranded cDNAs were recovered by centrifugation at 13,000 *x g* and blunt-ended in 30 µl of T4 DNA polymerase buffer (20 mM Tris-acetate, pH 7.9, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM DTT) containing 0.5 mM of each dNTP and 5 units of T4 DNA polymerase (New England Biolabs, Ipswich, MA, USA) by incubating the reaction mixture at 16°C for 1 hour. The reaction was stopped by addition of EDTA to a final concentration of 20 mM, followed by phenol and chloroform extractions, and precipitation for 12 hours at -20°C by adding 2 volumes of 96% ethanol and 0.1 volume of 3 M sodium acetate pH 5.2. After the fill-in reaction the cDNAs were recovered by centrifugation at 13,000 *x g*, washed in 70% ethanol, and dried.

Example 3: *Myceliophthora thermophila* CBS 202.75 and *Myceliophthora thermophila* CBS 117.65 genomic DNA extraction

Myceliophthora thermophila CBS 202.75 and *Myceliophthora thermophila* CBS 117.65 strains were grown in 100 ml of YEG medium in a baffled shake flask at 45°C and 200 rpm for 2 days. Mycelia were harvested by filtration using MIRACLOTH® (Calbiochem, La Jolla, CA, USA), washed twice in deionized water, and frozen under liquid nitrogen. Frozen mycelia were ground, by mortar and pestle, to a fine powder, and total DNA was isolated using a DNEASY® Plant Maxi Kit (QIAGEN Inc., Valencia, CA, USA).

Example 4: Molecular screening of a Family 61 gene from *Myceliophthora thermophila*

Degenerate primers were designed, as shown below, based upon peptide sequences obtained through tandem mass spectrometry as described in Example 1.

Primer 061562 (CI61A sense):

5'-GCCTCCAACTCGCCCGTCACNGAYGTNAC-3' (SEQ ID NO: 10)

Primer 061563 (CI61A anti):

5'-GAGGTAGTCGCCGGANGGGATRTCNGCNGG-3' (SEQ ID NO: 11)

Fifty picomoles each of CI61A sense and CI61A anti primers were used in a PCR reaction composed of 100 ng of *Myceliophthora thermophila* CBS 117.65 cDNA pool, or *Myceliophthora thermophila* CBS 117.65 genomic DNA, 1X ADVANTAGE® GC-Melt LA Buffer (Clontech Laboratories, Inc., Mountain View, CA, USA), 0.4 mM each of dATP, dTTP, dGTP, and dCTP, and 1.25 units of ADVANTAGE® GC Genomic Polymerase Mix (Clontech Laboratories, Inc., Mountain View, CA, USA) in a final volume of 25 µl. The amplifications were performed using an EPPENDORF® MASTERCYCLER® 5333 (Eppendorf Scientific, Inc., Westbury, NY, USA) programmed for 1 cycle at 94°C for 1 minute; and 30 cycles each at 94°C for 30 seconds, 56.5°C for 30 seconds, and 72°C for 30 seconds, followed by a final extension of 5 minutes at 72°C.

The reaction products were fractionated by 1% agarose gel electrophoresis in 40 mM Tris base-20 mM sodium acetate-1 mM disodium EDTA (TAE) buffer and bands of greater than 400 bp were excised, purified using a MINELUATE® Gel Extraction Kit (QIAGEN Inc., Valencia, CA, USA) according to the manufacturer's instructions, and subcloned using a TOPO® TA Kit (Invitrogen, Carlsbad, CA, USA). Plasmid DNA was extracted from a number of *E. coli* transformants and sequenced. Sequence analysis of the *E. coli* clones showed that the sequences contained a coding region of a Family 61 gene (*gh61a*).

A second *gh61* gene was isolated in a separate PCR reaction performed under different conditions. Thirty picomoles each of Cl61A sense and Cl61A anti primers were used in a PCR reaction composed of 200 ng of *Myceliophthora thermophila* CBS 202.75 genomic DNA, 1X THERMOPOL® Buffer (New England BioLabs, Ipswich, MA, USA), 0.28 mM each of dATP, dTTP, dGTP, and dCTP, and 1.0 unit of *Taq* DNA polymerase (New England BioLabs, Ipswich, MA) in a final volume of 30 µl. The amplifications were performed using a ROBOCYCLER® 40 (Strategene, La Jolla, CA, USA) programmed for 1 cycle at 96°C for 3 minutes; 1 cycle at 72°C for 3 minutes during which DNA polymerase was added; 30 cycles each at 94°C for 50 seconds, 52°C for 50 seconds, and 72°C for 90 seconds, followed by a final extension of 7 minutes at 72°C.

The reaction products were fractionated by 1% agarose gel electrophoresis in 40 mM Tris base-20 mM sodium acetate-1 mM disodium EDTA (TAE) buffer and a band of greater than 400-500 bp was excised, purified using a QIAEX II® Gel Extraction Kit (QIAGEN Inc., Valencia, CA, USA) according to the manufacturer's instructions, and subcloned using a TOPO® TA Kit (Invitrogen, Carlsbad, CA, USA). Plasmid DNA was extracted from a number of *E. coli* transformants and sequenced. Sequence analysis revealed several clones containing a coding region of one Family 61 gene designated *gh61f*.

Example 5: Isolation of a full-length Family 61 gene (*gh61a*) from *Myceliophthora thermophila* CBS 202.75

A full-length Family 61 gene (*gh61a*) from *Myceliophthora thermophila* CBS 202.75 was isolated using a GENOMEWALKER™ Universal Kit (Clontech Laboratories, Inc., Mountain View, CA, USA) according to the manufacturer's instructions. Briefly, total genomic DNA from *Myceliophthora thermophila* CBS 202.75 was digested separately with four different restriction enzymes (*Dra* I, *Eco* RV, *Pvu* II, and *Stu* I) that leave blunt ends. Each batch of digested genomic DNA was then ligated separately to the GENOMEWALKER™ Adaptor (Clontech Laboratories, Inc., Mountain View, CA, USA) to create four libraries. These libraries were then employed as templates in PCR reactions using four gene-specific primers for *Myceliophthora thermophila* Family 61 *gh61a* gene. The primers shown below were designed based on the partial Family *gh61a* gene sequences obtained in Example 4.

Upstream Region Primers:

MtCel61A-R1: 5'-CCGTTCGGGCCGTCTTGGTAGATCTTGAACC-3' (SEQ ID NO: 12)

MtCel61A-R2: 5'-CCGATGGAGGGATCCACGCTGAAGGTGAATT-3' (SEQ ID NO: 13)

Downstream Region Primers:

MtCel61A-F1: 5'-CAGGTCAAGGCGGGCTCCCAATTCACCTT-3' (SEQ ID NO: 14)

MtCel61A-F2: 5'-ACGGCACGGGAGCCGTGTGGTTCAAGATCTA-3' (SEQ ID NO: 15)

Two primary PCR amplifications were performed, one to isolate the upstream region and the other the downstream region of the *Myceliophthora thermophila gh61a* gene. Each PCR amplification (25 µl) was composed of 1 µl (approximately 6 ng) of each library as template, 0.4 mM each of dATP, dTTP, dGTP, and dCTP, 10 pmol of Adaptor Primer 1 (Clontech Laboratories, Inc., Mountain View, CA, USA), 10 pmol of primer MtCel61A-R1 or primer MtCel61A-F1, 1X ADVANTAGE® GC-Melt LA Buffer, and 1.25 units of ADVANTAGE® GC Genomic Polymerase Mix. The amplifications were performed using an EPPENDORF® MASTERCYCLER® 5333 programmed for pre-denaturing at 94°C for 1 minute; 7 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation at 72°C for 5 minutes; and 32 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation 67°C for 5 minutes, followed by a final extension of 7 minutes at 67°C.

The secondary amplifications were composed of 1 µl of each primary PCR product as template, 0.4 mM each of dATP, dTTP, dGTP, and dCTP, 10 pmol of Adaptor Primer 2 (Clontech Laboratories, Inc., Mountain View, CA, USA), 10 pmol of nested primer MtCel61A-R2 or MtCel61A-F2, 1X ADVANTAGE® GC-Melt LA Buffer, and 1.25 units of ADVANTAGE® GC Genomic Polymerase Mix in a final volume of 25 µl. The amplifications were performed using an EPPENDORF® MASTERCYCLER® 5333 programmed for pre-denaturing at 94°C for 1 minute; 5 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation at 72°C for 5 minutes; and 20 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation at 67°C for 5 minutes, followed by a final extension of 7 minutes at 67°C.

The reaction products were isolated by 1.0% agarose gel electrophoresis in TAE buffer where a 1.7 kb product band (upstream region) from the *Eco* RV library and a 1.6 kb band (downstream region) from the *Stu* I library were excised from the gel, purified using a MINELUTE® Gel Extraction Kit (QIAGEN Inc., Valencia, CA, USA) according to the manufacturer's instructions. The PCR products were sequenced directly or subcloned using a TOPO® TA Kit and then sequenced.

Example 6: Characterization of the *Myceliophthora thermophila* CBS 202.75 genomic sequence encoding a Family GH61A polypeptide having cellulolytic enhancing activity

DNA sequencing of the PCR fragment was performed with a Perkin-Elmer Applied Biosystems Model 377 XL Automated DNA Sequencer (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA, USA) using dye-terminator chemistry (Giesecke *et*

al., 1992, *Journal of Virology Methods* 38: 47-60) and primer walking strategy. Nucleotide sequence data were scrutinized for quality and all sequences were compared to each other with assistance of PHRED/PHRAP software (University of Washington, Seattle, WA, USA).

A gene model for the *Myceliophthora thermophila* GH61A polypeptide having cellulolytic enhancing activity was constructed based on similarity of the encoded protein to homologous glycoside hydrolase Family 61 proteins from *Thielavia terrestris* (accession numbers GENESEQP:ADM97933, GENESEQP:AEB90517), *Chaetomium globosum* (UNIPROT:Q2HGH1, UNIPROT:Q2GW98) and *Neurospora crassa* (UNIPROT:Q7S439). To verify the sequence information obtained for the *Myceliophthora thermophila gh61a* gene, a further PCR reaction was carried out using a pair of gene specific primers (shown below), which encompass the complete gene.

Primer MtGH61A-F3:

5'-ACTGGATTTACC**ATGAAGTTCACCTCGTCCCTCGCT**-3' (SEQ ID NO: 16)

Primer MtGH61A-R3:

5'-TCACCTCTAGTTAATTAA**TTAGCAAGAGACGGGGGCCG**-3' (SEQ ID NO: 17)

Bold letters represent coding sequence. The remaining sequence is homologous to the insertion sites of pAILo2 (WO 2004/099228).

The PCR consisted of fifty picomoles of forward and reverse primers in a PCR reaction composed of 100 ng of *Myceliophthora thermophila* CBS 202.75 genomic DNA, *Pfx* Amplification Buffer (Invitrogen, Carlsbad, CA, USA), 0.4 mM each of dATP, dTTP, dGTP, and dCTP, 1 mM $MgCl_2$, and 2.5 units of *Pfx* DNA polymerase (Invitrogen, Carlsbad, CA, USA) in a final volume of 50 µl. The amplification was performed using an EPPENDORF® MASTERCYCLER® 5333 programmed for 1 cycle at 98°C for 3 minutes; and 30 cycles each at 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute, , followed by a final extension of 15 minutes at 72°C. The heat block then went to a 4°C soak cycle.

The reaction products were isolated by 1.0% agarose gel electrophoresis in TAE buffer and purified using a MINELUTE® Gel Extraction Kit according to the manufacturer's instructions. In order to clone the PCR fragments into pCR®2.1-TOPO® vector (Invitrogen, Carlsbad, CA, USA), addition of 3' A-overhangs was performed using *Taq* DNA polymerase (New England Biolabs, Ipswich, MA, USA).

A 958 bp *Myceliophthora thermophila gh61a* gene fragment was cloned into pCR®2.1-TOPO® vector (Invitrogen, Carlsbad, CA, USA) using a TOPO® TA Cloning Kit to generate pSMai190 (Figure 2).

The *Myceliophthora thermophila gh61a* insert was confirmed by DNA sequencing. *E. coli* pSMai190 was deposited with the Agricultural Research Service

Patent Culture Collection, Northern Regional Research Center, Peoria, IL, USA, on December 5, 2007, and assigned accession number B-50083.

The nucleotide sequence (SEQ ID NO: 1) and deduced amino acid sequence (SEQ ID NO: 2) of the *Myceliophthora thermophila* GH61A polypeptide having cellulolytic enhancing activity are shown in Figure 1. The genomic polynucleotide encodes a polypeptide of 232 amino acids, interrupted by 2 introns of 88 and 137 bp. The % G+C content of the full-length coding sequence and the mature coding sequence are 61.1% and 66.5%, respectively. Using the SignalP software program (Nielsen *et al.*, 1997, *Protein Engineering* 10: 1-6), a signal peptide of 17 residues was predicted. The predicted mature protein contains 215 amino acids with a molecular mass of 22.6 kDa.

A comparative pairwise global alignment of amino acid sequences was determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in the Needle program of EMBOSS with gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 matrix. The alignment showed that the deduced amino acid sequence of the *Myceliophthora thermophila* GH61A mature polypeptide shared 76.6% identity (excluding gaps) to the deduced amino acid sequence of a Family 61 glycoside hydrolase protein from *Thielavia terrestris* (GeneSeqP accession numbers ADM97933).

Example 7: Isolation of a full-length Family 61 gene (*gh61f*) from *Myceliophthora thermophila* CBS 202.75

A full-length Family 61 gene (*gh61f*) from *Myceliophthora thermophila* CBS 202.75 was isolated using a GENOMEWALKER™ Universal Kit (Clontech Laboratories, Inc., Mountain View, CA, USA) according to the manufacturer's instructions. Briefly, total genomic DNA from *Myceliophthora thermophila* CBS 202.75 was digested separately with four different restriction enzymes (*Dra* I, *Eco* RV, *Pvu* II, and *Stu* I) that leave blunt ends. Each batch of digested genomic DNA was then ligated separately to the GENOMEWALKER™ Adaptor (Clontech Laboratories, Inc., Mountain View, CA, USA) to create four libraries. These libraries were then employed as templates in PCR reactions using gene-specific primers for the *Myceliophthora thermophila* Family 61 gene (*gh61f*). The primers shown below were designed based on the partial Family 61 *gh61f* gene sequences obtained in Example 4.

Upstream Region Primers:

MtGH61F-R1: 5'-CCCTTGTGGCTGGCGTCCATGACATCGTC-3' (SEQ ID NO: 18)

MtGH61F-R2: 5'-GTGCCTCCAGATGGCCTTGACCGTGGTG-3' (SEQ ID NO: 19)

Downstream Region Primers:

MtGH61F-F6: 5'-GGCGGCGAGCACTACATGTGAGCCATTCCT-3' (SEQ ID NO: 20)

MtGH61F-F7: 5'-TGACGATCTCGCTGACCCGTGCAACAAGTG-3' (SEQ ID NO: 21)

Two primary PCR amplifications were performed, one to isolate the upstream region and the other to isolate the downstream region of the *Myceliophthora thermophila gh61f* gene. Each PCR amplification (25 µl) was composed of 1 µl (approximately 6 ng) of each library as template, 0.4 mM each of dATP, dTTP, dGTP, and dCTP, 10 pmol of Adaptor Primer 1 (Clontech Laboratories, Inc., Mountain View, CA, USA), 10 pmol of primer MtGH61F-R1 or primer MtGH61F-F6, 1X ADVANTAGE® GC-Melt LA Buffer, and 1.25 units of ADVANTAGE® GC Genomic Polymerase Mix. The amplifications were performed using an EPPENDORF® MASTERCYCLER® 5333 programmed for pre-denaturing at 94°C for 1 minute; 7 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation at 72°C for 5 minutes; and 32 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation 67°C for 5 minutes, followed by a final extension of 7 minutes at 67°C.

The secondary amplifications were composed of 1 µl of each primary PCR product as template, 0.4 mM each of dATP, dTTP, dGTP, and dCTP, 10 pmol of Adaptor Primer 2 (Clontech Laboratories, Inc., Mountain View, CA, USA), 10 pmol of nested primer MtGH61F-R2 or MtGH61F-F7, 1X ADVANTAGE® GC-Melt LA Buffer, and 1.25 units of ADVANTAGE® GC Genomic Polymerase Mix in a final volume of 25 µl. The amplifications were performed using an EPPENDORF® MASTERCYCLER® 5333 programmed for pre-denaturing at 94°C for 1 minute; 5 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation at 72°C for 5 minutes; and 20 cycles each at a denaturing temperature of 94°C for 30 seconds; annealing and elongation at 67°C for 5 minutes, followed by a final extension of 7 minutes at 67°C.

The reaction products were isolated by 1.0% agarose gel electrophoresis in TAE buffer where a 1.3 kb PCR product (upstream region) from the *Puv* II library and a 1.2 kb PCR product (upstream region) from the *Puv* II library were excised from the gel, purified using a MINELUTE® Gel Extraction Kit (QIAGEN Inc., Valencia, CA, USA) according to the manufacturer's instructions, and the PCR products were sequenced directly or subcloned using a TOPO® TA Kit and then sequenced.

Example 8: Characterization of the *Myceliophthora thermophila* genomic sequence encoding a Family GH61F polypeptide having cellulolytic enhancing activity

DNA sequencing of the PCR fragments was performed with a Perkin-Elmer Applied Biosystems Model 377 XL Automated DNA Sequencer (Perkin-Elmer/Applied Biosystems, Inc., Foster City, CA, USA) using dye-terminator chemistry (Giesecke *et*

al., 1992, *Journal of Virology Methods* 38: 47-60) and primer walking strategy. Nucleotide sequence data were scrutinized for quality and all sequences were compared to each other with assistance of PHRED/PHRAP software (University of Washington, Seattle, WA, USA).

A gene model for the *Myceliophthora thermophila* GH61F polypeptide having cellulolytic enhancing activity was constructed based on similarity of the encoded protein to homologous glycoside hydrolase Family 61 proteins from *Thielavia terrestris* (accession numbers GENESEQP:ADM97933, GENESEQP:AEB90517), *Chaetomium globosum* (UNIPROT:Q2HGH1, UNIPROT:Q2GW98) and *Neurospora crassa* (UNIPROT:Q7S439). To verify the sequence information obtained for the *Myceliophthora thermophila gh61f* gene, a further PCR reaction was carried out using gene specific primers (shown below), which encompass the complete gene.

Primer MtGH61F-F8:

5'-ACTGGATTTACC**ATGAAGGCCCTCTCTCTCCTTGCG**-3' (SEQ ID NO: 22)

Primer MtGH61F-R3:

5'-TCACCTCTAGTTAATTAA**CTAGCACTTGAAGACGGGCG**-3' (SEQ ID NO: 23)

Bold letters represent coding sequence. The remaining sequence is homologous to the insertion sites of pAILo2 (WO 2004/099228).

The PCR consisted of 50 picomoles of forward and reverse primers in a PCR reaction composed of 100 ng of *Myceliophthora thermophila* CBS 202.75 genomic DNA, *Pfx* Amplification Buffer (Invitrogen, Carlsbad, CA, USA), 0.4 mM each of dATP, dTTP, dGTP, and dCTP, 1 mM $MgCl_2$, and 2.5 units of *Pfx* DNA polymerase (Invitrogen, Carlsbad, CA, USA) in a final volume of 50 µl. The amplification were performed using an EPPENDORF® MASTERCYCLER® 5333 programmed for 1 cycle at 98°C for 3 minutes; and 30 cycles each at 98°C for 30 seconds, 60°C for 30 seconds, and 72°C for 1 minute, followed by a final extension of 15 minutes at 72°C.. The heat block then went to a 4°C soak cycle.

The reaction products were isolated by 1.0% agarose gel electrophoresis in TAE buffer and purified using a MINELUTE® Gel Extraction Kit according to the manufacturer's instructions. In order to clone the PCR fragments into pCR®2.1-TOPO® vector (Invitrogen, Carlsbad, CA, USA), addition of 3' A-overhangs was performed using *Taq* DNA polymerase (New England Biolabs, Ipswich, MA, USA).

An 884 bp *Myceliophthora thermophila gh61f* gene fragment was cloned into pCR®2.1-TOPO® vector using a TOPO® TA Cloning Kit to generate pSMai192 (Figure 2).

The *Myceliophthora thermophila gh61f* insert was confirmed by DNA sequencing. *E. coli* pSMai192 was deposited with the Agricultural Research Service

Patent Culture Collection, Northern Regional Research Center, 1815 University Street, Peoria, IL, USA, on December 5, 2007, and assigned accession number B-50085.

The nucleotide sequence (SEQ ID NO: 3) and deduced amino acid sequence (SEQ ID NO: 4) of the *Myceliophthora thermophila* GH61F polypeptide having cellulolytic enhancing activity are shown in Figure 1. The genomic polynucleotide encodes a polypeptide of 235 amino acids, interrupted by 2 introns of 62 and 84 bp. The % G+C content of the full-length coding sequence and the mature coding sequence are 64.1% and 65.4%, respectively. Using the SignalP software program (Nielsen *et al.*, 1997, *Protein Engineering* 10: 1-6), a signal peptide of 15 residues was predicted. The predicted mature protein contains 220 amino acids with a molecular mass of 23 kDa.

A comparative pairwise global alignment of amino acid sequences was determined using the Needleman-Wunsch algorithm (Needleman and Wunsch, 1970, *J. Mol. Biol.* 48: 443-453) as implemented in the Needle program of EMBOSS with gap open penalty of 10, gap extension penalty of 0.5, and the EBLOSUM62 matrix. The alignment showed that the deduced amino acid sequence of the *Myceliophthora thermophila* GH61F mature polypeptide shared 83.8% identity (excluding gaps) to the deduced amino acid sequence of a Family 61 glycoside hydrolase protein from *Chaetomium globosum* (UniProt accession number Q2HGH1).

Example 9: Construction of an *Aspergillus oryzae* expression vector containing *Myceliophthora thermophila* CBS 202.75 genomic sequence encoding a Family GH61A polypeptide having cellulolytic enhancing activity

The same 958 bp *Myceliophthora thermophila gh61a* PCR fragment generated in Example 6 was cloned into *Nco* I and *Pac* I digested pAILo2 (WO 2004/099228) using an Infusion Cloning Kit (BD Biosciences, Palo Alto, CA, USA) resulting in pSMai185 (Figure 5) in which transcription of the *Myceliophthora thermophila gh61a* gene was under the control of a hybrid of promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase (NA2-tpi promoter). The ligation reaction (50 µl) was composed of 1X InFusion Buffer (BD Biosciences, Palo Alto, CA, USA), 1X BSA (BD Biosciences, Palo Alto, CA, USA), 1 µl of Infusion enzyme (diluted 1:10) (BD Biosciences, Palo Alto, CA, USA), 100 ng of pAILo2 digested with *Nco* I and *Pac* I, and 50 ng of the *Myceliophthora thermophila gh61a* purified PCR product. The reaction was incubated at room temperature for 30 minutes. One µl of the reaction was used to transform *E. coli* XL10 SOLOPACK® Gold Supercompetent cells (Stratagene, La Jolla, CA, USA). An *E. coli* transformant containing pSMai185 was detected by restriction digestion and plasmid DNA was prepared using a BIOROBOT® 9600 (QIAGEN Inc., Valencia, CA, USA). The *Myceliophthora*

thermophila gh61a insert in pSMai185 was confirmed by DNA sequencing.

Example 10: Construction of an *Aspergillus oryzae* expression vector containing *Myceliophthora thermophila* CBS 202.75 genomic sequence encoding a Family GH61F polypeptide having cellulolytic enhancing activity

The same 884 bp *Myceliophthora thermophila gh61f* PCR fragment generated in Example 8 was cloned into *Nco* I and *Pac* I digested pAlLo2 (WO 2004/099228) using an Infusion Cloning Kit (BD Biosciences, Palo Alto, CA, USA) resulting in pSMai198 (Figure 6) in which transcription of the *Myceliophthora thermophila gh61f* gene was under the control of a hybrid of promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase (NA2-tpi promoter). The ligation reaction (50 µl) was composed of 1X InFusion Buffer (BD Biosciences, Palo Alto, CA, USA), 1X BSA (BD Biosciences, Palo Alto, CA, USA), 1 µl of Infusion enzyme (diluted 1:10) (BD Biosciences, Palo Alto, CA, USA), 100 ng of pAlLo2 digested with *Nco* I and *Pac* I, and 50 ng of the *Myceliophthora thermophila gh61f* purified PCR product. The reaction was incubated at room temperature for 30 minutes. One µl of the reaction was used to transform *E. coli* XL10 SOLOPACK® Gold Supercompetent cells (Stratagene, La Jolla, CA, USA). An *E. coli* transformant containing pSMai198 was detected by restriction digestion and plasmid DNA was prepared using a BIOROBOT® 9600 (QIAGEN Inc., Valencia, CA, USA). The *Myceliophthora thermophila gh61f* insert in pSMai198 was confirmed by DNA sequencing.

Example 11: Expression of the *Myceliophthora thermophila* Family 61 glycosyl hydrolase genes (*gh61a* and *gh61f*) individually in *Aspergillus oryzae* JaL355

Aspergillus oryzae JaL355 (WO 2002/40694) protoplasts were prepared according to the method of Christensen *et al.*, 1988, *Bio/Technology* 6: 1419-1422. Three µg of pSMai185 (*gh61a*) or pSMai198 (*gh61f*) were transformed individually into *Aspergillus oryzae* JaL355.

Twenty transformants were isolated to individual Minimal medium plates from each transformation experiment.

Confluent Minimal Medium plates of each of the transformants were washed with 5 ml of 0.01% TWEEN® 20 and inoculated separately into 25 ml of M410 medium in 125 ml glass shake flasks and incubated at 34°C, 250 rpm. After 5 days incubation, 5 µl of supernatant from each culture were analyzed on CRITERION® 8-16% Tris-HCl gels with a CRITERION® Cell (Bio-Rad Laboratories, Inc., Hercules, CA, USA), according to the manufacturer's instructions. The resulting gels were stained with BIO-SAFE™ Coomassie Stain. SDS-PAGE profiles of the cultures showed that the majority

of the transformants had the expected band sizes: 23 KDa for GH61A and 23 KDa for GH61F.

One of each high protein expressing GH61A and GH61F transformants were washed with 10 ml of 0.01% TWEEN® 20 and inoculated into a 2 liter Fernbach containing 500 ml of M410 medium to generate broth for characterization of the proteins. The cultures were harvested on day 5 and filtered using a 0.22 µm EXPRESS™ Plus Membrane (Millipore, Bedford, MA, USA).

Example 12: Effect of *Myceliophthora thermophila* GH61A and GH61F polypeptides on enzymatic hydrolysis of pretreated corn stover

Culture broth was prepared as decribed in Example 11 and concentrated approximately 20-fold using an Amicon ultrafiltration device (Millipore, Bedfored, MA., 10 kDa polyethersulfone membrane, 40 psi, 4°C). Protein concentration was estimated by densitometry following SDS-PAGE and Coomassie blue staining. Corn stover was pretreated and prepared as an assay substrate as described in WO 2005/074647 to generate pretreated corn stover (PCS). The base cellulase mixture used to assay enhancing activity was prepared from *Trichoderma reesei* strain SMA135 (WO 2008/057637).

Hydrolysis of PCS was conducted using 1.6 ml deep-well plates (Axygen, Santa Clara, CA.) using a total reaction volume of 1.0 ml and a PCS concentration of 50 mg/ml in 1 mM manganese sulfate-50 mM sodium acetate, pH 5.0. The *M. thermophila* polypeptides (GH61A and GH61F) were added to the base cellulase mixture at concentrations ranging from 0 to 25% of the protein concentration of the base cellulase mixture. Incubation was at 50°C for 168 hours. Assays were performed in triplicate. Aliquots were centrifuged, and the supernatant liquid was filtered by centrifugation (MULTISCREEN® HV 0.45 µm, Millipore, Billerica, MA, USA) at 3000 rpm for 10 minutes using a plate centrifuge (SORVALL® RT7, Thermo Fisher Scientific, Waltham, MA, USA). When not used immediately, filtered hydrolysate aliquots were frozen at -20°C. Sugar concentrations of samples diluted in 0.005 M H_2SO_4 with 0.05% w/w benzoic acid were measured after elution by 0.005 M H_2SO_4 with 0.05% w/w benzoic acid at a flow rate of 0.6 ml/minute from a 4.6 x 250 mm AMINEX® HPX-87H column (Bio-Rad Laboratories, Inc., Inc., Hercules, CA, USA) at 65°C with quantitation by integration of glucose and cellobiose signal from refractive index detection (CHEMSTATION®, AGILENT® 1100 HPLC, Agilent Technologies, Santa Clara, CA, USA) calibrated by pure sugar samples (Absolute Standards Inc., Hamden, CT, USA). The resultant equivalents were used to calculate the percentage of cellulose conversion for each reaction. The degree of cellulose conversion to glucose plus cellobiose sugars

(conversion, %) was calculated using the following equation:

Conversion (%) = (glucose+cellobiose x 1.053) (mg/ml) x 100 x 162 / (Cellulose (mg/ml) x 180) = (glucose+cellobiose x 1.053) (mg/ml) x 100 / (Cellulose (mg/ml) x 1.111)

In this equation the factor 1.111 reflects the weight gain in converting cellulose to glucose, and the factor 1.053 reflects the weight gain in converting cellobiose to glucose. Cellulose in PCS was determined by a limit digest of PCS to release glucose and cellobiose.

The results of adding increasing amounts of *Myceliopthora thermophila* polypeptides to the base cellulase mix are shown in Figure 7. Addition of the *M. thermophila* GH61A polypeptide provided a stimulation factor of 1.26 at the 25% addition level. At the same addition percentage, *M. thermophila* GH61F provided a stimulation factor of 1.13. Stimulation factor is defined as the ratio of conversion observed in the presence of added GH61 protein versus conversion in the absence of added GH61 protein.

Deposits of Biological Material

The following biological materials have been deposited under the terms of the Budapest Treaty with the Agricultural Research Service Patent Culture Collection (NRRL), Northern Regional Research Center, 1815 University Street, Peoria, Illinois, 61604, USA, and given the following accession numbers:

Deposit	Accession Number	Date of Deposit
E. coli pSMai190	NRRL B-50083	December 5, 2007
E. coli pSMai192	NRRL B-50085	December 5, 2007

The strains have been deposited under conditions that assure that access to the cultures will be available during the pendency of this patent application to one determined by foreign patent laws to be entitled thereto. The deposits represent substantially pure cultures of the deposited strains. The depositse available as required by foreign patent laws in countries wherein counterparts of the subject application, or its progeny are filed. However, it should be understood that the availability of a deposit does not constitute a license to practice the subject invention in derogation of patent rights granted by governmental action.

The present invention is further described by the following numbered paragraphs:

[1] An isolated polypeptide having cellulolytic enhancing activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence having at least 60% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4;

(b) a polypeptide encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii);

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

(d) a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[2] The polypeptide of paragraph 1, comprising an amino acid sequence having at least 60% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[3] The polypeptide of paragraph 2, comprising an amino acid sequence having at least 65% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[4] The polypeptide of paragraph 3, comprising an amino acid sequence having at least 70% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[5] The polypeptide of paragraph 4, comprising an amino acid sequence having at least 75% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[6] The polypeptide of paragraph 5, comprising an amino acid sequence having at least 80% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[7] The polypeptide of paragraph 6, comprising an amino acid sequence having at least 85% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[8] The polypeptide of paragraph 7, comprising an amino acid sequence having at least 90% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[9] The polypeptide of paragraph 8, comprising an amino acid sequence having at least 95% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[10] The polypeptide of paragraph 1, comprising or consisting of the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4; or a fragment thereof having cellulolytic enhancing activity.

[11] The polypeptide of paragraph 10, comprising or consisting of the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4.

[12] The polypeptide of paragraph 10, comprising or consisting of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[13] The polypeptide of paragraph 1, which is encoded by a polynucleotide that

hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii).

[14] The polypeptide of paragraph 13, which is encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii).

[15] The polypeptide of paragraph 14, which is encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii).

[16] The polypeptide of paragraph 1, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 60% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[17] The polypeptide of paragraph 16, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 65% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[18] The polypeptide of paragraph 17, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 70% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[19] The polypeptide of paragraph 18, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 75% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[20] The polypeptide of paragraph 19, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 80% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[21] The polypeptide of paragraph 20, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 85% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[22] The polypeptide of paragraph 21, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 90% identity to the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[23] The polypeptide of paragraph 22, which is encoded by a polynucleotide comprising a nucleotide sequence having at least 95% identity to the mature

polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[24] The polypeptide of paragraph 1, which is encoded by a polynucleotide comprising or consisting of the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 3; or a subsequence thereof encoding a fragment having cellulolytic enhancing activity.

[25] The polypeptide of paragraph 24, which is encoded by a polynucleotide comprising or consisting of the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[26] The polypeptide of paragraph 24, which is encoded by a polynucleotide comprising or consisting of the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3.

[27] The polypeptide of paragraph 1, wherein the polypeptide is a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

[28] The polypeptide of paragraph 1, which is encoded by the polynucleotide contained in plasmid pSMai190 which is contained in *E. coli* NRRL B-50083 or plasmid pSMai192 which is contained in *E. coli* NRRL B-50085.

[29] The polypeptide of any of paragraphs 1-28, wherein the mature polypeptide is amino acids 18 to 232 of SEQ ID NO: 2 or SEQ ID NO: 4.

[30] The polypeptide of any of paragraphs 1-29, wherein the mature polypeptide coding sequence is nucleotides 52 to 921 of SEQ ID NO: 1 or nucleotides 46 to 851 of SEQ ID NO: 3.

[31] An isolated polynucleotide comprising a nucleotide sequence that encodes the polypeptide of any of paragraphs 1-30.

[32] The isolated polynucleotide of paragraph 31, comprising at least one mutation in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, in which the mutant nucleotide sequence encodes the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4, respectively.

[33] A nucleic acid construct comprising the polynucleotide of paragraph 31 or 32 operably linked to one or more (several) control sequences that direct the production of the polypeptide in an expression host.

[34] A recombinant expression vector comprising the nucleic acid construct of paragraph 33.

[35] A recombinant host cell comprising the nucleic acid construct of paragraph 33.

[36] A method of producing the polypeptide of any of paragraphs 1-30, 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.

[37] A method of producing the polypeptide of any of paragraphs 1-30, comprising: (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.

[38] A method of producing a mutant of a parent cell, comprising disrupting or deleting a nucleotide sequence encoding the polypeptide of any of paragraphs 1-30, which results in the mutant producing less of the polypeptide than the parent cell.

[39] A mutant cell produced by the method of paragraph 38.

[40] The mutant cell of paragraph 39, further comprising a gene encoding a native or heterologous protein.

[41] A method of producing a protein, comprising: (a) cultivating the mutant cell of paragraph 40 under conditions conducive for production of the protein; and (b) recovering the protein.

[42] The isolated polynucleotide of paragraph 31 or 32, obtained by (a) hybridizing a population of DNA under at least high stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii); and (b) isolating the hybridizing polynucleotide, which encodes a polypeptide having cellulolytic enhancing activity.

[43] The isolated polynucleotide of paragraph 42, wherein the mature polypeptide coding sequence is nucleotides 52 to 921 of SEQ ID NO: 1 or nucleotides 46 to 851 of SEQ ID NO: 3.

[44] A method of producing a polynucleotide comprising a mutant nucleotide sequence encoding a polypeptide having cellulolytic enhancing activity, comprising: (a) introducing at least one mutation into the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, wherein the mutant nucleotide sequence encodes a polypeptide comprising or consisting of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4; and (b) recovering the polynucleotide comprising the mutant nucleotide sequence.

[45] A mutant polynucleotide produced by the method of paragraph 44.

[46] A method of producing a polypeptide, comprising: (a) cultivating a cell comprising the mutant polynucleotide of paragraph 45 encoding the polypeptide under conditions conducive for production of the polypeptide; and (b) recovering the polypeptide.

[47] A method of producing the polypeptide of any of paragraphs 1-30, 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.

[48] A transgenic plant, plant part or plant cell transformed with a polynucleotide encoding the polypeptide of any of paragraphs 1-30.

[49] A double-stranded inhibitory RNA (dsRNA) molecule comprising a subsequence of the polynucleotide of paragraph 31 or 32, wherein optionally the dsRNA is a siRNA or a miRNA molecule.

[50] The double-stranded inhibitory RNA (dsRNA) molecule of paragraph 49, which is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more duplex nucleotides in length.

[51] A method of inhibiting the expression of a polypeptide having cellulolytic enhancing activity in a cell, comprising administering to the cell or expressing in the cell a double-stranded RNA (dsRNA) molecule, wherein the dsRNA comprises a subsequence of the polynucleotide of paragraph 31 or 32.

[52] The method of paragraph 51, wherein the dsRNA is about 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or more duplex nucleotides in length.

[53] A nucleic acid construct comprising a gene encoding a protein operably linked to a nucleotide sequence encoding a signal peptide comprising or consisting of amino acids 1 to 17 of SEQ ID NO: 2 or amino acids 1 to 15 of SEQ ID NO: 4, wherein the gene is foreign to the nucleotide sequence.

[54] A recombinant expression vector comprising the nucleic acid construct of paragraph 53.

[55] A recombinant host cell comprising the nucleic acid construct of paragraph 53.

[56] A method of producing a protein, comprising: (a) cultivating the recombinant host cell of paragraph 55 under conditions conducive for production of the protein; and (b) recovering the protein.

[57] A method for degrading or converting a cellulosic material, comprising: treating the cellulosic material with a cellulolytic enzyme composition in the presence of the polypeptide having cellulolytic enhancing activity of any of paragraphs 1-30, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity.

[58] The method of paragraph 57, wherein the cellulosic material is pretreated.

[59] The method of paragraph 57 or 58, wherein the cellulolytic enzyme composition comprises one or more cellulolytic enzymes are selected from the group consisting of a cellulase, endoglucanase, cellobiohydrolase, and beta-glucosidase.

[60] The method of any of paragraphs 57-59, further comprising treating the cellulosic material with one or more enzymes selected from the group consisting of a hemicellulase, esterase, protease, laccase, and peroxidase.

[61] The method of any of paragraphs 57-60, further comprising recovering the degraded cellulosic material.

[62] The method of paragraph 61, wherein the degraded cellulosic material is a sugar.

[63] The method of paragraph 62, wherein the sugar is selected from the group consisting of glucose, xylose, mannose, galactose, and arabinose.

[64] A method for producing a fermentation product, comprising:

(a) saccharifying a cellulosic material with a cellulolytic enzyme composition in the presence of the polypeptide having cellulolytic enhancing activity of any of paragraphs 1-20, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity;

(b) fermenting the saccharified cellulosic material of step (a) with one or more fermenting microorganisms to produce the fermentation product; and

(c) recovering the fermentation product from the fermentation.

[65] The method of paragraph 64, wherein the cellulosic material is pretreated.

[66] The method of paragraph 64 or 65, wherein the cellulolytic enzyme composition comprises one or more cellulolytic enzymes selected from the group consisting of a cellulase, endoglucanase, cellobiohydrolase, and beta-glucosidase.

[67] The method of any of paragraphs 64-66, further comprising treating the cellulosic material with one or more enzymes selected from the group consisting of a hemicellulase, esterase, protease, laccase, and peroxidase.

[68] The method of any of paragraphs 64-67, wherein steps (a) and (b) are performed simultaneously in a simultaneous saccharification and fermentation.

[69] The method of any of paragraphs 64-68, wherein the fermentation product is an alcohol, organic acid, ketone, amino acid, or gas.

[70] A method of fermenting a cellulosic material, comprising: fermenting the cellulosic material with one or more fermenting microorganisms, wherein the cellulosic material is hydrolyzed with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity of any of paragraphs 1-30 and the presence of the polypeptide having cellulolytic enhancing activity increases the hydrolysis of the cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity.

[71] The method of paragraph 70, wherein the fermenting of the cellulosic

material produces a fermentation product.

[72] The method of paragraph 71, further comprising recovering the fermentation product from the fermentation.

[73] The method of any of paragraphs 70-72, wherein the cellulosic material is pretreated before saccharification.

[74] The method of any of paragraphs 70-73, wherein the cellulolytic enzyme composition comprises one or more cellulolytic enzymes selected from the group consisting of a cellulase, endoglucanase, cellobiohydrolase, and beta-glucosidase.

[75] The method of any of paragraphs 70-74, wherein the cellulolytic enzyme composition further comprises one or more enzymes selected from the group consisting of a hemicellulase, esterase, protease, laccase, and peroxidase.

[76] The method of any of paragraphs 70-75, wherein the fermentation product is an alcohol, organic acid, ketone, amino acid, or gas.

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

Claims

What is claimed is:

1. An isolated polypeptide having cellulolytic enhancing activity, selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence having at least 60% identity to the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4;

(b) a polypeptide encoded by a polynucleotide that hybridizes under at least medium stringency conditions with (i) the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, (ii) the cDNA sequence contained in the mature polypeptide coding sequence of SEQ ID NO: 1 or SEQ ID NO: 3, or (iii) a full-length complementary strand of (i) or (ii);

(c) a polypeptide encoded by a polynucleotide comprising a nucleotide sequence having at least 60% identity to the mature polypeptide coding sequence of SEQ ID NO: 1; and

(d) a variant comprising a substitution, deletion, and/or insertion of one or more (several) amino acids of the mature polypeptide of SEQ ID NO: 2 or SEQ ID NO: 4.

2. The polypeptide of claim 1, comprising or consisting of the amino acid sequence of SEQ ID NO: 2 or SEQ ID NO: 4; or a fragment thereof having cellulolytic enhancing activity.

3. The polypeptide of claim 1, which is encoded by the polynucleotide contained in plasmid plasmid pSMai190 which is contained in *E. coli* NRRL B-50083 or plasmid pSMai192 which is contained in *E. coli* NRRL B-50085.

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.

6. A recombinant host cell comprising the nucleic acid construct of claim 5.

7. A method of producing the polypeptide of any of claims 1-3, 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.

8. A method of producing the polypeptide of any of claims 1-3, comprising: (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.

9. A method of producing a mutant of a parent cell, comprising disrupting or deleting a nucleotide sequence encoding the polypeptide of any of claims 1-3, which results in the mutant producing less of the polypeptide than the parent cell.

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.

11. A transgenic plant, plant part or plant cell transformed with a polynucleotide encoding the polypeptide of any of claims 1-3.

12. A double-stranded inhibitory RNA (dsRNA) molecule comprising a subsequence of the polynucleotide of claim 4, wherein optionally the dsRNA is a siRNA or a miRNA molecule.

13. A method of inhibiting the expression of a polypeptide having cellulolytic enhancing activity in a cell, comprising administering to the cell or expressing in the cell a double-stranded RNA (dsRNA) molecule, wherein the dsRNA comprises a subsequence of the polynucleotide of claim 4.

14. A nucleic acid construct comprising a gene encoding a protein operably linked to a nucleotide sequence encoding a signal peptide comprising or consisting of amino acids 1 to 17 of SEQ ID NO: 2 or amino acids 1 to 15 of SEQ ID NO: 4, wherein the gene is foreign to the nucleotide sequence.

15. A recombinant host cell comprising the nucleic acid construct of claim 14.

16. A method of producing a protein, comprising: (a) cultivating the recombinant host cell of claim 15 under conditions conducive for production of the protein; and (b) recovering the protein.

17. A method for degrading or converting a cellulosic material, comprising: treating the cellulosic material with a cellulolytic enzyme composition in the presence of the polypeptide having cellulolytic enhancing activity of any of claims 1-3, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity.

18. The method of claim 17, further comprising recovering the degraded cellulosic material.

19. A method for producing a fermentation product, comprising:

(a) saccharifying a cellulosic material with a cellulolytic enzyme composition in the presence of the polypeptide having cellulolytic enhancing activity of any of claims 1-3, wherein the presence of the polypeptide having cellulolytic enhancing activity increases the degradation of cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity;

(b) fermenting the saccharified cellulosic material of step (a) with one or more fermenting microorganisms to produce the fermentation product; and

(c) recovering the fermentation product from the fermentation.

20. A method of fermenting a cellulosic material, comprising: fermenting the cellulosic material with one or more fermenting microorganisms, wherein the cellulosic material is hydrolyzed with a cellulolytic enzyme composition in the presence of a polypeptide having cellulolytic enhancing activity of any of claims 1-3 and the presence of the polypeptide having cellulolytic enhancing activity increases the hydrolysis of the cellulosic material compared to the absence of the polypeptide having cellulolytic enhancing activity.

```
   M  K  F  T  S  S  L  A  V  L  A  A  A  G  A  Q  A  H  Y  T  F  P  R  A
 1ATGAAGTTCACCTCGTCCCTCGCTGTCCTGGCCGCTGCCGGCGCCCAGGCTCACTATACCTTCCCTAGGGCC
   G  T  G  G  S  L  S  G  E  W  E  V  V  R  M  T  E  N  H  Y  S  H  G  P
73GGCACTGGTGGCTCGCTCTCTGGCGAGTGGGAGGTGGTCCGCATGACCGAGAACCATTACTCGCACGGCCCG
   V  T  D  V  T  S  P  E  M  T  C  Y  Q  S  G  V  Q  G  A  P  Q  T  V  Q
145GTCACCGATGTCACCAGCCCCGAGATGACCTGCTATCAGTCCGGCGTGCAGGGTGCGCCCCAGACCGTCCAG
   V  K  A  G  S  Q  F  T  F  S  V  D  P  S  I  G  H  P  G  P  L  Q  F  Y
217GTCAAGGCGGGCTCCCAATTCACCTTCAGCGTGGATCCCTCGATCGGCCACCCCGGCCCTCTCCAGTTCTAC
   M  A  K  V  P  S  G  Q  T  A  A  T  F  D  G  T  G  A  V  W  F  K  I  Y
289ATGGCTAAGGTGCCGTCGGGCCAGACGGCCGCCACCTTTGACGGCACGGGAGCCGTGTGGTTCAAGATCTAC
   Q  D  G  P  N  G  L  G  T  D  S  I  T  W  P  S  A  G  K  T  E  V  S  V
361CAAGACGGCCCGAACGGCCTCGGCACCGACAGCATTACCTGGCCCAGCGCCGGCAAAACCGAGGTCTCGGTC
   T  I  P  S  C  I  D  D  G  E  Y  L  L  R  V  E  H  I  A  L  H  S  A  S
433ACCATCCCCAGCTGCATCGATGATGGCGAGTACCTGCTCCGGGTCGAGCACATCGCGCTCCACAGCGCCAGC
   S  V  G  G  A  Q  F  Y  I  A  C  A  Q  L  S  V  T  G  G  S  G  T  L  N
505AGCGTGGGCGGCGCTCAGTTCTACATTGCCTGCGCCCAGCTCTCCGTCACCGGCGGCTCCGGCACCCTCAAC
   T  G  S  L  V  S  L  P  G  A  Y  K  A  T  D  P  G  I  L  F  Q  L  Y  W
577ACGGGCTCGCTCGTCTCCCTGCCCGGCGCCTACAAGGCCACCGACCCGGGCATCCTCTTCCAGCTCTACTGG
   P  I  P  T  E  Y  I  N  P  G  P  A  P  V  S  C  *
649CCCATCCCGACCGAGTACATCAACCCCGGCCCGGCCCCCGTCTCTTGCTAA
```

Fig. 1

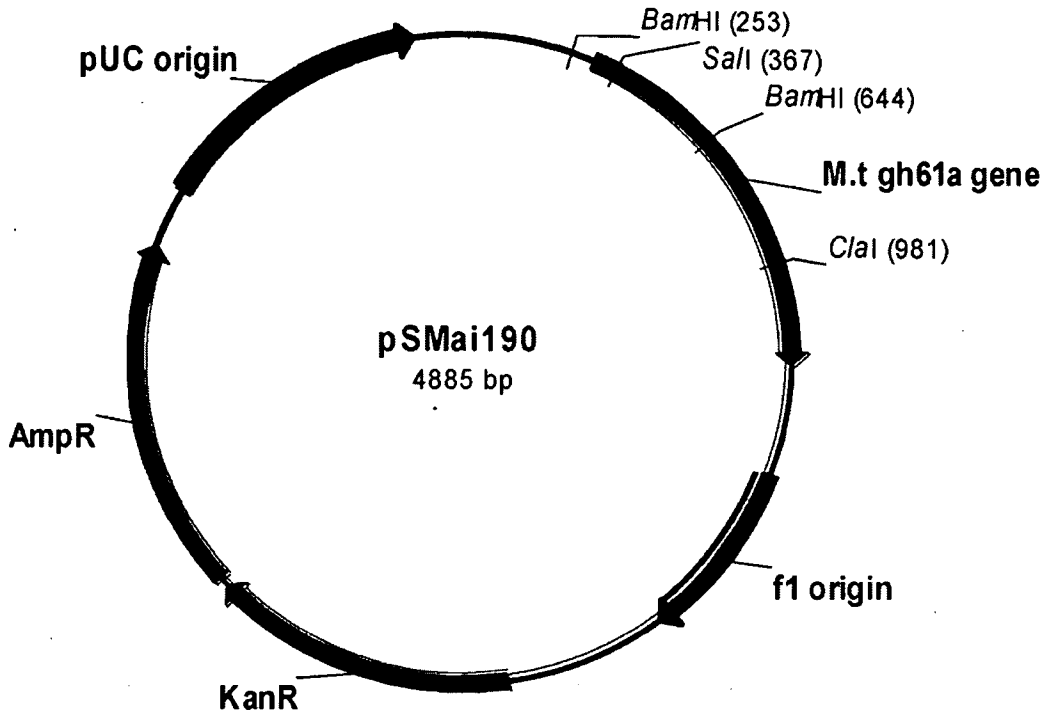
pUC origin
BamHI (253)
SalI (367)
BamHI (644)
M.t gh61a gene
ClaI (981)
pSMai190
4885 bp
AmpR
f1 origin
KanR

Fig. 2

```
    M  K  A  L  S  L  L  A  A  A  S  A  V  S  A  H  T  I  F  V  Q  L  E  A
  1 ATGAAGGCCCTCTCTCTCCTTGCGGCTGCCTCGGCAGTCTCTGCGCATACCATCTTCGTCCAGCTCGAAGCA
    D  G  T  R  Y  P  V  S  Y  G  I  R  D  P  S  Y  D  G  P  I  T  D  V  T
 73 GACGGCACGAGGTACCCGGTCTCGTACGGGATCCGGGACCCAAGCTACGACGGCCCCATCACCGACGTCACA
    S  N  D  V  A  C  N  G  G  P  N  P  T  T  P  S  S  D  V  I  T  V  T  A
145 TCCAACGACGTTGCTTGCAACGGCGGGCCGAACCCGACGACCCCCTCCAGCGACGTCATCACCGTCACCGCG
    G  T  T  V  K  A  I  W  R  H  T  L  Q  S  G  P  D  D  V  M  D  A  S  H
217 GGCACCACGGTCAAGGCCATCTGGAGGCACACCCTCCAATCCGGCCCGGACGATGTCATGGACGCCAGCCAC
    K  G  P  T  L  A  Y  L  K  K  V  G  D  A  T  K  D  S  G  V  G  G  G  W
289 AAGGGCCCGACCCTGGCCTACCTCAAGAAGGTCGGCGATGCCACCAAGGACTCGGGCGTCGGCGGTGGCTGG
    F  K  I  Q  E  D  G  Y  N  N  G  Q  W  G  T  S  T  V  I  S  N  G  G  E
361 TTCAAGATTCAGGAGGACGGCTACAACAACGGCCAGTGGGGCACCAGCACCGTTATCTCCAACGGCGGCGAG
    H  Y  I  D  I  P  A  C  I  P  E  G  Q  Y  L  L  R  A  E  M  I  A  L  H
433 CACTACATTGACATCCCGGCCTGCATCCCCGAGGGTCAGTACCTCCTCCGCGCCGAGATGATCGCCCTCCAC
    A  A  G  S  P  G  G  A  Q  L  Y  M  E  C  A  Q  I  N  I  V  G  G  S  G
505 GCGGCCGGGTCCCCCGGCGGTGCCCAGCTCTACATGGAATGTGCCCAGATCAACATCGTCGGCGGCTCCGGC
    S  V  P  S  S  T  V  S  F  P  G  A  Y  S  P  N  D  P  G  L  L  I  N  I
577 TCGGTGCCCAGCTCGACCGTCAGCTTCCCCGGCGCGTACAGCCCCAACGACCCGGGTCTCCTCATCAACATC
    Y  S  M  S  P  S  S  S  Y  T  I  P  G  P  P  V  F  K  C  *
649 TATTCCATGTCGCCCTCGAGCTCGTACACCATCCCGGGCCCGCCCGTCTTCAAGTGCTAG
```

Fig. 3

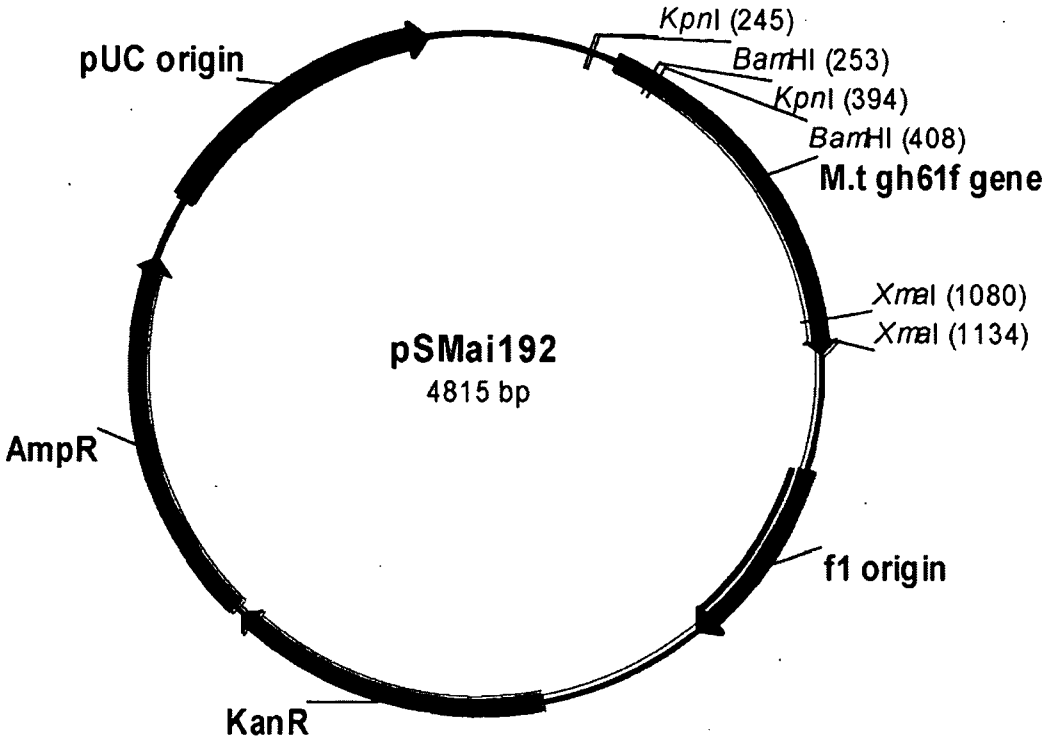
KpnI (245)
BamHI (253)
KpnI (394)
BamHI (408)
M.t gh61f gene
XmaI (1080)
XmaI (1134)
pSMai192
4815 bp
pUC origin
AmpR
f1 origin
KanR

Fig. 4

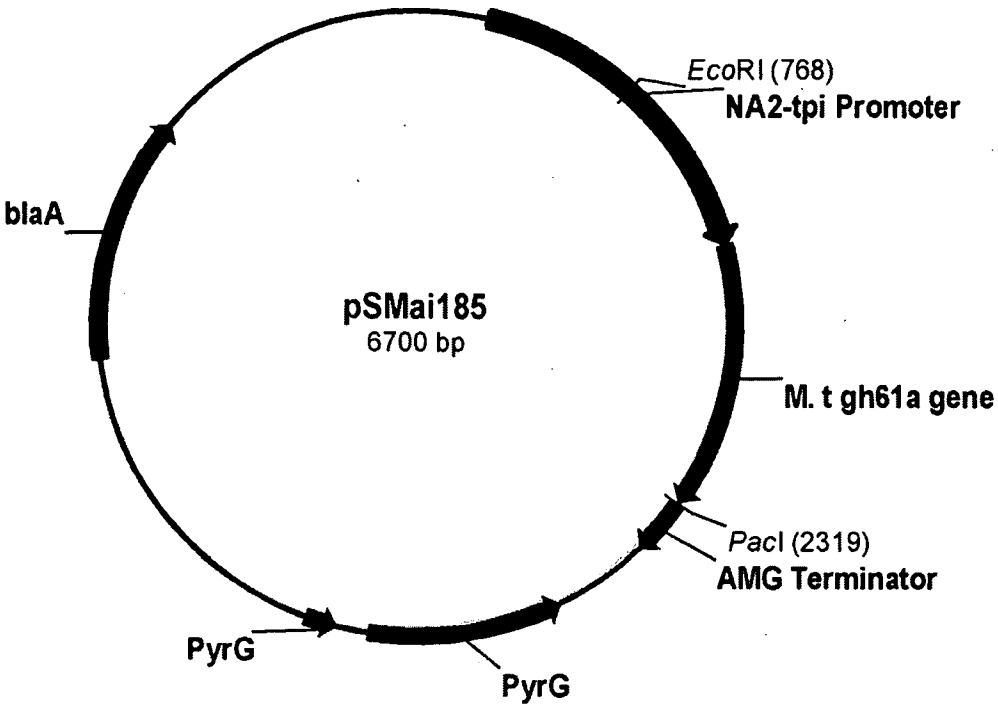
EcoRI (768)
NA2-tpi Promoter
blaA
pSMai185
6700 bp
M. t gh61a gene
PacI (2319)
AMG Terminator
PyrG
PyrG

Fig. 5

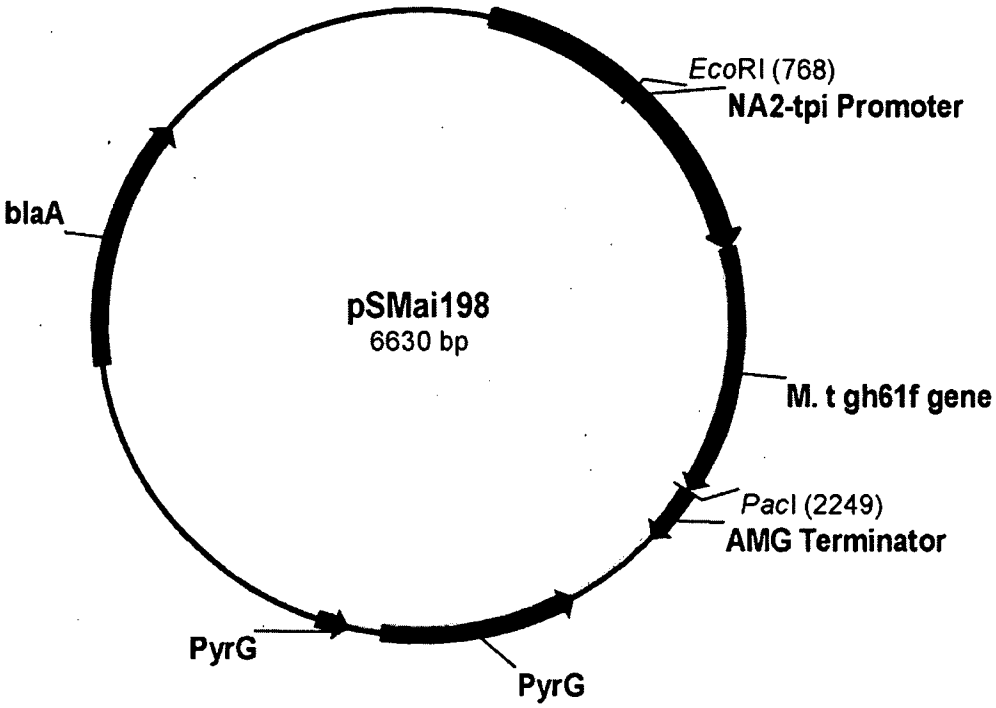
EcoRI (768)
NA2-tpi Promoter
blaA
pSMai198
6630 bp
M. t gh61f gene
PacI (2249)
AMG Terminator
PyrG
PyrG

Fig. 6

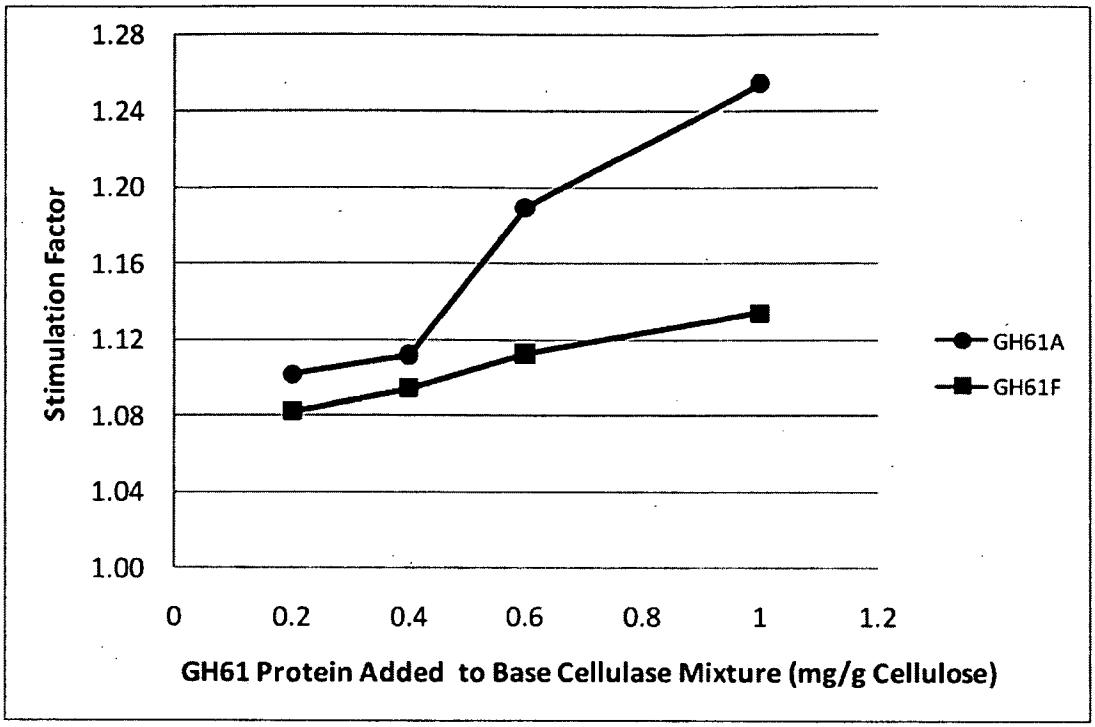
1.28
1.24
1.20
1.16
1.12
1.08
1.04
1.00
Stimulation Factor
0
0.2
0.4
0.6
0.8
1
1.2
GH61 Protein Added to Base Cellulase Mixture (mg/g Cellulose)
GH61A
GH61F

Fig. 7