A novel beta-glucosidase and nucleic acids encoding the beta-glucosidase. Also disclosed are cells, compositions, and methods relating to using the beta-glucosidase to convert ligocellullosic material to fermentable sugars.
**C. raphigera** D2 bgl genomic DNA region (2851 bps with 12 introns) for encoding BGL-Cr-D2 enzyme

**Fig. 1**
Fig. 5
Fig. 6
Fig. 7
Fig. 8
NOVEL BETA-GLUCOSIDASE AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the priority of U.S. Provisional Application No. 61/299,007, filed on Jan. 28, 2010. The contents of the application are hereby incorporated by reference in its entirety.

BACKGROUND

[0002] Developments of alternative energy resources (e.g., solar energy, biofuel, hydropower, wind power etc.) have been encouraged for years to cope with arising energy problems related to vast consumption of fossil fuels. Among many possible resources, plant biomass is of particular interest as it is renewable. Plant mass contains a high amount of cellulose, a starting material for making biofuel. To convert cellulose to biofuel, it is first degraded to fermentable sugars, such as cellobiose and glucose, by the cellulolytic system of microorganisms. This system includes three major types of hydrolases, i.e., endoglucanases (EC 3.2.1.4), exonuclease (EC 3.2.1.91), and β-glucosidases (EC 3.2.1.21). Although many glucosidases have been isolated from various microorganisms, their efficiencies are not satisfactory. Thus, there is a need for a high efficient glucosidase.

SUMMARY

[0003] This invention relates to a novel β-glucosidase, BGL-Cr-D2, which is isolated from a Taiwan-indigenous fungus Chaetomella raphigera. Shown in FIG. 2 are the sequences of the mature BGL-Cr-D2 protein (722 a.a. residues/SEQ ID NO: 1), the corresponding cggl Cr-D2 cDNA (2166 bps/SEQ ID NO: 2), the signal peptide sequence (SEQ ID NO: 7), the corresponding cDNA (SEQ ID NO: 8), the BGL-Cr-D2 protein with its signal peptide (SEQ ID NO: 4), and the corresponding cDNA (SEQ ID NO: 5). Shown in FIG. 1 is the genomic DNA region (2851 bps, SEQ ID NO: 3) in Chaetomella raphigera encoding the BGL-Cr-D2 enzyme, which contains 12 intron regions, and the genomic DNA region corresponding to SEQ ID NO: 4 (SEQ ID NO: 6).

[0004] Accordingly, one aspect of this invention features an isolated polypeptide containing a sequence exhibiting at least 70% (e.g., 80%, 90%, 95%, or 99%) amino acid identity to SEQ ID NO:1, as determined by the BLAST algorithm. In one example, the polypeptide has a β-glucosidase activity.

[0005] This invention also encompasses (i) an isolated nucleic acid including a nucleotide sequence that encodes the polypeptide described above, and (ii) a host cell containing such an isolated nucleic acid. In one example, the nucleotide sequence contains the sequence of SEQ ID NO: 2, 3, 5, or 6. The isolated nucleic acid of this invention can be an expression vector, in which the nucleotide sequence is operably linked to a suitable promoter sequence (i.e., a sequence capable of initiating transcription in a host cell).

[0006] The term “isolated polypeptide” or “isolated nucleic acid” used herein refers to a polypeptide or nucleic acid substantially free from naturally associated molecules, i.e., the naturally associated molecules constituting at most 20% by dry weight of a preparation containing the polypeptide or nucleic acid. Purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, and HPLC.

[0007] Also within the scope of this invention are (A) a composition containing the polypeptide described above, an exo-glucanase, an endo-glucanase, and/or a β-glucosidase, and (B) a method of producing a fermentable sugar from a lignocellulosic material. The method includes (i) providing the just-described composition, and (ii) contacting this composition with a lignocellulosic material to produce a fermentable sugar, e.g., glucose, xylose, arabinose, galactose, mannose, rhamnose, surose, fructose, lactose, maltose, trehalose, or cellobiose. The fermentable sugar can be converted to a fermentation product, such as alcohol, by microbial fermentation or enzyme treatment. Examples of the lignocellulosic material used in this method include, but are not limited to, orchard prunings, chaparral, mill waste, urban wood waste, municipal waste, logging waste, forest thinnings, short-rotation woody crops, industrial waste, wheat, wheat straw, oat straw, rice straw, barley straw, rye straw, flax straw, soy hulls, rice hulls, oat hulls, sugar cane, corn, corn stover, corn stalks, corn gluten feed, corn cobs, corn husks, corn kernel, fiber from kernels, prairie grass, garbaggrass, foxtail, sugar beet pulp, citrus fruit pulp, seed hulls, cellulosic animal wastes, lawn clippings, cotton, seaweed, trees, shrubs, grasses, sugar cane bagasse, products and by-products from wet or dry milling of grains, municipal solid waste, waste paper, yard waste, herbaceous material, agricultural residues, forestry residues, municipal solid wastes, waste paper, pulp, paper mill residues, branches, bushes, canes, corn, corn husks, energy crops, forests, fruits, flowers, grains, grasses, herbacius crops, leaves, bark, needles, logs, roots, saplings, shrubs, switch grasses, trees, vegetables, fruit peels, vines, sugar beet pulp, wheat middlings, oat hulls, hard and soft woods, organic waste materials generated from agricultural processes, forestry wood waste, or combinations thereof. If the fermentation product is combustible, the method further contains a step of combustizing the combustible fermentation product to produce energy.

[0008] This invention also features a bioreactor containing a lignocellulosic material and a composition, including the polypeptide described above, an exo-glucanase, an endo-glucanase, and/or a β-glucosidase.

[0009] The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

DESCRIPTION OF DRAWINGS

[0010] FIG. 1 is a diagram showing the genomic DNA region (2851 bps, SEQ ID NO: 3) in Chaetomella raphigera encoding the BGL-Cr-D2 enzyme, which contains 12 intron sections (as indicated with “White” sections in the diagram and “Small Caps/Gray/Strokesthrough” sections in the sequence). A region encoding signal peptide is indicated with “Black” in the diagram and “Underline” in the sequence.

[0011] FIG. 2 is a diagram showing the bgl-Cr-D2 cDNA (2166 bps/SEQ ID NO: 2) encoding signal sequence indicated with “Underline”/SEQ ID NO: 8) and its encoding amino acid sequences (722 a.a. residues/SEQ ID NO: 1, excluding signal peptide/SEQ ID NO: 7) for a Taiwan-indigenous fungus Chaetomella raphigera.

[0012] FIG. 3 is a diagram of the phylogenetic tree based on amino acid sequences of beta-glucosidase (BGL) from glycoside hydrolase (GH) family 3, showing that the BGL-Cr-D2-Pp secreted from reconstructed Pichia pastoris (strain
SM1168) was related to the beta-D-glucoside glucosylhydro-
dase (i.e., BGL) of T. stipitatus, A. terreus NIH2624, and A. 
flavipes A1163 (i.e., 64-65% similarity in AA sequences).

[0013] FIG. 4 is a diagram of a vector map of reconstructed 
pGAPZxC, showing that the bgl-Cr-D2 gene was cloned for 
further transformation to P. pastoris expression system. (In 
the simplified nucleotide sequence showed below the map, 
the lengths of the region, i.e., GMP promoter, alpha-factor 
signal sequence, and bgl-Cr-D2, do not correlate to the real 
sizes as scaled in the diagram map.)

[0014] FIGS. 5A and 5B are diagrams showing growth and 
PNPase activity versus time for the colonies of (A) Pichia 
pastoris strain SM1168 and (B) P. pastoris strain SM1168 
carrying bgl-Cr-D2; they show similar growth profile but 
not significant PNPase activity only for the colony cloned with 
bgl-Cr-D2.

[0015] FIGS. 6A-C are photographs showing enzymes sizes 
and activities of native PAGE (A), MUG zymogram (B) and 
SDS PAGE (C). (P: Pichia pastoris; Cr-D2: Chaetomella 
raphignosa fungus strain D2; BGL: beta-glucosidase.)

[0016] FIG. 7 is a diagram showing that the recombinant 
enzyme BGL-Cr-D2-Pp was stable under a wide pH range 
(4-9) based on 20-hour pH stability assessment (incubated 
the enzyme at 4°C under each of the particular pH levels for 20 
hours prior to performing standard PNPase activity 
measurement, i.e., pH 15 and 55°C for 10 mins). The optimal 
PPI level for PNPase activity of the BGL-Cr-D2-Pp was around 
5 (measured PNPase activity of the enzyme at each of the 
particular pH levels).

[0017] FIG. 8 is a diagram showing that the recombinant 
enzyme BGL-Cr-D2-Pp was stable under a temperature 
range of 4-55°C based on 4-hour thermostability assessment 
(incubated the enzyme under each of the particular tempera-
ture for 4 hours prior to performing standard PNPase activity 
measurement). Optimal temperature for PNPase activity 
of the BGL-Cr-D2-Pp was around 70°C. (measured PNPase 
activity of the enzyme at each of the particular temperature; 
but, at this high temperature, the enzyme would become 
inactive after longer incubation time (e.g., ≥2 hrs). FIG. 9.

[0018] FIG. 9 is a 3D diagram showing that thermostabil-
ity of the recombinant β-glucosidase BGL-Cr-D2-Pp under 
temperature ranged from 4°C to 75°C over 9 days period. 
Under a temperature ≤50°C, the BGL-Cr-D2-Pp remained 
stable (relative activity >80%) over 9 days.

DETAILED DESCRIPTION

[0019] This invention is based, at least in part, on the unex-
pected discoveries of a novel beta-glucosidase and its high 
cytoplastic activity or efficiency. Beta-glucosidases (or beta-
D-glucosidase glucosylhydrodase, BGL) are a group of the key 
enzymes for processing lignocellulosic saccharification. 
This type of enzyme can liberate D-glucose units from the cellu-
lose-hydrolyzing media such as cellulose, cello-oligosac-
charides, and other glucoses after the actions of endo-glu-
canases (EG) (mainly degrading the amorphous parts of 
cellulose fiber) and exo-glucanases (or cellulobiodyrosides, 
CBH) (mainly degrading the crystalline cellulose). For indus-
trial-scale production of biofuel from lignocelluloses, finding 
low-cost cellulases (i.e., CBHI, EG, and BGL) with high ac-

tivity and productivity is essential.

[0020] This invention discloses new discoveries in 
polypeptides with beta-glucosidase activity, polynucleotides 
encoding the polypeptides, and enhancement in beta-glucosi-
dase specific activity for recombinant beta-glucosidase 
expressed from yeast. Explicitly, a Taiwan-indigenous fun-
gus, i.e., Chaetomella raphignosa strain D2, was found 
capable of secreting an active beta-1,4-glucosidase (BGL-Cr-
D2). The gene (bgl-Cr-D2, size=2,166 bp) encoding this 
enzyme was cloned from reverse-transcribed cDNA and has 
been successfully expressed in a yeast expression system 
using Pichia pastoris strain SM1168 as the host. Recombi-
nant enzyme BGL-Cr-D2-Pp secreted from P. pastoris had a 
higher β-glucosidase activity (based on PNPase activity 
measurement) than the native enzyme (approx. 3x) and 
the benchmark Novozyme-188 (approx. 17x). This invention 
provides a new recombinant β-glucosidase enzyme that has 
potential to be utilized together with other cellulases for cel-
lulolic saccharification.

1. Biofuels from Lignocellulose

[0021] Developments of renewable power resources (e.g., 
solar energy, biofuel, hydropower, wind power etc.) have 
been encouraged for years to cope with arising energy 
problems related to vast consumptions of fossil fuels. Among 
the currently-exploited renewable energy forms/sources, bio-
fuel has a unique advantage in reducing agricultural/city 
and industrial organic wastes while generating cleanliness 
and energy-rich fuels (e.g., methane, ethanol, butanol, and 
hydrogen). Thus, it has been listed as one of important sus-
tainable energies (Antoni et al., 2007, Appl Microbiol Bio-
Not only in Taiwan, production of bioethanol from lignocell-
ulose is urgent worldwide as about 5×10^{12} tons of cellulose-
rich agricultural wastes are produced on the earth every 
year and supplement of ethanol in gasoline is ongoing or on-
its way in many countries. Prior to further fermentation for 
ethanol production, various enzyme systems (e.g., ligninase, 
cellulases, and hemicellulases) are required for complete 
degradation of lignocelluloses (containing about 5-30% of 
lignin, approximately 35-50% of cellulose, and approxi-
ately 20-35% of hemicellulose) to release fermentable glu-

2. Cellulases

[0022] For bioethanol production from cellulose, efficient 
saccharification (degradation of polymeric cellulose fibers to 
release smaller glucose molecules) by cellulases is crucial. 
Cellulases are a group of enzymes responsible for complex 
cellulose hydrolysis. Multiple cellulases have been classified 
into three groups, i.e., (i) exoglucanases (including 1,4-β-
d-glucan cellobiodyrosides, CBHI, EC 3.2.1.91 and 1,4-β-
d-glucan glucanohyrodases, EC 3.2.1.74), (ii) endogluca-

nases (EG) (EC 3.2.1.4), and (iii) beta-glucosidases (or β-D-glucosi-
dide glucohydrolase, BGL) (EC 3.2.1.21) based on structural 
and functional properties of the enzymes (Coughlan et al., 
1988, In Biochemistry and Genetics of Cellulose Degradation, 
The CBHI mainly act on the crystalline sections at reducing 
or nonreducing end of cellulose fiber while EG randomly attack 
the amorphous parts of cellulose. More specifically, these 
cellulases can cleave the β-1,4-glucan or β-D-glucosidic link-
ages between glucose residues within celluloses to form 
short-chain media including cellobiose (i.e., disaccharides), 
other cellooligomers (i.e., cello-oligosaccharides), and 
glucoses (e.g., alcohol, cyanogenic, or phenolic glycosides). 
These short-chain intermediates can be further hydrolyzed to 
glucose (i.e., monosacharide) by BGL that primarily cataly-
izes the transfer of glycosyl group between oxygen nucleo-
philes (Bhatia et al., 2002, Crit Rev Biotechnol 22(4):375-407). For cellulose saccharification, insufficient BGL activity can result in not only shortage of glucose but also accumulation of cellobiose that is a potent inhibitor for cellulose hydrolysis by CBH and EG (Harris et al., 2007, U.S. Pat. No. 7,244,605B2). Thus, BGL plays an important role in achieving high ethanol yield from cellulose (Lynd et al., 2002, Microbiol Mol Biol Rev 66:506-577; and Hong et al., 2007, Appl Microbiol Biotechnol 73:1331-1339).

3. Sources of β-glucosidases (BGLs)

[0023] The enzyme BGL is ubiquitous and has been discovered in all the living kingdoms from microbes, insects, and plants to highly-evolved mammals (Esen, 1993, in Biochemistry and Molecular Biology, American Chemical Society, 1-14; Bhatia et al., 2002, Crit Rev Biotechnol 22(4):375-407). One of major BGL-producing sources is from bacteria including some strains of genus Agrobacterium, Bacillus, Butyrivibrio, Cellovibrio, Clostridium, Erwinia, Pseudomonas, Pyrococcus, Ruminococcus, Streptomyces, Thermotoga, Thermoflava etc. (Hashimoto et al., 1998, Arch Biochem Biophys 360:1-9; Srivastava et al., 1999, Biotechnol Lett 21:293-297; and Yam et al., 2001, Biosci Biotechnol Biochem 65(9):2028-2032). Fusarial strains (molds or yeasts) with BGL-producing ability include those from genus Aspergillus, Candida, Humicola, Penicillium, Pichia, Saccharomyces, Talaromyces, Trichoderma etc. (Dan et al., 2000, J Biol Chem 275(7):4973-4980; and Dunn-Coleman et al., 2008, U.S. Patent Publication 2008/0095893A1). Among these microbes, fungal species Aspergillus niger, Aspergillus fumigates, and Trichoderma reesei are the most well-known and efficient BGL producers making them the major sources for commercial BGLs in the world (Dan et al., 2000, J Biol Chem 275(7):4973-4980; Harris et al., 2007, U.S. Pat. No. 7,244,605B2; and Dunn-Coleman et al., 2008, U.S. Patent Publication 2008/0095893A1). In addition, few plants like barley, Costus, and Maize have been found containing BGL as well.

[0024] Beside above-mentioned native sources, some recombinant BGLs have been expressed in different host (e.g., E. coli, B. subtilis, P. pastoris, S. cerevisiae, A. awamochii, and T. reesei) (Pundey et al., 1995, J Ferment Bioeng 80(5):440-453; Munney et al., 2004, Protein Expr Purif 38:248-257; and Roy et al., 2005, Biochem Biophys Res Co 336:299-308). For example, bgl genes from Bacillus sp. had been successfully cloned and expressed in E. coli system (Hashimoto et al., 1998, Arch Biochem Biophys 360:1-9; and Srivastava et al., 1999, Biotechnol Lett 21:293-297). In addition, an Aspergillus niger BGL coding gene, bgl1, had been expressed in P. pastoris and S. cerevisiae (Dan et al., 2000, J Biol Chem 275(7):4973-4980). Clone BGL genes from diverse sources and then express in different hosts has been a practical approach for gaining higher BGL productivity (over-expression) and/or activity (advantageous post-translational modification improve kinetic features of the enzymes).

4. Properties of BGLs

[0025] The properties of BGLs are rather diverse by far. In general, BGLs from different orders and kingdoms appear to differ in many characteristics for BGLs such as (aromatic, aliphatic, or amino-linked) to the glycosyl group (Bhatia et al., 2002, Crit Rev Biotechnol 22(4):375-407). They have been found to be intracellular, extracellular, cytosol-associated, membrane-bound, or periplasmically-localized. Their catalytic functionalities on hydrolyzing short-chain cellobiose and oligosaccharides are quite similar in cellulase microorganisms although some other various functions have been noticed in different organisms (e.g., synthesis of glycosyl-bond between different molecules in some microbes, releasing cyanides from cyanoglucoside precursors in insects and plants, and hydrolysis of glycosylglyceramides for treating Gaucher’s disease in humans) (Bhatia et al., 2002, Crit Rev Biotechnol 22(4):375-407). For cellulose hydrolysis, BGLs are able to hydrolyze their substrates with net retention of anomer configuration that occurs presumably via a two-step, double-displacement mechanism involving two carboxylate acid residues at key active site (Davies et al., 1998, Comprehensive Biological Catalysis Vol. 1, pp. 119-208; and Withers, 2001, Carbohydr Polym 44:325-357). It has been noted that most of BGLs have different substrate specificity and can act on a wide spectrum of substrates (e.g., pNPβG, cellobiose, salicin, MUG, arbutin, aryl-, alkyl-, and methyl-glucosides). In addition, most of BGLs have optimum pH within weak-acid levels and optimum temperature in mesophilic scopes despite some strong-acid and/or thermophilic properties have been reported. The BGL monomers from diverse organism sources have a very broad molecular size range (i.e., about 13-137 KD) (Gibbsberger et al., 1993, Appl Microbiol Biotechnol 40:44-52; Iwahata et al., 1999, Appl Environ Microbiol 65:5546-5553; Srivastava et al., 1999, Biotechnol Lett 21:293-297; and Dan et al., 2000, J Biol Chem 275(7):4973-4980). Amino acid sequence similarities between BGLs of different organism species also vary widely (10s-90s%). Nevertheless, amino acid bases of aspartic acid (Asp, D) and glutamine acid (Glu, E) have been recognized as catalytic nucleophile and catalytic proton donor of BGLs.

5. Classification of BGLs

[0026] Based on chemical reactions they catalyze, all BGLs have been assigned to EC 3.2.1.21 category (i.e., Enzyme Commission number of 3-hydrolyses, 2-glycosylases, 1-glycosidases, and 21-glucosidase) according to numerical classification scheme for more than 3,000 enzymes published by the IUBMB (International Union of Biochemistry and Molecular Biology) (Webb, 1992, Enzyme nomenclature: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes, ISBN 0-12-277164-5; and ExPASy, 2009, www.expasy.org/enzyme/). Additionally, BGLs are versatile hydrolytic enzymes that have been further classified according to various criteria (e.g., substrate specificity and sequence identity). At first, BGLs had been grouped into Type I and II (Beguin et al., 1990, Ann Rev Microbiol 44:219-248) or subfamily A and B (Rojas et al., 1995, Biochem Mol Biol Int 35:1223-1231); but lately, these earlier classifications have been mostly replaced by the currently-acceptable scheme based on amino acid and/or nucleotide sequence identity (Henrisaat et al., 1996, Biochim J 316:695-696).

[0027] Based on amino acid sequence and folding similarities, the BGLs have been placed in either family 1 or 3 among 115 up-to-date (November 2009) defined GH (glycoside-hydrolases) families with an exception of the glycosyltransferases (i.e., acid β-glucosidase placed in family 30) (Henrisaat, 1991, Biochem J 280:309-316; Henrisaat et al., 1993, Biochem J 293:781-788; Henrisaat et al., 1996, Biochem J 316:695-696; Hong et al., 2007, Appl Microbiol Biotechnol 73:1331-1339; and Cantarel et al., 2009, Nucleic Acids Res. 37:D233-238). Family 1 (GH1) contains β-glucosidases from some archaea, bacteria, fungi, plants, and mammals whereas
family 3 (GH3) comprises β-glucosidases of some bacterial, fungal, and plant origins (Henrissat et al., 1996, Biochem J 316:695-696; and Cantarel et al., 2009, Nucleic Acids Res. 37:D233-238).

6. Applications

[0028] Beside cellbiose/cellomedextrins hydrolysis function used for cellulose ethanol production, BGLs also can be broadly used in the fields of medicine, agriculture, and food industry wherein many reactions engaging cleavage or synthesis of glycosidic bonds are required. Applications based on hydrolytic activities of BGLs include (1) removal of bitterness from citrus fruit juices, (2) manufacture of low-viscosity food gels, (3) detoxification of cassava (the 3′-4′ largest source of calories in the tropics), (4) enhancement of aroma release for benefitting winemaking process, (5) feed additive for enhancing nutrient utilization in single-stomached animals, (6) hydrolysis of genistin to genistein as an anti-tumor agent, (7) production of melanin from phloridzin for reducing the risk of skin cancer and promoting dark color of hair, (8) production of hydroxytyrosol from oleuropein for preventing coronary heart disease and cancer, (9) hydrolysis of laminarin for the production of yeast extract and the conversion of algal biomass to fermentable sugars, (10) making pigments as natural food dyes in confectionary products, etc. (Bhatia et al., 2002, Crit Rev Biotechnol 22(4):375-407).

[0029] In addition, synthetic activities of BGLs have been widely used in the manufacture of pharmaceuticals, fine chemicals and food ingredients (as summarized by Bhatia et al., 2002, Crit Rev Biotechnol 22(4):375-407). For instance, synthesis of an aromatic n-alkyl glucoside ester that was effective in the treatment of fever, rheumatism, headache, and other ailments (Otto et al., 1998, Biotechnol Lett 20:437-440). Other applications of BGLs also have been reported such as zero-diagnosing histoplasmosis or post-diagnosing hepatic ischemia-reperfusion injury and recovery. Therefore, productions of high-quality BGLs from diverse sources have been fundamental tasks for these multipurpose applications.

[0030] Described herein is an isolated polypeptide including an amino acid sequence at least 70% identical to that of BGL-Cr-D2 (SEQ ID NO:1) or a part thereof having at least 20 (e.g., 30, 50, 80, 100, 150, 200, 250, 300, and 350) contiguous amino acids. This polypeptide has a high beta-glucosidase activity.

[0031] The "percent identity" of two amino acid sequences is determined using the algorithm of Karlin and Altschul Proc Natl Acad Sci USA 87:2264-68, 1990, modified as in Karlin and Altschul Proc Natl Acad Sci USA 90:5873-77, 1993. Such an algorithm is incorporated into the BLAST programs (version 2.0) of Altschul, et al. J. Mol. Biol. 215:403-10, 1990. BLAST protein searches can be performed with the XBLAST program, score=50, wordlength=3 to obtain amino acid sequences homologous to the protein molecules of the invention. Where gaps exist between two sequences, Gapped BLAST can be utilized as described in Altschul et al., Nucleic Acids Res. 25(17):3389-3402, 1997. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

[0032] The isolated polypeptide can be prepared by purification from a suitable microorganism, e.g., Chaetomella raphigera. It also can be prepared via conventional recombinant technology. An example follows. A DNA fragment encoding BGL-Cr-D2 can be prepared by polymerase chain reaction from Chaetomella raphigera cells and cloned into an expression vector. Upon insertion, the BGL-Cr-D2-encoding fragment is operably linked to a suitable promoter contained in the expression vector. The resultant DNA construct is then introduced into suitable host cells (e.g., E. coli, yeast cells, insect cells, and mammalian cells) for expression of BGL-Cr-D2, which can be purified from the cells by conventional methods. One example of yeast cells is Pichia pastoris, e.g., Pichia pastoris strain SMD1168.

[0033] To make a functional equivalent of BGL-Cr-D2, which is also within the scope of this invention, one or more conservative amino acid substitutions can be introduced into SEQ ID NO:1 without disrupting its β-glucosidase activity. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in SEQ ID NO: 1 is preferably replaced with another amino acid residue from the same side chain family. Alternatively, mutations can be introduced randomly along all or part of SEQ ID NO: 1, such as by saturation mutagenesis, and the resultant mutants can be screened for the β-glucosidase activity to identify mutants that retain the activity as described below in the Example section below.

[0034] Fusion protein technology can be applied to improve expression efficiency and facilitate purification of the polypeptide of this invention. To prepare a fusion protein containing BGL-Cr-D2, a DNA fragment encoding this β-glucosidase can be linked to another DNA fragment encoding a fusion partner, e.g., glutathione-s-transferase (GST), 6x-His epitope tag, or M13 Gene 3 protein. The resultant fusion nucleic acid expresses in suitable host cells a fusion protein that can be isolated by methods known in the art. The isolated fusion protein can be further treated, e.g., by enzymatic digestion, to remove the fusion partner and obtain the recombinant polypeptide of this invention.

[0035] Also described herein is an isolated nucleic acid encoding the polypeptide of this invention. A nucleic acid refers to a DNA molecule (e.g., a cDNA or genomic DNA), an RNA molecule, or a DNA/RNA analog which can be synthesized from nucleotide analogs. In one example, the nucleic acid of this invention is an expression vector in which a DNA fragment encoding the polypeptide is operably linked to a suitable promoter.

[0036] As used herein, the term "promoter" refers to a nucleotide sequence containing elements that initiate the transcription of an operably linked nucleic acid sequence in a desired host microorganism. In addition, a promoter contains an RNA polymerase binding site. It can further contain one or more enhancer elements which, by definition, enhance transcription, or one or more regulatory elements that control the on/off status of the promoter. When E. coli is used as the host microorganism, representative E. coli promoters include, but are not limited to, the β-lactamase and lactose
promoter systems (see Chung et al., *Nature* 275:615-624, 1978), the Sp6, T3, T5, and T7 RNA polymerase promoters (Studier et al., *Meth. Enzymol.* 185:60-89, 1990), the lambda promoter (Elvin et al., *Gene* 87:123-126, 1990), the tr promoter (Nichols and Yanofsky, *Meth. in Enzymology* 101:155-164, 1983), and the Tac and Trc promoters (Russell et al., *Gene* 20:231-245, 1982). When yeast is used as the host microorganism, exemplary yeast promoters include 3-phosphoglycerate kinase promoter, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) promoter, galactokinase (GAL1) promoter, galactose-1-phosphate uridyl transferase promoter, and alcohol dehydrogenase (ADH1) promoter. Promoters suitable for driving gene expression in other types of cells are also well known in the art.

[0037] A vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. The vector can be capable of autonomous replication or integrate into a host DNA. Examples of the vector include a plasmid, cosmid, or viral vector. The vector of this invention includes a nucleotide sequence encoding BGL-CrD2 in a form suitable for expression of the nucleic acid in a host cell. Preferably the vector includes one or more regulatory sequences operatively linked to the encoding sequence. A “regulatory sequence” includes promoters, enhancers, and other expression control elements (e.g., polyadenylation signals). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence, as well as tissue-specific regulatory and/or inducible sequences. The design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vector can be introduced into host cells to produce the polypeptide of this invention.

[0038] Also within the scope of this invention is a host cell that contains the above-described nucleic acid. Examples include *E. coli* cells, insect cells (e.g., using baculoviruses expression vectors), yeast cells, plant cells, or mammalian cells. See e.g., Goeddel, (1990) *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, Calif. To produce a polypeptide of this invention, one can culture a host cell in a medium under conditions permitting expression of the polypeptide encoded by a nucleic acid of this invention, and purify the polypeptide from the cultured cell or the medium of the cell. Alternatively, the nucleic acid of this invention can be transcribed and translated in vitro, for example, using T7 promoter regulatory sequences and T7 polymerase.

[0039] Further described herein is a method of converting lignocellulosic material to fermentable products (e.g., fermentable sugars) using a multi-enzyme composition containing the B-glucosidase described herein and other cellulosytic enzymes, such as exo-glucanase and endo-glucosidase. See, e.g., US Application Nos. 20070238155 and 20070259061. The term “cellulolytic enzyme” refers to an enzyme that hydrolyzes cellulose (a polysaccharide consisting of glucose units) into smaller sugar units. See Gilbert H J, Hazelwood G P, 1993 *J Gen Microbiol* 139:187-194; Oulfina K et al. 1995 *Biotechnol Genet Eng Rev.* 14:365-414. See also, e.g., US Application 2007018085. This multi-enzyme composition can be obtained from, e.g., a microalgal, a plant, or a combination thereof, and will contain enzymes capable of degrading lignocellulosic material. In addition to the cellulolytic enzymes mentioned above, it can further include cellubiohydrolases, endoglucanase, beta-glucosidases), hemicellulases (such as xylanases, including endoxylanases, oxoxygenase, and beta-xylidase), ligninases, amylases, alpha-arabinofuranosidases, alpha-glucuronidases, alpha-glucuronidases, arabinases, glucuronidases, proteases, esterases (including ferulic acid esterase and acetylxylan esterase), lipases, glucoammanases, or xyloligase.

[0040] As used herein the term “lignocellulosic material” refers to materials containing cellulose and/or hemicellulose. Generally, these materials also contain xylan, lignin, protein, and carbohydrates, such as starch and sugar. Lignocellulose is found, for example, in the stems, leaves, husks, and cobs of plants or leaves, branches, and wood of trees. The process of converting a complex carbohydrate (such as starch, cellulose, or hemicellulose) into fermentable sugars is also referred to herein as “saccharification.” Fermentable sugars, as used herein, refer to simple sugars, such as glucose, xylose, arabinose, galactose, mannose, rhamnose, surose, fructose, lactose, maltose, trehalose, or celllobiose. Lignocellulosic material can include virgin plant biomass and/or non-virgin plant biomass such as agricultural biomass, commercial organics, construction and demolition debris, municipal solid waste, waste paper, and yard waste. Common forms of lignocellulosic material include trees, shrubs and grasses, wheat, wheat straw, sugar cane bagasse, corn, corn husks, corn kernel including fiber from kernels, products and by-products from milling of grains such as corn, rice, wheat, and barley (including wet milling and dry milling), as well as municipal solid waste, waste paper, and yard waste. The lignocellulosic material can also be, but is not limited to, herbaceous material, agricultural residues, forestry residues, and paper mill residues. “Agricultural biomass” includes branches, bushes, canes, corn and corn husks, energy crops, forests, fruits, flowers, grains, grasses, herbaceous crops, leaves, bark, needles, logs, roots, saplings, short rotation woody crops, shrubs, switch grasses, trees, vegetables, fruit peels, vines, sugar beet pulp, wheat middlings, oat hulls, hard and soft woods (not including woods with deleterious materials), organic waste materials generated from agricultural processes including farming and forestry activities, specifically including forestry wood waste, or a mixture thereof.

[0041] The fermentable sugar produced in the method described above can be converted to useful value-added fermentation products via the fermentation treatment or biorefinery operation. Examples of the fermentation product include, but are not limited to amino acids, vitamins, pharmaceuticals, animal feed supplements, specialty chemicals, chemical feedstocks, plastics, solvents, fuels, or other organic polymers, lactic acid, and ethanol, including fuel ethanol. Specific value-added fermentation products that may be produced by the methods of the invention include, but not limited to, biofuels (including ethanol and butanol); lactic acid; plastics; specialty chemicals; organic acids, including citric acid, succinic acid and maleic acid; solvents; animal feed supplements; pharmaceuticals; vitamins; amino acids, such as lysine, methionine, tryptophan, threonine, and aspartic acid; chemical feedstocks. The fermentable sugar can also be used for culturing microbes that produce fermentation products, e.g., industrial enzymes, such as proteases, cellulases, amylases, glucanases, lactases, lipases, lyses, oxidoreductases, transferases and xylanases.

[0042] The invention also provides a method of producing energy from lignocellulosic material. This method includes providing the multi-enzyme composition described above; contacting the composition with the lignocellulosic material
to produce a fermentable product; fermenting the fermentable product to produce a combustible fermentation product, and combusting the combustible fermentation product to produce energy. This method can be performed in a bioreactor that contains all necessary components and may preferably be configured for anaerobic growth of microorganisms. Methods for making and using bioreactors are known in the art. See, e.g., US Application 2009/0131938.

[0043] The polypeptide and composition described above can also be used in the paper and pulp industry. For example, the polypeptide, which has a high β-glucosidase activity can be used in the deinking and refining of recycled paper. In this application, utilizing of them could reduce the amount of paper scrap, partially, and reduce the time of exposure to the enzyme needed to increase the brightness of the paper. Reducing the concentration of enzyme and the time of exposure to the enzyme in the refining process, correspondingly and desirably reduces the reaction of the cellulose on the fibrils that drag costs.

[0044] The polypeptide of this invention has additional industrial applications. Given its high β-glucosidase activity, this polypeptide can function at a lower amount. The polypeptide, in combination with other enzymes, can be used, with enhanced yields, in extracting juice from fruits, or extracting juice or soup flavorings from vegetables. In combination with protease, it can be used to dissociate dried seaweed, which is then fermented with alcohol to produce vinegar. The polypeptide, mixed with other enzymes, can also serve as a dough conditioner in the baking industry. See, e.g., U.S. Pat. No. 6,602,700.

[0045] Moreover, the polypeptide of this invention can also be used in the textile industry. It can be used to brighten and soften cotton fabrics by removing microfibers on the surface, which causes a dull look of clothes. More specifically, it can be included as an additive in formulating enzyme-containing detergents for soil removal, fabric softening, and color brightening. For example, it can be used as a replacement to pumice in producing blue jeans having a “stone-washed” effect. Enzyme treatment causes less damage to the jean fabric than lengthy exposure to pumice. See U.S. Pat. Nos. 5,232,851, 5,677,151, 6,451,060, and 7,226,773.

[0046] In another aspect, the present invention provides a transgenic plant, the genome of which is augmented with a recombinant polynucleotide encoding a polypeptide of this invention operably linked to a promoter sequence. The polynucleotide is optimized for expression in the plant and the polypeptide is produced at a level greater than 5% total soluble protein, greater than 10% total soluble protein or greater than 20% total soluble protein. The polypeptide may be expressed constitutively or tissue-specifically. For example, it may be expressed in a plant tissue selected from the roots, stems, leaves, flowers, or seeds. The plant may be a monocotyledonous plant or a dicotyledonous plant. In certain embodiments, the plant is a crop plant. The plant may be selected from the group consisting of corn, switchgrass, sorghum, miscanthus, sugarcane, poplar, pine, wheat, rice, soy, cotton, barley, turf grass, tobacco, bamboo, rape, sugar beet, sunflower, willow, and eucalyptus.

Methods for making transgenic plants are well known in the art.

[0047] The specific examples below are to be construed as merely illustrative, and not limitations of the remainder of the disclosure in any way whatsoever. Without further elaboration, it is believed that one skilled in the art can, based on the description herein, utilize the present invention to its fullest extent. All publications cited herein are hereby incorporated by reference in their entirety. Further, any mechanism proposed below does not in any way restrict the scope of the claimed invention.

I. Materials and methods

1. Cultivation of Native D2 Fungal Strain

[0048] The native fungal strain (i.e., Chaetomella rigida D2) was cultured with MR (Mendels-Reese) medium (pH 5.0) to harvest fungal cells for DNA/RNA extractions and their crude enzymes for property analyses (e.g., amino acid sequences, cellulolytic activities, and enzyme concentrations). Each one-liter MR medium contains 1 g soy peptone, 1.4 g (NH4)2SO4, 0.3 g urea, 2.0 g K2HPO4, 0.34 g CaCl2, 0.3 g MgSO4.7H2O, 5.0 mg FeSO4.7H2O, 1.6 mg MnSO4.7H2O, 1.4 mg ZnSO4.7H2O, 2 mg CoCl2.6H2O, and 0.72 g cellulose. Fungal culture was grown in a temperature of 30°C and a mixing rate of 125 rpm for 4 days prior to gather fungal cells and their crude enzymes.

2. Nucleic Acid Extraction and Crude Enzyme Collection

[0049] Genomic DNA of the C. rigida D2 fungus was extracted using Wizard® Genomic DNA Purification Kit (Promega, USA) and fungal RNA was extracted using Plant Total RNA Miniprep Purification Kit (GeneMark, Taiwan). Concentrations of extracted nucleic acid were measured by a spectrophotometer (NanoDrop ND-1000, Thermo Scientific, USA). The DNA and RNA extracts were stored at −20°C prior to further treatments (e.g., PCR and RT-PCR) were conducted.

[0050] Crude enzymes were collected by filtering supernatant of 4-day fungal culture through Whatman No.1 filter paper (Whatman/GE Healthcare, USA), then 0.45 μm Supor® Membrane Disc Filter (PALL, USA), and followed by 30K NMWL (Nominal Molecular Weight Limit) Amicon® Ultra-15 Centrifugal Filter (Millipore, USA). Enzyme concentrations were determined based on Bio-Rad Protein Assay using BSA (bovine serum albumin) as calibration standard (Bio-Rad, USA). The filtered enzymes were stored at 4°C until further analyses were conducted.

3. Enzyme Electrophoresis and N-Terminal Sequencing

[0051] Enzyme electrophoreses including native PAGE (polyacrylamide gel electrophoresis) and SDS (sodium dodecylsulfate) PAGE were performed based on the protocols suggested by Hoefer’s Protein Electrophoresis Applications Guide (Hoefer Scientific Instruments, 1994). Before Coomassie blue staining for native PAGE, zymograms were carried out by immersing gel with 0.5 mM MUG (4-Methylumbelliferyl β-D-glucopyranoside) solution and incubating at 50°C for 15 minutes and then followed by observation under UV light (Benoit et al., 1995, Curr Microbiol 30: 305-312). Additionally, instead of coomassie blue staining for SDS-PAGE, silver staining using PlusOne Silver Staining Kit (GE Healthcare, USA) was also performed for observing some enzymes with low concentration.

[0052] For N-terminal sequence analysis, the target enzyme (i.e., BGL with MUG activity) was transferred from SDS-PAGE gel to PVDF membrane (Blot Gel Transfer Stcks, Invitrogen, USA) and analyzed by Applied Biosystems Procise Protein Sequencer model 494 (Applied Biosystems, USA) based on Edman degradation chemistry. Part of the revealed N-terminal sequence, PGDGGWA (SEQID NO: 9), was used to design a degenerate primer D2-bgl-NT: CCN
GGN GAY GGN GAY TGG GC (SEQ ID NO: 10) for further amplifying target bgl gene from genomic DNA and/or reverse transcribed cDNA.

4. Cloning and Sequencing for D2-bgl Gene

[0053] The bgl-Cr-D2 cDNA was first reverse transcribed (RT) from RNA sample using a poly-T primer: GTG TCT TGC CAC AGT CAC TTT TTT TTT TTT TTT TTT (SEQ ID NO: 11) and SuperScript® III Reverse Transcriptase (Invitrogen, USA); and, subsequently amplified using a primer set of D2-bgl-N1 and poly-T anchor primer: GTG TCT TGC CAC AGT CAC TAC (SEQ ID NO: 12) and a DNA polymerase (TaKaRa Ex Taq™, TaKaRa Bio Inc, Japan). PCR thermo-cycle conditions were 94°C for 4 minutes, followed by 30 cycles at 94°C for 1 minute, 58°C for 30 seconds, and 72°C for 3 minutes, and then a final elongation at 72°C for 5 minutes. The RT-PCR product (i.e., amplicon of bgl-Cr-D2 cDNA) was cloned in pGEM®-T Easy vector (Promega, USA) and transformed into E. coli cells (strain DH5α) for preservation and further sequencing. Besides, the bgl-Cr-D2 gene (containing introns) was amplified from genomic DNA sample using a primer set of D2-bgl-e: CCT GCT GAT GAT TGG OG3 GC (SEQ ID NO: 13) and D2-bgl-r: ATG GCC ACC TTC CGG AT ACC TGG GC (SEQ ID NO: 14) and TaKaRa Ex Taq™. The PCR product was also cloned in pGEM®-T Easy vector for sequencing. Intron sections were determined by comparing bgl-Cr-D2 cDNA (without introns) and bgl-Cr-D2 genomic DNA (with introns) sequences.

[0054] Additionally, signal sequence upstream of the bgl-

Cr-D2 gene was revealed by randomly digesting genomic DNA with SacI restriction enzyme (NEB, New England Biolabs Inc., USA) and self-ligation with T4 DNA ligase (Promega, USA); and, then, followed by inverse PCR using a primer set of D2-bgl233r: CGT TTC GTC CAA AAT GTA ACA GCA T (SEQ ID NO: 15) and D2-bgl232r: GAT GCT TTC ACC GTC AGT TCA (SEQ ID NO: 16). The PCR program was as follow: 5 minutes at 95°C, followed by 25 cycles consisting of 1 minute at 95°C, 1 minute at 55°C, and 1 minute at 72°C, and a final elongation consisting of 10 minutes at 72°C. The inverse PCR product (should contain a sequence in the order of D2-bgl233f+partial D2-bgl+SacI+partial D2 genome+signal sequence of D2-bgl+partial D2-bgl+232r) was cloned in pGEM®-T Easy vector for further sequencing.

[0055] The putative amino acid sequence of BGL-Cr-D2 from revealed bgl-Cr-D2 cDNA was aligned with other GH3 family BGLs using Clustal X software (Thompson et al., 1997, Nucleic Acids Research 25(24): 4876-4882) to analyze phylogenetic relationships. The alignment result was used to build a phylogenetic tree that was visualized in TreeView (Page, 1996, Computer Applications in the Biosciences 12:357-358).

5. Expression in Pichia pastoris

[0056] The bgl-Cr-D2 cDNA (cloned in the pGEM®-T Easy Vector) was further constructed in pGAPZc A vector for constitutive expression of recombinant BGL-Cr-D2-Pp in Pichia yeast (Invitrogen, USA). Briefly, the bgl-Cr-D2 cDNA in the Easy vector was PCR amplified with a primer set of D2-bgl-f:GCG CTG ATG TAG CCT GGT GGT GGT TGG (SEQ ID NO: 17) and D2-bgl-r:NotI: TCG AAGCGGGCGCAT GTC CAC CTT TCC GAA TAC C (SEQ ID NO: 18). This PCR product was joined into the pGAPZc A vector by double digesting ampiclon and vector with EcoRI and NotI (NEB, USA) and then ligation with T4 DNA ligase. The constructed pGAPZc A vector carrying bgl-Cr-D2 cDNA was transformed into E. coli (strain DH5α) for preservation. Prior to subsequently transforming into Pichia host, the pGAPZc A vector with D2-bgl cDNA was purified with Plasmid Miniprep Purification Kit (GeneMark, Taiwan) and linearized with B sphi (NEB, USA). The linear plasmid DNA was transformed into Pichia pastoris (strains GS115 and SMD1168) via homologous recombination based on protocol for Pichia Expression Kit (Invitrogen, USA). The reconstructed Pichia pastoris yeast cells were grown in YPD medium, medium and the conditions (e.g., 30°C and 200 rpm) suggested in the Invitrogen protocol. Pichia cells and their enzymes were collected at different time points (0-14 days) to assess their growth and enzyme activity/quantity.

6. Activity Tests

[0057] Cellulolytic activities of crude enzymes from native D2 strain and recombinant enzymes from Pichia pastoris were tested using different substrates (p-nitrophenyl β-D-glucopyranoside, pNP; cellobiose; carboxymethylcellulose, CM; xylan; avicel; and, filter paper). The pNPase activities of the enzymes were measured based on pNP releasing rate from 2 mM pNP at 55°C in 5 minutes. The pNP concentrations were calibrated based on spectrophotometric absorbance at OD405. The cellobiase activities were assessed by cellobiose reducing rate and/or glucose producing rate where cellobiose and glucose concentrations were determined by HPLC (high performance liquid chromatography) analyses. Additionally, CMCase (for endo-glucanase activity), xylanase (for xylan hydrolyzing activity), avicelase (for exo-glucanase activity), and FPase (representing total cellulase activity) were measured based on production rate of reducing sugars that were detected by DNS (dimethylglyoxal acid) method.

II. Results

[0058] In this invention, a Taiwan-indigenous fungus Chaetomorpha raphigena strain D2 was first found capable of secreting crude enzyme with significant β-glucosidase activity based on assessment using pNP (p-nitrophenyl-β-D-glucopyranoside) and cellobiose as substrates. The gene (bgl-

Cr-D2) encoding the target β-glucosidase (BGL-Cr-D2) was derived from genomic DNA and reverse-transcribed cDNA. It showed that the bgl-Cr-D2 cDNA had a size of 2166 bps (not including signal sequence and stop codon) and the bgl-

Cr-D2 genomic DNA has a size of 2851 bps (FIGS. 1 & 2). Twelve introns were identified after comparing the bgl-Cr-D2 genomic DNA with cDNA. The bgl-Cr-D2 gene would encode 722 bases of amino acid to form the BGL-Cr-D2 (FIG. 2). Based on the comparison of the amino acid sequences for the β-glucosidases derived from glycoside hydrolase (GH) family 3, the newly-found BGL-Cr-D2 was homologous to the β-D-glucoside glucose hydrolase of Talaromyces stipitatus (65% similarity) and Aspergillus terreus (64% similarity) (FIG. 3). Based on the phylogenetic analyses, C. raphigena D2, T. stipitatus, and A. terreus formed a sub-group that could distinguish from other fungal and bacterial groups (e.g., Penicillium, Saccharomyces, Clostridium, and Escherichia coli).

[0059] The bgl-Cr-D2 gene was further cloned into pGAPZa C vector (FIG. 4) and transformed to a yeast expression system using Pichia pastoris strain SMD1168 as host cells. When compared the growth and pNPase activity for the Pichia pastoris strain SMD1168 and that carrying bgl-Cr-

D2, significant pNPase activity was observed only for the colony cloned with bgl-Cr-D2 while the growth profiles were similar for the both colonies (FIG. 5). The growth and activity reached plateau levels after incubating at 30°C, for around 4-6 days. Furthermore, the recombinant β-glucosidase (BGL-
Cr-D2-Pp) had a molecular size of around 95 kDa (FIG. 6C) (larger than theoretical size of 79 kDa due to post-translational glycosylation); and, showed high specific activity based on assessments of sucharifying rate (Table 1) and observations from native PAGE (polyacrylamide gel electrophoresis) (FIG. 6A). MUG (4-methylumbelliferyl-1-D-glucopyranoside) zymogram (FIG. 6B).

[0060] The recombinant BGL-Cr-D2-Pp remained active after 20-hour incubation under a wide pH range (i.e., pH4-9) (FIG. 7). However, the pNPase activity of this enzyme significantly decreased at pH levels higher than 5.5. Besides, the BGL-Cr-D2-Pp had highest pNPase activity at a temperature of 75°C; but, at this high temperature, the activity dropped tremendously after short-term incubation (i.e., 2 hours) (FIGS. 8 & 9). Thus, the pNPase activities of the recombinant BGL-Cr-D2-Pp were performed at pH 5 and a temperature of 55°C for 10 minutes.

**TABLE 1**

Comparison of enzyme activities for native BGL-Cr-D2, Pichia pastoris strain SMD1168 secretions, recombinant BGL-Cr-D2-Pp secreted from *P. pastoris*, and Novozyme-188 for different substrates (all reactions were performed at pH 5 and 55°C).

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Reaction time (minutes)</th>
<th>BGL-Cr-D2</th>
<th>Pp</th>
<th>BGL-Cr-D2-Pp</th>
<th>N-188</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 mM pNP</td>
<td>5</td>
<td>—</td>
<td></td>
<td>1,029.9</td>
<td>41.1</td>
</tr>
<tr>
<td>2 mM pNP</td>
<td>5</td>
<td>—</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Cellulose</td>
<td></td>
<td></td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

*Activities were performed at pH 5 and 55°C.

**Other Embodiments**

[0061] All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

[0062] A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.
Pro Asp Pro Tyr Leu Asn Gly Leu Ala Met Ser Asn Thr Ile Thr Gly
130 135 140
Met Gln Asp Ala Gly Val Gln Ala Cys Ala Lys His Phe Ile Gly Asn
145 150 155 160
Glu Gln Thr Thr Arg Asp Thr Met Ser Ser Asn Ile Asp Asp Arg
165 170 175
Thr Phe His Glu Leu Tyr Leu Trp Pro Phe Ala Asp Ala Ile Lys Ala
180 185 190
Asn Val Ala Ser Ile Met Cys Ser Tyr Asn Lys Phe Asn Glu Thr Tyr
195 200 205
Ala Cys Glu Asn Asn Phe Leu Thr Ile Leu Lys Gly Glu Leu Asp
210 215 220
Phe Gln Gly Phe Val Ser Asp Trp Ala Ala Gln His Thr Thr Ile
225 230 235 240
Gly Ser Ala Asn Ala Gly Leu Asp Val Ala Met Pro Gly Asp Asn Phe
245 250 255
Gly Asp Asn Tyr Tyr Leu Trp Gly Ser Asn Leu Leu Ala Ala Ile Ser
260 265 270
Asn Gly Thr Val Ala Gln Ser Arg Leu Asp Asp Met Val Thr Arg Ile
275 280 285
Leu Ala Ser Trp Tyr Phe Val Gly Gln Asp Gly Gly Tyr Pro Ala Val
290 295 300
Thr Trp Ser Ser Ser Trp Gly Gly Leu Gly Gly Pro Asn Val Gln Ala
305 310 315 320
Asp His Lys Gln Val Ala Arg Ala Ile Ala Arg Asp Gly Ile Val Leu
325 330 335
Leu Thr Asn Lys Asn Ala Leu Pro Leu Lys Pro Ala Ser Leu
340 345 350
Ala Ile Ile Gly Gln Asp Ala Ile Asp Asn Pro Ala Gly Ile Asn Ser
355 360 365
Cys Ser Asp Arg Gly Cys Asp Thr Gly His Leu Ala Met Gly Trp Gly
370 375 380
Ser Gly Thr Ala Asp Phe Pro Tyr Leu Val Ala Pro Leu Asp Ala Ile
395 390 395 400
Thr Pro Leu Ala Gln Ala Gly Lys Thr Leu Val Leu Ser Thr Thr
405 410 415
Asp Ser Thr Ser Ala Ala Ala Ala Ala Ala Ala Glu Thr Ala
420 425 430
Ile Val Phe Ile Thr Ala Asp Ser Gly Glu Gly Tyr Ile Thr Val Asp
435 440 445
Gly Gln Leu Gly Asp Arg Asn Ser Leu Ala Pro Trp Asn Asn Gly Thr
450 455 460
Ala Leu Val Gln Ala Val Ala Ser Ala Ser Lys Asn Val Ile Val Val
465 470 475 480
Ile Asn Ser Val Gly Pro Leu Ile Leu Gly Asp Ile Leu Ala Leu Ser
485 490 495
Ser Val Lys Ala Ile Val Trp Ala Gly Val Ser Gly Gln Glu Ser Gly
500 505 510
Asn Gly Leu Ala Asp Ile Leu Tyr Gly Ser Val Ser Pro Ser Gly Lys
515 520 525
Leu Pro Tyr Thr Ile Ala Lys Glu Ala Ser Asp Tyr Gly Thr Ala Ile
530 535 540
Val Pro Gly Asp Asp Arg Phe Pro Gly Leu Phe Val Asp Tyr Arg
545 550 555 560
His Phe Asp Gln Ala Asn Ile Gln Arg Phe Gly Phe Gly Tyr Gly
565 570 575
Leu Ser Tyr Thr Thr Phe Gln Tyr Ser Gly Leu Thr Ala Lys Tyr Ser
580 585 590
Asp Thr Ser Ala Gly Ser Ser Thr Leu Ala Pro Gly Gly Pro Gly Pro Lys Gly
595 600 605
Leu Tyr Asp Ile Val Ala Thr Val Thr Ala Lys Val Thr Asn Ser Gly
610 615 620
Thr Val Ser Gly Ala Gln Val Ala Gln Leu Tyr Ile Gly Leu Pro Gly
625 630 635 640
Ser Ala Pro Ala Ser Pro Pro Lys Gin Gin Leu Arg Gly Phe Asp Lys Ile
645 650 655
Ser Leu Lys Pro Gly Lys Ser Gly Thr Val Thr Phe Asn Leu Arg Arg
660 665 670
Lys Asp Leu Ser Tyr Trp Asp Thr Ala Ser Ala Gin Trp Val Thr Pro
675 680 685
Thr Ser Gly Leu Phe Ser Leu Tyr Val Gly Ala Ser Ser Arg Asp Ile
690 695 700
Arg Leu Gin Gin Ser Leu Lys Cys Ser Gly Gin Gly Ile Arg Lys Gly
705 710 715 720

Gly His

<210> SEQ ID NO 2
<211> LENGTH: 2166
<212> TYPE: DNA
<213> ORGANISM: Chaetomella raphigera

<400> SEQUENCE: 2

ctcgctgatg tgtatggggc agcggttata aaaaaaagta cccgtgtctt gcaaaaaattg 60
acgaacacc acaaaaagca tagtattct ggcgtggtgt ggaaaaaggg tctttgggt 120
gggtaactg cttcagatg atccaatggc cttccagagc tttgctatac agatgagcct 180
tctggccatc gtgttgcctc aatagtcacg gatcttctaa cggatttaca gaccttttcc 240
acgctggaga ctagttagct ctatcagttc ggtttagctt taggtcagca gggcaagggc 300
cctggcatca atgtctcggt cgtccacggat gttggtccca ttgtttatat tcagagggc 360
gttcgaacg gggaggtt cttctccagct ccatacgctg atgtctgadg aatgtcgaat 420
acaccacgg gatgtcagga tggctggctga gagcctttcg caagagcctt tattggcagc 480
gagcagagaa cgaatctcga tacaatgagc tccaacagc atgatcgagc tttccatgag 540
ttctgctcg cgcctttcgg agatgccaacg aagggcagcg tgtcatacag aatgtgctcg 600
taccaaggt tcaagctgag aatgtttgct cagaaaga ctctggccag tatttcacag 660
ggcaagcttg atttccagcg cttcgctgag aagcagctgg cttgtagctg atctactatt 720
sgaagctgca aatgttggtc ggtatgttcg tattttcgtg acaatcttttg cgaagcact 780	atctttgg ggcaagacatt cttcggcagc ctctccagcg gaacctgtgc ccaatctcgcg 840
cctcgatcga tgtcatatcgct ctcctcgctgc tttgtgtact tgttgtctca ggacacaggg 900
tatcagctct gcagctggtgc cttcggagat ggtatgtttc gcctgtagat 960
gatcataag agtgcagcag ccgcctcgcc cggagatgcg tttgctcatc cagaaacaag 1020
-continued

```
aacaaagccc tgcocctgaa gaagcgcgct agtttagcca tcatttgctc agtgtgcatc 
1080
gacaatcct cagacctaca ctctgctcct gcagcggctt gtatgcgggt gcacactgtg 
1140
taggtggttg gctgagggac agcaacacct tcattctgctt ttggggocgt gcagcgcgact 
1200
acacccctag ctacaggcoca gggaaacaca cttgtattgt cgacgcgagg cagactctct 
1260
gcgcctgcta gtgtgctgcgc gcgcgcagac cagaaaaacttc ttttcaagac tgcacaatgg 
1320
ggagagggat accaaccagt gtgggcccag tttggtgtgc gcaactactc tgcctactgg 
1380
aatagagct cttcgctgtc cttacggtgta gcagcggcaca gcacaatctgt ccattgctgtg 
1440
atcacaagct tcggccagat gactttgctg gcactttctg ctcctccagcg cggaaaaagca 
1500
atctgctggc gctgccggtc gccaggagaa tcgggcatag gcattgtgct ctattttctac 
1560
gttcagcat cttccccctcg gcattctccaa tcacaaaaca tgacagctga cagactact 
1620
ggcacagcc tttgccccgg ttgggtcgtc ccagactataa acatttttgc cagaccttcg 
1680
cattcagct gaaaggaact ccagcccggt tttgaaattg gtcatgtgctc ctctcatagc 
1740
acctctcaag actcagcgct taccgcaag tttcagcata cttccggccag cgcctcactct 
1800
cctgoccgct gggcaccgca gggctctgtat gatattgtgt agaagctgcc agcaagtctg 
1860
acaacacagcg gcacgctgct gcacagacta tatacattg cttcggggtg gcgtcctgcc 
1920
ttgccggtc catcctcacta gcactgatga cgcgttgttg gcacaatctcg cctccacgag 
1980
ggcacagct gcwgtggctg cttccwctcg gcggcacwag gccwctcagt cgwggwattgc 
2040
gctgggctgagccttatgg cggagctgtt ccttgagat gcwgtgctagc 
2100
tcgcagatgtcaggtcagct gggagtctgg ctattttctgcag ttcggggagct cgggagatctg 
2160
ggcagct 
2166
```

<210> SRQ_ID NO 3
<211> LENTHTH: 2951
<212> TYPE: DNA
<213> ORGANISM: Chaetomella zaphigera
<400> SEQUENCE: 3

cctggtgatg gctgggtgcc agcgtggcat tataaaaaa caatgtctgc ccgaaaaagt 
60
agcacaacgc aacaagacag tattggtact gcagcgggtg gggaaaaggg tcttggctgt 
120
gggatacct cttcagtagc accaatgaaga cttccacagc ttactgtaca agatggacac 
180
tcggagtgct cttcaggtta tttgactgct caaactctac gcagtttccag ctcgctttct 
240
cacaagatg caaaaaatttc ctacaggtat ctacaggtct ctcagcggct ccctaatgctt 
300
gctatcccg cggagtctgct cttctctagga gcctcgtgtg gcggaggagt gacagct 
360
gagtagatg ctcaggagca aggcatgct gcattctgct ccagcggct gtattgtctg 
420
gtccctgctg taaaatccaa gaggttggtg tgtgctggac gcggagatgt gcgggtgtac 
480
ctatgctgag gcgcgcagta ctggccagac ttttctttca gcattatac tgaattggtct 
540
agcattgcct aaagtgtgtgc gcattttaaaa ccccttagaa aagctatgca tgggtattag 
600
cattacgctg acgaaaggt gctgctgcct gcagttggtg gcacatctctt ctaaggtggt 
660
ttctagggcg gctctatctg cttccagatt gcacagcggag cagaaagcag 
720
cgcgatacag ctgagttcag cagcattctcc cttctggatt tttataagct 
780
aacacagat actttaggtg atacagatag ttgactgtc gccttttccc gcagctgcatc 
840
aagggcaccag ttgcatcat aatgtgctca tacaacaagt tcaatgagac atacgcgttg 900
gagaacaact ttcttgaacac tgtaarctct aaggtcgctg catcaaatct tcggatagtgc 960
aattactctc tcgcatcttc aagggcgacg tggagtcctca aaggtcgctg gcggcggtta 1020
tctggttctc tatactgtgc tggaaattgc cagccagcta acgcgatttg gcggagtacct 1080
gggggtgctca gcagctactc atcggaagtgc caatgctcgg tctgtgaatct tgaacaact 1140
ggcagcataa tatacgtgctg agatcctgat agcagctgta aactgaagctg tttcaaggtg 1200
gccggctgcctggtgcgac tttgagccag aacatatatct ttagggccag cacatctcttg 1260
ggggactct ccaagtaagt tctggctctg caaattgctt tgaagctgtgc gatactgctg 1320
cctcctacaag ggccgacttg ggccacacct ccgtgctgat gaactctctg cgcattcttg 1380
cctgctgctta ttggtgatgt cagggccacag gcacctcagc tcggtacatgc ttccattgga 1440
actggtgatt gggagctgcc aatgtgcaag ctgacataaa gcaggtgcca cgccgcatcg 1500
cccgagataag catcgtctca ttcacgaaac aaaaacagcg cctgctccctg aagaagccg 1560
cctgctgcttatc tatacgtgctg caatgctcgg tggagttgctg aatgtgctc 1620
cctgctgctgattg ctcagcagctg gccggagtgg ccctggtgctg atcagcagctg 1680
attaccactc ctgtgcggct cttgtgcgctct tctttacttg ggccgactgg acagcagctg 1740
agttgtcttt gtcgagcgac gcacatgatt tctgccgctgc attggtgctg gcgcgtcgacg 1800
agaagcagcct gcaggctccat atgcgctgctg cagggactgg atccatgctctgc 1860
ggcgctgcgcgt gcagctgctg acatctgctg aagccgctgc gcgggtactg cggccagact 1920
gggtgtctgc ctcagcagctg ctgacacggaa tcggaacctg ctggctccctg aagcgtttc 1980
cagcctggcag aaaaaagttc tgcagctctg cacggtacg ttcgagctgctg ttccctcttg 2040
ccttctcgcc ctggtgatgt cagggccacttg cttgctgctg aatgtgctc 2100
gggcagctgc ctggtgatgt tctttacttg ttcggtatct gcgaggctgg aatctccataa 2160
cacatgctgc gcacagcttg aagccgctgc gcgcgtcgacg gcgctctctc cgggctgtttc 2220
tcgcggagct ttttgctgact cagctgctgctg aacgcacccaa aacaaatctc acgcgctgt 2280
tgagctcccct ccgagctgctg cggagccgctg cttctttccct ccagacatag 2340
gtsagcggagct tgcagctgctg gaagaggtgc gccttctttta gcggctgtgta tggagctgc 2400
catacaacgc tcaacgcgct cctactctcg cagggctatgc agcgtttttgc cagacgtactc 2460
cgccacgc gcgcgcgctg cttgtgtgctg cttgctgctg aatgtgctc 2520
tggttcagct gcgcgcgcgc gcgcgcgctg ctcgctgctg gcggcctgctg 2580
acagctcttcc atgtgttggcg ggccgctgctg cggagctgcct gcgcctcagcg gcgctgctg 2640
tenctctgacc atgcctccct ggcggcgcgc gcgcgcgtgg gcgcgcgcgc gcgcgcgcgc 2700
agaggccgctgc agcgtttttgc ttcggtatct gcgcctcagcg gcggcgttgg gcgcgcgcgc 2760
gttttctcct gcaggt cgctgctg agcgtttttgc gcggcgcgc gcgcgcgcgc gcgcgcgcgc 2820
tggtggccgctgc gcggcgcgc gcgcgcgcgc gcgcgcgcgc gcgcgcgcgc 2880

Met Tyr Ser Arg Ala Phe Thr Phe Ala Ala Leu Ala Val Pro Ile Leu
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ala</td>
<td>Ala</td>
<td>Pro</td>
<td>Gly</td>
</tr>
<tr>
<td>2</td>
<td>Ala</td>
<td>Ala</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>3</td>
<td>Ala</td>
<td>Ala</td>
<td>Gly</td>
<td>Asp</td>
</tr>
<tr>
<td>4</td>
<td>Ala</td>
<td>Ala</td>
<td>Asp</td>
<td>Trp</td>
</tr>
<tr>
<td>5</td>
<td>Ala</td>
<td>Ala</td>
<td>Ala</td>
<td>Ala</td>
</tr>
<tr>
<td>6</td>
<td>Ala</td>
<td>Ala</td>
<td>Tyr</td>
<td>Lys</td>
</tr>
<tr>
<td>7</td>
<td>Lys</td>
<td>Ala</td>
<td>Ala</td>
<td>Thr</td>
</tr>
<tr>
<td>10</td>
<td>Ala</td>
<td>Ala</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td>11</td>
<td>Ala</td>
<td>Leu</td>
<td>Ala</td>
<td>Lys</td>
</tr>
<tr>
<td>12</td>
<td>Leu</td>
<td>Ser</td>
<td>Asn</td>
<td>Thr</td>
</tr>
<tr>
<td>13</td>
<td>Asp</td>
<td>Lys</td>
<td>Ala</td>
<td>Ser</td>
</tr>
<tr>
<td>14</td>
<td>Ile</td>
<td>Val</td>
<td>Ala</td>
<td>Thr</td>
</tr>
<tr>
<td>15</td>
<td>Gly</td>
<td>Val</td>
<td>Gly</td>
<td>Trp</td>
</tr>
<tr>
<td>16</td>
<td>Gly</td>
<td>Trp</td>
<td>Glu</td>
<td>Lys</td>
</tr>
<tr>
<td>17</td>
<td>Gly</td>
<td>Pro</td>
<td>Cys</td>
<td>Val</td>
</tr>
<tr>
<td>18</td>
<td>Gly</td>
<td>Pro</td>
<td>Gly</td>
<td>Asn</td>
</tr>
<tr>
<td>19</td>
<td>Thr</td>
<td>Ala</td>
<td>Ala</td>
<td>Val</td>
</tr>
<tr>
<td>20</td>
<td>Gly</td>
<td>Ala</td>
<td>Leu</td>
<td>Gly</td>
</tr>
<tr>
<td>21</td>
<td>Ala</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>22</td>
<td>Ala</td>
<td>Leu</td>
<td>Gly</td>
<td>Ile</td>
</tr>
<tr>
<td>23</td>
<td>Arg</td>
<td>Phe</td>
<td>Val</td>
<td>Gln</td>
</tr>
<tr>
<td>24</td>
<td>Gly</td>
<td>Asn</td>
<td>Val</td>
<td>Thr</td>
</tr>
<tr>
<td>25</td>
<td>Ala</td>
<td>Phe</td>
<td>Pro</td>
<td>Thr</td>
</tr>
<tr>
<td>26</td>
<td>Gly</td>
<td>Ile</td>
<td>Gln</td>
<td>Thr</td>
</tr>
<tr>
<td>27</td>
<td>Ala</td>
<td>Ser</td>
<td>Thr</td>
<td>Thr</td>
</tr>
<tr>
<td>28</td>
<td>Thr</td>
<td>Thr</td>
<td>Asp</td>
<td>Ile</td>
</tr>
<tr>
<td>29</td>
<td>Ser</td>
<td>Leu</td>
<td>Ile</td>
<td>Tyr</td>
</tr>
<tr>
<td>30</td>
<td>Ser</td>
<td>Arg</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td>31</td>
<td>Ala</td>
<td>Leu</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Gly</td>
<td>Glu</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>33</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td>34</td>
<td>Leu</td>
<td>Gly</td>
<td>Ile</td>
<td>Asn</td>
</tr>
<tr>
<td>35</td>
<td>Val</td>
<td>Gly</td>
<td>Leu</td>
<td>Gly</td>
</tr>
<tr>
<td>36</td>
<td>Pro</td>
<td>Val</td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Ala</td>
<td>Gly</td>
<td>Pro</td>
<td>Ile</td>
</tr>
<tr>
<td>38</td>
<td>Gly</td>
<td>Ile</td>
<td>Pro</td>
<td>Glu</td>
</tr>
<tr>
<td>39</td>
<td>Ala</td>
<td>Gly</td>
<td>Arg</td>
<td>Asn</td>
</tr>
<tr>
<td>40</td>
<td>Arg</td>
<td>Trp</td>
<td>Glu</td>
<td>Gly</td>
</tr>
<tr>
<td>41</td>
<td>Phe</td>
<td>Ser</td>
<td>Pro</td>
<td>Asp</td>
</tr>
<tr>
<td>42</td>
<td>Pro</td>
<td>Tyr</td>
<td>Leu</td>
<td>Asn</td>
</tr>
<tr>
<td>43</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Met</td>
</tr>
<tr>
<td>44</td>
<td>Ser</td>
<td>Thr</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>Thr</td>
<td>Gly</td>
<td>Met</td>
<td>Gln</td>
</tr>
<tr>
<td>46</td>
<td>Asp</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
</tr>
<tr>
<td>47</td>
<td>Cys</td>
<td>Ala</td>
<td>Lys</td>
<td>His</td>
</tr>
<tr>
<td>48</td>
<td>Phe</td>
<td>Ile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>Gly</td>
<td>Asn</td>
<td>Glu</td>
<td>Thr</td>
</tr>
<tr>
<td>50</td>
<td>Asp</td>
<td>Arg</td>
<td>Thr</td>
<td>His</td>
</tr>
<tr>
<td>51</td>
<td>Gly</td>
<td>Leu</td>
<td>Tyr</td>
<td>Thr</td>
</tr>
<tr>
<td>52</td>
<td>Pro</td>
<td>Phe</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>53</td>
<td>Ala</td>
<td>Ile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>Lys</td>
<td>Ala</td>
<td>Asn</td>
<td>Val</td>
</tr>
<tr>
<td>55</td>
<td>Ala</td>
<td>Ser</td>
<td>Ile</td>
<td>Met</td>
</tr>
<tr>
<td>56</td>
<td>Cys</td>
<td>Ser</td>
<td>Tyr</td>
<td>Asn</td>
</tr>
<tr>
<td>57</td>
<td>Lys</td>
<td>Phe</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td>58</td>
<td>Glu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>Thr</td>
<td>Tyr</td>
<td>Ala</td>
<td>Cys</td>
</tr>
<tr>
<td>60</td>
<td>Glu</td>
<td>Asn</td>
<td>Phe</td>
<td>Leu</td>
</tr>
<tr>
<td>61</td>
<td>Thr</td>
<td>Thr</td>
<td>Ile</td>
<td>Leu</td>
</tr>
<tr>
<td>62</td>
<td>Lys</td>
<td>Gly</td>
<td>Glu</td>
<td></td>
</tr>
<tr>
<td>63</td>
<td>Leu</td>
<td>Asp</td>
<td>Phe</td>
<td>Gly</td>
</tr>
<tr>
<td>64</td>
<td>Phe</td>
<td>Val</td>
<td>Ser</td>
<td>Asp</td>
</tr>
<tr>
<td>65</td>
<td>Trp</td>
<td>Ala</td>
<td>Ala</td>
<td>Gln</td>
</tr>
<tr>
<td>66</td>
<td>His</td>
<td>Thr</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>Thr</td>
<td>Ile</td>
<td>Gly</td>
<td>Ser</td>
</tr>
<tr>
<td>68</td>
<td>Ala</td>
<td>Asm</td>
<td>Ala</td>
<td>Gly</td>
</tr>
<tr>
<td>69</td>
<td>Leu</td>
<td>Asp</td>
<td>Val</td>
<td>Ala</td>
</tr>
<tr>
<td>70</td>
<td>Met</td>
<td>Pro</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>Asp</td>
<td>Gly</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>72</td>
<td>Asm</td>
<td>Leu</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>Leu</td>
<td>Thr</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>74</td>
<td>Ser</td>
<td>Asm</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>Leu</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>76</td>
<td>Ala</td>
<td>Ala</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>Ala</td>
<td>Ser</td>
<td>Arg</td>
<td>Leu</td>
</tr>
<tr>
<td>78</td>
<td>Asp</td>
<td>Met</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>Thr</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>Arg</td>
<td>Ile</td>
<td>Leu</td>
<td>Ala</td>
</tr>
<tr>
<td>81</td>
<td>Ser</td>
<td>Trp</td>
<td>Tyr</td>
<td>Phe</td>
</tr>
<tr>
<td>82</td>
<td>Val</td>
<td>Gly</td>
<td>Gln</td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>Gly</td>
<td>Gln</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>84</td>
<td>Tyr</td>
<td>Pro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>Ala</td>
<td>Val</td>
<td>Thr</td>
<td>Trp</td>
</tr>
<tr>
<td>86</td>
<td>Ser</td>
<td>Ser</td>
<td>Trp</td>
<td></td>
</tr>
<tr>
<td>87</td>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>88</td>
<td>Leu</td>
<td>Gly</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>89</td>
<td>Ala</td>
<td>Val</td>
<td></td>
<td></td>
</tr>
<tr>
<td>90</td>
<td>Gly</td>
<td>His</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>91</td>
<td>Val</td>
<td>Ala</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>Ala</td>
<td>Arg</td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>93</td>
<td>Gly</td>
<td>Ile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>94</td>
<td>Val</td>
<td>Leu</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>Ala</td>
<td>Leu</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>Ala</td>
<td>Pro</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>Lys</td>
<td>Leu</td>
<td></td>
<td></td>
</tr>
<tr>
<td>98</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>Ala</td>
<td>Ser</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>Gly</td>
<td>Ala</td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>Leu</td>
<td>Ala</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>102</td>
<td>Asp</td>
<td>Gly</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>Ala</td>
<td>Asp</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>104</td>
<td>Tyr</td>
<td>Leu</td>
<td>Asn</td>
<td></td>
</tr>
<tr>
<td>105</td>
<td>Gly</td>
<td>Ala</td>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>106</td>
<td>Ser</td>
<td>Thr</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>107</td>
<td>Thr</td>
<td>Gly</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td>108</td>
<td>Ala</td>
<td>Asm</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>109</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>110</td>
<td>Met</td>
<td>Pro</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>111</td>
<td>Asp</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>112</td>
<td>Arg</td>
<td>Asm</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>113</td>
<td>Tyr</td>
<td>Leu</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>114</td>
<td>Ser</td>
<td>Asm</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>115</td>
<td>Leu</td>
<td>Ala</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>116</td>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>117</td>
<td>Ile</td>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>118</td>
<td>Asn</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>119</td>
<td>Thr</td>
<td>Val</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>120</td>
<td>Gln</td>
<td>Ser</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>Leu</td>
<td>Asp</td>
<td>Met</td>
<td></td>
</tr>
<tr>
<td>122</td>
<td>Val</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>Arg</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>124</td>
<td>Ile</td>
<td>Leu</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>125</td>
<td>Ser</td>
<td>Trp</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>126</td>
<td>Phe</td>
<td>Val</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>127</td>
<td>Gln</td>
<td>Gln</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>128</td>
<td>Tyr</td>
<td>Pro</td>
<td></td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>Ala</td>
<td>Val</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>Thr</td>
<td>Ser</td>
<td>Trp</td>
<td></td>
</tr>
<tr>
<td>131</td>
<td>Ala</td>
<td>Gly</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>Leu</td>
<td>Gly</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>133</td>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>134</td>
<td>Gln</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>135</td>
<td>Ala</td>
<td>Asp</td>
<td>His</td>
<td></td>
</tr>
<tr>
<td>136</td>
<td>Lys</td>
<td>Gln</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td>137</td>
<td>Ala</td>
<td>Arg</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>138</td>
<td>Ile</td>
<td>Ala</td>
<td>Arg</td>
<td></td>
</tr>
<tr>
<td>139</td>
<td>Asp</td>
<td>Gly</td>
<td>Ile</td>
<td></td>
</tr>
<tr>
<td>140</td>
<td>Val</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>141</td>
<td>Leu</td>
<td>Leu</td>
<td>Thr</td>
<td></td>
</tr>
<tr>
<td>142</td>
<td>Ala</td>
<td>Asn</td>
<td>Lys</td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>Ala</td>
<td>Pro</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td>144</td>
<td>Lys</td>
<td>Leu</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>145</td>
<td>Ala</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>146</td>
<td>Ser</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>147</td>
<td>Ala</td>
<td>Ile</td>
<td>Gly</td>
<td></td>
</tr>
<tr>
<td>148</td>
<td>Asp</td>
<td>Ala</td>
<td>Ile</td>
<td>Asp</td>
</tr>
<tr>
<td>149</td>
<td>Ala</td>
<td>Asp</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>Ala</td>
<td>Gly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>Gly</td>
<td>Ser</td>
<td></td>
<td></td>
</tr>
<tr>
<td>152</td>
<td>Thr</td>
<td>Ala</td>
<td>Asp</td>
<td></td>
</tr>
<tr>
<td>153</td>
<td>Phe</td>
<td>Pro</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td>154</td>
<td>Leu</td>
<td>Val</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td>155</td>
<td>Ala</td>
<td>Pro</td>
<td>Leu</td>
<td></td>
</tr>
</tbody>
</table>
 Ala Ile Thr Pro Leu Ala Glu Ala Glu Gly Thr Lys Leu Val Leu Ser
  420 425 430
Thr Thr Asp Ser Thr Ser Ala Ala Ser Ala Ala Ala Ala Ala Glu
  435 440 445
Thr Ala Ile Val Phe Ile Thr Ala Asp Ser Gly Gly Gly Tyr Ile Thr
  450 455 460
Val Asp Gly Glu Gly Asp Arg Asn Ser Leu Ala Pro Trp Asn Asn
  465 470 475 480
Gly Thr Ala Leu Val Glu Ala Val Ala Ser Ala Ser Lys Asn Val Ile
  485 490 495
Val Val Ile Asn Ser Val Gly Pro Leu Ile Leu Glu Asp Ile Leu Ala
  500 505 510
Leu Ser Ser Val Lys Ala Ile Val Trp Ala Gly Val Ser Gly Gin Glu
  515 520 525
Ser Gly Asn Gly Leu Ala Asp Ile Leu Tyr Gly Ser Val Ser Pro Ser
  530 535 540
Gly Lys Leu Pro Tyr Thr Ile Ala Lys Gin Ala Ser Asp Tyr Gly Thr
  545 550 555 560
Ala Ile Val Pro Gly Asp Asp Asn Phe Pro Glu Gly Leu Phe Val Asp
  565 570 575
Tyr Arg His Phe Asp Gin Ala Asn Ile Gin Pro Arg Phe Glu Phe Gly
  580 585 590
Tyr Gly Leu Ser Tyr Thr Thr Phe Gin Tyr Ser Gin Leu Thr Ala Lys
  595 600 605
Tyr Ser Asp Thr Ser Ala Gly Ser Ser Thr Leu Ala Pro Gly Gly Pro
  610 615 620
Lys Gly Leu Tyr Asp Ile Val Ala Thr Val Thr Ala Val Thr Asn
  625 630 635 640
Ser Gly Thr Val Ser Gly Ala Val Ala Gin Leu Tyr Ile Gly Leu
  645 650 655
Pro Gly Ser Ala Pro Ala Ser Pro Pro Lys Gin Leu Arg Gly Phe Asp
  660 665 670
Lys Ile Ser Leu Lys Pro Gly Lys Ser Gly Thr Val Thr Phe Asn Leu
  675 680 685
Arg Arg Lys Asp Leu Ser Tyr Thr Trp Asp Thr Ala Ser Ala Gin Trp Val
  690 695 700
Thr Pro Thr Ser Gly Glu Phe Ser Leu Tyr Val Gly Ala Ser Ser Arg
  705 710 715 720
Asp Ile Arg Leu Gin Gly Ser Leu Lys Cys Ser Gly Gin Gly Ile Arg
  725 730 735
Lys Gly Gly His
  740

<210> STRID NO 5
<211> LENGTH: 2220
<212> TYPE: DNA
<213> ORGANISM: Chaetomella raphigera
<400> SEQUENCE: 5
atgatccct cagccttta cttcgcagcc ctatgcgtac caactctcg tcgtcctgtg
  60
gacaggtatt gggcagggc ttataaaaa getacgcgtg ctotgcgaa aatgagcaac
  120
acccgacaa gatgatgtcg tcagtcggtg ggttgtggaa aaggtctcgg cyttgggaa
  180
-continued

acgctgcag tagsatatat tggacttcca gacgttgtct atcaagagtg aacccactgg 240
atcgttctcc tccaaatatt aacgcattt cctacaggtta ttcaagatgc tctcaatggt 300
gcgtactgt tgaatccgcc ctagctttta gttatatgtc agagggcaca agcacttgcc 360
atcactgtcc agcgctggcc agtgctgtcg cccatgtgta aattaatgca ggcgcgttgcc 420
aacctggagaag gtctttctcc agatcatata ctaagacgcc tggagtctggt 480
acggaagtc aaggtgtgct gtcacaagtt tgcaggattt gaacagacaa 540
aagacgatac cggctcaacac cagaccttga atcgtatgac ggcactttcct ctaggttgat 600
cctctgtcct ttgcagatgc cttcaagggcc aacgtagtat ccataatgtg ctctcataac 660
aagttcaatg agacataaggt tggtagaacc aacctttgga caacattact caagggcagag 720
ctttgctttt aaggtggtttgt ctcagcagcc tgggtgtctg agatactacac tattggaact 780
gcaatgctgt tctgtagttgt ggtgcaactt tctggaacaa cttattacta 840
ttgccagcag atctttgctgc ggcgtatcct aacgggcatg ttcgcaacac tcgctgttag 900
gacctgctc cctcgcctgt ggtctttggt tccgtttcag gtcgcaacac agcagatcaca 960
gctgtggact ggtctttggt gaaggtgtggc ttggagagtc ccaagttgca ggcgtgatac 1020
aagctgagtgc caagctgcgtc ccggccggat ggcggatgct tacgacagca caaasaacaaa 1080
gcgcggtcgc tgaagacagc cggtagttta gcaatatatg gcggatagtc gctagttatca 1140
cctgctccagc tcatctgtgct ctcagccgct ggtgtagata ggtgatgggt 1200
tgggtgtcgc gcaccagca cttctctata cttcggttgc cgggtgatgc cttcagacccc 1260
ttagctcagc ccattggagca aaaaattgtt tgcgacgct cggacagatgc ttctgcgcct 1320
gctgtggtgc cccggcgctgcc agagacggca acatcaagtc tccatgctgcc tccagggag 1380
gacatcactc cttggtggccc ccaatgatgg cttgcgcaatt cagctttctcc 1440
ggcgttcgct tcttcccagct gttgagcagc gcgcggccaaa atgtctttgt cttgatcaac 1500
agtgctgctgc cattggtttc gcggactcatt cttcgatttt ccaggctggag acaatagttg 1560
tgggtgtcgc tccggagcag aagatgctgc aacgtgctgc atggatctct ttaagcttca 1620
gtaccttaca gttggaatac ccatacatac atgcggaaaa agccagcgac ttaagacaac 1680
gcctttggccc cgcgtagctgcc taacctttgc gaaggatgtt tgttagcata ctgcctttcc 1740
gacacagaca acatccagcgc ggcgttttggc tttggtttag gcgtccttota ttaagacatt 1800
catacctgc agttacgagc aagacttcac gatatttcccc cagggatgtc catctttgc 1860
cctgacgcgg ccagcgggtc gtagttatgt gcgcgaagcg tattcgcgctt gttcagacac 1920
agcggccagct taagtgccgcc ctagctctca tttgtcttgc gggtgctgctgc 1980
ccctcaatcc cccacaagcc gtaacggctt gcgtcttatt gcgtctttcc 2040
tgacgctgcag tctcggagt ccggagctgcc agcacccctt ctgcgtctggc 2100
gtactggtgg tttccagagc cagggccagagt cttgcctggt atgtggtgtc tagctgtaga 2160
gataaaggt taaaagttcgt cttgccacag gcgggccag gcgtgtttctt 2220
-continued

gatgggtatt ggcaaggggc tttatataaa gcatacctcg ctctgacaan attgtgcacc 120
aceccaaag ccagattttg taactgctgg ggttggaaaa aggtcttttg ggtttggagt 180
acgtgcagc tagatcaaat ttggcttctaa gacgtctggc ctctgacaat attgtgcacc 240
taggtctgct ggcttttcttg ctctgacaac tgcggtctga aagcactgttg ctctgacaat 300
tgcggaacag tttcttctac gttatgcttc gttcttctac gttcttctac gttgtaacct 360
cagtctgcacct ttcttcttct tctctctcttt cattttgctg aagctattctg ctctgacaat 420
ttaggtgcggc gcgcaagtct caagctgtgtgt gctctgtttc ggttggaaaa 480
tggtttaaat tctcgactgt tggtagatct gcggtttttc ggttggaaaa 540
acggtctgct gcaagcttgc agggcgtttc tctctctcttt ctctgacaan attgtgcacc 600
tgctataagtg gatctgtggt tctcctctgtg caagcttgc gcggctgaat 660
cgcagatgtg gcgcaagtct cttcttttc gcaagcttgc gcgctcctctg cttcctctgtg 720
gcggcgctc agcaagttgc agaagcttgc gcgctcctctg cttcctctgtg ctctgacaan 780
ccagatcagt gctgatacag ctcttctctct cattttgctg ggttggaaaa 840
agacatcgg gactgtgctc catagctgc catagctgc catagctgc catagctgc 900
aagctctgtc cattctgtgct cattctgtgct cattctgtgct cattctgtgct cattctgtgct 960
acactctgcag ccctgtgctc ctcttctctt cttcttttc gcaagcttgc gcgctcctctg 1020
tgctataagtg gatctgtggt tctcctctgtg caagcttgc gcgctcctctg cttcctctgtg 1080
ctctactgtc cattttgctg ggttggaaaa 1140
tgctataagtg gatctgtggt tctcctctgtg caagcttgc gcgctcctctg cttcctctgtg 1120
ccctgtgctc ctcttctctt cttcttttc gcaagcttgc gcgctcctctg cttcctctgtg 1180
acactctgcag ccctgtgctc ctcttctctt cttcttttc gcaagcttgc gcgctcctctg 1240
ctctgtgctc ctcttctctt cttcttttc gcaagcttgc gcgctcctctg cttcctctgtg 1300
atggtgcttg gcggcgtttc cattttgctg ggttggaaaa 1360
gatcgtgctg gcggcgtttc cattttgctg ggttggaaaa 1420
atggtgcttg gcggcgtttc cattttgctg ggttggaaaa 1480
gatcgtgctg gcggcgtttc cattttgctg ggttggaaaa 1540
gatcgtgctg gcggcgtttc cattttgctg ggttggaaaa 1600
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 1660
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 1720
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 1780
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 1840
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 1900
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 1960
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 2020
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 2080
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 2140
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 2200
tcagctgtgct gctctctcttt gcggcgtttc cattttgctg ggttggaaaa 2260
tggctatggc ctctgttaag ctccccgccgc gacggtacct ctccccgccgc gtctagatca gacgctaaca 2400
gagatgtgat atgtgaagag aagctagcct ctcagcgcct gtctagatca gacgctaaca 2460
aacggctacca gctctacttg cttttcgctgt ctggcgatct acggccaaagt acctccgatac 2520
ttccccggg cagcctcact ttgccccctg gggacccagag gggtgtgatag atatggtttgc 2580
aacggctacca gcaagaggtga cagcagcaggg caagcagcaggt ggggtgccag tggcaaccagct 2640
atatatgcttg ctcgctgctt cggcgctcgct atctcaccct cagcagctac ggggttttga 2700
caaaatcaggccccggaagcgaggttccgag cagcgtgaagag tttacccgct gcgcagaggg 2760
ccctcagctactgttgctatg ctcgccacag cttgggttgcag tggctattaca cggacccaggg ggcagttttg 2820
ccttgatggtgtcatcagttggtatcag cggacccaggg ggcagttttg 2880
ccagttgatc gggaagggag cacat 2905

<210> SEQ ID NO 7
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Chaetomella raphigera

<400> SEQUENCE: 7
Net Tyr Ser Arg Ala Phe Thr Phe Ala Ala Leu Ala Val Pro Ile Leu
1   5   10   15

Ala Ala

<210> SEQ ID NO 8
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Chaetomella raphigera

<400> SEQUENCE: 8
atgtatcccc gacgctcagc tttgccagcc ctagctgctc caaactcagc tgct 54

<210> SEQ ID NO 9
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Artificial sequence

<400> SEQUENCE: 9
Pro Gly Asp Gly Asp Trp Ala
1   5

<210> SEQ ID NO 10
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence

<400> SEQUENCE: 10
-continued

cmgngnggng ngnynynngcc 20

<210> SEQ ID NO 11
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic primer
<400> SEQUENCE: 11

ggtcctgcc acagtacga ctttcccc tttttttt 39

<210> SEQ ID NO 12
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic primer
<400> SEQUENCE: 12

ggtcctgcc acagtacga c 21

<210> SEQ ID NO 13
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic primer
<400> SEQUENCE: 13

cctgggtgg ggtgttgg cc 23

<210> SEQ ID NO 14
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic primer
<400> SEQUENCE: 14

atgtccacct ttccgaatct cttgcc 26

<210> SEQ ID NO 15
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic primer
<400> SEQUENCE: 15

cgttctgacc aaaaattgaac agcat 25

<210> SEQ ID NO 16
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic primer
<400> SEQUENCE: 16

gatgtttca tgtgaagtct tga 23

<210> SEQ ID NO 17
What is claimed is:

1. An isolated polypeptide comprising a sequence at least 70% identical to SEQ ID NO: 1, wherein the polypeptide has a β-glucosidase activity.

2. The isolated polypeptide of claim 1, wherein the sequence is at least 80% identical to SEQ ID NO: 1.

3. The isolated polypeptide of claim 2, wherein the sequence is at least 90% identical to SEQ ID NO: 1.

4. The isolated polypeptide of claim 3, wherein the sequence is at least 95% identical to SEQ ID NO: 1.

5. The isolated polypeptide of claim 4, wherein the sequence is at least 99% identical to SEQ ID NO: 1.

6. The isolated polypeptide of claim 5, wherein the polypeptide comprises the sequence of SEQ ID NO: 1.

7. An isolated nucleic acid encoding the polypeptide of claim 1.

8. The nucleic acid of claim 7, wherein the nucleic acid comprises the sequence of SEQ ID NO: 2, 3, 5, or 6.

9. An expression vector comprising a nucleic acid of claim 7.

10. A host cell comprising a nucleic acid of claim 7.

11. A method of producing a polypeptide, comprising culturing in a medium a host cell that contains a nucleic acid encoding the polypeptide of claim 1 under conditions permitting expression of the polypeptide encoded by the nucleic acid, and purifying the polypeptide from the cultured cell or the medium of the cell.

12. The method of claim 11, wherein the polypeptide has a sequence that is 90% identical to SEQ ID NO: 1.

13. The method of claim 11, wherein the polypeptide has a sequence that is 95% identical to SEQ ID NO: 1.

14. The method of claim 11, wherein the polypeptide has the sequence of SEQ ID NO: 1.


16. The composition of claim 15, wherein the composition further comprises an exoglucanase and an endoglucanase.

17. A method of producing fermentable sugars from lignocellulosic material, comprising providing a composition of claim 15, and contacting the composition with lignocellulosic material to produce fermentable sugars.

18. The method of claim 17, wherein the fermentable sugar is selected from the group consisting of glucose, xylose, arabinose, galactose, mannose, rhamnose, sucrose, fructose, lactose, maltose, trehalose, and cellobiose.

19. The method of claim 17, wherein the polypeptide has the amino acid sequence of SEQ ID NO: 1.

20. The method of claim 17, wherein the lignocellulosic material is selected from the group consisting of cellulosic animal waste, municipal solid waste, waste paper, yard waste, an agricultural residue, and a forestry residue.

21. The method of claim 17, further comprising converting the fermentable sugar to a fermentation product.

22. The method of claim 21, wherein the converting step is performed by microorganism fermentation or enzyme treatment.

23. A method of producing energy from lignocellulosic material, comprising providing a composition of claim 15, contacting the composition with the lignocellulosic material to produce a fermentable sugar, converting the fermentable sugar to produce a combustible fermentation product, and combusting the combustible fermentation product to produce energy.

24. A bioreactor containing a lignocellulosic material and a composition of claim 15.

* * * * *