



HU000027643T2

(19) **HU****MAGYARORSZÁG**
Szellemi Tulajdon Nemzeti Hivatala(11) Lajstromszám: **E 027 643**(13) **T2****EURÓPAI SZABADALOM**
SZÖVEGÉNEK FORDÍTÁSA(21) Magyar ügyszám: **E 06 847449**(22) A bejelentés napja: **2006. 10. 06.**(51) Int. Cl.: **C12P 19/04** (2006.01)**C12N 9/14** (2006.01)

(96) Az európai bejelentés bejelentési száma:

EP 20060847449

(86) A nemzetközi (PCT) bejelentési szám:

PCT/US 06/027347

(97) Az európai bejelentés közzétételi adatai:

EP 2041294 A2 **2008. 01. 17.**

(87) A nemzetközi közzétételi szám:

WO 08008070

(97) Az európai szabadalom megadásának meghirdetési adatai:

EP 2041294 B1 **2016. 04. 13.**

(30) Elsőbbségi adatok:

487547 **2006. 07. 13.** **US**

(73) Jogosult(ak):

Danisco US Inc., Palo Alto (US)

(72) Feltaláló(k):

GUSAKOV, Alexander V., Jupiter, FL 33477-5094 (US)**SALANOVICH, Tatyana N., Jupiter, FL 33477-5094 (US)****ANTONOV, Alexey I., Jupiter, FL 33477-5094 (US)****USTINOV, Boris B., Jupiter, FL 33477-5094 (US)****OKUNEV, Oleg N., Jupiter, FL 33477-5094 (US)****BURLINGAME, Richard, Jupiter, FL 33477-5094 (US)****EMALFARB, Mark, Jupiter, FL 33477-5094 (US)****BAEZ, Marco, Jupiter, FL 33477-5094 (US)****SINITSYN, Arkady P., Jupiter, FL 33477-5094 (US)**

(74) Képviselő:

Danubia Szabadalmi és Jogi Iroda Kft.,
Budapest

(54)

Nagy hatékonyságú celluláz készítmény létrehozása cellulóz enzimatis hidrolizisére

Az európai szabadalom ellen, megadásának az Európai Szabadalmi Közlönyben való meghirdetésétől számított kilenc hónapon belül, felszólalást lehet benyújtani az Európai Szabadalmi Hivatalnál. (Európai Szabadalmi Egyezmény 99. cikk(1))

A fordítást a szabadalmas az 1995. évi XXXIII. törvény 84/H. §-a szerint nyújtotta be. A fordítás tartalmi helyességét a Szellemi Tulajdon Nemzeti Hivatala nem vizsgálta.



(11) **EP 2 041 294 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
13.04.2016 Bulletin 2016/15

(21) Application number: **06847449.3**

(22) Date of filing: **06.10.2006**

(51) Int Cl.:
C12P 19/04 (2006.01) C12N 9/14 (2006.01)

(86) International application number:
PCT/US2006/027347

(87) International publication number:
WO 2008/008070 (17.01.2008 Gazette 2008/03)

(54) **CONSTRUCTION OF HIGHLY EFFICIENT CELLULASE COMPOSITIONS FOR ENZYMATIC HYDROLYSIS OF CELLULOSE**

KONSTRUKTION HOCHEFFIZIENTER CELLULASEZUSAMMENSETZUNGEN ZUR ENZYMATISCHEN HYDROLYSE VON CELLULOSE

CONSTRUCTION DE COMPOSITIONS DE CELLULASE HAUTEMENT EFFICACES POUR UNE HYDROLYSE ENZYMATIQUE DE LA CELLULOSE

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IS IT LI LT LU LV MC NL PL PT RO SE SI SK TR

(30) Priority: **13.07.2006 US 487547**

(43) Date of publication of application:
01.04.2009 Bulletin 2009/14

(73) Proprietor: **Dyadic International (USA), Inc.**
Jupiter, FL 33477-5094 (US)

(72) Inventors:

- **GUSAKOV, Alexander V.**
Jupiter, FL 33477-5094 (US)
- **SALANOVICH, Tatyana N.**
Jupiter, FL 33477-5094 (US)
- **ANTONOV, Alexey I.**
Jupiter, FL 33477-5094 (US)
- **USTINOV, Boris B.**
Jupiter, FL 33477-5094 (US)
- **OKUNEV, Oleg N.**
Jupiter, FL 33477-5094 (US)
- **BURLINGAME, Richard**
Jupiter, FL 33477-5094 (US)
- **EMALFARB, Mark**
Jupiter, FL 33477-5094 (US)
- **BAEZ, Marco**
Jupiter, FL 33477-5094 (US)
- **SINITSYN, Arkady P.**
Jupiter, FL 33477-5094 (US)

(74) Representative: **Swinkels, Bart Willem**
Nederlandsch Octrooibureau
P.O. Box 29720
2502 LS Den Haag (NL)

(56) References cited:

US-A- 6 015 707 US-A1- 2003 109 011
US-A1- 2004 002 136 US-A1- 2004 053 373
US-A1- 2004 197 890 US-A1- 2005 274 308
US-A1- 2006 053 514 US-B1- 6 214 164
US-B1- 6 573 086

- **BUKHTOJAROV F E ET AL: "CELLULASE COMPLEX OF THE FUNGUS CHRYSOSPORIUM LUCKNOWENSE: ISOLATION AND CHARACTERIZATION OF ENDOGLUCANASES AND CELLOBIOHYDROLASES", BIOCHEMISTRY, MAIK NAUKA - INTERPERIODICA, RU, vol. 69, no. 5, 1 May 2004 (2004-05-01), pages 542-551, XP008044624, ISSN: 0006-2979, DOI: DOI:10.1023/B:BIRY.0000029853.34093.13**
- **DATABASE UniProt [Online] 15 March 2005 (2005-03-15), "SubName: Full=Beta-glucosidase-like protein;", XP002612671, retrieved from EBI accession no. UNIPROT:Q5EMW3 Database accession no. Q5EMW3**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 2 041 294 B1

- DATABASE UniProt [Online] 21 March 2006 (2006-03-21), "SubName: Full=Putative uncharacterized protein;", XP002612672, retrieved from EBI accession no. UNIPROT:Q2H8N3 Database accession no. Q2H8N3
- GUSAKOV ALEXANDER V ET AL: "Design of highly efficient cellulase mixtures for enzymatic hydrolysis of cellulose", BIOTECHNOLOGY AND BIOENGINEERING, WILEY & SONS, HOBOKEN, NJ, US, vol. 97, no. 5, 1 August 2007 (2007-08-01) , pages 1028-1038, XP002565359, ISSN: 0006-3592, DOI: DOI:10.1002/BIT.21329 [retrieved on 2007-01-12]
- GRISHUTIN S G ET AL: "Specific xyloglucanases as a new class of polysaccharide-degrading enzymes", BIOCHIMICA ET BIOPHYSICA ACTA - GENERAL SUBJECTS, ELSEVIER SCIENCE PUBLISHERS, NL, vol. 1674, no. 3, 1 November 2004 (2004-11-01), pages 268-281, XP004634222, ISSN: 0304-4165, DOI: DOI:10.1016/J.BBAGEN.2004.07.001
- GUSAKOV ALEXANDER V ET AL: "Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*", ENZYME AND MICROBIAL TECHNOLOGY, STONEHAM, MA, US, vol. 36, no. 1, 6 January 2005 (2005-01-06), pages 57-69, XP002608583, ISSN: 0141-0229, DOI: DOI:10.1016/J.ENZMICTEC.2004.03.025 [retrieved on 2004-11-17]
- DATABASE EMBL [Online] BIRREN, ET AL.: 'Chaetomium globosum CBS 148.51', XP008107880 Retrieved from EBI Database accession no. (CH408033)
- DATABASE EMBL [Online] February 2006 BIRREN, ET AL.: 'Chaetomium globosum CBA 148.51', XP008107881 Database accession no. CH408030
- DATABASE EMBL [Online] February 2006 BIRREN ET AL: 'Chaetomium globosum CBS 148.51', XP008107882 Database accession no. (CH408035)
- DATABASE GENBANK [Online] 27 December 2008 TSANG: 'StheSEQ13790', XP008107883 Database accession no. (EB054411)
- DATABASE GENBANK [Online] 27 December 2008 DEAN: 'mgct001xa16f', XP008107885 Database accession no. (BG809574)

Description

FIELD OF THE INVENTION

[0001] This invention relates to compositions and methods for producing bioenergy or other value-added products from lignocellulosic biomass or cellulosic materials. In particular, the invention provides enzyme compositions capable of converting a variety of cellulosic substrates or lignocellulosic biomass into a fermentable sugar. The invention also provides methods for using such enzyme compositions.

INTRODUCTION

[0002] Bioconversion of renewable lignocellulosic biomass to a fermentable sugar that is subsequently fermented to produce alcohol (e.g., ethanol) as an alternative to liquid fuels has attracted an intensive attention of researchers since 1970s, when the oil crisis broke out because of decreasing the output of petroleum by OPEC (Bungay H.R., "Energy: the biomass options". NY: Wiley; 1981; Olsson L, Hahn-Hägerdal B. "Fermentation of lignocellulosic hydrolysates for ethanol production", Enzyme Microb Technol 1996;18:312-31; Zaldivar J, Nielsen J, Olsson L. "Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration", Appl Microbiol Biotechnol 2001;56:17-34; Galbe M, Zacchi G., "A review of the production of ethanol from softwood", Appl Microbiol Biotechnol 2002; 59:618-28). Ethanol has been widely used as a 10% blend to gasoline in the USA or as a neat fuel for vehicles in Brazil in the last two decades. The importance of fuel bioethanol will increase in parallel with skyrocketing prices for oil and gradual depletion of its sources. Additionally, fermentable sugars are being used to produce plastics, polymers and other biobased products and this industry is expected to grow substantially therefore increasing the demand for abundant low cost fermentable sugars which can be used as a feed stock in lieu of petroleum based feedstocks (e.g. see article "The Rise Of Industrial Biotech" published in Forbes July 24, 2006)

[0003] The major polysaccharides comprising different lignocellulosic residues, which may be considered as a potential renewable feedstock, are cellulose and hemicelluloses (xylans). The enzymatic hydrolysis of these polysaccharides to soluble sugars, for example glucose, xylose, arabinose, galactose, mannose, and other hexoses and pentoses occurs under the action of different enzymes acting in concert. Endo-1,4- β -glucanases (EG) and exo-cellobiohydrolases (CBH) catalyze the hydrolysis of insoluble cellulose to cellobiosaccharides (cellobiose as a main product), while β -glucosidases (BGL) convert the oligosaccharides to glucose. Xylanases together with other accessory enzymes (non-limiting examples of which include α -L-arabinofuranosidases, feruloyl and acetyl xylan esterases, glucuronidases, and β -xylosidases) catalyze the hydrolysis of hemicelluloses.

[0004] DATABASE UniProt 2006 "SubName: Full=Putative uncharacterized protein;", accession no. Q2H8N3, discloses a protein as putative uncharacterized protein with hydrolase activity.

[0005] Regardless of the type of cellulosic feedstock, the cost and hydrolytic efficiency of enzymes are major factors that restrict the commercialization of the biomass bioconversion processes. The production costs of microbially produced enzymes are tightly connected with a productivity of the enzyme-producing strain and the final activity yield in the fermentation broth. The hydrolytic efficiency of a multienzyme complex in the process of lignocellulose saccharification depends both on properties of individual enzymes, the synergies between them, and their ratio in the multienzyme cocktail.

[0006] *Chrysosporium lucknowense* is a fungus that is known to produce a wide variety of cellulases, hemicellulases, and possibly other accessory enzymes. *C. lucknowense* also secretes at least five different endoglucanases, the EG II (51 kDa, Cel5A) being the most active.

[0007] Bukhtojarov et al., Biochemistry (Moscow), 69(5):666-677, describes pH and temperature optimum of the enzymes belonging to the *C. lucknowense* cellulase complex. US 6,015,707 A relates to the industrial use of a cellulase complex from *C. lucknowense*.

[0008] Moreover, *C. lucknowense* mutant strains (including UV18-25) have been developed to produce enzymes for textile, pulp and paper, detergent and other applications, but not for the enzymatic saccharification of cellulose; these strains can also be used for a high-level production of homologous and heterologous proteins. The best *C. lucknowense* mutant strains secrete at least 50-80 g l⁻¹ of extracellular protein in low viscosity fermentations. The full fungal genome of the *C. lucknowense* has been sequenced in 2005 (see http://WorldWideWeb.dyadic-group.com/wt/dy-ad/pr_1115654417), and now the genome annotation is being carried out.

[0009] The crude *C. lucknowense* multienzyme complex demonstrates modest results in cellulose saccharification, with only a fraction of the cellulose being converted to glucose under the conditions tested. Two cellobiohydrolases of *C. lucknowense*, belonging to families 7 and 6 of glycoside hydrolases: CBH Ia (Cel7A) and CBH IIa (Cel6A), have been previously isolated and studied. CBH Ia was previously referred to as CBH I, 70(60) kD protein in US Patent No. 6,573,086. CBH Ia exists in the culture broth as a full size enzyme (observed molecular mass 65 kDa, SDS-PAGE data), consisting of a core catalytic domain and cellulose-binding module (CBM) connected by a flexible peptide linker, and its truncated form (52 kDa), representing the enzyme catalytic domain. CBH I (Cel7A) of *C. lucknowense* appears to be slightly less

effective in hydrolysis of crystalline cellulose but more thermostable than the CBH I of *T. reesei*. CBH IIa was previously thought to be an endoglucanase and has been referred to as 43 kD Endo and EG6. See, e.g., US Patent No. 6,573,086. CBH Ua (43 kDa) has no CBM, i.e. its molecule contains only the catalytic domain.

[0010] In spite of the continued research of the last few decades to understand enzymatic lignocellulosic biomass degradation and cellulase production, it remains desirable to discover or to engineer new highly active cellulases and hemicellulases. It would also be highly desirable to construct highly efficient enzyme compositions capable of performing rapid and efficient biodegradation of lignocellulosic materials.

SUMMARY OF THE INVENTION

[0011] The invention is defined by the appended claims.

[0012] Disclosed are several newly identified and isolated enzymes from *C. lucknowense*. The new enzymes include two new cellobiohydrolases (CBH Ib and IIb, or Cel7B and Cel6B), an endoglucanase (EG VI), (not to be confused with CBH IIa, which was previously referred to as EG 6) a β -glucosidase (BGL), and a xylanase (Xyl II). The CBH IIb has a high activity against Avicel and cotton and displayed a pronounced synergism with other *C. lucknowense* cellulases. Using these new enzymes, this invention provides highly effective enzyme compositions for cellulose hydrolysis.

[0013] Disclosed is an enzyme formulation that includes at least one isolated cellobiohydrolase obtained from *C. lucknowense*. The isolated cellobiohydrolase may be either CBH Ib and IIb. The enzyme formulation may optionally contain an endoglucanase and/or a β -glucosidase. Furthermore, the enzyme formulation may optionally contain a hemicellulase.

[0014] Also disclosed is a method for producing glucose from cellulose. The method includes producing an enzyme formulation that contains at least one isolated cellobiohydrolase obtained from *C. lucknowense*, which can be CBH Ib or IIb. Optionally, the enzyme formulation may contain an endoglucanase and/or a β -glucosidase. The enzyme formulation is applied to cellulose to form glucose.

[0015] Yet another aspect of this disclosure is a method of producing ethanol. The method includes providing an enzyme formulation that contains at least one isolated cellobiohydrolase obtained from *C. lucknowense*, which can be CBH Ib or IIb. The enzyme formulation optionally may contain an endoglucanase and/or a β -glucosidase. Furthermore, the enzyme formulation may optionally contain a hemicellulase. The method further includes applying the enzyme formulation to cellulose to produce glucose and subsequently fermenting the glucose to produce ethanol.

[0016] Disclosed is also a method of producing energy from ethanol. The method includes providing an enzyme formulation that contains at least one isolated cellobiohydrolase obtained from *C. lucknowense*, which can be CBH Ib or IIb. The enzyme formulation optionally may contain an endoglucanase and/or a β -glucosidase. Furthermore, the enzyme formulation may optionally contain a hemicellulase. The method further includes applying the enzyme formulation to cellulose to produce glucose, fermenting the glucose to produce ethanol, and combusting said ethanol to produce energy.

[0017] Disclosed is a mutant *Chrysosporium lucknowense* strain capable of expressing at least one cellobiohydrolase and at least one endo-1,4- β -glucanase at higher levels than the corresponding non-mutant strain under the same conditions. The cellobiohydrolase is selected from the group consisting of CBH Ia, CBH IIa, CBH Ib, and CBH IIb; and the endo-1,4- β -glucanase is selected from the group consisting of EG II, EG V, and EG VI.

[0018] Disclosed are proteins exhibiting at least 65% amino acid identity as determined by the BLAST algorithm with the CBH Ib, CBH IIb, EG VI, BGL, and Xyl II amino acid sequences of SEQ ID NOs. 2, 4, 16, 12, and 18, respectively, or a part thereof having at least 20 contiguous amino acids. This invention also contemplates the corresponding nucleic acid sequences that encode such a protein.

[0019] One aspect of this invention provides an enzyme formulation comprising at least the enzyme BGL.

[0020] Another aspect of this invention provides a method of producing fermentable sugars from lignocellulosic material. The method comprises (a) providing an enzyme formulation comprising at least one enzyme the enzyme BGL; and (b) applying the enzyme formulation to lignocellulosic material to produce fermentable sugars.

[0021] The invention also provides a method of producing a fermentation product or a starting material for a fermentation product from a fermentable sugar. This method comprises (a) providing an enzyme formulation, wherein the enzyme formulation contains at least the enzyme BGL; (b) applying the enzyme formulation to lignocellulosic material to produce a fermentable sugar; and (c) fermenting said fermentable sugar to produce a fermentation product.

[0022] In another aspect, the invention provides a method of producing energy from a fermentable sugar. The method comprises (a) providing an enzyme formulation, wherein the enzyme formulation comprises at least the enzyme BGL; (b) applying the enzyme formulation to lignocellulosic material to produce a fermentable sugar; (c) fermenting the fermentable sugar to produce a combustible fermentation product; and (d) combusting said combustible fermentation product to produce energy.

[0023] Disclosed is a mutant *Chrysosporium lucknowense* strain capable of expressing at least one cellobiohydrolase and at least one endo-1,4- β -glucanase at higher levels than the corresponding non-mutant strain under the same con-

ditions. The cellobiohydrolase is selected from the group consisting of CBH Ia, CBH Ib, CBH **Ila** and CBH IIb; and the endo-1,4- β -glucanase is selected from the group consisting of EG II, EG V, and EG VI.

[0024] Also disclosed is a protein exhibiting at least 65% amino acid identity as determined by the BLAST algorithm with the CBH Ib, IIb, EG VI, BGL, Xyl II amino acid sequences as defined herein or a part thereof having at least 20 contiguous amino acids.

[0025] Another aspect of this disclosure is a nucleic acid sequence having at least 80% homology with the nucleic acid sequence encoding CBH Ib, CBH IIb, EG II, EG VI, BGL, or Xyl II, as defined herein.

[0026] The invention also provides a method for degrading a lignocellulosic material to fermentable sugars. The method includes contacting the lignocellulosic material with an effective amount of a multi-enzyme product derived from a microorganism, to produce at least one fermentable sugar. At least one enzyme in the multi-enzyme product is BGL.

[0027] In another aspect, disclosed is a microorganism or plant capable of expressing one or more of an enzyme selected from the group consisting of CBH Ia, CBH Ib, CBH **Ila**, CBH IIb, EG II, EG V, EG VI, BGL, and Xyl II.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028]

FIG. 1: SDS/PAGE (A) and isoelectrofocusing (B) of purified cellobiohydrolases from *C. lucknowense*. Lanes: 1, markers with different molecular masses; 2 and 5, CBH Ib; 3 and 6, CBH IIb; 4, markers with different *pI*.

FIG. 2: Progress kinetics of Avicel (5 mg ml⁻¹) hydrolysis by purified cellobiohydrolases (0.1 mg ml⁻¹) in the presence of purified *A. japonicus* BGL (0.5 U ml⁻¹), 40°C, pH 5.0.

FIG. 3: Synergism between CBH IIb and other *C. lucknowense* purified enzymes during hydrolysis of cotton cellulose (5 mg ml⁻¹) in the presence of purified *A. japonicus* BGL (0.5 U ml⁻¹), 40°C, pH 5.0. The CBH and EG concentration was 0.15 and 0.05 mg ml⁻¹, respectively. Experimental data for the pairs of enzymes are shown with open symbols (continuous curves); the theoretical sums of glucose concentrations obtained under the action of individual enzymes are shown with filled symbols (dotted lines).

FIG. 4: Progress kinetics of cotton (25 mg ml⁻¹) hydrolysis by combination #1 of purified *C. lucknowense* enzymes and NCE L-600, a commercial *C. lucknowense* multienzyme cellulase preparation at protein loading of 0.5 mg ml⁻¹, 50°C, pH 5.0 (see text and Table 4 for details).

FIG. 5: Progress kinetics of Avicel (50 mg ml⁻¹) hydrolysis by combination #1 of purified *C. lucknowense* enzymes and NCE-L, a commercial *C. lucknowense* multienzyme cellulase preparation at protein loading of 0.5 mg ml⁻¹, 50°C, pH 5.0 (see text and Table 4 for details).

FIG. 6: Progress kinetics of hydrolysis of pretreated Douglas fir wood (50 mg ml⁻¹) by combination #1 of purified *C. lucknowense* enzymes and NCE-L 600, a commercial *C. lucknowense* at protein loading of 0.5 mg ml⁻¹, 50°C, pH 5.0 (see text and Table 4 for details).

FIG. 7: Progress kinetics of hydrolysis of pretreated Douglas fir wood (50 mg ml⁻¹) by different combinations of purified *C. lucknowense* enzymes at protein loading of 0.5 mg ml⁻¹, 50°C, pH 5.0 (see text and Table 5 for details).

FIG. 8: *cbh2* gene encoding CBH IB.

FIG. 9: *cbh4* gene encoding CBH IIb

FIG. 10: *cbh1* gene encoding CBH Ia

FIG.11: EG6 gene encoding CBH **Ila**

FIG.12: *eg2* gene encoding EG II

FIG.13: *bgl1* gene encoding BCL

FIG.14: *eg5* gene encoding EGV

FIG. 15: eg7 gene encoding EG VI

FIG. 16: xyl2 gene encoding Xyl II

DETAILED DESCRIPTION OF THE INVENTION

[0029] The invention is defined by the appended claims.

[0030] The present invention provides methods and compositions for the conversion of plant biomass to fermentable sugars that can be converted to useful products. The methods include methods for degrading lignocellulosic material using -enzyme mixtures to liberate sugars. The compositions of the invention include enzyme combinations that break down lignocellulose. As used herein the terms "biomass" or "lignocellulosic material" includes materials containing cellulose and/or hemicellulose. Generally, these materials also contain xylan, lignin, protein, and carbohydrates, such as starch and sugar. Lignocellulose is generally found, for example, in the stems, leaves, hulls, 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, refers to simple sugars, such as glucose, xylose, arabinose, galactose, mannose, rhamnose, sucrose and fructose.

[0031] Biomass can include virgin biomass and/or non-virgin biomass such as agricultural biomass, commercial organics, construction and demolition debris, municipal solid waste, waste paper and yard waste. Common forms of biomass 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, wheat and barley (including wet milling and dry milling) as well as municipal solid waste, waste paper and yard waste. The biomass can also be, but is not limited to, herbaceous material, agricultural residues, forestry residues, municipal solid wastes, waste paper, and pulp 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 midlings, oat hulls, and hard and soft woods (not including woods with deleterious materials). In addition, agricultural biomass includes organic waste materials generated from agricultural processes including farming and forestry activities, specifically including forestry wood waste. Agricultural biomass may be any of the aforesaid singularly or in any combination or mixture thereof.

[0032] The fermentable sugars can be converted to useful value-added fermentation products, non-limiting examples of which include 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 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; industrial enzymes, such as proteases, cellulases, amylases, glucanases, lactases, lipases, lyases, oxidoreductases, transferases and xylanases; and chemical feedstocks.

[0033] As used herein, a multi-enzyme product can be obtained from or derived from a microbial, plant, or other source or combination thereof, and will contain enzymes capable of degrading lignocellulosic material. Examples of enzymes comprising the multi-enzyme products of the invention include cellulases (such as cellobiohydrolases, endoglucanase, β -glucosidases, hemicellulases (such as xylanases, including endoxylanases, exoxylanase, and β -xylosidase), ligninases, amylases, α -arabinofuranosidases, α -glucuronidases, α -glucuronidases, arabinases, glucuronidases, proteases, esterases (including ferulic acid esterase and acetylxytan esterase), lipases, glucomannanases, and xyloglucanases.

[0034] In some embodiments, the multi-enzyme product comprises a hemicellulase. Hemicellulose is a complex polymer, and its composition often varies widely from organism to organism, and from one tissue type to another. In general, a main component of hemicellulose is beta-1,4-linked xylose, a five carbon sugar. However, this xylose is often branched as beta-1,3 linkages, and can be substituted with linkages to arabinose, galactose, mannose, glucuronic acid, or by esterification to acetic acid. Hemicellulose can also contain glucan, which is a general term for beta-linked six carbon sugars. Those hemicelluloses include xyloglucan, glucomannan, and galactomannan.

[0035] The composition, nature of substitution, and degree of branching of hemicellulose is very different in dicotyledonous plants (dicots, i.e., plant whose seeds have two cotyledons or seed leaves such as lima beans, peanuts, almonds, peas, kidney beans) as compared to monocotyledonous plants (monocots; i.e., plants having a single cotyledon or seed leaf such as corn, wheat, rice, grasses, barley). In dicots, hemicellulose is comprised mainly of xyloglucans that are 1,4-beta-linked glucose chains with 1,6-beta-linked xylosyl side chains. In monocots, including most grain crops, the principal components of hemicellulose are heteroxylans. These are primarily comprised of 1,4-beta-linked xylose backbone polymers with 1,3-beta linkages to arabinose, galactose and mannose as well as xylose modified by ester-linked acetic acids. Also present are branched beta glucans comprised of 1,3- and 1,4-beta-linked glucosyl chains. In monocots,

cellulose, heteroxylans and beta glucans are present in roughly equal amounts, each comprising about 15-25% of the dry matter of cell walls.

[0036] Hemicellulolytic enzymes, i.e. hemicellulases, include both exohydrolytic and endohydrolytic enzymes, such as xylanase, β -xylosidase and esterases, which actively cleave hemicellulosic material through hydrolysis. These xylanase and esterase enzymes cleave the xylan and acetyl side chains of xylan and the remaining xylo-oligomers are unsubstituted and can thus be hydrolysed with β -xylosidase only. In addition, several less known side activities have been found in enzyme preparations which hydrolyse hemicellulose. While the multi-enzyme product may contain many types of enzymes, mixtures comprising enzymes that increase or enhance sugar release from biomass are preferred, including hemicellulases. In one embodiment, the hemicellulase is a xylanase, an arabinofuranosidase, an acetyl xylan esterase, a glucuronidase, an endo-galactanase, a mannanase, an endo arabinase, an exo arabinase, an exo-galactanase, a ferulic acid esterase, a galactomannanase, a xylogluconase, or mixtures of any of these. In particular, the enzymes can include glucoamylase, β -xylosidase and/or β -glucosidase. The enzymes of the multi-enzyme product can be provided by a variety of sources. In one embodiment, the enzymes can be produced by growing microorganisms or plants which produce the enzymes naturally or by virtue of being genetically modified to express the enzyme or enzymes. In another embodiment, at least one enzyme of the multi-enzyme product is commercially available.

[0037] One embodiment of the present invention relates to an isolated enzyme for catalyzing the conversion of ligno-cellulosic material to fermentable sugars as described herein. Also included in the invention are isolated nucleic acid molecules encoding any of such proteins. According to the present invention, an isolated protein or polypeptide is a protein that has been removed from its natural milieu (i.e., that has been subject to human manipulation) and can include purified proteins, partially purified proteins, recombinantly produced proteins, and synthetically produced proteins, for example. As such, "isolated" does not reflect the extent to which the protein has been purified. Preferably, an isolated protein of the present invention is produced recombinantly. An isolated peptide can be produced synthetically (e.g., chemically, such as by peptide synthesis) or recombinantly. An isolated protein can also be provided as a crude fermentation product, or a protein preparation that has been partially purified or purified (e.g., from a microorganism) using protein purification procedures known in the art. In addition, and solely by way of example, a protein referenced as being derived from or from a particular organism, such as a "*Chrysosporium lucknowense* cellulase and/or hemicellulase" refers to a cellulase and/or hemicellulase (generally including a homologue of a naturally occurring cellulose and/or hemicellulase) from a *Chrysosporium lucknowense* microorganism, or to a cellulase and/or hemicellulase that has been otherwise produced from the knowledge of the structure (e.g., sequence), and perhaps the function, of a naturally occurring cellulase and/or hemicellulase from *Chrysosporium lucknowense*. In other words, general reference to a *Chrysosporium lucknowense* cellulase and/or hemicellulase or a cellulase and/or hemicellulase derived from *Chrysosporium lucknowense* includes any cellulase and/or hemicellulase that has substantially similar structure and function of a naturally occurring cellulase and/or hemicellulase from *Chrysosporium lucknowense* or that is a biologically active (i.e., has biological activity) homologue of a naturally occurring cellulase and/or hemicellulase from *Chrysosporium lucknowense* as described in detail herein. As such, a *Chrysosporium lucknowense* cellulase and/or hemicellulase can include purified, partially purified, recombinant, mutated/modified and synthetic proteins. The same description applies to reference to other proteins or peptides described herein and to other microbial sources for such proteins or peptides.

[0038] One embodiment of the present invention relates to isolated nucleic acid molecules comprising, or consisting of nucleic acid sequences that encode a protein with the amino acid sequence of SEQ ID No 12.

[0039] In accordance with the present invention, an isolated nucleic acid molecule is a nucleic acid molecule that has been removed from its natural milieu (i.e., that has been subject to human manipulation), its natural milieu being the genome or chromosome in which the nucleic acid molecule is found in nature. As such, "isolated" does not necessarily reflect the extent to which the nucleic acid molecule has been purified, but indicates that the molecule does not include an entire genome or an entire chromosome in which the nucleic acid molecule is found in nature. An isolated nucleic acid molecule can include a gene. An isolated nucleic acid molecule that includes a gene is not a fragment of a chromosome that includes such gene, but rather includes the coding region and regulatory regions associated with the gene, but no additional genes that are naturally found on the same chromosome. An isolated nucleic acid molecule can also include a specified nucleic acid sequence flanked by (i.e., at the 5' and/or the 3' end of the sequence) additional nucleic acids that do not normally flank the specified nucleic acid sequence in nature (i.e., heterologous sequences). Isolated nucleic acid molecule can include DNA, RNA (e.g., mRNA), or derivatives of either DNA or RNA (e.g., cDNA). Preferably, an isolated nucleic acid molecule of the present invention is produced using recombinant DNA technology (e.g., polymerase chain reaction (PCR) amplification, cloning) or chemical synthesis. A nucleic acid molecule homologue can be produced using a number of methods known to those skilled in the art (see, for example, Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press (1989)). For example, nucleic acid molecules can be modified using a variety of techniques including, but not limited to, classic mutagenesis techniques and recombinant DNA techniques, such as site-directed mutagenesis, chemical treatment of a nucleic acid molecule to induce mutations, restriction enzyme cleavage of a nucleic acid fragment, ligation of nucleic acid fragments, PCR amplification and/or mutagenesis of selected regions of a nucleic acid sequence, synthesis of oligonucleotide mixtures and ligation of mixture

groups to "build" a mixture of nucleic acid molecules and combinations thereof. Nucleic acid molecule homologues can be selected from a mixture of modified nucleic acids by screening for the function of the protein encoded by the nucleic acid and/or by hybridization with a wild-type gene.

[0040] Disclosed is a recombinant nucleic acid molecule comprising a recombinant vector and a nucleic acid sequence encoding protein or peptide having at least one enzymatic activity useful for catalyzing the conversion of lignocellulosic material to fermentable sugars.

[0041] Disclosed is a recombinant vector is an engineered (i.e., artificially produced) nucleic acid molecule that is used as a tool for manipulating a nucleic acid sequence of choice and for introducing such a nucleic acid sequence into a host cell. The recombinant vector is therefore suitable for use in cloning, sequencing, and/or otherwise manipulating the nucleic acid sequence of choice, such as by expressing and/or delivering the nucleic acid sequence of choice into a host cell to form a recombinant cell. Such a vector typically contains heterologous nucleic acid sequences, that is nucleic acid sequences that are not naturally found adjacent to nucleic acid sequence to be cloned or delivered, although the vector can also contain regulatory nucleic acid sequences (e.g., promoters, untranslated regions) which are naturally found adjacent to nucleic acid molecules of the present invention or which are useful for expression of the nucleic acid molecules of the present invention (discussed in detail below). The vector can be either RNA or DNA, either prokaryotic or eukaryotic, and typically is a plasmid. The vector can be maintained as an extrachromosomal element (e.g., a plasmid) or it can be integrated into the chromosome of a recombinant organism (e.g., a microbe or a plant). The entire vector can remain in place within a host cell, or under certain conditions, the plasmid DNA can be deleted, leaving behind the nucleic acid molecule of the present invention. The integrated nucleic acid molecule can be under chromosomal promoter control, under native or plasmid promoter control, or under a combination of several promoter controls. Single or multiple copies of the nucleic acid molecule can be integrated into the chromosome. A recombinant vector can contain at least one selectable marker.

[0042] Typically, a recombinant nucleic acid molecule includes at least one nucleic acid molecule of the present invention operatively linked to one or more expression control sequences. According to the present invention, the phrase "operatively linked" refers to linking a nucleic acid molecule to an expression control sequence (e.g., a transcription control sequence and/or a translation control sequence) in a manner such that the molecule can be expressed when transfected (i.e., transformed, transduced, transfected, conjugated or conducted) into a host cell. Transcription control sequences are sequences that control the initiation, elongation, or termination of transcription. Particularly important transcription control sequences are those that control transcription initiation, such as promoter, enhancer, operator and repressor sequences.

[0043] Suitable transcription control sequences include any transcription control sequence that can function in a host cell or organism into which the recombinant nucleic acid molecule is to be introduced.

Enzymes and NucleicAcids Encoding the Enzymes

[0044] As described in the examples, disclosed are several purified enzymes, including two cellobiohydrolases, (CBH Ib, SEQ ID NO. 2; CBH IIb, SEQ ID NO. 4), an endoglucanase (EG VI, SEQ ID NO. 16), a β -glucosidase (BGL, SEQ ID NO. 12), and a xylanase (Xyl II, SEQ ID NO. 18). Disclosed are also variants of such enzymes, including variants having amino acid sequence with at least 65%, 70%, or 75% amino acid identity with these enzymes, as determined by the conventionally used BLAST algorithm.

[0045] Additionally, disclosed are the nucleic acids that encode these sequences, including gene *cbh2* (SEQ ID NO. 1, encoding CBH Ib), gene *cbh4* (SEQ ID NO. 3, encoding CBH IIb); gene *eg7* (SEQ ID NO. 15, encoding EG VI), gene *bgII* (SEQ ID NO. 11, encoding BGL), and gene *xyl2* (SEQ ID NO. 17, encoding Xyl II).

[0046] Disclosed are variants of these nucleic acids, including variants that have at least 80%, 85% or 90% homology with these nucleic acids.

[0047] As described herein, the newly identified and isolated enzymes can be used in conjunction with at least one other enzyme that promotes saccharification of cellulosic materials. This additional enzyme is derived from *C. lucknowense*. For example, the enzyme may be CBH Ia (SEQ ID NO. 6), CBH IIa (SEQ ID NO. 8), EG II (SEQ ID NO. 10) or EG V (SEQ ID NO. 14). Note however, CBH Ia, CBH IIa EG II, and EG V may be obtained by genetically modifying a microorganism or plant to express *cbh1* (SEQ ID NO. 5, encoding CBH Ia), EG6 (SEQ ID NO. 7, encoding CBH IIa), *eg2* (SEQ ID NO. 9, encoding EG II), and/or EG5 (SEQ ID NO. 13, encoding EG V). One particularly useful combination for saccharification is CBH Ia, CBH Ib, CBH IIb, EG II, EG V, BGL, and Xyl II.

[0048] The polynucleotides and polypeptides of the disclosure can be evolved using molecular evolution techniques to create and to identify novel variants with desired structural, functional, and/or physical characteristics. Molecular evolution techniques can be "DNA Shuffling", or "sexual PCR" (WPC, Stemmer, PNAS, 91:10747, (1994)), also referred to as "directed molecular evolution", "exon-shuffling", "directed enzyme evolution", "in vitro evolution" and "artificial evolution". Such reference terms are known in the art and are encompassed by the invention. Characteristics such as activity, the protein's enzyme kinetics, the protein's K_i , K_{cat} , K_m , V_{max} , K_d , thermostability, pH optimum, and the like can

be modified.

[0049] The polynucleotides and/or polypeptides of the disclosure may be evolved to confer properties that are advantageous for *in situ* enzymatic saccharification and fermentation. For example, enzymes may be evolved to perform optimally in an environment which is suitable for fermentation of sugars. In one example, the enzymes are evolved to have maximum activity in an environment with elevated temperature and high ambient alcohol content, such as an environment where an organism such as yeast is fermenting sugars. In this way, saccharification of lignocellulose and fermentation occurs in a single process step. In another example, the enzymes are evolved to resist harsh chemical or thermal environments, such as those that may be experienced during lignocellulosic pretreatments, as described herein. It is not necessary to chemically or thermally pretreat the lignocellulose prior to adding enzymes. Rather, the treatment and enzymatic saccharification can be performed simultaneously. Of course, disclosed are also processes involving multiple steps to produce sugars from lignocellulose, such as those where evolved enzymes first saccharify lignocellulose, which is subsequently fermented by an organism, such as yeast, for example.

[0050] The ability to enhance specific characteristics of a protein may also be applicable to changing the characterized activity of an enzyme to an activity completely unrelated to its initially characterized activity. Other desirable enhancements of the disclosure would be specific to each individual protein, and would thus be well known in the art and contemplated by the disclosure.

Expression of Enzymes

[0051] The microorganisms useful in the present invention and/or as a source of enzymes useful in the present invention include any microorganism producing an enzyme capable of degrading lignocellulosic material, including bacteria, yeast, and filamentous fungi. For simplicity and convenience, filamentous fungal microorganisms will be discussed herein; however, one skilled in the art will recognize that other microorganisms will be useful in the present invention. Filamentous fungi have been widely used in industry for the production of proteins. These fungi are uniquely adapted for the production and secretion of proteins owing to their biological niche as microbial scavengers. In environments rich in biological polymers, such as forest floors, the fungi compete by secreting enzymes that degrade those polymers, producing monomers that can be readily utilized as nutrients for growth. The natural ability of fungi to produce proteins has been widely exploited, mainly for the production of industrial enzymes. Levels of protein production in natural isolates can be increased in improved strains by orders-of-magnitude; production yields of tens of grams of protein per liter of fermentation culture are commonplace.

[0052] Fungal strains, including, but not limited to, various species of *Talaromyces*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Penicillium*, *Fusarium*, *Humicola*, *Myceliophthora*, *Corynascus*, *Chaetomium*, *Tolypocladium*, *Thielavia*, *Acremonium*, *Sporotrichum*, *Thermoascus*, and *Chrysosporium*, are contemplated in the present invention. These are a few of many possible genera of fungi that will be useful sources of enzymes and/or would be suitable as host organisms for producing such enzymes mixtures. Such fungi can be obtained, for instance from various depositories such as the American Type Culture Collection (ATCC), the All Russian Collection of Microorganisms of the Russian Academy of Sciences (VKM), and Centraalbureau voor Schimmelcultures.

Mutant Strains of *C. lucknowense*

[0053] Particular strains of *Chrysosporium* express proteins in extremely large amounts and natural expression regulating sequences from these strains are of particular interest. These strains have been designated as *Chrysosporium* strain C1, strain UV13-6, strain NG7C-19 and strain UV18-25. They have been deposited in accordance with the Budapest Treaty with the All Russian Collection (VKM) depository institute in Moscow. The wild type C1 strain was deposited in accordance with the Budapest Treaty with the number VKM F-3500 D, deposit date August 29, 1996, C1 UV13-6 mutant was deposited with number VKM F-3632 D, and deposit date February 9, 1998, C1 NG7c-19 mutant was deposited with number VKM F-3633 D and deposit date February 9, 1998 and C1 UV18-25 mutant was deposited with number VKM F-3631 D and deposit date February 9, 1998.

[0054] Preferably an expression-regulating region enabling high expression in the selected host is applied. This can also be a high expression-regulating region derived from a heterologous host, such as are well known in the art. Specific examples of proteins known to be expressed in large quantities and thus providing suitable expression regulating sequences for the invention are without being limited thereto hydrophobin, protease, amylase, xylanase, pectinase, esterase, beta-galactosidase, cellulase (e.g. endo-glucanase, cellobiohydrolase) and polygalacturonase. The high production has been ascertained in both solid state and submerged fermentation conditions. Assays for assessing the presence or production of such proteins are well known in the art.

[0055] Heterologous expression-regulating sequences also work efficiently in *Chrysosporium* as native *Chrysosporium* sequences. This allows well known constructs and vectors to be used in transformation of *Chrysosporium* as well as offering numerous other possibilities for constructing vectors enabling good rates of expression in this novel expression

and secretion host. As extremely high expression rates for cellulase have been ascertained for *Chrysosporium* strains, the expression regulating regions of such proteins are particularly preferred.

[0056] A nucleic acid construct comprising a nucleic acid expression regulatory region from *Chrysosporium lucknowense* or a derivative thereof is the mutant *Chrysosporium* strain comprising such regions operably linked to a gene encoding a polypeptide to be expressed.

[0057] Such a nucleic acid construct will be an expression regulatory region from *Chrysosporium* associated with cellobiohydrolase, endoglucanase, β -glucosidase, and/or xylanase expression.

[0058] The disclosure also covers genetically engineered *Chrysosporium* strains wherein the sequence that is introduced can be of *Chrysosporium* origin. Such a strain can, however, be distinguished from natively occurring strains by virtue of for example heterologous sequences being present in the nucleic acid sequence used to transform or transfect the *Chrysosporium*, by virtue of the fact that multiple copies of the sequence encoding the polypeptide of interest are present or by virtue of the fact that these are expressed in an amount exceeding that of the non-engineered strain under identical conditions or by virtue of the fact that expression occurs under normally non-expressing conditions. The latter can be the case if an inducible promoter regulates the sequence of interest contrary to the non-recombinant situation or if another factor induces the expression than is the case in the non-engineered strain. The invention as defined in the preceding embodiments is not intended to cover naturally occurring *Chrysosporium* strains. The disclosure is directed at strains derived through engineering either using classical genetic technologies or genetic engineering methodologies.

[0059] A method of production of a recombinant microorganism or plant is also disclosed. The method comprises stably introducing a nucleic acid sequence encoding a heterologous or homologous polypeptide into a microbial strain or plant, the nucleic acid sequence being operably linked to an expression regulating region. Such procedures are for transforming filamentous fungi have been previously reported. The mutant *Chrysosporium lucknowense* may be derived from UV18-25 (Acc. No. VKM F-3631 D) that has been engineered to overexpress the Xyl II gene.

Genetically Modified Organisms

[0060] As used herein, a genetically modified microorganism can include a genetically modified bacterium, yeast, fungus, or other microbe. Such a genetically modified microorganism has a genome which is modified (i.e., mutated or changed) from its normal (i.e., wild-type or naturally occurring) form such that a desired result is achieved (e.g., increased or modified activity and/or production of a least one enzyme or a multi-enzyme product for conversion of lignocellulosic material to fermentable sugars). Genetic modification of a microorganism can be accomplished by using classical strain development and/or molecular genetic techniques. Such techniques known in the art and are generally disclosed for microorganisms, for example, in Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Labs Press.

[0061] A genetically modified microorganism can include a microorganism in which nucleic acid molecules have been inserted, deleted or modified (i.e., mutated; e.g., by insertion, deletion, substitution, and/or inversion of nucleotides), in such a manner that such modifications provide the desired effect within the microorganism.

[0062] In one aspect of the disclosure a genetically modified microorganism can endogenously contain and express an enzyme or a multi-enzyme product for the conversion of lignocellulosic material to fermentable sugars, and the genetic modification can be a genetic modification of one or more of such endogenous enzymes, whereby the modification has some effect on the ability of the microorganism to convert lignocellulosic material to fermentable sugars.

[0063] In another aspect of the disclosure a genetically modified microorganism can endogenously contain and express an enzyme or a multi-enzyme product for the conversion of lignocellulosic material to fermentable sugars, and the genetic modification can be an introduction of at least one exogenous nucleic acid sequence (e.g., a recombinant nucleic acid molecule), wherein the exogenous nucleic acid sequence encodes at least one additional enzyme useful for the conversion of lignocellulosic material to fermentable sugars and/or a protein that improves the efficiency of the enzyme or multi-enzyme product for the conversion of lignocellulosic material to fermentable sugars. In this aspect of the disclosure, the microorganism can also have at least one modification to a gene or genes comprising its endogenous enzyme(s) for the conversion of lignocellulosic material to fermentable sugars.

[0064] In yet another aspect of the disclosure the genetically modified microorganism does not necessarily endogenously (naturally) contain an enzyme or a multi-enzyme product for the conversion of lignocellulosic material to fermentable sugars, but is genetically modified to introduce at least one recombinant nucleic acid molecule encoding at least one enzyme, a multiplicity of enzymes, or a multi-enzyme product for the conversion of lignocellulosic material to fermentable sugars. Such a microorganism can be used in a method of the disclosure or as a production microorganism for crude fermentation products, partially purified recombinant enzymes, and/or purified recombinant enzymes, any of which can then be used in a method of the present disclosure

Genetically Modified Plants

[0065] The disclosure also contemplates genetically modified plants comprising such genes. The plants may be used for production of the enzymes, or as the lignocellulosic material used as a substrate in the methods of the invention. Methods to generate recombinant plants are known in the art. For instance, numerous methods for plant transformation have been developed, including biological and physical transformation protocols. See, for example, Miki et al., "Procedures for Introducing Foreign DNA into Plants" in *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pp. 67 -88. In addition, vectors and in vitro culture methods for plant cell or tissue transformation and regeneration of plants are available. See, for example, Gruber et al., "Vectors for Plant Transformation" in *Methods in Plant Molecular Biology and Biotechnology*, Glick, B.R. and Thompson, J.E. Eds. (CRC Press, Inc., Boca Raton, 1993) pp. 89-119.

[0066] Disclosed are genetically modified plants that express the enzymes of this invention are obtained by introducing an expression vector into plants based on the natural transformation system of *Agrobacterium*. See, for example, Horsch et al., *Science*, 227:1229 (1985). *A. tumefaciens* and *A. rhizogenes* are plant pathogenic soil bacteria which genetically transform plant cells. The Ti and Ri plasmids of *A. tumefaciens* and *A. rhizogenes*, respectively, carry genes responsible for genetic transformation of the plant. See, for example, Kado, C.I., *Crit. Rev. Plant. Sci.* 10:1 (1991). Descriptions of *Agrobacterium* vector systems and methods for *Agrobacterium*-mediated gene transfer are provided by numerous references, including Gruber et al., *supra*, Miki et al., *supra*, Moloney et al., *Plant Cell Reports* 8:238 (1989), and U.S. Patents Nos. 4,940,838 and 5,464,763.

[0067] Genetically modified plants are obtained by microprojectile-mediated transformation wherein DNA is carried on the surface of microprojectiles. The expression vector is introduced into plant tissues with a biolistic device that accelerates the microprojectiles to speeds sufficient to penetrate plant cell walls and membranes. Sanford et al., *Part. Sci. Technol.* 5:27 (1987), Sanford, J.C., *Trends Biotech.* 6:299 (1988), Sanford, J.C., *Physiol. Plant* 79:206 (1990), Klein et al., *Biotechnology* 10:268 (1992).

[0068] Another method for physical delivery of DNA to plants is sonication of target cells. Zhang et al., *Bio Technology* 9:996 (1991). Alternatively, liposome or spheroplast fusion have been used to introduce expression vectors into plants. Deshayes et al., *EMBO J.*, 4:2731 (1985), Christou et al., *Proc Natl. Acad. Sci. USA* 84:3962 (1987). Direct uptake of DNA into protoplasts using CaCh precipitation, polyvinyl alcohol or poly-L-ornithine have also been reported. Hain et al., *Mol. Gen. Genet.* 199:161 (1985) and Draper et al., *Plant Cell Physiol.* 23:451 (1982). Electroporation of protoplasts and whole cells and tissues have also been described. Donn et al., In *Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC*, A2-38, p. 53 (1990); D'Halluin et al., *Plant Cell* 4:1495-1505 (1992) and Spencer et al., *Plant Mol. Biol.* 24:51-61 (1994).

Methods of Using the Enzymes and Mutant Strains of *C. lucknowense*

[0069] This invention also provides methods of enzymatic saccharification of cellulosic materials. Any cellulose containing material can be treated by the enzymes of this invention, non-limiting examples of which include orchard prunings, chaparral, mill waste, urban wood waste, yard waste, municipal waste, logging waste, forest thinning, short-rotation woody crops, industrial waste, wheat straw, oat straw, rice straw, barley straw, rye straw, flax straw, sugar cane, corn stover, corn stalks, corn cobs, corn husks, prairie grass, gamagrass, foxtail; sugar beet pulp, citrus fruit pulp, seed hulls, cellulosic animal wastes, lawn clippings, cotton, and seaweed.

[0070] In certain preferred embodiments, the lignocellulosic materials are pretreated before being exposed to the enzymes or enzyme mixtures of the invention. Generally speaking, the pretreatment can be any procedure that makes the subsequent enzymatic saccharification of the lignocellulosic materials more efficient (i.e., either less time-consuming or less costly). For example, the lignocellulosic material may be pretreated by methods including, but not limited to, exposure to acids, bases, solvents, heat, peroxides, ozone, or some combination thereof prior to enzymatic saccharification. These pretreatments can also be combined with other forms of processing, such as mechanical shredding, grinding, milling, or rapid depressurization (e.g. steam explosion).

[0071] Generally, enzymatic saccharification involves using CBH Ia, CBH Ib, EG VI, BGL, Xyl II, or mixtures thereof. One or more of these enzymes may be further combined with other enzymes capable of promoting enzymatic saccharification, which may be derived from *C. lucknowense*, a mutant strain, or another organism. The enzymatic saccharification involves an enzyme mixture comprising BGL. Disclosed is an enzymatic mixture containing cellobiohydrolase, which may be CBH Ia, CBH Ib, CBH IIa, CBH IIb, and mixtures thereof, with a β -glucosidase such as BGL.

[0072] In certain embodiments, the enzyme compositions are artificial enzyme compositions that contain purified forms of

[0073] BGL. The purified forms of this enzyme may be used alone or mixed together. In certain preferred embodiments, the purified enzyme is present in higher relative amounts than would be the case for the enzyme secretions of the wild type *C. lucknowense*.

[0074] In certain aspects, disclosed is a mutant strain of *C. lucknowense* that is capable of expressing CBH Ia, CBH Ib, CBH IIa, CBH IIb, EG II, EG V, EG VI, BGL, or Xyl II, or mixtures thereof in proportions higher than found in the enzyme secretions of the wild-type organism. The secreted enzymes of such a mutant strain of *C. lucknowense* may serve as a raw source from which purified forms of BGL can be produced. Alternatively, the secreted enzymes of such a mutant strain may also be applied directly to the cellulosic materials to be saccharified.

[0075] The cellulosic materials are exposed directly to the mutant strain of *C. lucknowense* in an environment conducive to the proliferation of the mutant strain of *C. lucknowense*, such as in a bioreactor. The in situ secretions of BGL, by the mutant strain of *C. lucknowense*, in proportions higher than found in the enzyme secretions of the wild-type organism, lead to enhanced in situ saccharification of the cellulosic material.

[0076] Following enzymatic treatment by the inventive enzymatic compositions of the invention, the fermentable sugar that is produced can be exposed to microorganisms, either naturally occurring or genetically engineered, that are capable of fermenting the sugar to produce ethanol or some other value-added fermentation product. Preferably, substantially all of the glucose is converted to ethanol, which may be subsequently used as a fuel, solvent, or chemical reactant. In preferred embodiments, the ethanol is used as a fuel for powering transportation vehicles, nonlimiting examples of which include cars, trucks, buses, mopeds and motorcycles. Other potential fermentation products from glucose include, but are not limited to, biofuels (including ethanol); 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; industrial enzymes, such as proteases, cellulases, amylases, glucanases, lactases, lipases, lyases, oxidoreductases, and transferases; and chemical feedstocks.

EXAMPLES

[0077] Examples not falling under the scope of the claims are for illustrative purposes only.

Example 1: Enzyme Isolation

[0078] Culture filtrates produced by the *C. lucknowense* mutant strains were used for isolation of individual enzymes. Commercial preparation of NCE-L600 (*C. lucknowense*) were from Dyadic International, Inc., USA.

[0079] Highly purified BGL (cellobiase) from *Aspergillus japonicus* was obtained from a commercial preparation, having specific cellobiase activity 50 U mg⁻¹ protein (pH 5.0, 40°C), and was used in the experiments on hydrolysis of insoluble cellulose.

Example 2: Enzyme purification

[0080] The enzyme purification was carried out by chromatography on a Pharmacia FPLC system (Sweden). Cellobiohydrolases and endoglucanases BGL and Xyl II were isolated from a *C. lucknowense* UV18-25 culture filtrate. BGL and Xyl II

[0081] (xylanase II) were isolated from culture filtrates produced by the *C. lucknowense* UV18ΔCbhI#10 and Xy12-18 mutant strains, respectively.

[0082] In all cases, the first purification stage was anion-exchange chromatography on a Source 15Q column (40 ml volume). The column was equilibrated with 0.02 M Bis-Tris-HCl buffer, pH 6.8. The initial culture filtrate was preliminarily desalted and transferred into the starting buffer by gel-filtration on Acrylex P4 (Reanal, Hungary). The sample (400 mg of protein) was applied to the Source 15Q column, and the elution was carried out with a gradient of 0-1 M NaCl at a flow rate of 10 ml min⁻¹.

[0083] The first protein fraction after the Source 15Q, eluted at 0.05 M NaCl and having high Avicelase activity, was subjected to hydrophobic interaction chromatography on a Source 15 Isopropyl column (Pharmacia, Sweden). The column was equilibrated with 1.7 M ammonium sulfate in 50 mM Na-acetate buffer, pH 5.0. Proteins were eluted with a reverse linear gradient of 1.7-0 M ammonium sulfate at a flow rate of 4 ml min⁻¹. The protein fraction with the highest activity against Avicel (eluting at a salt concentration of 0.30-0.35 M) contained the homogeneous protein with a molecular mass of 70 kDa (CBH IIb, see Fig. 1).

[0084] The protein fraction after the Source 15Q, eluted at 0.22 M NaCl and having the activity against Avicel and p-NP-β-D-cellobioside, was further purified by chromatofocusing on a Mono P HR 5/20 column (Pharmacia, Sweden). The column was equilibrated with 0.025 M Na-formate buffer, pH 4.0. Proteins were eluted with a gradient of pH 4.5-3.0 (using Polybuffer 74) at a flow rate of 0.5 ml min⁻¹. Homogeneous 60 kDa CBH Ib was obtained as a result of chromatofocusing (FIG. 1).

[0085] The two newly isolated cellobiohydrolases are homogeneous according to the data of SDS-PAGE and isoelectrofocusing (FIG. 1), their molecular masses were found to be 60 and 70 kDa, pI 3.8 and 5.6, respectively. Peptide mass fingerprinting using MALDI-TOF mass spectrometry (data not shown) indicated that these proteins were different

from the above-mentioned cellobiohydrolases (Cel6A and Cel7A) as well as from other *C. lucknowense* enzymes previously isolated. Subsequent *de novo* sequencing of tryptic peptides from the new cellobiohydrolases, using tandem TOF/TOF mass spectrometry (MS/MS), followed by the BLAST search in the SWISS-PROT (UniProtKB) database showed that the 60 kDa and 70 kDa proteins display sequence similarity to cellobiohydrolases from the GH families 7 and 6 (Table 1, see classification into families in HyperTextTransferProtocol://afmb.cnrs-mrs.fr/CAZY/). So, they were classified as Cel7B (CBH Ib) and Cel6B (CBH IIb), respectively. Thus, the *C. lucknowense* fungus secretes at least four cellobiohydrolases encoded by different genes, two of them belonging to the glycosyl hydrolase family 6 (GH6) and two other enzymes - to the GH7 family (Table 2). The molecules of the CBH Ia (Cel7A) and CBH IIb (Cel6B) represent typical cellulases consisting of a catalytic domain and CBM connected by a flexible peptide linker. The molecules of CBH Ib (Cel7B) and CBH IIa (Cel6A) consist of only the catalytic domains (they lack CBM). It should be noted that the most studied fungus *T. reesei* has only two cellobiohydrolases: I (Cel7A) and II (Cel6A). Other fungi, such as *Humicola insolens*, also secrete two cellobiohydrolases (Cel7A and Cel6A), while *Phanerochaete chrysosporium* produces at least seven different cellobiohydrolases, of which six enzymes belong to the GH7 family. All the enzymes mentioned, except for the *P. chrysosporium* CBH 1-1 (Cel7A), possess CBM.

[0086] The BGL was isolated from the protein fraction after the Source 15Q (eluted at 0.10 M NaCl) containing the highest activity against *p*-NP- β -D-glucopyranoside and cellobiose. The fraction was subjected to hydrophobic interaction chromatography as described above, the homogeneous BGL with a molecular mass of 106 kDa and *pI* 4.8 was eluted at 1.3 M of ammonium sulfate. The specific activity of the BGL toward *p*-NP- β -D-glucopyranoside and cellobiose was found to be 11 and 26 U mg⁻¹ of protein, respectively (40°C, pH 5.0). Purified BGL had optimum activity at pH 4.0 and retained >50% of activity in the range of pH 2.5-6.5. The temperature optimum was 40°C. After heating for three hours, the enzyme retained 10% activity at 60°C, 64% at 50°C, and 100% at 40°C. The enzyme was highly active against cellobiose, gentiobiose, and laminarobiose as substrates. Weak activity was also observed using sophorose, cellotriose, cellotetraose, cellopentaose, and cellohexaose as substrates. No activity was observed with lactose or trehalose as substrates.

[0087] The homogeneous Xyl II (24 kDa, *pI* 7.9) was obtained after anion-exchange chromatography followed by hydrophobic interaction chromatography as described above and gel-filtration on a Superose 12 HR 10/30 column (Pharmacia, Sweden). Elution at the last chromatographic stage was performed with 0.1 M Na-acetate buffer, pH 5.0, at a flow rate of 0.3 ml min⁻¹. The Xyl II had specific xylanase activity of 395 U mg⁻¹ of protein (50°C, pH 5.0, birchwood xylan as a substrate). The enzyme had a pH optimum of 6.0 and a temperature optimum of 70°C. Xyl II was highly specific for xylan as substrate, with no activity against carboxymethylcellulose (CMC) or barley β -glucan.

[0088] The *C. lucknowense* CBH Ia (65 kDa), CBH IIa (43 kDa), EG n (51 kDa), EG V (25 kDa), EG VI (47 kDa) were purified as described elsewhere (see, Gusakov AV, Sinitsyn AP, Salanovich TN, Bukhtojarov FE, Markov AV, Ustinov BB, van Zeijl C, Punt P, Burlingame R. "Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*" Enzyme Microb Technol 2005;36:57-69; Bukhtojarov FE, Ustinov BB, Salanovich TN, Antonov AI, Gusakov AV, Okunev ON, Sinitsyn AP. "Cellulase complex of the fungus *Chrysosporium lucknowense*: isolation and characterization of endoglucanases and cellobiohydrolases", Biochemistry (Moscow) 2004;69:542-51.

[0089] The enzyme purity was characterized by SDS-PAGE and isoelectrofocusing. SDS-PAGE was carried out in 12% gel using a Mini Protean II equipment (Bio-Rad Laboratories, USA). Isoelectrofocusing was performed on a Model 111 Mini IEF Cell (Bio-Rad Laboratories, USA). Staining of protein was carried out with Coomassie Blue.

Example 3: MALDI-TOF and tandem TOF/TOF mass spectrometry of peptides

[0090] The in-gel tryptic digestion of the protein bands after the SDS-PAGE was carried out essentially as described by Smith (Smith BE. Protein sequencing protocols. Totowa: Humana Press; 1997). Trypsin (Promega, modified, 5 μ g/mL) in 50 mM NH₄HCO₃ was used for a protein digestion. The resulting peptides were extracted from a gel with 20% aqueous acetonitrile containing 0.1% trifluoroacetic acid and subjected to MALDI-TOF MS (see, James P. (Ed.) Proteome research: mass spectrometry. Berlin: Springer-Verlag; 2001.) Selected peptides from the mass spectra of the tryptic digests of the CBH Ib and IIb were analyzed by tandem mass spectrometry in order to determine their sequences *de novo*. Ultraflex TOF/TOF mass spectrometer (Bruker Daltonik GmbH, Germany) was used in the MS experiments.

Example 4: Enzyme activity assays

[0091] CMCase activity was measured by assaying reducing sugars released after 5 min of enzyme reaction with 0.5% carboxymethylcellulose (CMC, medium viscosity, Sigma, USA) at pH 5.0 and 50°C (Sinitsyn AP, Chemoglazov VM, Gusakov AV. "Methods of investigation and properties of cellulolytic enzymes" (in Russian), Biotechnology Series, v. 25. Moscow: VINITI Press; 1990). Enzyme activities against barley β -glucan (Megazyme, Australia) and birchwood xylan (Sigma, USA) were determined in the same way as the CMCase activity, except the incubation time was 10 min.

Avicelase activity was determined by analysing reducing sugars released after 60 min of enzyme reaction with 5 mg ml⁻¹ Avicel PH 105 (Serva, Germany) at pH 5.0 and 40°C. Reducing sugars were analysed by the Somogyi-Nelson method (Sinitsyn AP, Chemoglazov VM, Gusakov AV, "Methods of investigation and properties of cellulolytic enzymes" (in Russian); Biotechnology Series, v. 25. Moscow: VINITI Press; 1990; Somogyi M., "Notes on sugar determination" J Biol Chem 1952;195:19-23.. Filter paper activity (FPA) was determined as recommended by Ghose (Ghose TK. "Measurement of cellulase activities", Pure Appl Chem 1987;59:257-68).

[0092] Activities against *p*-NP-β-D-glucopyranoside, *p*-NP-β-D-cellobioside and *p*-NP-β-D-lactoside (Sigma, USA) were determined at pH 5.0 and 40°C as described elsewhere (Gusakov AV, Sinitsyn AP, Salanovich TN, Bukhtjarov FE, Markov AV, Ustinov BB, van Zeijl C, Punt P, Burlingame R. "Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*", Enzyme Microb Technol 2005;36:57-69).

[0093] Cellobiase activity was assayed at pH 5.0 and 40°C by measuring the initial rate of glucose release from 2 mM cellobiose by the glucose oxidase - peroxidase method (Sinitsyn AP, Chernoglazov VM, Gusakov AV, "Methods of investigation and properties of cellulolytic enzymes" (in Russian), Biotechnology Series, v. 25. Moscow: VINITI Press; 1990)..

[0094] All activities were expressed in International Units, i. e. one unit of activity corresponded to the quantity of enzyme hydrolysing one μmol of substrate or releasing one μmol of reducing sugars (in glucose equivalents) per one minute.

Example 5: Enzymatic hydrolysis of cellulosic substrates

[0095] The enzymatic hydrolysis of cellulosic substrates was carried out at pH 5.0 under magnetic stirring. Avicel PH 105 (Serva, Germany), cotton pretreated with acetone-ethanol mixture (1:1) for two days in order to remove wax from the surface of cellulose fibres, and Douglas fir wood pretreated by organosolv were used as substrates.

[0096] The experiments on progress kinetics of Avicel hydrolysis by purified individual cellobiohydrolases and experiments on synergistic interaction between *C. lucknowense* cellulases (with cotton as a substrate) were carried out at 40°C. The substrate concentration in those experiments was 5 mg ml⁻¹. In order to eliminate the effect of product (cellobiose) inhibition on the kinetics and to convert all celooligosaccharides to glucose, the hydrolysis was carried out in the presence of purified BGL (cellobiase) from *A. japonicus*, which was extra added to the reaction system in excessive quantity (0.5 U ml⁻¹).

[0097] The experiments on enzymatic saccharification of Avicel, cotton, and pretreated Douglas fir wood by combinations of purified *C. lucknowense* enzymes and crude multienzyme preparations were carried out at 50°C. The concentration of Avicel and pretreated wood in those experiments was 50 mg ml⁻¹, while the concentration of cotton was 25 mg ml⁻¹.

[0098] A typical experiment was carried out in the following way. A weighed amount of dry cellulosic substrate was placed into a 2-ml plastic test tube, then 0.5-1 ml of 0.05 M Na-acetate buffer, containing 1 mM NaN₃ to prevent microbial contamination, was added, and the substrate was soaked in the buffer for 1 h. Then, the tube was placed into a thermostated water bath, located on a magnetic stirrer, and suitably diluted enzyme solution in the same buffer was added to the substrate suspension in order to adjust the total volume of the reaction system to 2 ml and to start the hydrolysis. The tube was hermetically closed with a lid, and the hydrolysis was carried out with magnetic stirring. At defined times in the reaction, an aliquot of the suspension (0.05-0.1 ml) was taken, diluted, centrifuged for 3 min at 15000 rpm, and the concentrations of glucose and reducing sugars in the supernatant were determined by the glucose oxidase - peroxidase and Somogyi-Nelson methods. In those cases, when glucose was a single product of the reaction, the degree of substrate conversion (for Avicel and cotton, which represented pure cellulosic substrates) was calculated using the following equation:

$$\text{Conversion (\%)} = \frac{\text{Glucose concentration (mg ml}^{-1}\text{)} \times 100\%}{\text{Initial substrate concentration (mg ml}^{-1}\text{)} \times 1.11}$$

[0099] The kinetic experiments were carried out in duplicates. Protein concentration was the measure of enzyme loading in the reaction system. In the case of purified enzymes, the protein concentration was calculated from the UV absorption at 280 nm using enzyme extinction coefficients predicted by the ProtParam tool (<http://WorldWideWeb.expasy.ch/tools/protparam.HyperTextMarkupLanguage>). For crude multienzyme preparations, the protein concentration was determined by the Lowry method using bovine serum albumin as a standard.

[0100] The CBH Ib and IIb displayed maximum activity at pH 4.7 and 5.0. Both enzymes were stable during 24 h

incubation at pH 5.0 and 50°C. Study of the enzyme adsorption on Avicel, carried out at pH 5.0 and 6°C, revealed that only the CBH IIb has CBM. After incubation of the CBH Ib and IIb (1 mg ml⁻¹) with Avicel (25 mg ml⁻¹) for 30 min on stirring the degree of protein adsorption was 65 and 99%, respectively. It should be noted that the adsorption degree of the catalytic domain of the *C. lucknowense* CBH Ia was 59% under the same conditions, while that for the full size *C. lucknowense* CBH Ia (an enzyme with CBM) was 89%.

[0101] The CBH IIb had a high activity against Avicel and very low CMCase activity, while the activity toward synthetic *p*-nitrophenyl derivatives of disaccharides was completely absent (Table 2). The CBH Ib displayed lower Avicelase activity, but hydrolysed *p*-NP-β-D-cellobioside and *p*-NP-β-D-lactoside, which is typical for family 7 cellulases. For a comparison, specific activities of previously isolated *C. lucknowense* cellobiohydrolases (now named as CBH Ia and CBH IIa) are also given in Table 2.

[0102] FIG. 2 shows the progress kinetics of Avicel hydrolysis by the all purified *C. lucknowense* cellobiohydrolases, where the enzymes were equalized by protein concentration (0.1 mg ml⁻¹). In order to eliminate the effect of product (cellobiose) inhibition on the kinetics, the hydrolysis was carried out in the presence of purified BGL (cellobiase) from *A. japonicus*, added to the reaction system in excessive quantity (0.5 U ml⁻¹).

[0103] The highest hydrolysis rate amongst a few cellobiohydrolases tested, including three other *C. lucknowense* enzymes (CBH Ia, Ib, IIa) was observed in the case of *C. lucknowense* CBH IIb: 3.2 mg ml⁻¹ of glucose, i.e. 58% cellulose conversion was achieved after 5 days of hydrolysis (see FIG. 2). The *C. lucknowense* CBH Ia (which has a CBM) was notably less effective (the yield of glucose after 5 days was 2.5 mg ml⁻¹, which corresponded to the cellulose conversion degree of 46%, respectively). As expected, the *C. lucknowense* cellobiohydrolases without CBM (CBH Ib and IIa) had the lowest ability to hydrolyse Avicel: only 23 and 21% cellulose conversion was achieved after the same time of reaction.

[0104] Both *C. lucknowense* cellobiohydrolases having a CBM (Ia and IIb) displayed a pronounced synergism with three major endoglucanases from the same fungus (EG II, EG V, EG VI) in hydrolysis of cotton as well as a strong synergy with each other (Table 3). In these studies, the concentration of cotton was 5 mg ml⁻¹, the CBH concentration was 0.15 mg ml⁻¹ in all cases, while the EG concentration was always 0.05 mg ml⁻¹. In order to eliminate the effect of product inhibition on the kinetics and to convert the intermediate oligosaccharides to glucose, the hydrolysis was carried out in the presence of purified BGL from *A. japonicus*, added to the reaction system in excessive quantity (0.5 U ml⁻¹). The experiments were carried out at pH 5.0 and 40°C for 140 h.

[0105] As seen from Table 3, individual cellobiohydrolases, CBH Ia and CBH IIb, and the individual endoglucanases, did not completely hydrolyze cotton under the conditions tested. The CBH IIb provided the highest glucose yield after 140 h of hydrolysis: 1.18 mg ml⁻¹, which corresponded to the substrate conversion degree of 21%. However, when either cellobiohydrolase was incubated with endoglucanase, a pronounced synergism was observed. The highest glucose yields (4.1-4.7 mg ml⁻¹) were achieved with combinations of CBH Ia or CBH IIb with EG II, the coefficient of synergism being varied in the range of 2.6-2.8. A strong synergism ($K_{syn} = 2.75$) was also observed between CBH Ia and CBH IIb. In fact, the combination of two cellobiohydrolases (1:1 by weight) with BGL provided practically complete conversion (98.6%) of cotton cellulose to glucose after 140 h of hydrolysis.

[0106] As an example, the progress kinetics of cotton hydrolysis by combinations of CBH IIb with other *C. lucknowense* enzymes are shown in FIG. 3, where real experimental data are shown with open symbols (continuous curves) while the theoretical sums of glucose concentrations obtained under the action of individual enzymes are shown with filled symbols (dotted lines). Glucose yields obtained after 140 h of cotton hydrolysis under the action of individual cellobiohydrolases and endoglucanases and their combinations are summarized in Table 3. The coefficient of synergism (K_{syn}) was calculated as a ratio of experimental glucose concentration (column 2 of Table 3) to the theoretical sum of glucose concentrations (column 3).

[0107] Using four purified *C. lucknowense* enzymes (CBH Ia and IIb, EG II, BGL), an artificial cellulase complex was constructed (*C.l.* combination #1) that demonstrated an extremely high ability to convert different cellulosic substrates to glucose (FIGS. 4-6). This multienzyme composition was notably more effective in hydrolysis of pure crystalline cellulose (cotton and Avicel) than the crude *C. lucknowense* multienzyme preparation NCE-L600. In 72-h hydrolysis of a ligno-cellulosic substrate (Douglas fir wood pretreated by organosolv), the *C.l.* combination #1 was also very effective in cellulose hydrolysis.

[0108] In *C. lucknowense* combination #1, the enzyme consisted of the two cellobiohydrolases CBH Ia and CBH Ib, and the endoglucanase EG II, the enzymes with strong adsorption ability on crystalline cellulose (the molecules of these enzymes have CBM). The activity of tightly adsorbed cellulases is gradually decreased during in the course of hydrolysis of insoluble cellulose as a result of the enzyme limited mobility along the substrate surface or unproductive binding (so called pseudoinactivation). Without wishing to be bound by theory, it is believed that there may exist a synergism between tightly and loosely adsorbed cellulases wherein loosely binding cellulases (enzymes without CBM) may destroy obstacles hindering the processive action of the tightly adsorbed cellobiohydrolases, thus helping them to move to the next cellulose reactive sites. The total protein concentration in the reaction system was 0.5 mg ml⁻¹. The composition of the multienzyme composition (*C.l.* combination #1) was the following: 0.2 mg ml⁻¹ of CBH Ia + 0.2 mg ml⁻¹ of CBH Ib + 0.08 mg ml⁻¹ of EG II + 0.02 mg ml⁻¹ of BGL. Avicel (50 mg ml⁻¹) and cotton (25 mg ml⁻¹) were used as substrates representing pure

crystalline cellulose in these experiments. Sample of Douglas fir wood pretreated by organosolv (50 mg ml⁻¹) was taken as an example of real lignocellulosic feedstock that may be used for bioconversion to ethanol. A crude *C. lucknowense* multienzyme cellulase preparation NCE L-600 (diluted so that the protein concentration in the reaction system would also be 0.5 mg ml⁻¹) was taken for a comparison in these studies. The hydrolysis experiments with them were carried out also in the presence of extra added *A. japonicus* BGL (0.5 U ml⁻¹).

[0109] The progress kinetics of cotton, Avicel and Douglas fir hydrolysis by different cellulase multienzyme preparations are shown in **FIGS. 4-6**. It should be noted that in all cases, the concentrations of glucose and reducing sugars after 24-72 h of hydrolysis in a concrete experiment were practically the same, i.e. glucose made up >96% of the total soluble sugars. So, the glucose yield can be taken as reliable criterion in comparison of the hydrolytic efficiency of different multienzyme samples.

[0110] In hydrolysis of cotton (**FIG. 4**), the combination #1 of purified *C. lucknowense* enzymes provided much higher glucose yield after 72 h of the reaction (23.4 mg ml⁻¹, i.e. 84% degree of substrate conversion) than the 4.2 mg ml⁻¹ exhibited by (NCE-L600). In hydrolysis of Avicel (**FIG. 5**), the *C.l.* combination #1 was also superior (45.0 mg ml⁻¹ of glucose, or 81% substrate conversion after 72 h of hydrolysis). In the case of pretreated Douglas fir (**FIG. 6**), the *C.l.* combination #1 was also effective (28.8 mg ml⁻¹ glucose, 63% conversion after 72 hours).

[0111] Unlike Avicel and cotton, the pretreated wood sample contained not only cellulose (~85%) but also lignin (13%) and hemicellulose (2%). The artificial *C. lucknowense* four-enzyme combination #1 was composed of only cellulases; all of them, except for the BGL, having CBM. All other multienzyme samples possessed not only cellulase but also xylanase and other types of carbohydrase activity, i.e. they contained non-cellulase accessory enzymes. This may explain relatively lower efficiency of the *C.l.* combination #1 on pretreated Douglas fir compared to the *P. verruculosum* #151 preparation (**FIG. 6**).

[0112] In one set of experiments (**FIG. 7**), the pretreated wood sample was hydrolysed by different compositions of purified *C. lucknowense* enzymes, to which cellulases lacking a CBM were included (EG V or EG V in combination with CBH Ib). The total protein concentration in the reaction system was maintained at the same level of 0.5 mg ml⁻¹ (Table 5). Indeed, two *C.l.* combinations (#3 and #4), containing weakly adsorbed enzymes, provided a notable enhancement of the glucose yield after 72 h of the enzymatic reaction in comparison with the *C.l.* combination #1.

[0113] In two experiments, the highly active *C. lucknowense* Xyl II (Xyn11A) was added to the above-mentioned four enzymes (*C.l.* combinations #2 and #4). Since a synergism between tightly and loosely adsorbed cellulases has been described [38], EG V or EG V together with CBH Ib (both enzymes have lack CBM) were used in the *C.l.* combinations #3 and #4.

[0114] As can be seen from **FIG. 7**, the initial rate of glucose formation decreased sequentially from *C.l.* combination #1 to combination #4, however the glucose yield after 2-3 days of hydrolysis increased in the same sequence. The Xyl II demonstrated only slight positive effect on the glucose yield, while the EG V or EG V together with CBH Ib provided a very notable increase in the product concentration after 72 h hydrolysis of wood (37 and 41 mg ml⁻¹, respectively) compared to the *C.l.* combination #1 (29 mg ml⁻¹), i.e. the combinations #3 and #4 performed much better than all crude multienzyme samples (**FIG. 6**).

[0115] The low performance of the crude *C. lucknowense* preparation (NCE-L600) in hydrolysis of different cellulosic substrates (**FIGS. 4-6**) deserves a special attention. Without wishing to be bound by theory, it may be explained by the low total content of different cellobiohydrolases in the NCE-L600 (35-40% of the total protein content). Moreover, two of four *C. lucknowense* cellobiohydrolases (Ib and IIa) lack CBM, while two other enzymes (CBH Ia and IIb) also partially lose the CBM during the course of fermentation. The CBM absence in major part of cellobiohydrolases from the NCE-L600 may lead to the lower activity of the crude preparation toward crystalline cellulose.

Table 1

Identification of peptides in the isolated *C. lucknowense* proteins using MALDI-TOF MS/MS

Enzyme	m/z	Peptide	BLAST identification ^b	UniProtKB No.
Protein 60 kDa	1133. 6	HEYGTNIGSR	118 HEYGTNIGSR 127 (cbh1.2 <i>Humicola grisea</i> -GH7)	094093
	1829. 9	MGNQDFYGPGLTVDT K	291 LGNTDFYGPGLTVDT 305 (cbhB <i>Aspergillus niger</i> -GH7)	Q9UVS8
	1061. 4	YPANDYYR	127 ANYYR 132 (Avicelase 2 <i>Humicola insolens</i> -GH6)	Q9C1S9
Protein 70 kDa	1990. 0	HYIEAFSPLLNSAGFPAR	367 KYIEAFSPLLNAAGFPA 383 (CBH II <i>Neurospora crassa</i> - GH6)	Q872J7

EP 2 041 294 B1

(continued)

Identification of peptides in the isolated *C. lucknowense* proteins using MALDI-TOF MS/MS

Enzyme	m/z	Peptide	BLAST identification ^b	UniProtKB No.
	2073.5	LWQPTGQQQWGDWCN VK	381 QPTGQQQWGDWCNV 394 (CBH II <i>T. reesei</i> - GH6)	P07987

^a Since the MS/MS can not distinguish between Leu and Ile residues (they have the same masses), there may be ambiguity in the appropriate positions of the identified peptides.

^bResidues conserved in the *C. lucknowense* enzymes are shown in bold.

Table 2

Specific activities (U mg⁻¹ of protein) of purified cellobiohydrolases from *C. lucknowense* toward different substrates at pH 5.0 and 40°C

Enzyme	Mol. mass (kDa)	Cat. domain designation	CBM presence	Avicel	CMC ^a	Barley β -glucan ^a	p-NP-P-D-cellobioside	p-NP- β -D-lactoside
CBH Ia	65	Cel7A	Yes	0.21	0.1	<0.1	0.021	0.12
CBH Ib	60	Cel7B	No	0.12	0.3	<0.1	0.020	0.09
CBH IIa	43	Cel6A	No	0.08	1.1	2.0	0	0
CBH IIb	70	Cel6B	Yes	0.22	0.2	0.2	0	0

^a Activity was determined at 50°C.

Table 3

Synergism between *C. lucknowense* cellulases in hydrolysis of cotton cellulose (5 mg ml⁻¹) at pH 5.0 and 40°C in the presence of 0.5 U ml⁻¹ of *A. japonicus* BGL. In all cases the CBH concentration was 0.15 mg ml⁻¹, the EG concentration was 0.05 mg ml⁻¹.

Enzyme	Glucose concentration after 140 h, experimental (mg ml ⁻¹)	Glucose concentration after 140 h, theoretical ^a (mg ml ⁻¹)	K_{syn}
CBH Ia	0.81	-	-
CBH IIb	1.18	-	-
EG II	0.64	-	-
EG V	0.70	-	-
EG VI	0.40	-	-
CBH Ia + EG II	4.05	1.45	2.79
CBH Ia + EG V	3.68	1.51	2.44
CBH Ia + EG VI	3.93	1.21	3.25
CBH IIb + EG n	4.72	1.82	2.59
CBH IIb + EG V	3.81	1.88	2.03
CBH IIb + EGVI	4.05	1.58	2.56
CBH Ia + CBH IIb	5.47	1.99	2.75

^a Calculated as a sum of glucose concentrations obtained under the action of individual enzymes.

Table 4

Specific activities (U mg⁻¹ of protein) of multienzyme preparations toward different substrates at pH 5.0 and 50°C

Preparation	Protein (mg ml ⁻¹ or mg g ⁻¹)	Filter paper	CMC	Xylan	Cellobiose ^a
NCE-L600	45	0.25	12.2	4.8	0.07
<i>C.l.</i> combination #1	1000	1.10	6.6	0	1.05

^a Activity was determined at 40°C.

Table 5

Composition of artificial multienzyme combinations based on purified *C. lucknowense* enzymes and yields of glucose after 72-h hydrolysis of pretreated Douglas fir wood (50 mg ml⁻¹), pH 5.0, 50°C. The total protein concentration in the reaction system was 0.5 mg ml⁻¹, the concentration of each component and glucose yields are given in mg ml⁻¹.

Combination	CBH Ia	CBH Ib	CBH IIb	EG II	EG V	BGL	Xyl II	Glucose yield
#1	0.2	0	0.2	0.08	0	0.02	0	28.8
#2	0.2	0	0.2	0.07	0	0.02	0.01	30.1
#3	0.2	0	0.2	0.04	0.04	0.02	0	37.3
#4	0.1	0.1	0.2	0.03	0.04	0.02	0.01	41.0

SEQUENCE LISTING

[0116]

<110> Dyadic International, Inc.
GUSAKOV, Alexander V.
SALANOVICH, Tatyana N.
ANTONOV, Alexey I.
USTINOV, Boris B.
OKUNEV, Oleg N.

BURLINGAME, Richard P.
 EMALFARB, Mark A.
 BAEZ, Marco
 SINITSYN, Arkady P.

<120> CONSTRUCTION OF HIGHLY EFFICIENT CELLULASE COMPOSITIONS FOR ENZYMATIC HYDROLYSIS OF CELLULOSE

<130> 3123-4008PC

<140> TBD

<141> 2006-07-13

<160> 18

<170> PatentIn version 3.3

<210> 1

<211> 6360

<212> DNA

<213> Chrysosporium lucknowense

<400> 1

```

ctcagattct aggggtaggg cgggagcaga ggcgaaaatt gggttgtaga atatgaggag      60
ctagggttgt taaactcaaa gaacttcttg ctcttggtct tagtcttctc tcctgggaaa      120
aggggggttt tccgaaagcg gcgctatacg aagccagagg ctactttcct tgctttggat      180
ggcccttgtc caccgttctt gtttcccgtt tgtcaattgc gacgttgccg gcaacctagg      240
tcctaataat taggtagata tttcggtaga ggtagtttaa ttatgcttca gtagagaaat      300
cgttgtctcc acgtctcgca accttgcgaa acttcgccac attgaagata gcattgtctg      360
agttgatttt aaccctttcc agagacgata taatagtgcg agtttctttg atcggaatca      420
tcgacattcg gattttccct taattatatg aagtattcgg cccacggaac cgggccccga      480
gcagggttgaa ccgcgcaaaa cctcaaccga gtcacctcgc gtccatgttt gtcattggaat      540
caggctccga atcccgtcag atcagtcagt tctggtggct atggacgcgg gagttacggc      600
cagtcgtccc gttgttctg ggggttgatc aacaggagga agagatctga gatcgaacta      660
cacccattga tttatcgacg cataatcaag ttttaataaaa accaaacagc gtgtttgggtg      720
ctaccaccga atgcgagatc cgggctagcc cgcggaagga tgatggccac agatctagcg      780
tcatgtatga ttattaccta tgcattctatc ttcgtatctg cctcgggttg gcaacacctg      840
accgagagac gactcgacaa cctgacactt ggcaaaagac atttcggttg acagcgggag      900
aactccagcg aggaagtcgc ccagagatgc ggatgagaag acaacgccga gacgtgccgg      960

```

	cgttggctct	ccacgaatcg	gagccgactc	ttccgtttgg	ccaatctccg	ggataaatcc	1020
	cagcggcggg	tcacgtcacg	tttcatgggg	aggcgcggac	agccatccca	gccaggccat	1080
5	ggaagagaac	aattcttggg	ggtagcgacc	gagccaaaag	gggggggggg	gaagcgggag	1140
	gggaagaagt	ggtattagag	cacgcacceg	aaaacgcatt	tgggcccttg	ccaacaaaca	1200
	ccacaccccc	cgtcctggga	gcaagacatc	caggatgcaa	cccagtaggg	gatgccaaaga	1260
10	agcatctacg	gcaccatctg	ccggcgcctc	gcctgttaga	gtcccggcac	ccgccaatgg	1320
	ggccgtgctg	ggccctgccc	ggcaatgctg	gcgagcggc	atcaacaaca	ttgctcgggg	1380
	agggggccga	ttttattgat	tagcaaaaaa	acaattaaat	tacccttcca	ttccagcaga	1440
15	gcttctcctc	cacgcggcgg	cgggaccgct	tgtggacggc	ggtacactac	aaccgcgggg	1500
	ctccagtctc	cgtgctgggc	gtgcagatca	cgacccggaa	gagaaatgat	cgcggtctga	1560
	cgccgggtac	ggagtactga	gccgccaaac	acagccgatg	gaccgtgata	tctcaatgcg	1620
20	ttcaagcaac	acagcaacac	cctggacgag	tctctcctcc	cctaccaccc	cctccccccc	1680
	tgccctggcc	gcgaacgggg	cgcgtagccc	agatttctac	tccgtactga	caccccaatc	1740
	tattcccgtc	ggcgtcgccc	agtctggggc	ggtccggcca	agactctcgg	tgacagatac	1800
25	cgcgacgaaa	tcggattaac	cgttggctga	tcaattccaa	gtcaagggag	aagtggatatg	1860
	gaaagtcggc	tcagttttcc	actgcccccg	acaggcaggt	tccggatctg	gacagcagtc	1920
	ttccgaatct	ttggcagaga	ctcatgataa	tataaaaagg	caaatgaggc	ggcgccttgg	1980
30	acaggtccat	tctcccaccg	ctcaaccagc	ctccaattcc	tcagaagtct	gttgctctct	2040
	cgcagtcgca	gtcaagatga	agcagtacct	ccagtacctc	gcggcgaccc	tgcccctggt	2100
	gggcctggcc	acggcccagc	aggcgggtaa	cctgcagacc	gagactcacc	ccaagctcac	2160
35	ttggtcgaag	tgacaggccc	cgggatcctg	ccaacaggtc	aacggcgagg	tcgtcatcga	2220
	ctccaactgg	cgctgggtgc	acgacgagaa	cgcgcagaac	tgctacgacg	gcaaccagtg	2280
	gaccaacgct	tgagctctg	ccaccgactg	cgccgagaat	tgcgcgctcg	agggtgccga	2340
40	ctaccagggc	acctatggcg	cctcgaccag	cggcaatgcc	ctgacgctca	ccttcgtcac	2400
	taagcacgag	tacggcacca	acattggttc	gcgctcttac	ctcatgaacg	gcgcgaaaca	2460
	gtaccagatg	ttcacctca	agggcaacga	gctggccttc	gacgtcgacc	tctcggccgt	2520
45	cgagtgcggc	ctcaacagcg	ccctctactt	cgtggccatg	gaggaggatg	gcggtgtgtc	2580
	gagctacccg	accaacacgg	ccggtgctaa	gttcggcact	ggggtaagtt	caacgacccg	2640
	agacgggtgc	ccttattatc	tgctgcgaaa	acggacggtc	cccttttgct	aactaccctc	2700
50	ctccaaacag	tactgcgacg	cccaatgcgc	acgcgacctc	aagtctgctg	gcggaagggg	2760
	caacatcgag	ggctggaagc	cgtccaccaa	cgatgccaat	gccggtgtcg	gtccttatgg	2820
	cgggtgctgc	gctgagatcg	acgtctggta	agttttgttg	cctgggcagc	aatggtatat	2880

55

	tagctcgagt	ggttcccgtc	gttgctgacc	ctctcttacc	agggagtcga	acaagtatgc	2940
	tttcgctttc	accccgcacg	gttgcgagaa	ccctaaatac	cacgtctgcg	agaccaccaa	3000
5	ctgcggtggc	acctactccg	aggaccgctt	cgctggtgac	tgcgatgccca	acggctgcga	3060
	ctacaacccc	taccgcatgg	gcaaccagga	cttctacggg	cccggcttga	cggtcgatac	3120
	cagcaagaag	ttcacgtgag	tacaccgtgc	ttgaagcccc	ctcccccccc	ccccccaaaa	3180
10	aaaaaaagaa	aaaagaagtc	aaatgattga	tgctaaccac	atcaaataac	agcgctcgta	3240
	gccagttcga	ggagaacaag	ctcaccaggt	tcttcgtcca	ggacggcaag	aagattgaga	3300
	tccccggccc	caaggctcgag	ggcatcgatg	cggacagcgc	cgctatcacc	cctgagctgt	3360
15	gcagtgccct	gttcaaggcc	ttcgatgacc	gtgaccgctt	ctcggagggt	ggcggcttcg	3420
	atgccatcaa	cacggccctc	agcactccca	tggtcctcgt	catgtccatc	tgggatgatg	3480
	tacgttacct	aacccccccc	cccttttttt	ttcccgcttc	tctccccgaa	actgccacta	3540
20	cttatatacg	tcccgcgctc	atgatgctta	ctttttctcc	ttccagcact	acgccaatat	3600
	gctctgggctc	gactcgagct	acccccctga	gaaggctggc	cagcctggcg	gtgaccgtgg	3660
	cccgtgtcct	caggactctg	gcgtcccggc	cgacgttgag	gctcagtacc	ctaatacgctg	3720
25	agtcgaaacc	gtaaaatgtc	gggcaaaaaa	aagatcgctc	aagctaacga	aataatatga	3780
	ttagcaagggt	catctggctc	aacatccgct	tcggccccc	cggctcgact	gtcaacgtct	3840
	aaactgcaac	ctgaccgggc	cttttctctc	cacccccacc	cctctcaagt	tctctctggt	3900
30	ggagccctcg	tgtccttctt	ttcctaggtt	cgcgaaacct	tgagcttggt	tatcgtaggg	3960
	tcattgtgta	catacacaaa	aacttaacat	ctgctaccaa	gatcttggcg	ctttgccagg	4020
	tcttctcaaa	cctcgaagca	ctgagccttt	gtcctccgag	tgaagtagga	tgactattta	4080
35	cgttgcaaga	ctacgcggta	aaggggacgg	agcagacctg	ccacagatat	tcgtttgggt	4140
	gcttgattta	tagcagagtc	cgaacgtaga	catggcccct	gaaggtgcca	accctagata	4200
	gccagaagcc	ttgtttttacg	aaaggggtgt	caaccaacgg	tgctcctcgc	tcagcgaatc	4260
40	tacccgcacg	caatgtatcg	taagaatgtg	aactaaaggg	aacgacgagg	catagggaaa	4320
	cgtcaatgtg	gcttgaataa	cagagttaaa	tacctaatag	aagaaattag	catgccaaaga	4380
	ttgagccagc	aacacatggt	agaatagcca	gcaaaggacg	cttgttcgct	tgatctcgaa	4440
45	ccgtccaacc	tgattcgaag	gaggagggaa	aagttgaaga	ataccggcaa	taattactcg	4500
	aggttcctat	gccctgcaga	gtctaattaa	tattaaaggc	accacccgca	tgattccgca	4560
	attataagca	taataagctc	gcgggcccc	cacgtgcctt	caccctccca	tgtgtataca	4620
50	atctgtacct	cgttattgtc	gaatcgctat	tccgatagcg	aaggtctggc	actcatcaga	4680
	taccgtgaca	tcgattgaga	tttggccggg	ccaccggtag	taagcgatga	gttggtcatc	4740

55

aattatcaac aatgcgctca atcagcgcata atcagcctat caaccgcgaa atcatacgcg 4800
 catcaacgaa ttgtccatca tgcacgtagc ttgtcggcag tgccgcatac cctccagagc 4860
 5 atcatagccg ggatagaaag ctgcctttca gccgtcccag agtccgagat gcaggtagca 4920
 agccttcaag accagttata tgtgacccgg gtaaaatact tgggtgagatg caatgggcgt 4980
 agcttcgggc acttataagc ttacttagat attatctcaa ggtttctttt tgaactcctc 5040
 10 ctagacattt actataaact accgagcttc aatgctagac gccctccttc tgttaaatag 5100
 tcttttcctt ctaagagcat ctgccttttt tcccttaggc ttagaggata gggcccctcc 5160
 atcttgctgc gacggcctta gccttgggga gtaattattg gtatccgcgt acctgtttcc 5220
 15 cagacagccg aagtttcgac gacaaaagtaa ttattgcgac aataccaccg ccatatgcta 5280
 ttccgagtgg gtgagccccg aaaacatcgc ttaccgcata gccatcccag acgacagagg 5340
 gcgactttga tgtcttgctc cagatcgccg cacctaacac ggtgggatgg gctggtatcg 5400
 20 tatgggacgg catcatggtc aacaaccccc tgacggggtg ttgggccaat ggaaacacca 5460
 ccgttgctc gagccggatc gcaaggtaag ccgaagagga caaatgacga tgagactttc 5520
 tttctttttt attttatttt tttttaaatt tcttttttaa gcgtaatgaa aagagctaca 5580
 25 tatctgtggg tcgttcctca atttcagcga cctctccacc gaagcatcgt caaataagaa 5640
 gttgtcggaa acaaagggtg tcagaagcta tagagcttct aaggatatta gccacataca 5700
 tgccatagct gtataaggct atttaacgct ttggccagct cttttgtcta taaatattag 5760
 30 tcgttttgct tcctttgtag ataattttta caaggcactc ttttccttta tatagccacc 5820
 tactatagac tgctttcaac gctcccggaa gcttattact acgttcggca gttataagcc 5880
 tggcgccttg actactcctc tgccgacgta tctttaatat tagtagtagc ttcttctatt 5940
 35 acgaactctc ttaccctgct ttaatacgct ttcgacgacg tgtctattat atctaagatc 6000
 ctagtcgaga cttctatatg ccttactagg cctagtctt agaacttgta gtatattaaa 6060
 ctatagttat aggctaaatt tgctagtata tagagatttg ttaaccttaa tagtaattat 6120
 40 aaactagatc tagaagtttt atagtgccta acctataaat aagctagaga taaccttatt 6180
 ttagcttcct aggagtaatt cctagaagga gtattacctt taatatctat agatttgata 6240
 ccttctaata tagctatcat agctaaattt atataattat aagattcctt ttataaaaat 6300
 45 attatatata ctatagatat tagtaagtag ataggatagc tataatacta gctagtatat 6360

<210> 2

<211> 450

<212> PRT

50 <213> *Chrysosporium lucknowense*

<400> 2

55 Met Lys Gln Tyr Leu Gln Tyr Leu Ala Ala Thr Leu Pro Leu Val Gly

	1		5		10		15									
5	Leu	Ala	Thr	Ala 20	Gln	Gln	Ala	Gly	Asn 25	Leu	Gln	Thr	Glu	Thr 30	His	Pro
	Lys	Leu	Thr 35	Trp	Ser	Lys	Cys	Thr 40	Ala	Pro	Gly	Ser	Cys 45	Gln	Gln	Val
10	Asn	Gly 50	Glu	Val	Val	Ile	Asp 55	Ser	Asn	Trp	Arg	Trp 60	Val	His	Asp	Glu
15	Asn 65	Ala	Gln	Asn	Cys	Tyr 70	Asp	Gly	Asn	Gln	Trp 75	Thr	Asn	Ala	Cys	Ser 80
	Ser	Ala	Thr	Asp	Cys 85	Ala	Glu	Asn	Cys	Ala 90	Leu	Glu	Gly	Ala	Asp 95	Tyr
20	Gln	Gly	Thr	Tyr 100	Gly	Ala	Ser	Thr	Ser 105	Gly	Asn	Ala	Leu	Thr 110	Leu	Thr
25	Phe	Val	Thr 115	Lys	His	Glu	Tyr	Gly 120	Thr	Asn	Ile	Gly	Ser 125	Arg	Leu	Tyr
	Leu	Met 130	Asn	Gly	Ala	Asn	Lys 135	Tyr	Gln	Met	Phe	Thr 140	Leu	Lys	Gly	Asn
30	Glu 145	Leu	Ala	Phe	Asp	Val 150	Asp	Leu	Ser	Ala	Val 155	Glu	Cys	Gly	Leu	Asn 160
35	Ser	Ala	Leu	Tyr	Phe 165	Val	Ala	Met	Glu	Glu 170	Asp	Gly	Gly	Val	Ser 175	Ser
	Tyr	Pro	Thr	Asn 180	Thr	Ala	Gly	Ala	Lys 185	Phe	Gly	Thr	Gly	Tyr 190	Cys	Asp
40	Ala	Gln	Cys 195	Ala	Arg	Asp	Leu	Lys 200	Phe	Val	Gly	Gly	Lys 205	Gly	Asn	Ile
45	Glu	Gly 210	Trp	Lys	Pro	Ser	Thr 215	Asn	Asp	Ala	Asn	Ala 220	Gly	Val	Gly	Pro
	Tyr 225	Gly	Gly	Cys	Cys	Ala 230	Glu	Ile	Asp	Val	Trp 235	Glu	Ser	Asn	Lys	Tyr 240
50	Ala	Phe	Ala	Phe	Thr 245	Pro	His	Gly	Cys	Glu 250	Asn	Pro	Lys	Tyr	His 255	Val

55

Cys Glu Thr Thr Asn Cys Gly Gly Thr Tyr Ser Glu Asp Arg Phe Ala
 260 265 270
 5 Gly Asp Cys Asp Ala Asn Gly Cys Asp Tyr Asn Pro Tyr Arg Met Gly
 275 280 285
 Asn Gln Asp Phe Tyr Gly Pro Gly Leu Thr Val Asp Thr Ser Lys Lys
 290 295 300
 10 Phe Thr Val Val Ser Gln Phe Glu Glu Asn Lys Leu Thr Gln Phe Phe
 305 310 315 320
 15 Val Gln Asp Gly Lys Lys Ile Glu Ile Pro Gly Pro Lys Val Glu Gly
 325 330 335
 Ile Asp Ala Asp Ser Ala Ala Ile Thr Pro Glu Leu Cys Ser Ala Leu
 340 345 350
 20 Phe Lys Ala Phe Asp Asp Arg Asp Arg Phe Ser Glu Val Gly Gly Phe
 355 360 365
 25 Asp Ala Ile Asn Thr Ala Leu Ser Thr Pro Met Val Leu Val Met Ser
 370 375 380
 Ile Trp Asp Asp His Tyr Ala Asn Met Leu Trp Leu Asp Ser Ser Tyr
 385 390 395 400
 30 Pro Pro Glu Lys Ala Gly Gln Pro Gly Gly Asp Arg Gly Pro Cys Pro
 405 410 415
 35 Gln Asp Ser Gly Val Pro Ala Asp Val Glu Ala Gln Tyr Pro Asn Ala
 420 425 430
 Lys Val Ile Trp Ser Asn Ile Arg Phe Gly Pro Ile Gly Ser Thr Val
 435 440 445
 40 Asn Val
 450

<210> 3

<211> 3900

<212> DNA

<213> Chrysosporium lucknowense

<400> 3

ccgcaagtga atatgtaatt actcaatgga agttctcgaa acggagtcca gaaatgatgt 60
 ggttctgtgg gaatgcggca agaggcgacg ttgccgtgaa tgcgtgaaca ttcccgctc 120
 55 ttctttcttct cgtcttcttc cttcttcttc tttcgggtcg cggatggttg acggccagcg 180

	tgcgcacggc	tgctgttat	cgagcgtcgg	tacgtctagc	caacatcccc	tagacacgac	240
	gaccaagcgt	cttgagaatg	caacaacgtc	tcggaacctg	gcacgcatct	tccgccgcag	300
5	gtcggcagac	gccgcctggg	caataccacc	cctgtccagg	ccctttcccc	gcaggcagag	360
	ccgcgctctt	cctttcatgg	ttattcagga	acgtggcttc	cgagattctc	gcctgttctc	420
	ccccagtcaa	cctgccgacc	gtaacccggg	tccaccaccg	cggactgtcc	gcaaaacctg	480
10	gttcgcccga	gattaatatg	ctatttccgg	actaagtgca	caacacacaa	gcaccccttc	540
	cgcctcgcgc	tctagaatct	gctttctaac	ccggttctcg	ggcccttccc	tttcgcgacg	600
	cctccgctct	ccttaccagg	caccatccgc	aataggttaag	gtagccaacc	gttttgagac	660
15	gtgattctgc	caaggaccgc	atccttgcat	tcgccatctg	gtcaaggacc	cctctttccc	720
	gtctcattct	ggtggctcta	tcgggacggc	gttccccatg	gctctccagg	agagtgatgt	780
	gcgagtctgg	agagccgggg	ttggcgtcac	gatgctgccc	acctagggcc	ggccagcccc	840
20	gactgcgct	cccgttgatc	cgtctatccc	cgtcaagagc	accagccccg	gcgctcgtga	900
	atcttcgact	tgcttcgactt	gctacaggtg	ataaagagga	tgacgcccgc	cctcgatcgg	960
	cctgtgtggt	ttctctccct	cgtgccaaac	cactcccacc	tcccgccccg	agatagttgc	1020
25	ttgtttcgct	ccgtgagagg	gacacacacc	aatggccaag	aagcttttca	tcaccgccgc	1080
	gcttgcggt	gccgtgttg	cggccccgt	cattgaggag	cgccagaact	gcggcgctgt	1140
	gtggttaagaa	agcccggtcc	gagtctccca	tgattttctc	gtcgagtaat	ggcataagg	1200
30	ccaccccttc	gactgaccgt	gagaatcgat	caaatccagg	actcaatgcg	gcggtaacgg	1260
	gtggcaagg	cccacatgct	gcgcctcggg	ctcgacctgc	gttgcgcaga	acgagtggta	1320
	ctctcagtgc	ctgcccaaca	gccaggtgac	gagttccacc	actccgtcgt	cgacttccac	1380
35	ctcgcagcgc	agcaccagca	cctccagcag	caccaccagg	agcggcagct	cctcctcctc	1440
	ctccaccacg	cccccgcccc	tctccagccc	cgtgaccagc	attcccgccg	gtgcgacctc	1500
	cacggcgagc	tactctggca	accctttctc	gggcgtccgg	ctcttcgcca	acgactacta	1560
40	caggtccgag	gtccacaatc	tcgccattcc	tagcatgact	ggtactctgg	cggccaaggc	1620
	ttccgccgtc	gccgaagtcc	ctagcttcca	gtggctcgac	cggaaagtca	ccatcgacac	1680
	cctgatggtc	cagactctgt	cccaggtccg	ggctctcaat	aaggccggtg	ccaatcctcc	1740
45	ctatgctggt	gagttacatg	gcgacttgcc	ttctcgtccc	ctacctttct	tgacgggatc	1800
	ggttacctga	cctggaggca	aaacaacaac	agcccaactc	gtcgtctacg	acctccccga	1860
	ccgtgactgt	gccgccgctg	cgtccaacgg	cgagttttcg	attgcaaacg	gcggcgccgc	1920
50	caactacagg	agctacatcg	acgctatccg	caagcacatc	attgagtact	cggacatccg	1980
	gatcatcctg	gttatcgagc	ccgactcgat	ggccaacatg	gtgaccaaca	tgaacgtggc	2040

55

EP 2 041 294 B1

caagtgcagc aacgccgcgt cgacgtacca cgagttgacc gtgtacgcgc tcaagcagct 2100
 gaacctgccc aacgtcgcca tgtatctcga cgccggccac gccggctggc tcggctggcc 2160
 5 cgccaacatc cagcccgcgc ccgagctggt tgccggcatc tacaatgatg ccggaagcc 2220
 ggctgccgtc cgccggcctgg ccactaacgt cgccaactac aacgcctgga gcatcgcttc 2280
 ggccccgtcg tacacgtcgc ctaaccctaa ctacgacgag aagcactaca tcgaggcctt 2340
 10 cagcccgtct ttgaactcgg ccggcttccc cgcacgcttc attgtcgaca ctggccgcaa 2400
 cggcaaacia cctaccggtg tgtttttttt tcttttgtct ctgtcccccc cttttctccc 2460
 ccttcagttg gcgtccacaa ggtctcttag tctgtcttca tctgtgacca acctcccccc 2520
 15 ccccggcacc gccacaacc gtttgactct atactcttgg gaatgggcgc cgaaactgac 2580
 cgttcacag gccacaaca gtgggggtgac tgggtcaatg tcaagggcac cggctttggc 2640
 gtgcgcccga cggccaacac gggccacgag ctggctcgatg cttttgtctg ggtcaagccc 2700
 20 gggggcgagt ccgacggcac aagcgacacc agcggccgcc gctacgacta cactgcggc 2760
 ctgtccgatg ccctgcagcc tgcccccgag gctggacagt ggttccaggc ctacttcgag 2820
 cagctgctca ccaacgcaa cccgcccttc taaacctcgt cataaagaga gagagatggc 2880
 25 gggcatgggc ctgattgggt tcattgacca tgcggctctt ctgggggtac atattttacc 2940
 tactaccta taaataaggc ggcctatcgg gctctcgctt cgtttattag gtacttgctc 3000
 ttgtacatac tttgtttata catacagcag ttagcatcca ctattcgttt cgacaaagcg 3060
 30 gaactttcca gaaaaaaaaa ggttgatcat aattagtctt taggcttcga ttctttgtgc 3120
 ctttcttttt ggtaaaaaaaa aaattttttt tgaggcatga ttaccttagg tacgttcgtc 3180
 gttgtattgg tccccctgca ttttggcgcg agagcagctc agccccttgc aaatccctca 3240
 35 acggggcgctt aattccctcc actcgggtct tcagcgagac cagccgtcca gagtatccca 3300
 gcgtgtagtt gccccacgaa ccagtcgtcc tcgtaagcct cgtcaaagtg tccaagagca 3360
 gtatagaagc aacgacctcc gtcaaaagtc tggcaccatg cgatcgggtg gtcctccccg 3420
 40 tgcgccccgc cctcgttaga cttctcatcc acgccaagga gcacgtgcag gccgtcggac 3480
 gtcgccccgc ggtgcgcctt gaagttgtac cattcgctct tccagacgcg ctccagctgc 3540
 gcctgcttgg gttcctgcgg ttctcgcggt tcctgcgctg gccggctcggc gccgcccgtc 3600
 45 tggtcacacg cccgcagcga catgactggg tgtttcgggt cgagcagctt gacgagcccg 3660
 acctgggggt ccgggtgggt gtcgaacacg gcgccaatga ggtggccgta ccattcggat 3720
 gactgcatgg cgaagctggc gcagtgtacc gccacgatcc cgccggccgc ctggacgaaa 3780
 50 cccgcaggg cgccagctg cgcgccgtcc aggaactcgc ccgagcactg caggaggacg 3840
 atgacgcgat acgccgagag ggagccgggg ctgaacacgg cgggatcctc gctgtcgtcc 3900

<210> 4

<211> 481

55 <212> PRT

<213> *Chrysosporium lucknowense*

<400> 4

Met Ala Lys Lys Leu Phe Ile Thr Ala Ala Leu Ala Ala Ala Val Leu
 1 5 10 15
 5 Ala Ala Pro Val Ile Glu Glu Arg Gln Asn Cys Gly Ala Val Thr Gln
 20 25 30
 10 Cys Gly Gly Asn Gly Trp Gln Gly Pro Thr Cys Cys Ala Ser Gly Ser
 35 40 45
 Thr Cys Val Ala Gln Asn Glu Trp Tyr Ser Gln Cys Leu Pro Asn Ser
 50 55 60
 15 Gln Val Thr Ser Ser Thr Thr Pro Ser Ser Thr Ser Thr Ser Gln Arg
 65 70 75 80
 Ser Thr Ser Thr Ser Ser Ser Thr Thr Arg Ser Gly Ser Ser Ser Ser
 85 90 95
 20 Ser Ser Thr Thr Pro Pro Pro Val Ser Ser Pro Val Thr Ser Ile Pro
 100 105 110
 25 Gly Gly Ala Thr Ser Thr Ala Ser Tyr Ser Gly Asn Pro Phe Ser Gly
 115 120 125
 Val Arg Leu Phe Ala Asn Asp Tyr Tyr Arg Ser Glu Val His Asn Leu
 130 135 140
 30 Ala Ile Pro Ser Met Thr Gly Thr Leu Ala Ala Lys Ala Ser Ala Val
 145 150 155 160
 35 Ala Glu Val Pro Ser Phe Gln Trp Leu Asp Arg Asn Val Thr Ile Asp
 165 170 175
 Thr Leu Met Val Gln Thr Leu Ser Gln Val Arg Ala Leu Asn Lys Ala
 180 185 190
 40 Gly Ala Asn Pro Pro Tyr Ala Ala Gln Leu Val Val Tyr Asp Leu Pro
 195 200 205
 45 Asp Arg Asp Cys Ala Ala Ala Ala Ser Asn Gly Glu Phe Ser Ile Ala
 210 215 220

Asn Gly Gly Ala Ala Asn Tyr Arg Ser Tyr Ile Asp Ala Ile Arg Lys
 225 230 235 240
 5 His Ile Ile Glu Tyr Ser Asp Ile Arg Ile Ile Leu Val Ile Glu Pro
 245 250 255
 Asp Ser Met Ala Asn Met Val Thr Asn Met Asn Val Ala Lys Cys Ser
 260 265 270
 10 Asn Ala Ala Ser Thr Tyr His Glu Leu Thr Val Tyr Ala Leu Lys Gln
 275 280 285
 15 Leu Asn Leu Pro Asn Val Ala Met Tyr Leu Asp Ala Gly His Ala Gly
 290 295 300
 Trp Leu Gly Trp Pro Ala Asn Ile Gln Pro Ala Ala Glu Leu Phe Ala
 305 310 315 320
 20 Gly Ile Tyr Asn Asp Ala Gly Lys Pro Ala Ala Val Arg Gly Leu Ala
 325 330 335
 25 Thr Asn Val Ala Asn Tyr Asn Ala Trp Ser Ile Ala Ser Ala Pro Ser
 340 345 350
 Tyr Thr Ser Pro Asn Pro Asn Tyr Asp Glu Lys His Tyr Ile Glu Ala
 355 360 365
 30 Phe Ser Pro Leu Leu Asn Ser Ala Gly Phe Pro Ala Arg Phe Ile Val
 370 375 380
 35 Asp Thr Gly Arg Asn Gly Lys Gln Pro Thr Gly Gln Gln Gln Trp Gly
 385 390 395 400
 Asp Trp Cys Asn Val Lys Gly Thr Gly Phe Gly Val Arg Pro Thr Ala
 405 410 415
 40 Asn Thr Gly His Glu Leu Val Asp Ala Phe Val Trp Val Lys Pro Gly
 420 425 430
 45 Gly Glu Ser Asp Gly Thr Ser Asp Thr Ser Ala Ala Arg Tyr Asp Tyr
 435 440 445
 His Cys Gly Leu Ser Asp Ala Leu Gln Pro Ala Pro Glu Ala Gly Gln
 450 455 460
 50 Trp Phe Gln Ala Tyr Phe Glu Gln Leu Leu Thr Asn Ala Asn Pro Pro
 465 470 475 480
 55 Phe

<210> 5

<211> 1648

EP 2 041 294 B1

<212> DNA

<213> Chyrsosporium lucknowense

<220>

5 <221> misc_feature

<222> (812)..(812)

<223> n is a, c, g, or t

<220>

10 <221> misc_feature

<222> (1162)..(1162)

<223> n is a, c, g, or t

<400> 5

15

```

atgtacgcca agttcgcgac cctcgccgcc cttgtggctg gcgccgctgc tcagaacgcc      60
tgcactctga ccgctgagaa ccacccctcg ctgacgtggt ccaagtgcac gtctggcggc      120
agctgcacca gcgtccaggg ttccatcacc atcgacgcca actggcggtg gactcaccgg      180
accgatagcg ccaccaactg ctacgagggc aacaagtggg atacttcgta ctgcagcgat      240
ggtccttctt gcgcctccaa gtgctgcata gacggcgctg actactcgag cacctatggc      300
atcaccacga gcggtaaact cctgaacctc aagttcgtca ccaagggcca gtactcgacc      360
aacatcggtc cgcttaccta cctgatggag agcgacacca agtaccagag taagttcctc      420
tcgcacccgg ccgccgggag atgatggcgc ccagcccgtg gacgcgaatg acacagtgtt      480
ccagctcctc ggcaacgagt tcacctcga tgtcgacgtc tccaacctcg gctgcggcct      540
caatggcgcc ctctacttcg tgtccatgga tgccgatggt ggcattgtcca agtactcggg      600
caacaaggca ggtgccaaagt acgggtaccg ctactgtgat tctcagtgcc ccgcgacct      660
caagttcatc aacggcgagg ccaacgtaga gaactggcag agctcgacca acgatgccaa      720
cgccggcacg ggcaagtacg gcagctgctg ctccgagatg gacgtctggg aggccaacaa      780
catggccgcc gccttcaact cccacccttg cncctgtgat ggccagtcgc gctgcgaggg      840
cgactcgtgc ggcggtacct acagcaccga ccgctatgcc ggcattctgcg accccgacgg      900
atgcgacttc aactcgtacc gccagggcaa caagaccttc tacggcaagg gcatgacggt      960
cgacacgacc aagaagatca cggtcgtcac ccagttcctc aagaactcgg ccggcgagct     1020
ctccgagatc aagcggttct acgtccagaa cggcaaggtc atccccaact ccgagtccac     1080
catcccgggc gtcgagggca actccatcac ccaggactgg tgcgaccgcc agaaggccgc     1140
cttcggcgac gtgaccgact tncaggacaa gggcggcatt gtccagatgg gcaaggccct     1200
cgcgggggccc atggtcctcg tcattgtccat ctgggacgac cagccgtca acatgctctg     1260

```

55

gctcgactcc acctggccca tcgacggcgc cggcaagccg ggcgccgagc gcggtgcctg 1320
 ccccaccacc tcgggcggtcc ccgctgaggt cgaggccgag gcccccaact ccaacgtcat 1380
 5 cttctccaac atccgcttcg gcccacatcg ctcaccgctc tccggcctgc ccgacggcgg 1440
 cagcggcaac cccaacccgc ccgtcagctc gtccaccccg gtcccctcct cgtccaccac 1500
 atcctccggt tcctccggcc cgactggcgg cacgggtgtc gctaagcact atgagcaatg 1560
 10 cggaggaatc gggttcactg gccctacca gtgcgagagc ccctacactt gcaccaagct 1620
 gaatgactgg tactcgcagt gcctgtaa 1648

<210> 6

15 <211> 526

<212> PRT

<213> Chrysosporium lucknowense

<220>

20 <221> misc_feature

<222> (249)..(249)

<223> Xaa can be any naturally occurring amino acid

<220>

25 <221> misc_feature

<222> (365)..(365)

<223> Xaa can be any naturally occurring amino acid

<400> 6

30

Met Tyr Ala Lys Phe Ala Thr Leu Ala Ala Leu Val Ala Gly Ala Ala
1 5 10 15

35

Ala Gln Asn Ala Cys Thr Leu Thr Ala Glu Asn His Pro Ser Leu Thr
20 25 30

Trp Ser Lys Cys Thr Ser Gly Gly Ser Cys Thr Ser Val Gln Gly Ser
35 40 45

40

Ile Thr Ile Asp Ala Asn Trp Arg Trp Thr His Arg Thr Asp Ser Ala
50 55 60

45

Thr Asn Cys Tyr Glu Gly Asn Lys Trp Asp Thr Ser Tyr Cys Ser Asp
65 70 75 80

Gly Pro Ser Cys Ala Ser Lys Cys Cys Ile Asp Gly Ala Asp Tyr Ser
85 90 95

50

Ser Thr Tyr Gly Ile Thr Thr Ser Gly Asn Ser Leu Asn Leu Lys Phe
100 105 110

55

Val Thr Lys Gly Gln Tyr Ser Thr Asn Ile Gly Ser Arg Thr Tyr Leu

	115		120		125
5	Met	Glu 130	Ser Asp Thr Lys Tyr	Gln Met Phe Gln	Leu 140
					Leu Gly Asn Glu
	Phe 145	Thr Phe Asp Val	Asp 150	Val Ser Asn Leu	Gly 155
10					Cys Gly Leu Asn Gly 160
	Ala Leu Tyr Phe	Val 165	Ser Met Asp Ala	Asp 170	Gly Gly Met Ser Lys Tyr 175
15	Ser Gly Asn	Lys 180	Ala Gly Ala Lys Tyr	Gly Thr Gly Tyr	Cys Asp Ser 190
	Gln Cys	Pro 195	Arg Asp Leu Lys Phe	Ile Asn Gly Glu	Ala 205
20					Asn Val Glu
	Asn Trp	Gln Ser Ser Thr	Asn 215	Asp Ala Asn Ala	Gly 220
					Thr Gly Lys Tyr
25	Gly 225	Ser Cys Cys Ser	Glu 230	Met Asp Val Trp	Glu 235
					Ala Asn Asn Met Ala 240
	Ala Ala Phe Thr	Pro 245	His Pro Cys Xaa	Val 250	Ile Gly Gln Ser Arg Cys 255
30					
	Glu Gly Asp	Ser 260	Cys Gly Gly Thr Tyr	Ser Thr Asp Arg Tyr	Ala Gly 270
35	Ile Cys	Asp 275	Pro Asp Gly Cys	Asp 280	Phe Asn Ser Tyr Arg Gln Gly Asn 285
	Lys Thr	Phe Tyr Gly Lys	Gly 295	Met Thr Val Asp	Thr Thr Lys Lys Ile 300
40					
	Thr Val Val Thr	Gln Phe Leu Lys Asn Ser	Ala 315	Gly Glu Leu Ser	Glu 320
45	Ile Lys Arg Phe	Tyr 325	Val Gln Asn Gly	Lys 330	Val Ile Pro Asn Ser Glu 335
	Ser Thr Ile	Pro 340	Gly Val Glu Gly	Asn 345	Ser Ile Thr Gln Asp Trp Cys 350
50	Asp Arg	Gln 355	Lys Ala Ala Phe	Gly 360	Asp Val Thr Asp Xaa Gln Asp Lys 365
55					

Gly Gly Met Val Gln Met Gly Lys Ala Leu Ala Gly Pro Met Val Leu
 370 375 380
 5 Val Met Ser Ile Trp Asp Asp His Ala Val Asn Met Leu Trp Leu Asp
 385 390 395 400
 Ser Thr Trp Pro Ile Asp Gly Ala Gly Lys Pro Gly Ala Glu Arg Gly
 405 410 415
 10 Ala Cys Pro Thr Thr Ser Gly Val Pro Ala Glu Val Glu Ala Glu Ala
 420 425 430
 15 Pro Asn Ser Asn Val Ile Phe Ser Asn Ile Arg Phe Gly Pro Ile Gly
 435 440 445
 Ser Thr Val Ser Gly Leu Pro Asp Gly Gly Ser Gly Asn Pro Asn Pro
 450 455 460
 20 Pro Val Ser Ser Ser Thr Pro Val Pro Ser Ser Ser Thr Thr Ser Ser
 465 470 475 480
 25 Gly Ser Ser Gly Pro Thr Gly Gly Thr Gly Val Ala Lys His Tyr Glu
 485 490 495
 Gln Cys Gly Gly Ile Gly Phe Thr Gly Pro Thr Gln Cys Glu Ser Pro
 500 505 510
 30 Tyr Thr Cys Thr Lys Leu Asn Asp Trp Tyr Ser Gln Cys Leu
 515 520 525

<210> 7

<211> 4376

<212> DNA

<213> Chrysosporium lucknowense

<400> 7

ggatccacac ctaccatacc ggatagtatg ctacccaagt gacatagggt tggtaaagta 60
 atacgagaac tcagagagca ctgcccatat ggctcgccaa tgacctcaag tgccagggtca 120
 45 gctttgcgag acagacctga gcgcgtcgga tgtgtgacat ggaacgcgcc ggatcgccctt 180
 gttgattaat tatagggaag tagcgaggaa ggtttcagca attgacgtga gcgtacatta 240
 aaagctgtat gatttcagga agacgagcca tggaccaggt ttcaaggctg aatggcttga 300
 50 cgacttaagc accgaacgag gaatgaaaga atgaaaagtg ggggatcatt ctggcccctc 360
 ctcgtatgtc gagtggtaaa gaaggcgggt ctacggagga cctaaagagc tccaatttgc 420
 tctgttgagc ttaagccaca tatctcaaga tgaatacatg tcaggcatag tcaccctgat 480
 55 cttgttcac cagtcacaca cttttcagtt cagcatgttg attcctcatc catatcactt 540

	tccattacta	tctctttatg	tccttgggtca	agactccaag	gaaccgatatg	gtgagcatcg	600
	gtgaggctcc	ctcaagggtac	caaagtagcc	atcatcaccg	aggtctggga	atggcgccgt	660
5	gcccgatctg	agtcctccaa	ctccacggta	cgacgacagc	acgtcacatt	gacgcaccac	720
	ggttgaacaa	gcagagaggg	acacgtcttg	ctacgcgaat	cctggcactg	gatggagacg	780
	cgtgtgagca	ggtttccgga	accatgacgg	cctgggtccgg	cttctcgaac	aaagaagtgg	840
10	aacacaaaaa	gaaccgaaac	ggaaacgcag	gcacggcatc	gacgaccgga	ttgtcccacg	900
	gggacctcgg	ccagtcaagc	gttgcccttg	ccgtcagctc	cctggcgacg	gggattcagc	960
	acatctcacg	ttataggcga	cctcatcccc	cttccgtctt	gtgcggtcgt	tgtctccgtgc	1020
15	cgagtaccca	ggcgtgccgg	ggccttttagc	cggggcgga	tcagagtcaa	gatgcggccg	1080
	aattggacgg	cagacgaagt	ttcgtagagg	gtcatgatcg	gcactgacga	caccaccccc	1140
	tgcgtgatcc	cgtggccctg	ggctgggaat	tgccggctaa	taatctacgg	cttaatagat	1200
20	atgcactttg	cacgcgggtgc	agataaataa	gctgtgggtt	caaacactgg	cctccgtact	1260
	ttaccaccca	actgccgctt	agcgccggga	cctgagtctt	gggagtgcgc	ggagcggcag	1320
	ccacctcggg	ttagcgtaca	cacgacggct	gcatgcgggg	atgccgcgtg	catggcttca	1380
25	tagtgtacga	cagaccgtca	agtcctaaatc	tgggtgatgc	ttgatgagat	gacagcgagc	1440
	cccgtcggcg	gcaccccggc	tatgcatcgc	gaattgacaa	cactctcagc	tctattgcga	1500
	cccatcggtat	aaaagaagaa	gaaaaaaaatg	gaccttgagt	acgggcgtca	gaaaccaaaa	1560
30	aaaaactccg	gaaccaata	tgctcgggcat	ggccgggggtg	aacgaccgct	actccccgtt	1620
	cccttcttcg	caaacagaac	gctacagagg	gttttctggt	ttgtcaaaga	gttcggagggt	1680
	cctctgctcc	gcgaatgcgt	ggtgaacca	ccagcagcca	ttgttcttgc	atgcgtggcg	1740
35	gaccgttagc	cgctgatcga	catggcgagc	ttccacctc	agacctggag	cagacggttg	1800
	cgaggagcaa	ggggctgccc	tccccctgac	ggtcggacce	caatgacttc	cccaaacggg	1860
	gacatcgagg	gtcgtgcatg	atgggtggaaa	gtagttgcag	tatgggaagt	accccgggtt	1920
40	gccaggaacc	gttggttcggc	cccccacatt	ttctctctgc	catgtcaact	gtgtgtcgtt	1980
	cgagagttcc	tggctccggc	cccccgcca	attccctaac	gggaccgcgg	ggcatcgcct	2040
	gtaactaact	tccaaatgaa	gccggatatg	agggagggag	attggatctg	gcaagccagc	2100
45	cattcgctgc	gatcggcact	cgctccgtcag	ccccgcagtc	catatcccca	aaggcaactg	2160
	ctcggcgcg	ctcaagtctt	cttcggaacg	tccagcccga	aggcgcgcgc	cagcaccggc	2220
	cctatgttcc	tgattgcgat	cctcgatctc	cagagacggg	tcacctcgcc	tcgaggacgg	2280
50	tgcaggggca	tcggcttcgc	ttcctagagc	tccgggctgt	gtgtgggtcaa	ggggagaagg	2340
	cggcggcgcc	aagggtgcgtc	tcggcgcaact	cacccatcgc	ctttaccccc	ctccccccca	2400

	gtatataaaa gatggccatc gtctcctcgt ctgcttgggga agaaaggatc tctcgaccat	2460
	gcaccacagc ctagctctaa cccagcttgt cgtgtgttgt tgcccagcat gaagttcgtg	2520
5	cagtccgcca ccctggcgtt cgccgccacg gccctcgtg cgccctcgcg cacgactccc	2580
	cagaagcccc gccaggcctc ggcgggctgc gcgtcggccg tgacgctcga tgccagcacc	2640
	aacgtgttcc agcagtacac gctgcacccc aacaacttct accgtgccga ggtcgaggct	2700
10	gccgccgagg ccatctccga ctcggcgctg gccgagaagg cccgcaaggc cgccgacgtc	2760
	ggtaccttcc tgtggctcga caccatcgag aacattggcc ggctggagcc cgcgctcgag	2820
	gacgtgccct gcgagaacat cgtgggtctc gtcattctacg acctcccggg ccgtgactgc	2880
15	gcggccaagg cctccaacgg cgagctcaag gtcggcgagc tcgacaggta caagaccgag	2940
	tacatcgaca gtgagttaac cttttgtggc cctttctttt cccccgagag agcgtctggt	3000
	tgagtggggt tgtgagagag aaaatggggc gagcttaaag actgacgtgt tggctcgcag	3060
20	agatcgccga gatcctcaag gcccaactcca acacggcctt cgccctcgtc atcgagcccg	3120
	actcgtctcc caacctggtc accaatagcg acctgcagac gtgccagcag agcgttccg	3180
	gctaccgcga ggggtgtcgc tatgccctca agcagctcaa cctccccaac gtggtcatgt	3240
25	acatcgatgc cggccacggc ggctggctcg gctgggacgc caacctcaag cccggcgccc	3300
	aggagctcgc cagcgtctac aagtctgctg gttcgccctc gcaagtccgc ggtatctcca	3360
	ccaacgtggc tggttggaac gcctggtaag aactctatg tccccctcgt cggatcaatgg	3420
30	cgagcggaat ggcgtgaaat gcatgggtgt gacctttgat cttttccccc tcctataggg	3480
	accaggagcc cgggtgagttc tcggacgcct cggtatgcca gtacaacaag tgccagaacg	3540
	agaagatcta catcaacacc tttggcgctg agctcaagtc tgccggcatg cccaaccacg	3600
35	ccatcatcga cactggccgc aacggtgtca ccggtctccg cgacgagtgg ggtgactggt	3660
	gcaacgtcaa cggcgccggc ttcggtgtgc gcccgactgc caacactggc gacgagctcg	3720
	ccgacgcctt cgtgtgggtc aagcccgggt gcgagtcgga cggcaccagc gactcgtcgg	3780
40	cggcgcgcta cgacagcttc tgcggcaagc ccgacgcctt caagcccagc cccgaggccg	3840
	gtacctggaa ccaggcctac ttcgagatgc tcctcaagaa cgccaacccg tccttctaag	3900
	ctcctcgacg gcttcttgct gtcagtcgct ctgacgggtg tgtgctggtg gtgcccctgc	3960
45	tcctgctgct gctgctccgc ggggagggga ggcaacgaaa atgaagtcct gcttcaaaac	4020
	aaaacagaaa caagcgaggc gcggtgcaat ggtcgtgctg tcgtcttttt tcatgttccc	4080
	ttctagtgtg gtagtttgat agtcgtacat aaggggtttc agaaccgtct ctctgtctcg	4140
50	gtctttttgc gagttgttgc gactcgtgat tatggccttt gttgctcgtt gcggcagagt	4200
	agaaccacag cgtgttgggg tagcagcttg ctccgtagga cgtagggaaa caacctgaga	4260
	ctctggaatt gcagtcagcc tgcgtcgccc ctctaggaaa cgaaggggag aaccagtagt	4320
55	ggctgcagct tacaaacgcg agcatggtga acatctccga gaaaaggag ggatcc	4376

<211> 395

<212> PRT

<213> Chrysosporium lucknowense

5 <400> 8

10

15

20

25

30

35

40

45

50

55

Met Lys Phe Val Gln Ser Ala Thr Leu Ala Phe Ala Ala Thr Ala Leu
 1 5 10 15

Ala Ala Pro Ser Arg Thr Thr Pro Gln Lys Pro Arg Gln Ala Ser Ala
 20 25 30

Gly Cys Ala Ser Ala Val Thr Leu Asp Ala Ser Thr Asn Val Phe Gln
 35 40 45

Gln Tyr Thr Leu His Pro Asn Asn Phe Tyr Arg Ala Glu Val Glu Ala
 50 55 60

Ala Ala Glu Ala Ile Ser Asp Ser Ala Leu Ala Glu Lys Ala Arg Lys
 65 70 75 80

Val Ala Asp Val Gly Thr Phe Leu Trp Leu Asp Thr Ile Glu Asn Ile
 85 90 95

Gly Arg Leu Glu Pro Ala Leu Glu Asp Val Pro Cys Glu Asn Ile Val
 100 105 110

Gly Leu Val Ile Tyr Asp Leu Pro Gly Arg Asp Cys Ala Ala Lys Ala
 115 120 125

Ser Asn Gly Glu Leu Lys Val Gly Glu Leu Asp Arg Tyr Lys Thr Glu
 130 135 140

Tyr Ile Asp Lys Ile Ala Glu Ile Leu Lys Ala His Ser Asn Thr Ala
 145 150 155 160

Phe Ala Leu Val Ile Glu Pro Asp Ser Leu Pro Asn Leu Val Thr Asn
 165 170 175

Ser Asp Leu Gln Thr Cys Gln Gln Ser Ala Ser Gly Tyr Arg Glu Gly
 180 185 190

Val Ala Tyr Ala Leu Lys Gln Leu Asn Leu Pro Asn Val Val Met Tyr
 195 200 205

Ile Asp Ala Gly His Gly Gly Trp Leu Gly Trp Asp Ala Asn Leu Lys
 210 215 220
 5 Pro Gly Ala Gln Glu Leu Ala Ser Val Tyr Lys Ser Ala Gly Ser Pro
 225 230 235 240
 Ser Gln Val Arg Gly Ile Ser Thr Asn Val Ala Gly Trp Asn Ala Trp
 245 250 255
 10 Asp Gln Glu Pro Gly Glu Phe Ser Asp Ala Ser Asp Ala Gln Tyr Asn
 260 265 270
 15 Lys Cys Gln Asn Glu Lys Ile Tyr Ile Asn Thr Phe Gly Ala Glu Leu
 275 280 285
 Lys Ser Ala Gly Met Pro Asn His Ala Ile Ile Asp Thr Gly Arg Asn
 290 295 300
 20 Gly Val Thr Gly Leu Arg Asp Glu Trp Gly Asp Trp Cys Asn Val Asn
 305 310 315 320
 25 Gly Ala Gly Phe Gly Val Arg Pro Thr Ala Asn Thr Gly Asp Glu Leu
 325 330 335
 Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Thr
 340 345 350
 30 Ser Asp Ser Ser Ala Ala Arg Tyr Asp Ser Phe Cys Gly Lys Pro Asp
 355 360 365
 35 Ala Phe Lys Pro Ser Pro Glu Ala Gly Thr Trp Asn Gln Ala Tyr Phe
 370 375 380
 Glu Met Leu Leu Lys Asn Ala Asn Pro Ser Phe
 385 390 395

<210> 9

<211> 5777

<212> DNA

<213> Chrysosporium lucknowense

<400> 9

tgctgctctg atgtgctgat gcacagcttc ccctcgcgat tgccggcagg atctccaacc 60
 50 ctctggatcg gagcagacga tcagcgggca caatggccag cttgccagcg ttcaactcca 120
 agttgaccgg cttttatcac gcccagctg gacatgcaca ggcttggtt ctcgtgttcc 180
 tacgatctgc acagtaggtt tgactgctga tcttcgcttt cctgtgcgcc ctccccctcc 240
 55 ctcacgggta ctttatcctt gcctgtaacc ccgcgttatg tcaaacttga gtttgaccaa 300

	tgctagcgca aaagtaccta catagtacta tgtaataagg taggtacata catcagtagg	360
	cgtttatcta gtaaattttg gctttttgaa actcaattgc tcctctcctc gcctccacct	420
5	ctgcttgga attgacaacc ctggctgtgc ctagaggtag catcgacgat caatcaaadc	480
	taaagtattc gagattgacc tttctgtctc aattatatta attatccgca caatgctgta	540
	gtcattgact ctcctttcaa gttgccttct cgtttatgta tgtacaatgg gcggctcatgc	600
10	ttcatgccaa cagatgggtc tatcggaaca atgtttgact ttctggtcgc cccgtcgaac	660
	tgttttgatt tcgcacggga agtgttctta ccaaagctaa gtcgactcgt ggagcttcgt	720
	aacggccagt gatcgttgat cgcttttgga ggagttgcga tggagcgaga ccggctacga	780
15	gcacgttcgc aaaggcagca cgatagacga ccctccgtgg cgccattcgg gagatgcaca	840
	tgacataagc atatcaatac tcacctgaac tcacggtccg atgcctcgca ggtagttaca	900
	agacatattt gtgtgggtat attatcccaa cccgtacctt tgtcgcgtca tttcgggtatg	960
20	tgctgatgcc tacttaggga gcaaagacgc ctctcctcac ctgcgggtta cttacttact	1020
	gtgcagcatg gccttatgtt ctcccgggtc ttgcttgccg gaatgaacaa aaacgcccga	1080
	agaaaagccg cttcttcgag ttgtgtctac ccgaacataa gaggttattg tcgcagaccg	1140
25	ccagcaaatg tcaacaacc acccacggcg ttccagaacc ttcgaaatat catctagttt	1200
	aagtttaaat gacggcccga gtcccagccg agattcccat attggccgat accagcgttc	1260
	ccttgttttt ccaagggtgt ctcgtaact ggcgcatctg cctacaacga gatataatta	1320
30	ccgttttctt ttgcaaaagg gcatgcatgg atgtatatta tttatgcctg cagaacgaga	1380
	agcaatcatg gtgtagggtt tgtgcggtat ggagctaata atattgaacg gatctctggt	1440
	ccgtcctaaa tcgttgaaac gctaggccca ggaggacctg ctcgacttg ggaacggaga	1500
35	tttccaggat gaaaggctcg aacatgtcca tccgcggcca gcctgaacac ttttgctcgt	1560
	ttccggacca tcgaccacg aaaacagtgc ggttgctggc acagtcagca ctacgatgg	1620
	cgatggtcca gcccgttccc gccgatgcc cacttgacg gcaactctcc ttcattcggc	1680
40	ggcccggcgg tgtctggcct attagtacga ttttggtatc cggttggtc gccgcccgg	1740
	tttttcttgg ccgatacggg aatctcgggt gtcccaactc cacctgggca cgctctggtg	1800
	ccaacatgga acttcgggat gccgctccgg gcacagtcaa gcgctttaaa atacgacttt	1860
45	accccacaag aatcgaggcg taaccggaa ttagggacac ctggacggcg caaccctgg	1920
	accgaagggc ctgctaacc gggttcctgg agccgatgc gcggctgccc gcttgccgc	1980
	tcttgagatg acacttctt tcagcgagg atggtcggg agggaaatga tgtattataa	2040
50	gaagcgagcc gattccgaag gactcgacc cctctctgc cctgtgtccg ccagctaatt	2100
	acagcactcc ttctcgactt gaaacggcg agatgaagtc ctccatctc gccagcgtct	2160

55

	tcgccacggg	cgccgtggct	caaagtggct	cgtggcagca	atgtggtggc	atcggatggc	2220
	aaggatcgac	cgactgtgtg	tcgggttacc	actgcgtcta	ccagaatgat	tggtacagcc	2280
5	agtgcgtgcc	tggcgcgggc	tcgacaacgc	tccagacatc	taccacgtcc	aggcccaccg	2340
	ccaccagcac	cgcctctccg	tcgtccacca	cctcgcctag	caagggcaag	ctcaagtggc	2400
	tcggcagcaa	cgagtcgggc	gccgagttcg	gggagggcaa	ctaccccgcc	ctctggggaa	2460
10	agcacttcat	cttcccgtcg	acttcggcga	ttcaggtacg	ggccaataat	aatatattat	2520
	tatagcaggc	aggagggagc	aggagaagaa	gggaggggca	ggtggccaac	aatcgggaaga	2580
	agaccgggag	gcactgaccg	ttgattcctt	tgtgtaatag	acgctcatca	atgatggata	2640
15	caacatcttc	cggatcgact	tctcgatgga	gcgtctggtg	cccaaccagt	tgacgtcgtc	2700
	cttcgacgag	ggctacctcc	gcaacctgac	cgaggtggtc	aacttcgtga	cgaacgcggg	2760
	caagtacgcc	gtcctggacc	cgcacaacta	cggccggtag	tacggcaacg	tcatcacgga	2820
20	cacgaacgcg	ttccggacct	tctggaccaa	cctggccaag	cagttcgcct	ccaactcgct	2880
	cgatcatctt	gacaccaaca	acgagtacaa	cacgatggac	cagaccctgg	tgctcaacct	2940
	caaccaggcc	gccatcgacg	gcatccgggc	cgcggcgcg	acctcgagc	acatcttcgt	3000
25	cgagggcaac	gcgtggagcg	gggcctggag	ctggaacacg	accaacacca	acatggccgc	3060
	cctgacggac	ccgcagaaca	agatcgtgta	cgagatgcac	cagtacctcg	actcggacag	3120
	ctcgggcacc	cacgccgagt	gcgtcagcag	caacatcggc	gccagcgcg	tcgtcggagc	3180
30	caccagtggt	ctccgcgcca	acggcaagct	cggcgtcctc	ggcgagttcg	ccggcggcgc	3240
	caacgccgtc	tgccagcagg	ccgtcaccgg	cctcctcgac	cacctccagg	acaacagcga	3300
	ggtctggctg	ggtgccctct	ggtgggccgc	cggcctctgg	tggggcgact	acatgtactc	3360
35	gttcggtaag	tttctccctt	gttcttggct	ttccccccag	taagggagtc	aggcaacatg	3420
	cccaagaccg	gctcggcttc	gttccaaggc	gttcgttgta	cacactgaag	agttccaact	3480
	tccaaccctg	ttcgtgtcct	ccgatcagct	tcgacggggg	gaagggggaa	gggatttggg	3540
40	agtgaggtgg	aggtcaaaaag	gagggatata	cccagatctc	cacaaacggc	cctgagccaa	3600
	caacagcctc	tgggggtcaaa	atgggcgcca	accatacggg	cattcactca	ggacacctgc	3660
	taacgcgtct	cttttttttg	tttccagagc	ctccttcggg	caccggctat	gtcaactaca	3720
45	actcgatcct	aaagaagtac	ttgccgtaag	gggcatgcag	caaggtcgag	cgagcattat	3780
	tcagggccat	ctgcttgtgt	cggcaggcat	cacgtcaacc	catcgaatcg	gacagcggaa	3840
	tgctccgaga	tgccatacac	taagtctggt	gatgacgtga	gaatgctggc	cctggtcggg	3900
50	ggttaccgcc	aacaaaaagc	accgggacgc	tgccgcgccc	ggataccatg	gtttcatgta	3960
	catattgggt	ctttgtcttc	ttacgggggg	gggggggggg	gggggctctg	cagcgttgct	4020
	gagcgattcg	tttccaagta	tatactttgt	ctggaattga	attttgagtg	acattgaccc	4080
55							

EP 2 041 294 B1

aatcaaccag ctcggtgtgc tcacctcccg ttaccccccc tcttctcccc ctgctcggct 4140
 tggctttcct ctccggtgtg gagcacggcc acggcgggtcc caatccatat aagatcgatg 4200
 5 gtatactatg gtatacacta gcttggaat aaactaatcc atacgctaac taatggacgg 4260
 attatcctaa gggtcaccgg ctcaccgttg gatataacac ctaggatacg ggagagctga 4320
 tagaaaggga tgtactccgt attgtactgt acaatacaaa gtacagatag cacacgaagt 4380
 10 acggtaggtg gtcccgcta gtccggacca acaatagaac atgcgttcct ggggacctgc 4440
 aggaagaag gggggggggg ttgccaagac gcccggggtt caaagaaagc cccgggccgc 4500
 cgatgagatg agacggacgc cggcccaagg agaggccggt ggtcgatcct gcaaatgcca 4560
 15 gcaaaaaaaaa tccataccat aatccagtca actttcgtca cactcctgtg aaacgagctg 4620
 gagggactgc tggaaagggt ttgcagggtta atcactgtat gtggagcatg ccgtacctac 4680
 tgtgttctgt taacagatag agttccagtt gaacacacaa agttctgccc cgcttgcag 4740
 20 acgtgaaaag aagctcctcc gggggagctt taggcaactg ggagggctct ctcccaggtt 4800
 catggtgtct gctcttcttc aaatttttat gctgccaccc catttgacag aggtgtgcac 4860
 accgttgcca ggtcttgcca tccggcaaaa agcagaaaag tcgacccatc gcctaagaaa 4920
 25 ggcggtcggg aggggatcgg atgctcattg cggcttagcg tctgcccatt ctgacgctgc 4980
 ccattgtttt gtgtcgcatt cgtcttcgga tgtcggatca agagtcccgg attttttccc 5040
 ctgtgcttcc agcctaattc gagcgggagc tggctcgggt tcgagtggag ttgccttggt 5100
 ggtggagcag caaccagcca attcactccc ccgcattttc gcggccgccc aggcatcccc 5160
 30 ggcattgcgt tgggcggtta ctactccgta ctggggtagg tgaaattggt tctcccgtcg 5220
 caggaggctc gtgtcggtc aggggagaac aaagtccaac tgctccttcc tggcaacaat 5280
 gagagggggt tctattgcca acgttgacg aaaggagcag ccacaaaacc caaagcagg 5340
 35 ttaccttact gtacctgagc ttgaacgtcg cgtagcattg gagctctcgt ctaccggcgg 5400
 cgtcacactc cattggcagg tcaaggcagt cagtggcagc gacccaacaa cgtcaatgct 5460
 tgttacccca gaattacccc gggctgcaac actgcagggg ccgccgccga tgttgatcac 5520
 40 cggttgatta cttctcggcc cgcaaccggg agatgagaag cagaactttg ttctcctttc 5580
 aaaaaggacc tgacttgagg ggaacgcact gccggcagtg gagtggatgc acgctagtta 5640
 tatgtttccc gccatcccca gtccgcccgt cgcgtccgtg aggtcagtt tggcttcccg 5700
 45 tgccgccgac aaacgagcgg tgcataatta catttcgctc catgtaccgt gcaccctccc 5760
 cgttcgcgac cgtagta 5777

<210> 10

<211> 389

<212> PRT

<213> *Chrysosporium lucknowense*

<400> 10

Met Lys Ser Ser Ile Leu Ala Ser Val Phe Ala Thr Gly Ala Val Ala
 1 5 10 15
 5 Gln Ser Gly Pro Trp Gln Gln Cys Gly Gly Ile Gly Trp Gln Gly Ser
 20 25 30
 Thr Asp Cys Val Ser Gly Tyr His Cys Val Tyr Gln Asn Asp Trp Tyr
 35 40 45
 10 Ser Gln Cys Val Pro Gly Ala Ala Ser Thr Thr Leu Gln Thr Ser Thr
 50 55 60
 Thr Ser Arg Pro Thr Ala Thr Ser Thr Ala Pro Pro Ser Ser Thr Thr
 65 70 75 80
 Ser Pro Ser Lys Gly Lys Leu Lys Trp Leu Gly Ser Asn Glu Ser Gly
 85 90 95
 20 Ala Glu Phe Gly Glu Gly Asn Tyr Pro Gly Leu Trp Gly Lys His Phe
 100 105 110
 Ile Phe Pro Ser Thr Ser Ala Ile Gln Thr Leu Ile Asn Asp Gly Tyr
 115 120 125
 25 Asn Ile Phe Arg Ile Asp Phe Ser Met Glu Arg Leu Val Pro Asn Gln
 130 135 140
 30 Leu Thr Ser Ser Phe Asp Glu Gly Tyr Leu Arg Asn Leu Thr Glu Val
 145 150 155 160
 Val Asn Phe Val Thr Asn Ala Gly Lys Tyr Ala Val Leu Asp Pro His
 165 170 175
 35 Asn Tyr Gly Arg Tyr Tyr Gly Asn Val Ile Thr Asp Thr Asn Ala Phe
 180 185 190
 40 Arg Thr Phe Trp Thr Asn Leu Ala Lys Gln Phe Ala Ser Asn Ser Leu
 195 200 205
 Val Ile Phe Asp Thr Asn Asn Glu Tyr Asn Thr Met Asp Gln Thr Leu
 210 215 220
 45 Val Leu Asn Leu Asn Gln Ala Ala Ile Asp Gly Ile Arg Ala Ala Gly
 225 230 235 240
 50
 55

Ala Thr Ser Gln Tyr Ile Phe Val Glu Gly Asn Ala Trp Ser Gly Ala
245 250 255

5 Trp Ser Trp Asn Thr Thr Asn Thr Asn Met Ala Ala Leu Thr Asp Pro
260 265 270

Gln Asn Lys Ile Val Tyr Glu Met His Gln Tyr Leu Asp Ser Asp Ser
275 280 285

Ser Gly Thr His Ala Glu Cys Val Ser Ser Asn Ile Gly Ala Gln Arg
290 295 300

15 Val Val Gly Ala Thr Gln Trp Leu Arg Ala Asn Gly Lys Leu Gly Val
305 310 315 320

Leu Gly Glu Phe Ala Gly Gly Ala Asn Ala Val Cys Gln Gln Ala Val
325 330 335

Thr Gly Leu Leu Asp His Leu Gln Asp Asn Ser Glu Val Trp Leu Gly
340 345 350

25 Ala Leu Trp Trp Ala Ala Gly Pro Trp Trp Gly Asp Tyr Met Tyr Ser
355 360 365

Phe Glu Pro Pro Ser Gly Thr Gly Tyr Val Asn Tyr Asn Ser Ile Leu
370 375 380

Lys Lys Tyr Leu Pro
385

35 <210> 11
 <211> 6060
 <212> DNA
 <213> *Chrysosporium lucknowense*

40 <400> 11

	ccggcctcca	gttccaggag	cttggtctctg	ccgacatact	gtgtacacta	ggaattctct	60
45	tatgcgggggt	gtgcgcggggg	aaatgtttggg	gaactcgagt	tgggtcatgt	ggacaagacc	120
	aatgggagct	gacatcattg	tgcgaccctg	taaaccggaa	gctacaacaa	cattctggat	180
	tctacactag	tggaagaggt	aagtaattga	cgacaagcaa	gaagcattgc	catgtttctgc	240
50	gaaggatgcg	gggtgtttttg	catgagcagg	aagctgtggc	tttttagtgc	tcctttgtgc	300
	tcgccggggcg	cgcagaacac	taccgaaacg	caggggactg	cgtgcctctg	gggtcgaatg	360
	ccgatcccca	tcttcacatt	cccaccatcg	tgttctgtta	acgaagccgg	agcggcggga	420
	actcgaagct	ccactacgta	tggatacttg	ggaccgtacg	gagtggtgtg	gtacggatgc	480

	ctgcacaagt	gttgtgcttc	ctacgaagac	gccaaccac	ataatacaca	aaagctgttg	540
	taagtcgagt	tacctcaggc	acgttcgggc	aactcgggca	acctgacgag	atttccccgc	600
5	cattccgcc	agaggccggc	gcctgccctg	attaggcagc	tcttggaaca	atactatgta	660
	gaatggaagc	tccatccata	gtcagctcca	ttggcgggtcc	cagtgatctc	gatggctgga	720
	tggtgctct	gtacggtaca	tacatagtaa	gttctcgcct	tgagagccca	attcgctgca	780
10	atagcatctt	tccccgcagt	gcgccggccg	ccctgggtcc	cgctccacaa	tgaccttgct	840
	tctggagctt	ctcgacgaac	agatcggccc	gtttcttctc	cacaccaatc	cgaaccagtc	900
	gggagcatgg	ctgcggatgc	gacgcagcct	tccttcgcgc	tgtacaaaca	gctccgggaa	960
15	cgtcgactgg	tatgtacgga	ctacagtaag	tacactacga	gtgcacatac	tgacgaatac	1020
	cggcctcaga	ggaacctggc	aggaccctac	cccacacgaa	accacagcga	gaaagcgcaa	1080
	tggtatcagta	actactgcga	agtaaccgtg	gtcccgggca	aaggatctga	gggccgatcg	1140
20	ctcgtggggc	tgcgaggcga	gggagagcaa	acaagccagt	cctcccgcga	acctggaaaa	1200
	tcacttataa	acacacgtca	ccggcgccgg	ggtgcgcgcc	atgtgtcacc	tccaggctcc	1260
	tcccgggcga	tgatctctgc	cggtgccatc	aatcatctcg	gttcgccgca	gctgcttctt	1320
25	tctgtgcagt	gaacgctctc	aaactgcaac	gacgctgtcc	gacatgaagg	ctgctgcgct	1380
	ttcctgcctc	ttcggcagta	cccttgccgt	tgaggcgcc	attgaatcga	gaaaggatatg	1440
	gacgggcttt	cgtcaaagac	tcgctccccg	atcaacttcc	cctttcatcc	agaccacccc	1500
30	aaccttccca	gtcctgcttc	gagcacgata	tcttcgggca	gcacccacc	cacatccact	1560
	cagattagcg	gcgacaccgt	tgactgttgc	aatccgcaat	cgacatgcaa	cttccagccg	1620
	cagcccaatg	gctgctcacg	cttcccgcga	aagcctcact	tgctgacaat	catcgtcagg	1680
35	ttcaccagaa	gcccctcgcg	agatctgaac	ctttttaccc	gtcgccatgg	atgaatccca	1740
	acgccgacgg	ctgggaggag	gcctatgccc	aggccaagtc	ctttgtctcc	caaatgactc	1800
	tgctagagaa	ggtcaacttg	accacgggag	tcgggtaagt	tttgtcattt	tgtccaggta	1860
40	acatgcaa	ggttctgcta	acaataactt	accgtagctg	gggggctgag	cagtgcgtcg	1920
	gccaagtggg	cgcatccct	cgccttgga	ttcgagctct	gtgcatgcat	gactcccctc	1980
	tcggcatccg	aggagccgac	tacaactcag	cggtcccctc	tgccagacc	gttgtgcta	2040
45	cctgggatcg	cggtctgatg	taccgtcgcg	gctacgcaat	gggccaggag	gccaaaggca	2100
	agggcatcaa	tgtccttctc	ggaccagtcg	ccggccccct	tgcccgcatg	cccgagggcg	2160
	gtcgtaactg	ggaaggcttc	gctccggatc	ccgtccttac	cgcatcggc	atgtccgaga	2220
	cgatcaaggg	cattcaggat	gctggcgta	tcgcttgctg	gaagcacttt	attggaaacg	2280
50	agcaggggtga	gtagtcaaag	acgggccgtc	tcggacccgc	ggcttcaagc	tgctgactct	2340
	gctgcagagc	acttcagaca	ggtgccagaa	gcccgaggat	acgggttaca	catcagcgaa	2400
55							

EP 2 041 294 B1

	accctctcct	ccaacattga	cgacaagacc	atgcacgagc	tctacctttg	gccgtttgcc	2460
	gatgccgtcc	gggccggcgt	cggtctgtgc	atgtgctcgt	accagcaggt	caacaactcg	2520
5	tacgcctgcc	agaactcgaa	gctgctgaac	gacctcctca	agaacgagct	tgggtttcag	2580
	ggcttcgtca	tgagcgactg	gcaggcacag	cacactggcg	cagcaagcgc	cgtggctggt	2640
	ctcgatatgt	ccatgccggg	cgacaccag	ttcaacactg	gcgtcagttt	ctggggcgcc	2700
10	aatctcacc	tcgccgtcct	caacggcaca	gtccctgcct	accgtctcga	cgacatggcc	2760
	atgcgcatca	tggccgccct	cttcaaggtc	accaagacca	cccacctgga	acccatcaac	2820
	ttctccttct	ggaccgacga	cacttatggc	ccgatccact	gggccgcca	gcatggctac	2880
15	cagaagatta	attcccacgt	tgacgtccgc	gccgaccacg	gcaacctcat	ccgggagatt	2940
	gccgccaagg	gtacggtgct	gctgaagaat	accggctctc	tacctctgaa	caagccaaag	3000
	ttcgtggccg	tcacgggcga	ggatgctggg	tcgagcccca	acgggcccga	cggtgcagc	3060
20	gaccgcggct	gtaacgaagg	cacgctcgcc	atgggctggg	gatccggcac	agccaactat	3120
	ccgtacctcg	tttccccga	cgccgcgctc	caggcccggg	ccatccagga	cggcacgagg	3180
	tacgagagcg	tcctgtccaa	ctacgccgag	gaaaagacaa	aggctctggt	ctcgcaggcc	3240
25	aatgcaaccg	ccatcgtctt	cgtcaatgcc	gactcaggcg	agggtacat	caacgtggac	3300
	ggtaacgagg	gcgaccgtaa	gaacctgact	ctctggaaca	acggtgatac	tctggtcaag	3360
	aacgtctcga	gctggtgcag	caacaccatc	gtcgtcatcc	actcggtcgg	cccggtcctc	3420
30	ctgaccgatt	ggtacgacaa	ccccaacatc	acggccattc	tctgggctgg	tcttcggggc	3480
	caggagtcgg	gcaactccat	caccgacgtg	ctttacggca	aggtaacccc	cgccgcccgc	3540
	tcgcccttca	cttggggcaa	gacccgcgaa	agctatggcg	cggacgtcct	gtacaagccg	3600
35	aataatggca	atggtgcgcc	ccaacaggac	ttcaccgagg	gcgtcttcat	cgactaccgc	3660
	tacttcgaca	aggttgacga	tgactcggtc	atctacgagt	tcggccacgg	cctgagctac	3720
	accaccttcg	agtacagcaa	catccgcgtc	gtcaagtcca	acgtcagcga	gtaccggccc	3780
40	acgacgggca	ccacggccca	ggccccgacg	tttggcaact	tctccaccga	cctcgaggac	3840
	tatctcttcc	ccaaggacga	gttcccctac	atctaccagt	acatctaccc	gtacctcaac	3900
	acgaccgacc	cccggagggc	ctcggccgat	ccccactacg	gccagaccgc	cgaggagttc	3960
	ctcccccccc	acgccaccga	tgacgacccc	cagccgtccc	tccggtcctc	gggcggaac	4020
45	tcccccggcg	gcaaccgcca	gctgtacgac	attgtctaca	caatcacggc	cgacatcacg	4080
	aatacgggct	ccgttgtagg	cgaggaggta	ccgcagctct	acgtctcgct	gggcggtccc	4140
	gaggatccca	agggtgcagct	gcgcgacttt	gacaggatgc	ggatcgaacc	cggcgagacg	4200
50	aggcagttca	ccggccgcct	gacgcgcaga	gatctgagca	actgggacgt	cacggtgcag	4260

55

gactgggtca tcagcaggta tccaagacg gcatatgttg ggaggagcag ccggaagttg 4320
 gatctcaaga ttgagcttcc ttgaatgagt ttcacaggg gctgcagagg gatggtaaca 4380
 5 cgttcttaat cagaagtatg atggagaaaa gcacttggca agttccggtg agcaaaaaga 4440
 aggcacttat taagtgtagg gcggtgttct atgtttaata ggtgctatgt ttacatatata 4500
 ttagtatata atgatttaat aattatgttt agcagttgct aatgtcgtaa atttcggcgt 4560
 10 gtgatgactg ctacaacact ggttctgtct tctagtcgcc attgttaatt atgaaggtta 4620
 ttgtctacaa tttctaatac cttatggatg attgcccagc tggtttcaaa ctcgttacgc 4680
 gcaaatggta cgattgaggt attattcatt gtaagtacct ccgtacagcg tccccaacta 4740
 15 tttccattca cgagatgcct cgcttttcgg tgctttcgga acagggctgg cagcggatca 4800
 tggcgcgac aaaacatggc gagcagctgt ccaggacgga ggacaggttg gggactgatg 4860
 cctcccgac gcattaaggt cagaagatag acacgtttta cacagcgttg agaccgacaa 4920
 20 gccacattag gcagcgccgg ttgcaccacc gccgtcacgg gcaacggttc aatcaatcga 4980
 caacagtgga agacaaagta ctgaagatca ggtattaata gtgtgagaga gaaacagacg 5040
 gtggaactag ggtgctaata tttctcttga tttcgggtgc catggtagta cagaacacaa 5100
 25 gaaaaagaag gaggagttag cggagaagga ggagggggaa gccagaaaaa agaacatgaa 5160
 aaagcataca cattggagtc ggtcagtcgg ttgattgggt tggtagagag cgaaaaagca 5220
 agcgtcacct gtaggattcg aacctacgt cccgaaggaa ctgcctaaga acgctaagca 5280
 30 aggttagcag ggcagcgcgt taaccactcc gccaaagtga ctgtcgttga tcatggtcga 5340
 attcaagtag cttataggag ttcaaccaga tcacaaatgc ataggtgctc gtagaacggt 5400
 ctaagtatga gttgattata agcaaccgaa tggctctcag cggcaacacc gtagctgaag 5460
 35 taacaaaacg cacctttggt tactttctga ctataaaaat gggatatttg gaaatgacca 5520
 cccgataagg tgtcaaattc taaatgactg tctgggtgtg aagatgttac tgtggttcca 5580
 ccacgaacca gtttttagtat ccgcatgctt cagtctctgc gcctcgacag gcggagggtg 5640
 40 tgtgttagat cagaatcgat gtgacgctgt gaccgcgagg ctctcgagcc taggtgcggt 5700
 agttctgttc aaaaagaagt gtgtggccgg gtttgggcgc cttatagcc taccatcctg 5760
 gctgtggttc ccgagcggga gccggttctc cgttttgggt ccgataaagt gtcatatctg 5820
 45 cctcccggtt tcgcatctaa tttctgactt cgttcgggac ctctggagac gtagggatag 5880
 gtatgggata tgcccgcat ttcgtaaatg tccatagtct ctttcgggac gaggcggcaa 5940
 gctctcagag ctatctaagc ttaaccaacc cctgatcctt aaccctcca gaccacacct 6000
 50 cctgggagaa taaaccgggc tccaagatcg aaatcgaaat cagtgcgcga acttgaaatc 6060

<210> 12

<211> 871

<212> PRT

<213> Chrysosporium lucknowense

<400> 12

Met Gln Leu Pro Ala Ala Ala Gln Trp Leu Leu Thr Leu Pro Ala Lys
 1 5 10 15
 5 Ala Ser Leu Ala Asp Asn His Arg Gln Val His Gln Lys Pro Leu Ala
 20 25 30
 Arg Ser Glu Pro Phe Tyr Pro Ser Pro Trp Met Asn Pro Asn Ala Asp
 35 40 45
 10 Gly Trp Ala Glu Ala Tyr Ala Gln Ala Lys Ser Phe Val Ser Gln Met
 50 55 60
 15 Thr Leu Leu Glu Lys Val Asn Leu Thr Thr Gly Val Gly Trp Gly Ala
 65 70 75 80
 Glu Gln Cys Val Gly Gln Val Gly Ala Ile Pro Arg Leu Gly Leu Arg
 85 90 95
 20 Ser Leu Cys Met His Asp Ser Pro Leu Gly Ile Arg Gly Ala Asp Tyr
 100 105 110
 25 Asn Ser Ala Phe Pro Ser Gly Gln Thr Val Ala Ala Thr Trp Asp Arg
 115 120 125
 Gly Leu Met Tyr Arg Arg Gly Tyr Ala Met Gly Gln Glu Ala Lys Gly
 130 135 140
 30 Lys Gly Ile Asn Val Leu Leu Gly Pro Val Ala Gly Pro Leu Gly Arg
 145 150 155 160
 35 Met Pro Glu Gly Gly Arg Asn Trp Glu Gly Phe Ala Pro Asp Pro Val
 165 170 175
 Leu Thr Gly Ile Gly Met Ser Glu Thr Ile Lys Gly Ile Gln Asp Ala
 180 185 190
 40 Gly Val Ile Ala Cys Ala Lys His Phe Ile Gly Asn Glu Gln Glu His
 195 200 205
 45 Phe Arg Gln Val Pro Glu Ala Gln Gly Tyr Gly Tyr Asn Ile Ser Glu
 210 215 220
 Thr Leu Ser Ser Asn Ile Asp Asp Lys Thr Met His Glu Leu Tyr Leu
 225 230 235 240

Trp Pro Phe Ala Asp Ala Val Arg Ala Gly Val Gly Ser Val Met Cys
 245 250 255
 5 Ser Tyr Gln Gln Val Asn Asn Ser Tyr Ala Cys Gln Asn Ser Lys Leu
 260 265 270
 10 Leu Asn Asp Leu Leu Lys Asn Glu Leu Gly Phe Gln Gly Phe Val Met
 275 280 285
 Ser Asp Trp Gln Ala Gln His Thr Gly Ala Ala Ser Ala Val Ala Gly
 290 295 300
 15 Leu Asp Met Ser Met Pro Gly Asp Thr Gln Phe Asn Thr Gly Val Ser
 305 310 315
 20 Phe Trp Gly Ala Asn Leu Thr Leu Ala Val Leu Asn Gly Thr Val Pro
 325 330 335
 Ala Tyr Arg Leu Asp Asp Met Ala Met Arg Ile Met Ala Ala Leu Phe
 340 345 350
 25 Lys Val Thr Lys Thr Thr His Leu Glu Pro Ile Asn Phe Ser Phe Trp
 355 360 365
 30 Thr Asp Asp Thr Tyr Gly Pro Ile His Trp Ala Ala Lys His Gly Tyr
 370 375 380
 Gln Lys Ile Asn Ser His Val Asp Val Arg Ala Asp His Gly Asn Leu
 385 390 395
 35 Ile Arg Glu Ile Ala Ala Lys Gly Thr Val Leu Leu Lys Asn Thr Gly
 405 410 415
 Ser Leu Pro Leu Asn Lys Pro Lys Phe Val Ala Val Ile Gly Glu Asp
 420 425 430
 40 Ala Gly Ser Ser Pro Asn Gly Pro Asn Gly Cys Ser Asp Arg Gly Cys
 435 440 445
 45 Asn Glu Gly Thr Leu Ala Met Gly Trp Gly Ser Gly Thr Ala Asn Tyr
 450 455 460
 50 Pro Tyr Leu Val Ser Pro Asp Ala Ala Leu Gln Ala Arg Ala Ile Gln
 465 470 475 480
 Asp Gly Thr Arg Tyr Glu Ser Val Leu Ser Asn Tyr Ala Glu Glu Lys

	485										490					495				
5	Thr	Lys	Ala	Leu	Val	Ser	Gln	Ala	Asn	Ala	Thr	Ala	Ile	Val	Phe	Val				
				500					505					510						
	Asn	Ala	Asp	Ser	Gly	Glu	Gly	Tyr	Ile	Asn	Val	Asp	Gly	Asn	Glu	Gly				
			515					520					525							
10	Asp	Arg	Lys	Asn	Leu	Thr	Leu	Trp	Asn	Asn	Gly	Asp	Thr	Leu	Val	Lys				
		530					535					540								
15	Asn	Val	Ser	Ser	Trp	Cys	Ser	Asn	Thr	Ile	Val	Val	Ile	His	Ser	Val				
	545					550					555					560				
	Gly	Pro	Val	Leu	Leu	Thr	Asp	Trp	Tyr	Asp	Asn	Pro	Asn	Ile	Thr	Ala				
					565					570					575					
20	Ile	Leu	Trp	Ala	Gly	Leu	Pro	Gly	Gln	Glu	Ser	Gly	Asn	Ser	Ile	Thr				
				580					585					590						
25	Asp	Val	Leu	Tyr	Gly	Lys	Val	Asn	Pro	Ala	Ala	Arg	Ser	Pro	Phe	Thr				
			595					600					605							
	Trp	Gly	Lys	Thr	Arg	Glu	Ser	Tyr	Gly	Ala	Asp	Val	Leu	Tyr	Lys	Pro				
		610					615					620								
30	Asn	Asn	Gly	Asn	Gly	Ala	Pro	Gln	Gln	Asp	Phe	Thr	Glu	Gly	Val	Phe				
	625					630					635					640				
35	Ile	Asp	Tyr	Arg	Tyr	Phe	Asp	Lys	Val	Asp	Asp	Asp	Ser	Val	Ile	Tyr				
					645					650					655					
	Glu	Phe	Gly	His	Gly	Leu	Ser	Tyr	Thr	Thr	Phe	Glu	Tyr	Ser	Asn	Ile				
				660					665					670						
40	Arg	Val	Val	Lys	Ser	Asn	Val	Ser	Glu	Tyr	Arg	Pro	Thr	Thr	Gly	Thr				
			675					680					685							
45	Thr	Ala	Gln	Ala	Pro	Thr	Phe	Gly	Asn	Phe	Ser	Thr	Asp	Leu	Glu	Asp				
		690					695					700								
	Tyr	Leu	Phe	Pro	Lys	Asp	Glu	Phe	Pro	Tyr	Ile	Tyr	Gln	Tyr	Ile	Tyr				
	705					710					715					720				
50	Pro	Tyr	Leu	Asn	Thr	Thr	Asp	Pro	Arg	Arg	Ala	Ser	Ala	Asp	Pro	His				
					725					730					735					

55

Tyr Gly Gln Thr Ala Glu Glu Phe Leu Pro Pro His Ala Thr Asp Asp
 740 745 750
 5 Asp Pro Gln Pro Leu Leu Arg Ser Ser Gly Gly Asn Ser Pro Gly Gly
 755 760 765
 Asn Arg Gln Leu Tyr Asp Ile Val Tyr Thr Ile Thr Ala Asp Ile Thr
 770 775 780
 10 Asn Thr Gly Ser Val Val Gly Glu Glu Val Pro Gln Leu Tyr Val Ser
 785 790 795 800
 Leu Gly Gly Pro Glu Asp Pro Lys Val Gln Leu Arg Asp Phe Asp Arg
 805 810 815
 Met Arg Ile Glu Pro Gly Glu Thr Arg Gln Phe Thr Gly Arg Leu Thr
 820 825 830
 20 Arg Arg Asp Leu Ser Asn Trp Asp Val Thr Val Gln Asp Trp Val Ile
 835 840 845
 Ser Arg Tyr Pro Lys Thr Ala Tyr Val Gly Arg Ser Ser Arg Lys Leu
 850 855 860
 Asp Leu Lys Ile Glu Leu Pro
 865 870

<210> 13
 <211> 998
 <212> DNA
 <213> Chrysosporium lucknowense

<400> 13

atggcccatc tctccgccac caccgggttc ctcgccctcc cggccctggc cctggcccag 60
 40 ctctcgggca gcggccagac gaccgggtac tgggactgct gcaagccgag ctgcgcctgg 120
 cccggcaagg gcccctcgtc tccggtgcag gcctgcgaca agaacgacaa cccgctcaac 180
 gacggcggtc ccacccggtc cggctgcgac gcggggcgga gcgcctacat gtgctcctcc 240
 45 cagagcccct gggccgtcag cgacgagctg tcgtacggct gggcggccgt caagctcgcc 300
 ggcagctccg agtcgcagtg gtgctgcgcc tgctacgagc tgaccttcac cagcgggccg 360
 gtcgcgggca agaagatgat tgtgcaggcg accaacaccg gtggcgacct gggcgacaac 420
 50 cactttgacc tggccgtgag ttgcctcccc ttctccccgg accgctcaga ttagatgaga 480
 ttagactttg ctcgtaaatc ggtccaagat tcccttgact gaccaacaaa catcatacgg 540
 gcagatcccc ggtggcggtg tcggtatttt caacggtaag ctggtgcccc cggaccctc 600
 55 cccggacccc tccccctttt cctccagcga gccgagttgg gatcgccgag atcgagaact 660

cacacaactt ctctctcgac agcctgcacc gaccagtacg gcgctcccc gaacggctgg 720
 ggcgaccgct acggcgcat ccattccaag gaagagtgcg aatccttccc ggaggccctc 780
 5 aagcccggct gcaactggcg cttcgactgg tacgttgctt tgacataccg gaacccaatt 840
 cctccaaccc cccccctttt ctcccccaac tccgggggta gtcggaatgt cgcgactgac 900
 cctatttcag gttccaaaac gccgacaacc cgtcggtcac cttccaggag gtggcctgcc 960
 10 cgtcggagct cacgtccaag agcggctgct cccgttaa 998

<210> 14

<211> 225

<212> PRT

15 <213> Chrysosporium lucknowense

<400> 14

20 Met His Leu Ser Ala Thr Thr Gly Phe Leu Ala Leu Pro Ala Leu Ala
 1 5 10 15
 Leu Ala Gln Leu Ser Gly Ser Gly Gln Thr Thr Arg Tyr Trp Asp Cys
 20 25 30
 25 Cys Lys Pro Ser Cys Ala Trp Pro Gly Lys Gly Pro Ser Ser Pro Val
 35 40 45
 30 Gln Ala Cys Asp Lys Asn Asp Asn Pro Leu Asn Asp Gly Gly Ser Thr
 50 55 60
 Arg Ser Gly Cys Asp Ala Gly Gly Ser Ala Tyr Met Cys Ser Ser Gln
 65 70 75 80
 35 Ser Pro Trp Ala Val Ser Asp Glu Leu Ser Tyr Gly Trp Ala Ala Val
 85 90 95
 40 Lys Leu Ala Gly Ser Ser Glu Ser Gln Trp Cys Cys Ala Cys Tyr Glu
 100 105 110
 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Ile Val Gln
 115 120 125
 45 Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala
 130 135 140
 50 Ile Pro Gly Gly Gly Val Gly Ile Phe Asn Ala Cys Thr Asp Gln Tyr
 145 150 155 160
 Gly Ala Pro Pro Asn Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser
 165 170 175

55

EP 2 041 294 B1

Lys Glu Glu Cys Glu Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn
180 185 190

5 Trp Arg Phe Asp Trp Phe Gln Asn Ala Asp Asn Pro Ser Val Thr Phe
195 200 205

Gln Glu Val Ala Cys Pro Ser Glu Leu Thr Ser Lys Ser Gly Cys Ser
210 215 220

10

Arg 225

<210> 15

<211> 4190

15

<212> DNA

<213> Chrysosporium lucknowense

<400> 15

20

gcgcttccgg cctgggagag taaaatgacg gaagccgggc cccgtccgac tgcgtttgtc 60

ccaactcgga agcaggcatc gttttttggg cgggaggaag cggtgcaaca cgcactatcg 120

25

ccaaggtgga ctcggcgcaa tctggagggt cggcccgcgg aggacggaat ccgggctgaa 180

tctgcgcaaa ggctgaccct gcgatggtgg gaaaatgtaa atatgtgaag ttataggcat 240

ataggactca gcgatgacat ggaaattgca gaggcattgt ggatttcagc gtttggcatg 300

cattggctcg atctctcgcc ttgtctgatg tgatcccgcc ggagggtgtt cgggtctctgg 360

30

ggaagggacc cccctgggcc cccacactgc cccgcattcat gcctcgccac gactcccgcg 420

cgccgaggaa gaacttcggg tctttgtgac gggagattcc actgagttag cattggccaa 480

ccaagcacac aattactccg tacatacaca gtacttctga ctccgtaaag taaaccgtgt 540

35

gtttcaaaga tcggtaatcc gtaacaggta ctccgtatct aaggtaaatt taccctgtgc 600

acggagcaga acctgaactt cttccccctt cttactcgag tagtcaccct actccaacca 660

gcggcttttc aactcgcaaa gtcttggtta taacagtgca tatacctgca tttcgtatct 720

40

cgctagtgtg aagacgacca cacgcggaca aagaaagaaa aatccaattg ccgatggct 780

cttagtttga ggacagcagc gaaggactac actgcgccgt agtgaccagg ccaagaaacg 840

cgaatcgtat attaacggca aatcaaaatg gatttatatgc catttcgctt ccgggttgcg 900

45

tgctcgtccg aagtctggtg ccgatcgatt gcgaaccccc ggaatcgcg gatgattcct 960

acagccgccg aaaggggggg ggggggaggg gggcttgagc gggacgtgca taacttcgaa 1020

tttctagaat attgcggatt gggttccctt cagccctgcg agcgcgcccc cttctggaac 1080

50

cgcacccttc accggttcca cacacagagg acatgggtgg aaatgtgtac ctgacggttg 1140

cccctttggg acagtggaga ggcggatggt cggataacca tccggagccg cagtgtcgac 1200

55

	caagatcttg	gcttaccatc	gacaccaaca	tgcggactcg	tccctcagtc	atggagcctt	1260
	ggctcgcgga	gcctccgttc	gaagcggcta	tcccgtcctg	ccagcggagg	atctcgtacc	1320
5	gcttccgcga	actgtgaatg	tcctgggtat	aagagcatgg	cgcgaccttg	tctcgtcagg	1380
	aacggggagg	aggagggctt	ggttagggtc	gcgttcgttt	ggagattgct	gagctctgag	1440
	ccttcgggtc	ttggatccct	gcgggtcccc	gtctcctctc	tctctctctc	tctctctctc	1500
10	tctctctctt	cttcccacgc	tcgttcgaca	gacgcctccc	cttcttcgct	ctcctttccc	1560
	tcgcacgtag	cacactaata	gtgcaccatg	cgcgtctcta	gtttggtcgc	ggcccttgct	1620
	accggtggtc	ttgtcgccgc	cacgcctaag	cccaaggggt	cgtcgcccc	tggggccgtg	1680
15	gacgcgaacc	ctttcaaggg	caagacgcag	ttcgtcaacc	cggcatgggc	ggccaagctg	1740
	gaacagacca	aaaaggcggt	cctggccagg	aacgacaccg	tcaatgccgc	caagacggag	1800
	aagggtccagc	agaccagctc	gttcgtctgg	gtctcgagga	tcgccgagct	ctccaacatc	1860
20	gacgacgcca	tcgcggctgc	ccgcaaggcg	cagaagaaga	cgggcaggag	gcagatcgtc	1920
	ggcctgggtg	tctacaacct	tccggaccgc	gactgcagcg	cgggcgagag	cgcgggcgag	1980
	ctcagcagcg	acaagaacgg	gctcgagatc	tacaagactg	agttcgtaa	gcccttcgcc	2040
25	gacaaggtgg	cggccgcaaa	ggacctcgac	ttcgccatcg	tcctggagcc	cgactcgctg	2100
	gccaacctgg	tcaccaacct	gggcatcgag	ttctgcgcca	acgccgcccc	cgtctaccgc	2160
	gagggcatcg	cctatgccat	ctccagcctt	cagcagccaa	acgtgcactt	gtacatcgat	2220
30	gctgcccacg	gcggctggct	cggctgggac	gacaacctgc	cgtggccgc	caaggagttt	2280
	gccgaggtgg	tcaagcttgc	cggcgagggc	aagaagatcc	gcggcttcgt	caccaacgtg	2340
	tccaactaca	accctttcca	cgccgtcgtg	cgcgagaact	ttaccgagtg	gagcaactcg	2400
35	tgggacgagt	ctcactacgc	ctcctcgctc	acaccgttcc	tcgagaaaga	ggggctgccg	2460
	gcacgcttca	tcgtcgacca	gggtcgcggt	gccctcccgg	gagcccgcaa	ggagtgggtga	2520
	gtttcgacca	gattgacct	cgacctatgc	gaccgagatt	gctgacgatt	gaattgcgtg	2580
40	tcccgtcccc	caggggtgaa	tggtgcaacg	tggcaccgc	cggatttggc	cccgcgcccc	2640
	cgaccagggg	caacaacacc	gtcgtcgatg	ctctcgtctg	ggtcaagcct	ggcggcgaga	2700
	gcgacggcga	gtgtggcttg	gctggcgccc	ccaaggccgg	ccagtggttc	gacgagtacg	2760
45	cccagatgct	ggtcgagaat	gcccacccgt	ctgtcgtcca	caagtggtag	ataaatTTTg	2820
	gagtccgaga	aggggtcccag	atagactttt	gttttaaaac	aaaatgcaag	gtgtcgacag	2880
	atactggctt	aacattaacc	aagcaccatg	aacatgactt	gtcaacatat	tgatacattc	2940
50	cgctgctttc	ccatacgtgc	tctcaggtct	cagggatcaa	atggataggt	cggtaatgca	3000
	aaacgatcca	ttggatatcc	agaagagaga	aaaaaaaaag	gacatgcatg	ccttgtctgt	3060

55

catcatgagg aaacaaagga aaaacaaacg atcgtcgtgt tccaacaagc tttccaagac 3120
 cacaagaccc atccaccaac acaaccaaac gacaagcaat acgatggacc gccgttggtc 3180
 5 catctctcaa gagctgacta aacgaacagt cgttgaaatc atcctacatg agtacgccgc 3240
 accacctgtt atcgtgtaaa ccaaaatggc tgttaaagtg catcatctct taggtatgat 3300
 cgtaagtacc ggtcacggtc acggatcagg gatggttctc aattcgtgtg tcgcgtagcc 3360
 10 gccgccgtat ctggacaaga cttcttgatg tgctccgaaa ccgcttttgc cgccctaata 3420
 atctgtagcc ttcttacctg gtggtgcctt gaaagacgcg gcaggcaaca cttcgagggt 3480
 ctgtggcgca ccagcaccag gctgtggtga tgccccggaa ccggtcgtcg acttgctcgc 3540
 15 ggtgtcctcg gctgggtggg atgggggtga tgagggttg gaggggtgtg ttgcgcccgc 3600
 aacatccggc tccggctccg gaccgtccac agacattgga cctgcgagca tgactcgtgc 3660
 cttcagccag accaaagcca tgccatcatc gcctctgccg acgctgttga gcgggaggct 3720
 20 gatgtttctc gccagaactg cgggctgtac ggccatgacc atgggctgtt cggctctggcc 3780
 gtcttgccgc ggtttctccc tgccagcttg ttgtgcgcgg tgcttgcgag attcgacttc 3840
 gacctgggcg tggcagaggg tgacgagggg cgttgacgcc ttgatctcct tgctccccat 3900
 25 gtccttcac ccgtacaggc ggacgggtgc catacgcgtc cacagcctgc acgagaacct 3960
 cagggcgctg tcaatgagtt ctgtcaactt gctctccagc ctctctatgc cgcgagcatc 4020
 ctgatcctgg agcagaaacc gtgccgagcc tccgaggaaa cgctccttca gcttccgcgc 4080
 30 gtagtttagg cgtgattcaa caaacgtccg gcgggactcg ttgttgcccc cagcagcgac 4140
 gtccttgatg ctgaagccgc cgtcggcgaa caggcgcac atctgggccc 4190

<210> 16

<211> 381

35 <212> PRT

<213> Chrysosporium lucknowense

<400> 16

40

Met Arg Val Ser Ser Leu Val Ala Ala Leu Ala Thr Gly Gly Leu Val
 1 5 10 15

45

Ala Ala Thr Pro Lys Pro Lys Gly Ser Ser Pro Pro Gly Ala Val Asp
 20 25 30

Ala Asn Pro Phe Lys Gly Lys Thr Gln Phe Val Asn Pro Ala Trp Ala
 35 40 45

50

Ala Lys Leu Glu Gln Thr Lys Lys Ala Phe Leu Ala Arg Asn Asp Thr
 50 55 60

55

Val Asn Ala Ala Lys Thr Glu Lys Val Gln Gln Thr Ser Ser Phe Val

	65		70		75		80									
5	Trp	Val	Ser	Arg	Ile 85	Ala	Glu	Leu	Ser	Asn 90	Ile	Asp	Asp	Ala	Ile 95	Ala
	Ala	Ala	Arg	Lys 100	Ala	Gln	Lys	Lys	Thr 105	Gly	Arg	Arg	Gln	Ile 110	Val	Gly
10	Leu	Val	Leu 115	Tyr	Asn	Leu	Pro	Asp 120	Arg	Asp	Cys	Ser	Ala 125	Gly	Glu	Ser
15	Ala	Gly 130	Glu	Leu	Ser	Ser	Asp 135	Lys	Asn	Gly	Leu	Glu 140	Ile	Tyr	Lys	Thr
	Glu 145	Phe	Val	Lys	Pro	Phe 150	Ala	Asp	Lys	Val	Ala 155	Ala	Ala	Lys	Asp	Leu 160
20	Asp	Phe	Ala	Ile	Val 165	Leu	Glu	Pro	Asp	Ser 170	Leu	Ala	Asn	Leu	Val 175	Thr
25	Asn	Leu	Gly	Ile 180	Glu	Phe	Cys	Ala	Asn 185	Ala	Ala	Pro	Val	Tyr 190	Arg	Glu
	Gly	Ile	Ala 195	Tyr	Ala	Ile	Ser	Ser 200	Leu	Gln	Gln	Pro	Asn 205	Val	His	Leu
30	Tyr	Ile 210	Asp	Ala	Ala	His	Gly 215	Gly	Trp	Leu	Gly	Trp 220	Asp	Asp	Asn	Leu
35	Pro 225	Leu	Ala	Ala	Lys	Glu 230	Phe	Ala	Glu	Val	Val 235	Lys	Leu	Ala	Gly	Glu 240
	Gly	Lys	Lys	Ile	Arg 245	Gly	Phe	Val	Thr	Asn 250	Val	Ser	Asn	Tyr	Asn 255	Pro
40	Phe	His	Ala	Val 260	Val	Arg	Glu	Asn	Phe 265	Thr	Glu	Trp	Ser	Asn 270	Ser	Trp
45	Asp	Glu	Ser 275	His	Tyr	Ala	Ser	Ser 280	Leu	Thr	Pro	Phe	Leu 285	Glu	Lys	Glu
	Gly	Leu 290	Pro	Ala	Arg	Phe	Ile 295	Val	Asp	Gln	Gly	Arg 300	Val	Ala	Leu	Pro
50	Gly 305	Ala	Arg	Lys	Glu	Trp 310	Gly	Glu	Trp	Cys	Asn 315	Val	Ala	Pro	Ala	Gly 320

55

Phe Gly Pro Ala Pro Thr Thr Arg Val Asn Asn Thr Val Val Asp Ala
325 330 335

5 Leu Val Trp Val Lys Pro Gly Gly Glu Ser Asp Gly Glu Cys Gly Leu
340 345 350

10 Ala Gly Ala Pro Lys Ala Gly Gln Trp Phe Asp Glu Tyr Ala Gln Met
355 360 365

15 Leu Val Glu Asn Ala His Pro Ser Val Val His Lys Trp
370 375 380

15 <210> 17
<211> 3000
<212> DNA
<213> Chrysosporium lucknowense

20 <400> 17

cgcgggcccg tctttgaacg cttgagaagc gcacgggtgaa gaaccatcaa ctccgattcc 60
gctcctcatc ctcccacgaa gccgattgaa atagccacag cggctatgta cggattactc 120
25 tgctccgttt gcacatccat acacagcgct attttttaaaa gttcaggacg gccaaagccc 180
gttcttgga a cggacgaccc ggattccgaa agctccagcg ctcaatgcgg tcagtcgtgg 240
cgctgatcct gctgatctgc tgatctcata aacccgcaac ttcaactttt cactttgaag 300
30 cgtatacacg cagcgctctt ttcaccggcg cattcatact cgcaaattaa ccgctaatat 360
cctcgcactt ggataatgtg tagccgacac ggaggagggg ggttgggggg gggttggggg 420
gagacatgat ggtctgcca acggatatta ttattttgtt gttttgtata attactgcgg 480
35 caacattctc aaagggggcg tgcctcgcgg cgggaaagcc catgacagag aattggacag 540
ctccaagctc gcgatatact ctaacaacgg cgtgactcgg caatgaaggc ctgccgctcg 600
agtgataggg cgaagtaaaa cggacgttac atgcggcact tagccggctg atgccggaga 660
40 atacgggatt caacgataca atcacacgat gcgacacacc tcggcgactt ggcgctctat 720
ggaagaaggc tgggttaaag ctggcgtaga ttttgcgctt cttggtttct taaccgggtt 780
atttctattt ctcatatgcc gcgagcgaat gcgggggtgca gagcgcccgg gagtcgatgg 840
45 tcctatcaga caagagcctg gccccggaac ctgggataat agaagccaaa ttaagccatg 900
ggagtatcgt ccgggggtag gaaccgcacg ggcaactaga ggaggaagaa tttgggtataa 960
agggaggacg gcggaacagg cttgatggac atgaatcaga agacgacact gggcaactaa 1020
50 acagcttgca gcagagtttt gtgccttgca taggccctcg atatcatggt ctcgttcact 1080
ctcctcctca cggtcacgc cgctgcggtg acgacggcca gccctctcga ggtggtcaag 1140
cgcggcaccc agccgggcac gggcaccac gaggggtact tctactcgtt ctggaccgac 1200
55 ggccgtggct cggtcgactt caaccccggg ccccgcggtt cgtacagcgt cacctggaac 1260

aacgtcaaca actgggttgg cggcaagggc tggaaccccg gcccgccgcg caagattgcg 1320
 tacaacggca cctggaacaa ctacaacgtg aacagctgtg cgttgtcctc ctctttctcc 1380
 5 ctttcgcttg ttttccttga tgattgggat ccattttaaa agagaaggaa aaaaaaaca 1440
 aaggaaaata gaagataact aacgccaagc tctggcagac ctcgccctgt acggctggac 1500
 tcgcaacccg ctggtcgagt attacatcgt ggaggcatac ggcacgtaca acccctcgtc 1560
 10 gggcacggcg cggctgggca ccatcgagga cgacggcggc gtgtacgaca tctacaagac 1620
 gacgcggtac aaccagccgt ccatcgaggg gacctccacc ttcgaccagt actggtccgt 1680
 ccgccgccag aagcgcgctc gcggcactat cgacacgggc aagcactttg acgagtggaa 1740
 15 gcgccagggc aacctccagc tcggcacctg gaactacatg atcatggcca ccgagggcta 1800
 ccagagctct ggttcggcca ctatcgaggt ccgggaggcc taaagaagcc aggcgccttt 1860
 cttttgtttt gcaggagggg gtagaggggg ggggggaggg aaaacgaaa gtagcaggg 1920
 20 gggttttatgc cggcagccgt gggccattcg agtgaacct gtatctctct ctctcccaag 1980
 tctccgggct ctttctcaga gaacttcaat atgtctgggg acaaaccacc ttgtgaaata 2040
 caacggtaat tatctaagtt tgagtgcctt atcgtatgct tctgaaaatt tcctgctcct 2100
 25 tgatacaagt cggtttgagc cgagccaatg agactgtgtc gattgataga ggcctgaag 2160
 gatcaagcgc gatgcaacaa ttaagcatga ctacgtgcct agctgcagat aaatggaagc 2220
 cactcaccaa ggtcaacccc gcatactggc acgtaagaac cttccgtgta caaggcccaa 2280
 ccgactcaca tatctatctg cttgggtttt gggatgcggt tttttacca caaaacaaat 2340
 30 ttgatacaat gctctgctgt gcccggttg ctgagaccaa gccgtaatca gcgggcaggg 2400
 aatcgagtag gtcacgcctg ttgcttggtc tagaacaac taatattaaa aagccttggtg 2460
 ctcggcacac atacagaact cgacctgagg catgttcttg gaaggcggct agccagtcaa 2520
 35 gtctggcacc aggccttggt ctcgtcgagg ataccgaggg cgaggaggat gaggaagacc 2580
 tctttctcgc ctcagatctc ttaggggacg aagaagacaa cgccggagcc acacaataat 2640
 taggtctcat atcagacgtt tcggcctggc cgagctaata tgtctaatta tgcccatcag 2700
 40 ccgtatgtcg aggcagggtg caccgatacg ctcgccgcgc gcctcattc atctccgact 2760
 gggcacaatg tcgccatctc ggccgtcaag gtggtgcaag atacctatta tgcaagcaga 2820
 ggatcagatg gcgggccgat acgagcggct gctccggctt gcgagaaagc cgcttcgcag 2880
 45 caaggtatcg tggcaggccg ccattttcgg ttgggtattc tttgtcttgt ttgcttcgta 2940
 attatgtcct ggctggcatt gtgggaaggg gcgaacctct tgatttccga tgggggtcga 3000

<210> 18

<211> 218

<212> PRT

<213> *Chrysosporium lucknowense*

<400> 18

Met Val Ser Phe Thr Leu Leu Leu Thr Val Ile Ala Ala Ala Val Thr
 1 5 10 15
 Thr Ala Ser Pro Leu Glu Val Val Lys Arg Gly Ile Gln Pro Gly Thr
 5 20 25 30
 Gly Thr His Glu Gly Tyr Phe Tyr Ser Phe Trp Thr Asp Gly Arg Gly
 10 35 40 45
 Ser Val Asp Phe Asn Pro Gly Pro Arg Gly Ser Tyr Ser Val Thr Trp
 50 55 60
 Asn Asn Val Asn Asn Trp Val Gly Gly Lys Gly Trp Asn Pro Gly Pro
 15 65 70 75 80
 Pro Arg Lys Ile Ala Tyr Asn Gly Thr Trp Asn Asn Tyr Asn Val Asn
 20 85 90 95
 Ser Tyr Leu Ala Leu Tyr Gly Trp Thr Arg Asn Pro Leu Val Glu Tyr
 100 105 110
 Tyr Ile Val Glu Ala Tyr Gly Thr Tyr Asn Pro Ser Ser Gly Thr Ala
 25 115 120 125
 Arg Leu Gly Thr Ile Glu Asp Asp Gly Gly Val Tyr Asp Ile Tyr Lys
 130 135 140
 Thr Thr Arg Tyr Asn Gln Pro Ser Ile Glu Gly Thr Ser Thr Phe Asp
 30 145 150 155 160
 Gln Tyr Trp Ser Val Arg Arg Gln Lys Arg Val Gly Gly Thr Ile Asp
 35 165 170 175
 Thr Gly Lys His Phe Asp Glu Trp Lys Arg Gln Gly Asn Leu Gln Leu
 180 185 190
 Gly Thr Trp Asn Tyr Met Ile Met Ala Thr Glu Gly Tyr Gln Ser Ser
 40 195 200 205
 Gly Ser Ala Thr Ile Glu Val Arg Glu Ala
 210 215

Claims

1. An enzyme composition for cellulose hydrolysis, said composition comprising a betaglucosidase (BGL) obtainable by the expression of a *bg1* gene having the sequence of SEQ ID NO: 11.
2. A method of producing fermentable sugars from lignocellulosic material, said method comprising
 - (a) providing an enzyme composition for cellulose hydrolysis, wherein said composition comprises a beta-glucosidase (BGL) obtainable by the expression of a *bg1* gene having the sequence of SEQ ID NO: 11;
 - (b) applying said composition to lignocellulosic material to produce fermentable sugars.
3. A method of producing a fermentation product or a starting material for a fermentation product from a fermentable

sugar, wherein said method comprises

- (a) providing an enzyme composition for cellulose hydrolysis, wherein said composition comprises a beta-glucosidase (BGL) obtainable by the expression of a *bgl1* gene having the sequence of SEQ ID NO: 11;
- (b) applying said composition to lignocellulosic material to produce a fermentable sugar; and
- (c) fermenting said fermentable sugar to produce a fermentation product.

4. The method according to claim 3, wherein said fermentation product is a biofuel, or is selected from the group consisting of lactic acid, plastics, specialty chemicals, organic acids, solvents, animal feed supplements, pharmaceuticals, vitamins, amino acids, industrial enzymes, and chemical feedstocks.

5. A method of producing energy from a fermentable sugar, said method comprising

- (a) providing an enzyme composition for cellulose hydrolysis, wherein said composition comprises a beta-glucosidase (BGL) obtainable by the expression of a *bgl1* gene having the sequence of SEQ ID NO: 11;
- (b) applying said composition to lignocellulosic material to produce a fermentable sugar,
- (c) fermenting said fermentable sugar to produce a combustible fermentation product;
- (d) combusting said combustible fermentation product to produce energy.

6. The method according to claim 2, 3 or 5, wherein the fermentable sugar is selected from the group consisting of glucose, xylose, arabinose, galactose, mannose, rhamnose, sucrose and fructose.

7. The method according to claim 2, 3 or 5, wherein the lignocellulosic material is selected from the group consisting of orchard prunings, chaparral, mill waste, urban wood waste, municipal waste, logging waste, forest thinnings, short-rotation woody crops, industrial waste, wheat straw, oat straw, rice straw, barley straw, rye straw, flax straw, soy hulls, corn gluten feed, oat hulls, sugar cane, corn stover, corn stalks, corn cobs, corn husks, prairie grass, gamagrass, foxtail, sugar beet pulp, citrus fruit pulp, seed hulls, cellulosic animal wastes, lawn clippings, cotton, seaweed, trees, shrubs, grasses, wheat, sugar cane bagasse, corn, corn kernel, fiber from kernels, products and by-products from wet or dry milling of grains, municipal solid waste, waste paper, yard waste, herbaceous material, agricultural residues, forestry residues, pulp, paper mill residues, branches, bushes, canes, energy crops, forests, fruits, flowers, grains, grasses, herbaceous crops, leaves, bark, needles, logs, roots, saplings, shrubs, switch grasses, trees, vegetables, fruit peels, vines, wheat middlings, hard and soft woods, organic waste materials generated from agricultural processes, forestry wood waste, and combinations thereof.

8. The method according to claim 5, wherein said combustible fermentation product is an alcohol.

9. The method according to claims 2, 3 or 5, wherein the lignocellulosic material is subjected to a pretreatment prior to being exposed to enzymes, wherein optionally the pretreatment comprises exposing the lignocellulosic biomass to an acid, base, solvent, heat, peroxide, ozone, mechanical shredding, grinding, milling, rapid depressurization, or a combination thereof, wherein optionally said solvent is an acetone/ethanol mixture.

10. A protein obtainable by the expression of a *bgl1* gene having the sequence of SEQ ID NO: 11.

11. A *bgl1* gene having the nucleic acid sequence of SEQ ID NO: 11.

Patentansprüche

1. Enzymzusammensetzung zur Cellulosehydrolyse, wobei die Zusammensetzung eine Betaglucosidase (BGL) umfasst, die durch die Expression eines *bgl1*-Gens mit der Sequenz von SEQ ID NO: 11 erhältlich ist.

2. Verfahren zum Herstellen fermentierbarer Zucker aus lignocellulosehaltigem Material, wobei das Verfahren umfasst

- (a) Bereitstellen einer Enzymzusammensetzung zur Cellulosehydrolyse, wobei die Zusammensetzung eine beta-Glucosidase (BGL) umfasst, die durch die Expression eines *bgl1*-Gens mit der Sequenz von SEQ ID NO: 11 erhältlich ist;
- (b) Anwenden der Zusammensetzung auf lignocellulosehaltiges Material, um fermentierbare Zucker herzustellen.

3. Verfahren zum Herstellen eines Fermentationsprodukts oder eines Ausgangsmaterials für ein Fermentationsprodukt aus einem fermentierbaren Zucker, wobei das Verfahren umfasst

- (a) Bereitstellen einer Enzymzusammensetzung zur Cellulosehydrolyse, wobei die Zusammensetzung eine beta-Glucosidase (BGL) umfasst, die durch die Expression eines *bgl1*-Gens mit der Sequenz von SEQ ID NO: 11 erhältlich ist;
- (b) Anwenden der Zusammensetzung auf lignocellulosehaltiges Material, um einen fermentierbaren Zucker herzustellen; und
- (c) Fermentieren des fermentierbaren Zuckers, um ein Fermentationsprodukt herzustellen.

4. Verfahren nach Anspruch 3, wobei das Fermentationsprodukt ein Biotreibstoff ist, oder ausgewählt ist aus der Gruppe bestehend aus Milchsäure, Kunststoffen, Spezialchemikalien, organischen Säuren, Lösungsmitteln, Tierfutterergänzungsmitteln, Pharmazeutika, Vitaminen, Aminosäuren, Industrieenzymen und chemischen Ausgangsmaterialien.

5. Verfahren zum Herstellen von Energie aus einem fermentierbaren Zucker, wobei das Verfahren umfasst

- (a) Bereitstellen einer Enzymzusammensetzung für Cellulosehydrolyse, wobei die Zusammensetzung eine beta-Glucosidase (BGL) umfasst, die durch Expression eines *bgl1*-Gens mit der Sequenz von SEQ ID NO: 11 erhältlich ist;
- (b) Anwenden der Zusammensetzung auf lignocellulosehaltiges Material, um einen fermentierbaren Zucker herzustellen;
- (c) Fermentieren des fermentierbaren Zuckers, um ein brennbares Fermentationsprodukt herzustellen;
- (d) Verbrennen des brennbaren Fermentationsprodukts, um Energie herzustellen.

6. Verfahren nach Anspruch 2, 3 oder 5, wobei der fermentierbare Zucker ausgewählt ist aus der Gruppe bestehend aus Glucose, Xylose, Arabinose, Galactose, Mannose, Rhamnose, Saccharose und Fructose.

7. Verfahren nach Anspruch 2, 3 oder 5, wobei das lignocellulosehaltige Material ausgewählt ist aus der Gruppe bestehend aus Obstgartenschnitten, Chaparral, Mühlenabfällen, städtischem Holzabfall, kommunalem Abfall, Abholzungsabfall, Forstausdünnung, Energieholz im Kurzumtrieb, Industrieabfall, Weizenstroh, Haferstroh, Reisstroh, Gerstenstroh, Roggenstroh, Flachsstroh, Sojabohnenhülsen, Maisglutenfutter, Haferspelzen, Zuckerrohr, Maiserntesteresten, Maisstängeln, Maiskolben, Maiskolbenblättern, Präriegras, *Tripsacum*-Gräsern, Foxtail, Zuckerrübenschnitzeln, Zitrusfrüchtetrester, Samenspelzen, cellulose-haltigen Tierexkrementen, Rasenschnitt, Baumwolle, Meeresalgen, Bäumen, Büschen, Gräsern, Weizen, Zuckerrohrbagasse, Mais, Maiskörnern, Fasern aus Körnern, Produkten und Nebenprodukten aus Nass- oder Trockenmahlen von Getreide, kommunalem Festabfall, Abfallpapier, Gartenabfall, Krautmaterial, landwirtschaftlichen Reststoffen, forstwirtschaftlichen Reststoffen, Pulpe, Papiermühlenreststoffen, Ästen, Gebüsch, Rohr, Energienutzpflanzen, Wäldern, Früchten, Blumen, Körnern, Gräsern, krautigen Nutzpflanzen, Blättern, Rinde, Nadeln, Holzstämmen, Wurzeln, Schösslingen, Sträuchern, Rutenhirsen, Bäumen, Gemüse, Obstschalen, Kletterpflanzen, Weizendunstmehl, Hart- und Weichhölzern, aus landwirtschaftlichen Verfahren erzeugten organischen Abfallmaterialien, forstwirtschaftlichem Holzabfall, und Kombinationen davon.

8. Verfahren nach Anspruch 5, wobei das brennbare Fermentationsprodukt ein Alkohol ist.

9. Verfahren nach Ansprüchen 2, 3, oder 5, wobei das lignocellulosehaltige Material einer Vorbehandlung unterzogen wird, bevor es Enzymen ausgesetzt wird, wobei gegebenenfalls die Vorbehandlung Aussetzen der lignocellulosehaltigen Biomasse einer Säure, einer Base, einem Lösungsmittel, Hitze, Peroxid, Ozon, mechanischem Schreddern, Reiben, Mahlen, schnellem Druckablassen, oder einer Kombination davon umfasst, wobei gegebenenfalls das Lösungsmittel ein Aceton/Ethanolgemisch ist.

10. Protein, erhältlich durch die Expression eines *bgl1*-Gens mit der Sequenz von SEQ ID NO: 11.

11. *bgl1*-Gen mit der Nukleinsäuresequenz von SEQ ID NO: 11.

Revendications

1. Composition enzymatique pour l'hydrolyse de la cellulose, ladite composition comprenant un bêta-glucosidase (BGL) pouvant être obtenu par l'expression d'un gène *bglI* présentant la séquence de SEQ ID NO: 11.
2. Procédé de production de sucres fermentescibles à partir de matière lignocellulosique, ledit procédé consistant à
 - (a) fournir une composition enzymatique pour l'hydrolyse de la cellulose, dans lequel ladite composition comprend un bêta-glucosidase (BGL) pouvant être obtenu par l'expression d'un gène *bglI* présentant la séquence de SEQ ID NO: 11 ;
 - (b) appliquer ladite composition à la matière lignocellulosique pour produire des sucres fermentescibles.
3. Procédé de production d'un produit de fermentation ou d'une matière de départ pour un produit de fermentation à partir d'un sucre fermentescible, dans lequel ledit procédé consiste à
 - (a) fournir une composition enzymatique pour l'hydrolyse de la cellulose, dans lequel ladite composition comprend un bêta-glucosidase (BGL) pouvant être obtenu par l'expression d'un gène *bglI* présentant la séquence de SEQ ID NO: 11 ;
 - (b) appliquer ladite composition à la matière lignocellulosique pour produire un sucre fermentescible ; et
 - (c) faire fermenter ledit sucre fermentescible pour produire un produit de fermentation.
4. Procédé selon la revendication 3, dans lequel ledit produit de fermentation est un biocarburant ou est choisi dans le groupe comprenant l'acide lactique, les plastiques, les produits chimiques de spécialité, les acides organiques, les solvants, les compléments alimentaires pour animaux, les vitamines, les acides aminés, les enzymes industriels et les matières premières chimiques.
5. Procédé de production d'énergie à partir d'un sucre fermentescible, ledit procédé consistant à
 - (a) fournir une composition enzymatique pour l'hydrolyse de la cellulose, dans lequel ladite composition comprend un bêta-glucosidase (BGL) pouvant être obtenu par l'expression d'un gène *bglI* présentant la séquence de SEQ ID NO: 11 ;
 - (b) appliquer ladite composition à la matière lignocellulosique pour produire un sucre fermentescible ;
 - (c) faire fermenter ledit sucre fermentescible pour produire un produit de fermentation combustible ;
 - (d) brûler le produit de fermentation combustible pour produire de l'énergie.
6. Procédé selon la revendication 2, 3 ou 5, dans lequel le sucre fermentescible est choisi dans le groupe comprenant le glucose, le xylose, l'arabinose, le galactose, le mannose, le rhamnose, le sucrose et le fructose.
7. Procédé selon la revendication 2, 3 ou 5, dans lequel la matière lignocellulosique est choisie dans le groupe comprenant les produits d'égavage de verger, le chaparral, les déchets de meunerie, les déchets de bois urbain, les déchets municipaux, les déchets d'exploitation forestière, les coupes d'éclaircie, les cultures de bois à rotation courte, les déchets industriels, la paille de blé, la paille d'avoine, la paille de riz, la paille d'orge, la paille de seigle, la paille de lin, les coques de soja, le gluten de maïs, les coques d'avoine, la canne à sucre, la paille de maïs, les tiges de maïs, les rafles de maïs, les spathes de maïs, l'herbe de prairie, le tripsacum dactyloïdes, la sétaire, la pulpe de betterave sucrière, la pulpe d'agrumes, les coques de graines, les déchets celluloseux d'origine animale, les déchets de tonte, le coton, les algues, les arbres, les arbustes, les herbes, le blé, la bagasse de canne à sucre, le maïs, les grains de maïs, les fibres des grains, les produits et sous-produits de mouture humide ou sèche de grains, les déchets municipaux solides, les vieux papiers, les déchets de jardin, les matières herbacées, les résidus agricoles, les résidus de sylviculture, la pulpe, les résidus de papeterie, les branches, les buissons, les cannes, les cultures énergétiques, les forêts, les fruits, les fleurs, les graines, les herbes, les cultures herbacées, les feuilles, l'écorce, les aiguilles, les rondins, les racines, les gaules, les arbustes, le panic érigé, les arbres, les légumes, les épiluchures de fruits, les vignes, les remoulages de blé, les bois durs et tendres, les déchets organiques produits par les procédés agricoles, les déchets de bois de sylviculture, et une combinaison de ceux-ci.
8. Procédé selon la revendication 5, dans lequel ledit produit de fermentation combustible est un alcool.
9. Procédé selon les revendications 2, 3 ou 5, dans lequel la matière lignocellulosique est soumise à un prétraitement avant d'être exposée aux enzymes, dans lequel le prétraitement comprend optionnellement l'exposition de la bio-

EP 2 041 294 B1

masse lignocellulosique à un acide, une base, un solvant, de la chaleur, le peroxyde, l'ozone, un déchiquetage mécanique, un meulage, un fraisage, une dépressurisation rapide, ou une combinaison de ceux-ci, dans lequel ledit solvant est optionnellement un mélange d'acétone et d'éthanol.

5 **10.** Protéine pouvant être obtenue par l'expression d'un gène *bgII* présentant la séquence de SEQ ID NO: 11.

11. Gène *bgII* présentant la séquence d'acide nucléique de SEQ ID NO: 11.

10

15

20

25

30

35

40

45

50

55

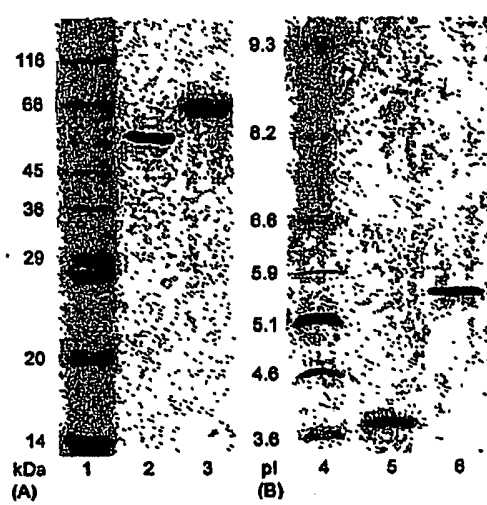


FIGURE 1

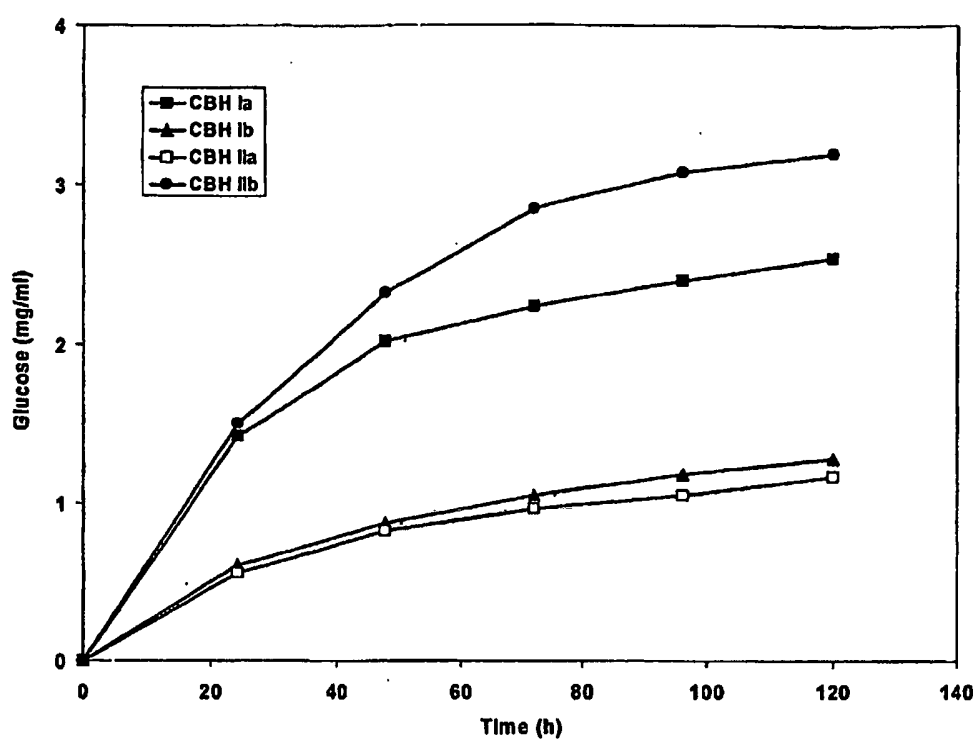


FIGURE 2

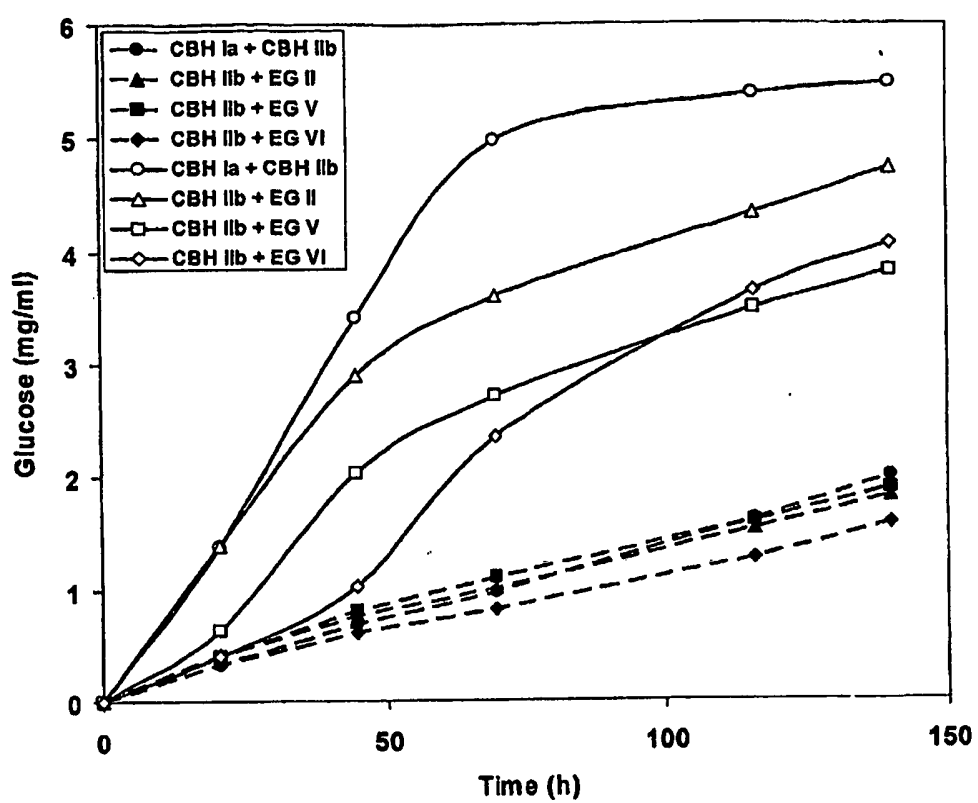


FIGURE 3

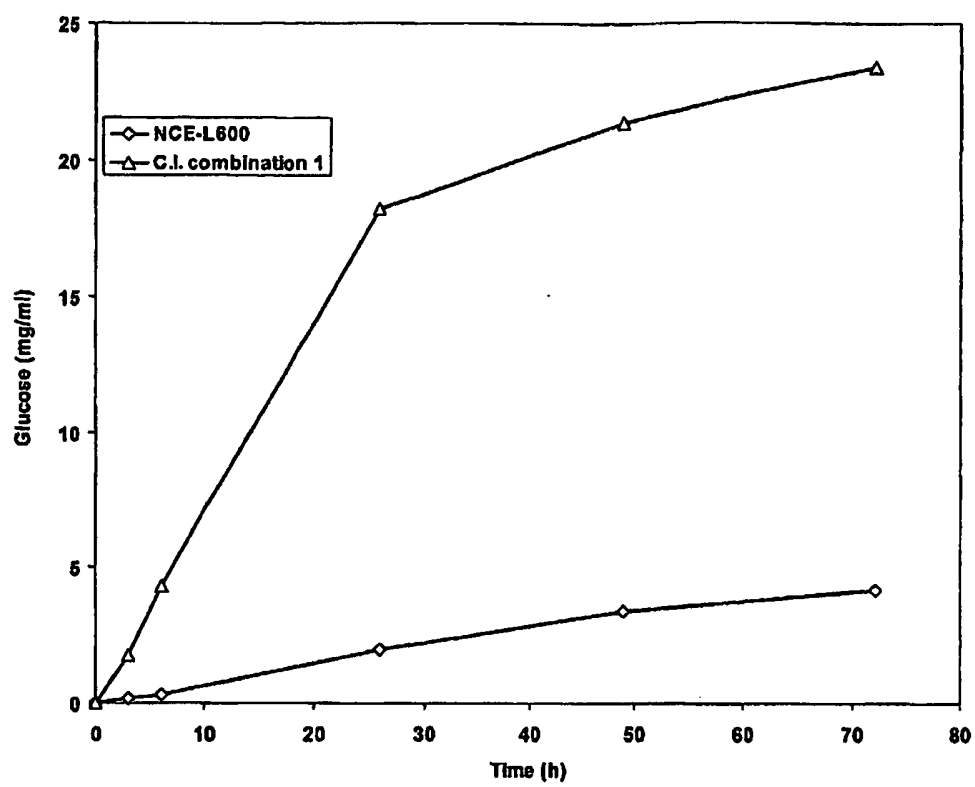


FIGURE 4

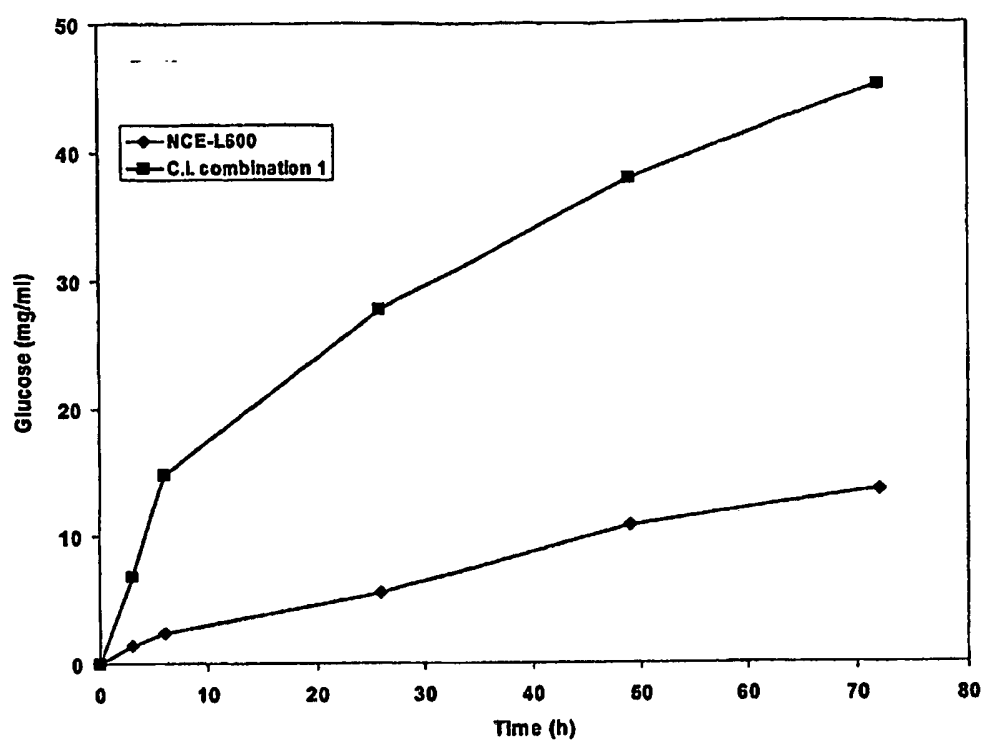


FIGURE 5

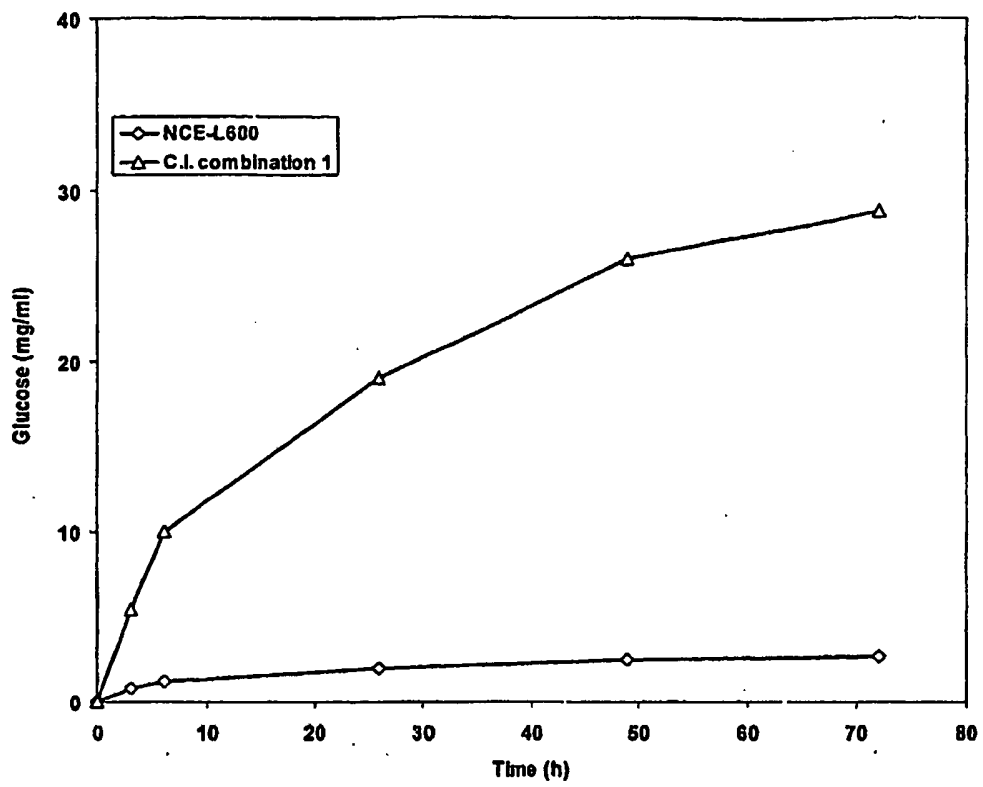


FIGURE 6

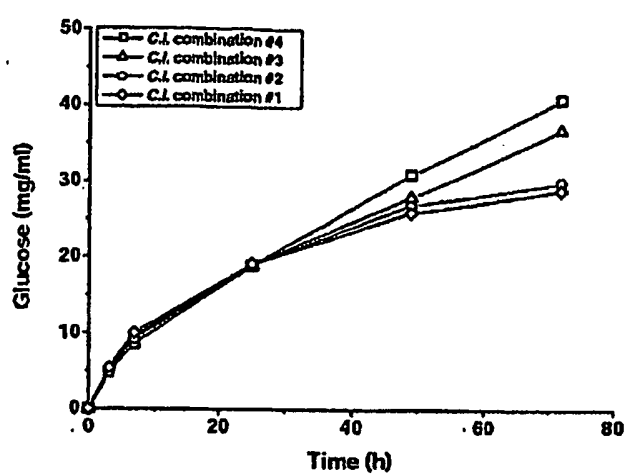


FIGURE 7

FIGURE 8

Translation of Contig 2370 14521-20840 cbh2(1-6360)

```

1      ctcagattctagggtagggcgaggagcagaggcgaaaattgggttagaataatgaggag
61     ctagggttgtaaaactcaagaactcttgcctctgtctcttagtctctctctctgggaaa
121    aggggggttttccgaaagcggcgctatagcagccagaggctacttctctgcttggat
181    ggcccttgccaccgttctgtttcccggttctgcaattgcgacgttgcggcgcaacctagg
241    tcctaataattaggtagataatttcggttagaggttagttaaattatgcttcagtagaanaat
301    cgttgcctccacgtctcgcgaaccttgcgaaccttcgccacattgaagatagcattgtctg
361    agttgattttaacccttccagagacgataataatagtgcaagttctcttgatcggaatca
421    tcgacattcggtatttcccttaattatataagatttcggccacggaacggggcccgga
481    gcaggttgaaaccgcgcaaaacctcaaccgagtcacctcggtccatgtttgtcatggaat
541    caggctccgaatcccgatcagtcagtcagttctgggtggctatggacgaggagttacggc
601    cagtcgctcccggtgttctgggggggtgatcaacaggaggagagatctgagatcgaaacta
661    caccattgatttatcgacgcataatcaagtttaataaaaccaaacagcgtgtttgggtg
721    ctaccaccgaatgcgagatccgggctagcccgcggaaggatgatggccacagatctagcg
781    tcatgtatgatttatcctatgcatctatcttctgctctggttcggcaaacctg
841    accgagagacgactcgcaacctgacacttggcaaaagacatttctggtgacagcgggag
901    aactccagcgagggaagtgcggcagagatgcggatgagaagacaacgcgagagcgtgcgg
961    cgttggctctccacgaatcggaagcgcactcttccgttggccaatctccgggataaatcc
1021   cagcggcggggtcacgtcacgtttcatggggaggcgcggaacagccatccagccaggccat
1081   ggaagagaacaattcttgggggttagcgacgcagcaaaagggggggggggaagcgggag
1141   gggaaagaagtggatttagagcacgcacccggaacacgcatttgggcccctgccaacaaaca
1201   ccacacccccgcgtcctgggagcaagacatccaggatgcaaacccagtaggggatgccaaga
1261   agcatctacggcaccatctgcccggcgctcgctgttagagtcctgggaccccgccaatgg
1321   ggccgtgctgggcccctgcgggcaatgctggcgagcgccatcaacaaacttgcctcgggg
1381   agggggcccgattttattgattagcaaaaaaacaattaaattacccttccattccagcaga
1441   gcttctcctccacgcggcgggcggaacgcgttctggacggcggtacactacaaccgcgggg
1501   ctccagtcctcggtgctggggtgcagatcacgaccgggaagagaatgatcgcggtctga
1561   cgccgggtacggagtagtgagccgcaacacacagccgagtggaacgtagatctcaatgcg
1621   ttcaagcaacacagcaacacccctggacgagtcctctcctccctaccaccccctcccccc
1681   tgccttggcgcggaacggggcggtaccccgagatttctactcgtactgacaccccacatc
1741   tattcccgctggcgctgcggcagtcggggcggtccggccaagactctcgtgcaagatac
1801   cggcagcaaatcggaattaacgctggctgatcaattccaagtcaagggaagtggtatg
1861   gaaagtcggctcagtttccactgcccccgcagggcaggttccggatctggacagcagtc
1921   tccgaatcttggcagagactcatgataatataaaaaggcaaatgaggcgcgcttgg
1981   acaggtcoattotcccacgcgtcaaccagcctccaattctcagaaagtctgtgctctct
2041   cgcagtcgcagtcgaagTGAAGCAGTACCTCCAGTACCTCGCGCGACCTCGCCCTCGT
      M K Q Y L Q Y L A A T L P L V
2101   GGGCCTGGCCACGCGCCAGCAGGCGGGTAACCTGCAGACGAGACTACCCCAAGCTCAC
      G L A T A Q Q A G N L Q T E T H P K L T
2161   TTGGTCGAAGTGACGCGCCCGGGATCCTGCCAACAGGTCAACGCGGAGTCTGTCATCGA
      W S K C T A P G S C Q Q V N G E V V I D
2221   CTCCAATGGCGCTGGGTGACGACGAGAAACGCGCAGAACTGCTACGACGCGCAACGAGTG
      S N W R W V R D E N A Q N C Y D G N Q W
2281   GACCAACGCTTGCGCTCTGCCACCGACTGCGCGGAGAAATTCGCGCTCGAGGGTGCCGA
      T N A C S S A T D C A E N C A L E G A D
2341   CTACCGGGCACCTATGGCGCTCGACCGCGCAATGCGCTGACGCTCACCTTCGTTCAC
      Y Q G T Y G A S T S G N A L T L T F V T
2401   TAAGCACGAGTACGGCACCAACATTGGTTCGCGCTCTACCTCATGAACGCGCGCAACAA
      K H E Y G T N I G S R L Y L M N G A N K
2461   GTACGAGATGTTACCTCAAGGGCAACGAGCTGGCTTCGACGTGACCTCTCGGCCGT
      Y Q M F T L K G N E L A F D V D L S A V
2521   CGAGTGGCGCTCAACAGCGCCCTCTACTTCGTGGCCATGGAGGAGGATGGCGGTGTGTC
      E C G L N S A L Y F V A M E E D G G V S
2581   GAGCTACCCGACCAACACGCGCGGTGCTAAGTTCGGCACTGGGgtaagttcaacgacccg
      S Y P T N T A G A K F G T G
2641   agacgggtgcccctattatctgctgcgaaaacggacgggtccccttttgcataactaccctc
2701   ctccaaacagTACTGCGACGCCCAATGCGCACGCGACCTCAAGTTCTGTCGGCGCAAGGG
      Y C D A Q C A R D L K F V G G K G
2761   CAACATCGAGGGCTGGAAGCGCTCCACCAACGATGCCAAATGCCGGTGTGGTCTTATGG
      N I E G W K P S T N D A N A G V G P Y G
2821   CGGGTGCTGCGCTGAGATCGACGTCTGgtaagtttgggtgcccggcagcaatgatatat
      G C C A E I D V W
2881   tagctcgagtggttcccgtegtgctgacccctctcttaccaggGAGTGAACAAATATGC
      E S N K Y A
2941   TTTCGCTTTCACCCCGCACGGTTGCGAGAACCTTAATACCAAGTCTGCGAGACCAACAA
      F A P T P H G C E N P K Y H V C E T T N

```

3001 CTGCGGTGGCACCTACTCCGAGGACCGCTTCGCTGGTGA CTGCGATGCCAACCGGCTGCGA
 C G G T Y S E D R F A G D C D A N G C D
 3061 CTACAACCCCTACCGCATGGGCAACAGGACTTCTACGGTCCCGGCTTGACGGTGGATAC
 Y N P Y R M G N Q D F Y G P G L T V D T
 3121 CAGCAAGAAGTTCACgtgagtaacacgtgcttggaagccccctcccccccccccccaaaa
 S K K P T
 3181 aaaaaaagaaaaaagaagtcaaatgattgatgctaaccaaatcaataaacagCGTCTGCTCA
 V V
 3241 GCCAGTTCGAGGAGAACAGCTCAACCACTTCTTCGTCAGGACGGCAAGAAGATTGAGA
 S Q P E E N K L T Q F F V Q D G K K I E
 3301 TCCCGGGCCCCAAGGTCGAGGGCATCGATGCGGACAGCGCGCTATCACCCCTGAGCTGT
 I P G P K V E G I D A D S A A I T P E L
 3361 GCAGTGCCTGTTCAGGCCTTCGATGACCGTGACCGCTTCTCGGAGGTTGGCGGCTTCG
 C S A L F K A P D D R D R F S E V G G P
 3421 ATGCCATCAACACGGCCCTCAGCACTCCCATGGTCTCTGTCATGTCCATCTGGGATGATg
 D A I N T A L S T P M V L V M S I W D D
 3481 tacgttacctaacccccccccccttttttttcccgcttctctccccgaaactgccacta
 3541 ottatatacgtcccgctccatgatgcttaccttttctccttccagCACTACGCCAATAT
 H Y A N M
 3601 GCTCTGGCTCGACTCGAGCTACCCCTGAGAAGGCTGGCCAGCCTGGCGGTGACCGTGG
 L W L D S S Y P P E K A G Q P G D R G
 3661 CCGGTGTCTCAGGACTCTGGCGTCCCGGCGAGCTTGAGGCTCAGTACCTAATGCGTG
 P C P Q D S G V P A D V E A Q Y P N A *
 3721 agtcgaaacccgtaaaatgtcgggcaaaaaaagatcgctcaagctaaccgaaataataga
 3781 ttagCAAGGTCTCTGGTCCAACATCCGCTTCGGCCCCATCGGCTCGACTCTCAACGTCT
 K V I W S N I R F G P I G S T V N V
 3841 AAactgcaacctgacccgggcccccttctctcacccecccccctctcaagttctctctggt
 3901 ggagccctcggtccttcttctcctaggttcgcaacctttgagcttggtatcgtaggg
 3961 tcattgtgtacatacacaacaaacttaacatctgctaccaagatcttggcgcttggcagg
 4021 tcttctcaaacctcgaagcaactgagcctttgtcctccgagtggaagttaggtgactattta
 4081 cgttgcaagactacgcggtaaagggaacggagcagacctgccacagatattcgtttgggt
 4141 gcttgatttatagcagagtcggaacgtagacatggccccctgaaggtgccaaaccttagata
 4201 gccagaagccttggtttacgaaggggtgggcaacccacgggtgctcctcgctcagcgaatc
 4261 taccgcacgcaatgtatcgtaagaatgtgaactaaagggaacgacgaggtatagggaaa
 4321 cgtcaatgtggcttgaataacagagttaaatacctaataagaagaattagcatgccaaaga
 4381 ttgagccagcaacacatggtagaatagccagcaaaaggacgcttgctcgcttgatctcgaa
 4441 ccgtccaacctgattcggaaggaggaggaaagttgaagaataccggcaataatctactcg
 4501 aggttccctatgcccctgcagagttcctaataatataaagggaaccccgcatgattccgcg
 4561 attataagcataataagctcgcgggccccacacgtgccttaccctcccattgtatata
 4621 atctgtacctcgcttattgtgaatcgctattccgataagcgaaggtctggcactcataga
 4801 taccgtgacatcgattgagatttggccggggccacgggtagtaagcgatgagttgggtcatc
 4681 aattatcaacaaatgcgctcaatcagcgataatcagcctatcaacccggaataatcagcgc
 4741 catcaacgaattgtccatcatgcacgtagcttgtcggcagtgccgcataccctccagagc
 4861 atcatagccgggatagaagctcgcttccagccgtccagagtcagagatgcaggtagca
 4921 agccttcaagaccagttatgtgacccgggtaaaatacttggtagatgcaatgggcgt
 4981 agcttccgggcacttataagctttactagatattatctcaaggtttcttttgaactcctc
 5041 ctagacatttactataaactaccgagcttcaatgctagacgcccctctctgtttaaatag
 5101 tcttttcttctaaagacatctgcttttttcccttaggcttagaggatagggccccctcc
 5161 atcttgcgtcgacggccttagccttggggagtaattattggtatccgcgtacctggttcc
 5221 cagacagccgaagtttcgacgacaaagtaattattgacgacaaatccacccgcatatgcta
 5281 ttccgagtggtgagccccgaacacatcgcttaccgcacgcatccagacgacagagg
 5341 ggcagcttgatgtcttgcctccagatcgccgcacctaacaacgggtgggagtggtggtatcg
 5401 tatgggacggcatcatggtcaacaacccccctgacgggggtgttgggccaatgggaaccca
 5461 ccgttgcctcgagccggatcgcaaggttaagccgaagaggacaaatgacgatgagacttcc
 5521 tttcttttttatttttttttttttttaatttcttttttaagcgtaattgaaaagagctaca
 5581 tatctgtggttcgttctcaatttcagcgacctctccacgaagcatcgtaataaagaa
 5641 gttgtcggaacaaaggggtgtcagaagctatagagcttctaaggatattagccacataca
 5701 tgccatagctgtataaggtatttaacgcttggccagctccttgcctataaatatttag
 5761 tcgttttgcctcctttgtagataattttaacaaggcactcttttcttataatagccacc
 5821 tactatagactgcttcaacgctcccggaagcttattactacgttcggcagttataagcc
 5881 tggcgcttgactactcctctgcccagctatctttaaattatagtagtagcttcttctatt
 5941 acgaactctcttaccctgctttaaactgcttgcagcagctgtctattatctcaagatc
 6001 ctagtgcagacttctatagccttactaggcctagttcttagaacttgtagtatattaaa
 6061 ctatagttataggctaaatttgcagtagtatagagatttgttaaccttaatagtaattat
 6121 aaactagatctagaagttttagtgcttaacctataaataagctagagataaccttatt
 6181 tttagcttccagggagtaattcttagaaggagtagttacctttaaattatctatagatttgata
 6241 ccttctaatatagctatcatagctaaatttatataattataagattcctttcataaaaaat

6301	attatatactatagatattagtaagtagataggatagctateatactagctagtatat
------	--

FIGURE 9

SEQ ID NOs: 3 and 4 cbh4 gene encoding CBH IIb

```

1      ccgcaagtgaatatgtaattactcaatggaaagttctcgaaacggagtcagaaaatgatgt
61     ggttctgtgggaatgcggcaggaggcgacgttgccgtgaatgcgtgaacaattcccgcctc
121    ttcttcttctcgtcttcttcttcttcttcttcttctcggtcgcggaatggttgacggccagcg
181    tgcgcacggctgcgtgttatcgagcgtcggtacgtctagccaacatcccgtagacacgac
241    gaccaagcgtcttgagaatgcaacaacgtctcggaacctggcacgcacatctccgcgcgag
301    gtccggcagacgccgcctgggcaataaccacccctgtccaggccctttccccgcaggcagag
361    ccgcgctcttcccttcatggttattcaggaacgtggcttccgagattctcgcctgttctc
421    ccccgagtaacctgccgaccgtaacccgggtccaccaccgcggactgtccgcaaacctg
481    gttcgcgcgagattaatatgctatttccggactaagtgcacaacacacaagcacccttc
541    cgctcgcgctctagaatctgctttctaaccgggttctcgggcccttccctttcgcgacg
601    cctccgctctccttaccaggcccatccgcaataggtaaggtagccaaccgttttgagc
661    gtgattctgccaaggaccgcatccttgcatctgccatctggtcaaggaccctctttccc
721    gctccattctgggtggctctatcgggacggcggttccccatggctctccaggagagtgatgt
781    gcgagctctggagagccggggttggcgctcacgatgctgccacctaagggccggccagcccg
841    gcaactgcgctcccggttgatccgtctatccccgtcaagagcaccagccccggcgctcgtga
901    attttcgacttggtcgacttgctacaggtgataaagaggatgcacgccgcctcgatcgg
961    cctgtgtgggtttctctccctcgtgcaaaccactcccacctcccgccccgagatagttgc
1021   ttgtttcgctccgctgagagggacacacaccaATGGCCAAGAAGCTTTTCATCACCGCCGC
                                         M A K K L F I T A A

1081   GCTTGCGGCTGCCGTGTTGGCGGCCCCCGTCAATTGAGGAGCGCCAGAACTGCGGCGCTGT
361       L A A A V L A A P V I E E R Q N C G A V

1141   gtggtaagaaagcccgggtccgagtcctccatgattttctcgtcgagtaatggcataaggg

1201   ccaccccttcgactgaccgtgagaatcgatcaaatccagGACTCAATGCGGCGGTAAACGG
                                           T Q C G G N G

1261   GTGGCAAGGTCCCACATGCTGCGCCTCGGGCTCGACCTGCGTTGCGCAGAACGAGTGGTA
        W O G P T C C A S G S T C V A O N E W Y

```

FIGURE 9 (CONT'D)

SEQ ID NOs: 3 and 4 cbh4 gene encoding CBH IIb

1321 CTCTCAGTGCCTGCCCAACAGCCAGGTGACGAGTTCCACCACTCCGTCGTCGACTTCCAC
 S Q C L P N S Q V T S S T T P S S T S T
 1381 CTCGCAGCGCAGCACCAGCACCTCCAGCAGCACCACCAGGAGCGGCAGCTCCTCCTCCTC
 S Q R S T S T S S S T T R S G S S S S S
 1441 CTCCACCACGCCCCCGCCCGTCTCCAGCCCCGTGACCAGCATTCCCGGCGGTGCGACCTC
 S T T P P P V S S P V T S I P G G A T S
 1501 CACGGCGAGCTACTCTGGCAACCCCTTCTCGGGCGTCCGGCTCTTCGCCAACGACTACTA
 T A S Y S G N P F S G V R L F ██████████
 1561 CAGGTCCGAGGTCCACAATCTCGCCATTCTAGCATGACTGGTACTCTGGCGGCCAAGGC
 ██████ S E V H N L A I P S M T G T L A A K A
 1621 TTCCGCCGTCGCCGAAGTCCCTAGCTTCCAGTGGCTCGACCGGAACGTCACCATCGACAC
 S A V A E V P S F Q W L D R N V T I D T
 1681 CCTGATGGTCCAGACTCTGTCCCAGGTCCGGGCTCTCAATAAGGCCGGTGCCAATCCTCC
 L M V Q T L S Q V R A L N K A G A N P P
 1741 CTATGCTGgtgagttacatggcgacttgcccttctcgtccctacctttcttgacgggata
 Y A
 1801 gggtacctgacctggaggcaaaacaacaacagCCCAACTCGTCGTCTACGACCTCCCCGA
 A Q L V V Y D L P D
 1861 CCGTGACTGTGCCCGCCGCTGCGTCCAACGGCGAGTTTTCGATTGCAAACGGCGGCGCCGC
 R D C A A A A S N G E F S I A N G G A A
 1921 CAACTACAGGAGCTACATCGACGCTATCCGCAAGCACATCATTGAGTACTCGGACATCCG

FIGURE 9 (CONT'D)

SEQ ID NOs: 3 and 4 cbh4 gene encoding CBH IIb

N Y R S Y I D A I R K H I I E Y S D I R
 1981 GATCATCTCTGGTTATCGAGCCCCGACTCGATGGCCAACATGGTGACCAACATGAACGTGGC
 I I L V I E P D S M A N M V T N M N V A
 2041 CAAGTGCAGCAACGCCGCGTCGACGTACCACGAGTTGACCGTGTACGCGCTCAAGCAGCT
 K C S N A A S T Y H E L T V Y A L K Q L
 2101 GAACCTGCCCAACGTCGCCATGTATCTCGACGCCGGCCACGCCGGCTGGCTCGGCTGGCC
 N L P N V A M Y L D A G H A G W L G W P
 2161 CGCCAACATCCAGCCCCGCCGCGAGCTGTTTGCCGGCATCTACAATGATGCCGGCAAGCC
 A N I Q P A A E L F A G I Y N D A G K P
 2221 GGCTGCCGTCCGCGGCCTGGCCACTAACGTCGCCAACTACAACGCCTGGAGCATCGCTTC
 A A V R G L A T N V A N Y N A W S I A S
 2281 GGCCCCGTGCTACACGTCGCCTAACCTAACTACGACGAGAAGCACTACATCGAGGCCTT
 A P S Y T S P N P N Y D E K [REDACTED]
 2341 CAGCCCCGTCTTTGAACTCGGCCGGCTTCCCCGCACGCTTCATTGTCGACACTGGCCGCAA
 [REDACTED] F I V D T G R N
 2401 CGGCAACAACCTACCGgtatgttttttttcttttgtctctgtcccccccttttctccc
 G K [REDACTED]
 2461 ccttcagttggcggtccacaaggtctcttagtctgtcttcctgtgaccaacctcccc
 2521 ccccggcacgcccacaaccgtttgactctatactcttgggaatggg'gcgcgaaactgac
 2581 cgttccacagGCCAACAACAGTGGGGTGACTGGTGCAATGTCAAGGGCACCGGCTTTGGC

FIGURE 9 (CONT'D)

SEQ ID NOs: 3 and 4 cbh4 gene encoding CBH IIb

G T G F G

2641 GTGCGCCCGACGGCCAACACGGGCCACGAGCTGGTCGATGCCTTTGTCTGGGTCAAGCCC
V R P T A N T G H E L V D A F V W V K P

2701 GGCGGCGAGTCCGACGGCACAAGCGACACCAGCGCCGCCCGCTACGACTACCACTGCGGC
G G E S D G T S D T S A A R Y D Y H C G

2761 CTGTCCGATGCCCTGCAGCCTGCCCCGAGGCTGGACAGTGGTTCCAGGCCTACTTCGAG
L S D A L Q P A P E A G Q W F Q A Y F E

2821 CAGCTGCTCACCAACGCCAACCCGCCCTTCTAAacctcgtcataaagagagagagatggc
Q L L T N A N P P F *

2881 gggcatgggcctgattgggttcattgaccatgcggctcttctgggggtacatattttacc
2941 tacctacctataaataaggcggcctatcgggctctcgcttcgtttatttaggtacttggtc
3001 ttgtacatactttgtttatacatagcagttagcatccactattcggtttcgacaaagcg
3061 gaactttccagaaaaaaaaagggtgtacataattagtccttttaggcttcgattcctttgtgc
3121 ctttctttttggtaaaaaaaaaattttttttgaggcatgattaccttaggtacgttcgtc
3181 gttgtattgggtccccctgcattttggcgcgagagcagctcagccccctgcaaatccctca
3241 acgggcggttcaattccctccactcgggtcttcagcgagaccagccgtccagagtatccca
3301 gcgtgtagttgccccacgaaccagtcgtcctcgtaagcctcgtaaaagtgtccaagagca
3361 gtatagaagcaacgacctccgtcaaaagtctggcaccatgcgatcgggtgggtcctccccg
3421 tgcgccccgcccctcgtaggaacttctcatccacgccaaggagcacgtgcaggccgtcggac
3481 gtcgccccggggtgcgccttgaagttgtaccattcgctcctccagacgcgctccagctgc
3541 gcctgcttgggttcctgcggttcctgcggttcctgcgctggcgcggtcggcgccgctct
3601 tggtcacacgcccgcagcgacatgactgggtgtttcggtcgagcagcttgacgagcccc
3661 acctggggttccgggtggttgtcgaacacggcgccaatgaggtggcggtaccattcggtat
3721 gactgcatggcgaagctggcgagtgtagccgacgatccccgccccgcctggacgaaa
3781 ccccgaggcgcccagctgcgcgcgctccaggaaactcgcccgagcactgcaggaggacg
3841 atgacgcgatacgccgagaggggagcggggctgaacacggcggggatcctcgctgtcgctcc

FIGURE 10

SEQ ID NOs: 5 and 6 cbh1 gene encoding CBH Ia

```

ATGTACGCCAAGTTCGCGACC 1800
M Y A K F A T

CTCGCCGCCCTTGTGGCTGGCGCCGCTGCTCAGAACGCCTGCACTCTGACCGCTGAGAAC 1860
L A A L V A G A A A Q N A C T L T A E N

CACCCCTCGCTGACGTGGTCCAAGTGCACGTCTGGCGGCAGCTGCACCAGCGTCCAGGGT 1920
H P S L T W S K C T S G G S C T S V Q G

TCCATCACCATCGACGCCAACTGGCGGTGGACTCACCGGACCGATAGCGCCACCAACTGC 1980
S I T I D A N W R W T H R T D S A T N C

TACGAGGGCAACAAGTGGGATACTTCGTACTGCAGCGATGGTCCTTCTTGCGCCTCCAAG 2040
Y E G N K W D T S Y C S D G P S C A S K

TGCTGCATCGACGGCGCTGACTACTCGAGCACCTATGGCATCACCACGAGCGGTAACCTCC 2100
C C I D G A D Y S S T Y G I T T S G N S

CTGAACCTCAAGTTCGTACCAAGGGCCAGTACTCGACCAACATCGGCTCGCGTACCTAC 2160
L N L K F V T K G Q Y S T N I G S R T Y

CTGATGGAGAGCGACACCAAGTACCAGagtaagttcctctcgccccggccgccccgaga 2220
L M E S D T K Y Q M

tgatggcgcccagcccgtgacgcgaatgacacaGTGTTCCAGCTCCTCGGCAACGAGTT 2280
F Q L L G N E F

CACCTTCGATGTCGACGTCTCCAACCTCGGCTGCGGCCTCAATGGCGCCCTCTACTTCGT 2340
T F D V D V S N L G C G L N G A L Y F V

GTCCATGGATGCCGATGGTGGCATGTCCAAGTACTCGGGCAACAAGGCAGGTGCCAAGTA 2400
S M D A D G G M S K Y S G N K A G A K Y

CGGTACCGGCTACTGTGATTCTCAGTGCCCCGCGACCTCAAGTTCATCAACGGCGAGGC 2460
G T G Y C D S Q C P R D L K F I N G E A

CAACGTAGAGAACTGGCAGAGCTCGACCAACGATGCCAACGCCGGCAGGGCAAGTACGG 2520
N V E N W Q S S T N D A N A G T G K Y G

CAGCTGCTGCTCCGAGATGGACGTCTGGGAGGCCAACCAACATGGCCGCCGCTTCACTCC 2580
S C C S E M D V W E A N N M A A A F T P

CCACCCTTGNCCTGATCGGCCAGTCGCGCTGCGAGGGCGACTCGTGCGGCGGTACCTA 2640
H P C ? V I G Q S R C E G D S C G G T Y

CAGCACCGACCGCTATGCCGGCATCTGCGACCCCGACGGATGCGACTTCAACTCGTACCG 2700
S T D R Y A G I C D P D G C D F N S Y R

CCAGGGCAACAAGACCTTCTACGGCAAGGGCATGACGGTCGACACGACCAAGAAGATCAC 2760
Q G N K T F Y G K G M T V D T T K K I T

```

FIGURE 10 (CONT'D)

SEQ ID NOs: 5 and 6 cbh1 gene encoding CBH Ia

```

GGTCGTCACCCAGTTCCTCAAGAACTCGGCCGGCGAGCTCTCCGAGATCAAGCGGTTCTA 2820
V V T Q F L K N S A G E L S E I K R F Y

CGTCCAGAACGGCAAGGTCATCCCCAACTCCGAGTCCACCATCCCGGGCGTCGAGGGCAA 2880
V Q N G K V I P N S E S T I P G V E G N

CTCCATCACCCAGGACTGGTGCAGACCGCCAGAAGGCCGCTTCGGCGACGTGACCGACTT 2940
S I T Q D W C D R Q K A A F G D V T D ?

NCAGGACAAGGGCGGCATGGTCCAGATGGGCAAGGCCCTCGCGGGGCCCATGGTCCTCGT 3000
Q D K G G M V Q M G K A L A G P M V L V

CATGTCCATCTGGGACGACCACGCCGTCAACATGCTCTGGCTCGACTCCACCTGGCCCAT 3060
M S I W D D H A V N M L W L D S T W P I

CGACGGCGCCGGCAAGCCGGGCGCCGAGCGCGGTGCCTGCCCCACCACCTCGGGCGTCCC 3120
D G A G K P G A E R G A C P T T S G V P

CGCTGAGGTGAGGCCGAGGCCGCCCAACTCCAACGTCATCTTCTCCAACATCCGCTTCGG 3180
A E V E A E A P N S N V I F S N I R F G

CCCCATCGGGCTCCACCGTCTCCGGCCTGCCCGACGGCGGCAGCGGCAACCCCAACCCGCC 3240
P I G S T V S G L P D G G S G N P N P P

CGTCAGCTCGTCCACCCCGGTCCCCTCCTCGTCCACCACATCCTCCGGTTCCTCCGGCCC 3300
V S S S T P V P S S S T T S S G S S G P

GACTGGCGGCACGGGTGTCGCTAAGCACTATGAGCAATGCGGAGGAATCGGGTTCAGTGG 3360
T G G T G V A K H Y E Q C G G I G F T G

CCCTACCCAGTGCAGAGAGCCCTACACTTGACCAAGCTGAATGACTGGTACTCGCAGTG 3420
P T Q C E S P Y T C T K L N D W Y S Q C

CCTGTAA
L *

```

FIGURE 11

SEQ ID NOs: 7 and 8 eg6 gene encoding CBHIIa

ggatccacac	ctaccatacc	ggatagtatg	ctacccaagt	gacatagggg	tggtaaagta	60
atacgagaac	tcagagagca	ctgcccata	ggctcgccaa	tgacctcaag	tgccagggtca	120
gctttgcgag	acagacctga	gcgcgtcgga	tgtgtgacat	ggaacgcgcc	ggatcgccctt	180
gttgattaat	tataggggaa	tagcgaggaa	ggtttcagca	attgacgtga	gcgtacatta	240
aaagctgtat	gatttcagga	agacgagcca	tggaccaggt	ttcaaggctg	aatggcttga	300
cgacttaagc	accgaacgag	gaatgaaaga	atgaaaagtg	ggggatcatt	ctggccccctc	360
ctcgtatgtc	gagtgttaaa	gaaggcgggt	ctacggagga	cctaaagagc	tccaattttgc	420
tctgttgagc	ttaagccaca	tatctcaaga	tgaatacatg	tcaggcatag	tcaccctgat	480
cttgttcac	agtcacaca	cttttcagtt	cagcatgttg	attcctcctc	catatcactt	540
tccattacta	tctctttatg	tccttgggtc	agactccaag	gaaccgatag	gtgagcatcg	600
gtgaggctcc	ctcaaggtag	caaagtagcc	atcatcaccg	aggtctggga	atggcgccgt	660
gcccgatctg	agtcctccaa	ctccacggta	cgacgacagc	acgtcacatt	gacgcaccac	720
ggttgaacaa	gcagagaggg	acacgtcttg	ctacgcgaat	cctggcactg	gatggagacg	780
cgtgtgagca	ggtttccgga	accatgacgg	cctgggtccg	cttctcgaac	aaagaagtgg	840
aacacaaaaa	gaaccgaaac	ggaaacgcag	gcacggcatc	gacgaccgga	ttgtcccacg	900
gggacctcgg	ccagtcaagc	gttgccctgg	ccgtcagctc	cctggcgacg	gggattcagc	960
acatctcacg	ttataggcga	cctcatcccc	cttccgtctt	gtgcggctgt	tgctccgtgc	1020
cgagtaccca	ggcgtgccgg	ggcctttago	cggggcgga	tcagagtcaa	gatgcggccg	1080
aattggacgg	cagacgaagt	ttcgtagagg	gtcatgatcg	gcactgacga	cacccacccc	1140
tgcgtgatcc	cgtggccctg	ggctgggaat	tgccggctaa	taatctacgg	cttaatatagat	1200
atgcactttg	cacgcggtgc	agataaataa	gctgtggttt	caaacactgg	cctccgtact	1260
ttaccaccca	actgccgctt	agcgccggga	cctgagtctt	gggagtgcgc	ggagcggcag	1320
ccacctcggg	ttagcgtaca	cacgacggct	gcacgagggg	atgcgcggtg	catggcttca	1380
tagtgtagca	cagaccgtca	agtcctcaatc	tgggtgatgc	ttgatgagat	gacagcgagc	1440
cccgtcggcg	gcaccccggc	tatgcatcgc	gaattgacaa	cactctcagc	tctattgcga	1500
cccacgggat	aaaagaagaa	gaaaaaaatg	gaaccttgagt	acgggcgtca	gaaacaaaaa	1560
aaaaactccg	gaaccaaata	tgctcgggcat	ggccgggggtg	aacgaccgct	actccccgtt	1620
cccttctctg	caaacagaac	gctacagagg	gttttctggt	ttgtcaaaga	gttcggagggt	1680
cctctgctcc	gcgaatgcgt	ggtgaaccca	ccagcagcca	ttgttcttgc	atgcgtggcg	1740
gaccgttagc	cgctgatcga	catggcgagc	ttccacctc	agacctggag	cagacgggtg	1800
cgaggagcaa	ggggctgccc	tccccctgac	ggtcggaccc	caatgacttc	cccaaacggg	1860
gacatcgagg	gtcgtgcatg	atgggtggaaa	gtagttgcag	tatgggaagt	accccggtt	1920
gccaggaacc	gttggtcggc	ccccacatt	ttctctctgc	catgtcaact	gtgtgtcgtt	1980
cgagagttcc	tggctccggc	ccccgtcca	attccctaac	gggaccgcgg	ggcatcgctt	2040

FIGURE 11 (CONT'D)

SEQ ID NOS: 7 and 8 eg6 gene encoding CBHIIa

gtaactaact tccaaatgaa gccggatatg agggagggag attggatctg gcaagccagc	2100
cattcgctgc gatcggcact cgtccgtcag ccccgagtc catatcccca aaggcaactg	2160
ctcggcgcggt ctcaagtctt cttcggaacg tccagccga aggcgcgcgc cagcaccggc	2220
cctatgttcc tgattgcgat cctcgatctc cagagacggg tcacctcgcc tcgaggacgg	2280
tgcaggggca tcggcttcgc ttcctagagc tccgggctgt gtgtgggtcaa ggggagaagg	2340
cggcgggcgcc aaggtgcgtc tcggcgact caccatcgc ctttaccccc ctcccccca	2400
gtatataaaa gatggccatc gtctcctcgt ctgcttggga agaaaggatc tctcgaccat	2460
gcaccacagc ctagctctaa cccagcttgt cgtgtgttgt tgcccagc atg aag ttc	2517
Met Lys Phe	
1	
gtg cag tcc gcc acc ctg gcg ttc gcc gcc acg gcc ctc gct gcg ccc	2565
Val Gln Ser Ala Thr Leu Ala Phe Ala Ala Thr Ala Leu Ala Ala Pro	
5 10 15	
tcg cgc acg act ccc cag aag ccc cgc cag gcc tcg gcg ggc tgc gcg	2613
Ser Arg Thr Thr Pro Gln Lys Pro Arg Gln Ala Ser Ala Gly Cys Ala	
20 25 30 35	
tcg gcc gtg acg ctc gat gcc agc acc aac gtg ttc cag cag tac acg	2661
Ser Ala Val Thr Leu Asp Ala Ser Thr Asn Val Phe Gln Gln Tyr Thr	
40 45 50	
ctg cac ccc aac aac ttc tac cgt gcc gag gtc gag gct gcc gcc gag	2709
Leu His Pro Asn Asn Phe Tyr Arg Ala Glu Val Glu Ala Ala Ala Glu	
55 60 65	
gcc atc tcc gac tcg gcg ctg gcc gag aag gcc cgc aag gtc gcc gac	2757
Ala Ile Ser Asp Ser Ala Leu Ala Glu Lys Ala Arg Lys Val Ala Asp	
70 75 80	
gtc ggt acc ttc ctg tgg ctc gac acc atc gag aac att ggc cgg ctg	2805
Val Gly Thr Phe Leu Trp Leu Asp Thr Ile Glu Asn Ile Gly Arg Leu	
85 90 95	
gag ccc gcg ctc gag gac gtg ccc tgc gag aac atc gtg ggt ctc gtc	2853
Glu Pro Ala Leu Glu Asp Val Pro Cys Glu Asn Ile Val Gly Leu Val	
100 105 110 115	
atc tac gac ctc ccg ggc cgt gac tgc gcg gcc aag gcc tcc aac ggc	2901
Ile Tyr Asp Leu Pro Gly Arg Asp Cys Ala Ala Lys Ala Ser Asn Gly	
120 125 130	

FIGURE 11 (CONT'D)

SEQ ID NOs: 7 and 8 eg6 gene encoding CBHIIa

```

gag ctc aag gtc ggc gag ctc gac agg tac aag acc gag tac atc gac a 2950
Glu Leu Lys Val Gly Glu Leu Asp Arg Tyr Lys Thr Glu Tyr Ile Asp
      135              140              145
gtgagttaac cctttgtggc cccttctttt ccccgagag agcgtctggt tgagtgggt 3010
tgtgagagag aaaatggggc gagcttaaag actgacgtgt tggctcgag ag atc 3065
                                Lys Ile

gcc gag atc ctc aag gcc cac tcc aac acg gcc ttc gcc ctc gtc atc 3113
Ala Glu Ile Leu Lys Ala His Ser Asn Thr Ala Phe Ala Leu Val Ile
150              155              160              165
gag ccc gac tcg ctc ccc aac ctg gtc acc aat agc gac ctg cag acg 3161
Glu Pro Asp Ser Leu Pro Asn Leu Val Thr Asn Ser Asp Leu Gln Thr
      170              175              180
tgc cag cag agc gct tcc ggc tac cgc gag ggt gtc gcc tat gcc ctc 3209
Cys Gln Gln Ser Ala Ser Gly Tyr Arg Glu Gly Val Ala Tyr Ala Leu
      185              190              195
aag cag ctc aac ctc ccc aac gtg gtc atg tac atc gat gcc ggc cac 3257
Lys Gln Leu Asn Leu Pro Asn Val Val Met Tyr Ile Asp Ala Gly His
      200              205              210
ggg ggc tgg ctc ggc tgg gac gcc aac ctc aag ccc ggc gcc cag gag 3305
Gly Gly Trp Leu Gly Trp Asp Ala Asn Leu Lys Pro Gly Ala Gln Glu
      215              220              225
ctc gcc agc gtc tac aag tct gct ggt tcg ccc tcg caa gtc cgc ggt 3353
Leu Ala Ser Val Tyr Lys Ser Ala Gly Ser Pro Ser Gln Val Arg Gly
      230              235              240              245
atc tcc acc aac gtg gct ggt tgg aac gcc tg gtaagacact ctatgtcccc 3405
Ile Ser Thr Asn Val Ala Gly Trp Asn Ala Trp
      250              255
ctcgtcggtc aatggcgagc ggaatggcgt gaaatgcatg gtgctgacct ttgatctttt 3465
ccccctccta tag g gac cag gag ccc ggt gag ttc tcg gac gcc tcg gat 3515
                                Asp Gln Glu Pro Gly Glu Phe Ser Asp Ala Ser Asp
                                260              265
gcc cag tac aac aag tgc cag aac gag aag atc tac atc aac acc ttt 3563
Ala Gln Tyr Asn Lys Cys Gln Asn Glu Lys Ile Tyr Ile Asn Thr Phe
      270              275              280

```

FIGURE 11 (CONT'D)

SEQ ID NOs: 7 and 8...eg6 gene encoding CBHIIa

```

ggc gct gag ctc aag tct gcc ggc atg ccc aac cac gcc atc atc gac      3611
Gly Ala Glu Leu Lys Ser Ala Gly Met Pro Asn His Ala Ile Ile Asp
285                290                295                300
act ggc cgc aac ggt gtc acc ggt ctc cgc gac gag tgg ggt gac tgg      3659
Thr Gly Arg Asn Gly Val Thr Gly Leu Arg Asp Glu Trp Gly Asp Trp
305                310                315
tgc aac gtc aac ggc gcc ggc ttc ggt gtg cgc ccg act gcc aac act      3707
Cys Asn Val Asn Gly Ala Gly Phe Gly Val Arg Pro Thr Ala Asn Thr
320                325                330
ggc gac gag ctc gcc gac gcc ttc gtg tgg gtc aag ccc ggt ggc gag      3755
Gly Asp Glu Leu Ala Asp Ala Phe Val Trp Val Lys Pro Gly Gly Glu
335                340                345
tcc gac ggc acc agc gac tcg tcg gcg gcg cgc tac gac agc ttc tgc      3803
Ser Asp Gly Thr Ser Asp Ser Ser Ala Ala Arg Tyr Asp Ser Phe Cys
350                355                360
ggc aag ccc gac gcc ttc aag ccc agc ccc gag gcc ggt acc tgg aac      3851
Gly Lys Pro Asp Ala Phe Lys Pro Ser Pro Glu Ala Gly Thr Trp Asn
365                370                375                380
cag gcc tac ttc gag atg ctc ctc aag aac gcc aac ccg tcc ttc      3896
Gln Ala Tyr Phe Glu Met Leu Leu Lys Asn Ala Asn Pro Ser Phe
385                390                395
taagctcctc gacggcttct tgctgtcagt cgctctgacg gtggtgtgct ggtggtgccc      3956
ctgctcctgc tgctgtctgct ccgcggggag gggaggcaac gaaaatgaag tcctgcttca      4016
aaacaaaaca gaaacaagcg aggcgcgggtg caatggtcgt gcgttcgtct tttttcatgt      4076
tcccttctag tgtagtagtt tgatagtcgt acataagggg tttcagaacc gtctctctgt      4136
ctcgggtcttt ttgcgagttg ttgcgactcg tgattatggc ctttgttget cgttgcgcca      4196
gagtagaacc acagcgtgtt ggggtagcag cttgctccgt aggacgtagg gaaacaacct      4256
gagactctgg aattgcagtc agcctgcgtc gccctcttag gaaacgaagg ggagaaccag      4316
tagtggtctc agcttacaaa cgcgagcatg gtgaacatct ccgagaaaag ggaggggatcc      4376

```

FIGURE 12

SEQ ID NOs: 9 and 10 eg2 gene encoding EGII

```

1      tgetgctctgatgtgctgatgcacagcttcccctcgcgattgccggcaggatctccaacc
61     ctctggatcggagcagacgatcagcgggcacaatggccagcttgccagcgttcaactcca
121    agttgacccgcttttatcacgcccagctggacatgcacaggcttggcttctcgtgttcc
181    tacgatctgcacagtaggtttgactgctgatcttcgctttcctgtgcgccctcccctcc
241    ctacgggtaccttatccttgctgtgaaccccgcttatgtcaaacttgagtttgaccaa
301    tgctagcgcgaaaagtacctacatagtagtactatgtaataaggtaggtagacatcatcagtagg
361    cgtttatctagtaaattttggctttttgaaactcaattgctcctctcctcgccctccacct
421    ctgcttggaattgacaacctggctgtgcctagaggtagcatcgacgatcaatcaaatc
481    taaagtattcgagattgacctttctgctctaattatattaattatccgcacaatgctgta
541    gtcattgactctcctttcaagttgccttctcgtttatgtatgtacaatgggcggtcatgc
601    tteatgccaacagatggttctatcggaacaatgtttgactttctggtcgccccgtcgaa
661    tgttttgatttcgcacgggaagtgttcttaccaaagctaagtcgactcgtggagcttcgt
721    aacggccagtgatcgttgatcgcttttgaggagttgcatggagcgagaccggctacga
781    gcacgttcgcaaaggcagcagcatagacgacctcgtggcgccattcgggagatgcaca
841    tgacataagcatatcaatactcacctgaactcatcgccgatgcctcgaggtagttaca
901    agacatatgtgtgggtatattatcccaaccgtagctttgtcgcgatcatttcggtatg
961    tgctgatgcctacttagggagcaaaagcgcctctcctcacctgcgggttacttacttact
1021   gtgcagcatggccttatgttctcccggtcttgcttgcgcgaaatgaacaaaaacgccga
1081   agaaaagcgcttcttcgagttgtgtctacccgaacataagaggttatgtcgagaccg
1141   ccagcaaatgtcaacaacccacccacggcgttcagaaccttcgaaatatcatctagttt
1201   aagtttaaatgacggcccgagtcacagccgagattcccatattggccgataccagcgttc
1261   ccttggttttccaaggttgctcgtcaactggcgcatctgcctacaacgagatataatta
1321   ccgttttctttgcaaaaggcatgcatggatgtatattatgatgcctgcagaacgaga
1381   agcaatcatggtgtaggttttgctcggtatggagctaataatattgaacggatctctggt
1441   ccgtcctaaatcgttgaaacgctaggcccaggaggacctgctcgacttggcgaacggaga
1501   tttccaggatgaaaggtcggaacatgtccatccgcggccagcctgaacacttttgctcgt
1561   ttccggaccatcgacccacgaaaacagtgcggttgctggcacagttagcactcacgatgg
1621   cgatggtccagcccgctcccgcccgatgccacttgacgcgaactctccttcattcggc
1681   ggccccggcggtgtctggcctattagtagcgttttgatgacggcttggtcgccgcccgg
1741   ttttcttgccgatgacgggaatctcggtggtcccaactccacctgggcacgctctggtg
1801   ccaacatggaaacttcgggatgcccgtccgggcacagtcaagcgctttaaatacgaactt
1861   accccacaagaatcgaggcgtaacccggaattagggacacctggacggcgcaacccctgg
1921   accgaagggcctcgctaaccgggttctcgagccgcatgcgcggctgcccgttgccgc
1981   tcttgagatgacacttcttttcagcgagggtggcgggcagggaaatgatgtattataa

```

FIGURE 12 (CONT'D)

SEQ ID NOs: 9 and 10 _eg2_gene.encoding_EGII

2041 gaagcgagccgattccgaaggactcgacccccctctctcgccctgtgtccgccagctaatt
2101 acagcactccttctcgacttgaaacgccccgagATGAAGTCCTCCATCCTCGCCAGCGTCT
700 M K S S I L A S V

2161 TCGCCACGGGCGCCGTGGCTCAAAGTGGTCCGTGGCAGCAATGTGGTGGCATCGGATGGC
720 F A T G A V A Q S G P W Q Q C G G I G W

2221 AAGGATCGACCGACTGTGTGTGGGTACCCTGCGTCTACCAGAATGATTGGTACAGCC
740 Q G S T D C V S G Y H C V Y Q N D W Y S

2281 AGTGCGTGCCTGGCGCGCGCTCGACAACGCTCCAGACATCTACCACGTCCAGGCCACCG
760 Q C V P G A A S T T L Q T S T T S R P T

2341 CCACCAGCACCGCCCCCTCCGTTCGTCCAACCTCGCCTAGCAAGGGCAAGCTCAAGTGGC
780 A T S T A P P S S T T S P S K G K L K W

2401 TCGGCAGCAACGAGTCGGGCGCCGAGTTCGGGGAGGGCAACTACCCCGGCCTCTGGGGAA
800 L G S N E S G A E F G E G N Y P G L W G

2461 AGCACTTCATCTTCCCGTCGACTTTCGGCGATTTCAGgtacgggccaataataatatattat
820 K H F I F P S T S A I Q

2521 tatagcaggcaggaggggagcaggagaagaagggagggggcaggtggccaacaatcgyaaga
2581 agaccgggaggcactgaccgttgattcctttgtgtaatagACGCTCATCAATGATGGATA
861 T L I N D G Y

2641 CAACATCTTCCGGATCGACTTCTCGATGGAGCGTCTGGTGCCCAACCAGTTGACGTCGTC
881 N I F R I D F S M E R L V P N Q L T S S

2701 CTTGACGAGGGCTACCTCCGCAACCTGACCGAGGTGGTCAACTTCGTGACGAACGCGGG
901 F D E G Y L R N L T E V V N F V T N A G

2761 CAAGTACGCCGTCTCTGGACCCGCACTACGGCCGGTACTACGGCAACGTCATCACGGA
921 K Y A V L D P H N Y G R Y Y G N V I T D

2821 CACGAAACGGCTTCCGGACCTTCTGGACCAACCTGGCCAAGCAGTTGCGCTCCAACCTCGCT

FIGURE 12 (CONT'D)

SEQ ID NOs: 9 and 10 eg2 gene encoding EGII

```

941      T N A F R T F W T N L A K Q F A S N S L

2881    CGTCATCTTCGACACCAACAACGAGTACAACACGATGGACCAGACCCCTGGTGCTCAACCT
961      V I F D T N N E Y N T M D Q T L V L N L

2941    CAACCAGGCCGCCATCGACGGCATCCGGGCCGCCGGCGCGACCTCGCAGTACATCTTCGT
981      N Q A A I D G I R A A G A T S Q Y I F V

3001    CGAGGGCAACGCGTGGAGCGGGGCCTGGAGCTGGAACACGACCAACACCAACATGGCCGC
1001     E G N A W S G A W S W N T T N T N M A A

3061    CCTGACGGACCCGCGAGAACAAGATCGTGTACGAGATGCACCACTACCTCGACTCGGACAG
1021     L T D P Q N K I V Y E M H Q Y L D S D S

3121    CTCGGGCACCCACGCCGAGTGCCTCAGCAGCAACATCGGGCGCCAGCGCGTCTCGGAGC
1041     S G T H A E C V S S N I G A Q R V V G A

3181    CACCCAGTGGCTCCGCGCCAACGGCAAGCTCGGCGTCCCTCGGCGAGTTCGCCGGCGGGCGC
1061     T Q W L R A N G K L G V L G E F A G G A

3241    CAACGCCGTCTGCCAGCAGGCCGTACCGGCCTCCTCGACCACCTCCAGGACAACAGCGA
1081     N A V C Q Q A V T G L L D H L Q D N S E

3301    GGTCTGGCTGGGTGCCCTCTGGTGGGCGCCCGGTCCCTGGTGGGGCGACTACATGTACTC
1101     V W L G A L W W A A G P W W G D Y M Y S

3361    GTTCGgtaagtttctcccttgttcttggctttccccccagtaaggagtcaggcaacat
1121     F

3421    gccaagaccggctcggcttcgcttcaaggcggttcggtgtacacactgaagagttccaac
3481    ttccaaccctgttcgtgtcctccgatcagcttcgacggggtgaaggggaagggatttgg
3541    gagtgagggtggaggtcaaaaggagggatatccccagatctccacaaacggccctgagcca
3601    acaacagcctctggggtcaaaatgggcyccaaccatacggtcattcactcaggacacctg
3661    ctaacgcgtctcttttttttgtttccagAGCCTCCTTCGGGCACCGGCTATGTCAACTAC
1221                                     E P P S G T G Y V N Y

3721    AACTCGATCCTAAAGAAGTACTTGCCGTAAGgggcatgcagcaagggtcgagcgagcatta

```

FIGURE 12 (CONT'D)

-SEQ-ID NOs: 9 and 10 -eg2-gene-encoding EGII

```

1241      N S I L K K Y L P *
3781      ttcagggccatctgcttgtgtcggcaggcatcacgtcaacccatcgaatcggacagcggg
3841      atgctccgagatgccatacactaagtctggtgatgacgtgagaatgctggccctggtcgg
3901      ggggtaccgccaacaaaaagcaccggacgctgccgcgccggataccatggtttcatgt
3961      acatattgggttcttctgtcttcttacgggggggggggggggggggggctctgcagcgttgc
4021      tgagcgttctgttccaagtatacttctgtctggaattgaattttgagtgcattgacc
4081      caatcaaccagctcgggtgtgtctcacctcccggtaccctccctctctccctctgctcggc
4141      ttggcttctctctcgggtgtggagcagggccacggcggtcccaatccatataagatcgat
4201      ggtatactatggtatacactagcttgggaataaactaatccatacgcctaactaatggacg
4261      gattatcctaagggtcaccgggtcaccggttggtataaacacctaggatacgggagagctg
4321      atagaaagggtgtactccgtattgtactgtacaatacaaaagtacagatagcacacgaag
4381      tacggtagggtggtccgcctagtccggaccaacaatagaacatgcgttccctggggacctg
4441      caggaaagaagggggggggggttgccaagacgcccggggttcaaagaagccccggggccg
4501      ccgatgagatgagacgggacggcgcccaaggagaggccgggtggtcgatcctgcaaatgcc
4561      agcaaaaaaatccataccataatccagtcacttctcgtcacactcctgtgaaacgagct
4621      ggagggactgctggaaagggttttgagggttaatactgtatgtggagcatgccgtacct
4681      ctgtgcttctgttaacagatagagttccagttgaacacacaaagttctgccccgcctgcc
4741      gacgtgaaaagaagctcctccgggggagctttaggcaactgggaggggtctctcccaggt
4801      tcatggtgtctgtctcttctcaaattttatgctgccacccatttgacagaggtgtgca
4861      caccgttgccaggtcttgccatccggcaaaaagcagaaaagtcgacccatcgccctaagaa
4921      aggcggtcggaaggggatcggtgctcattgagggttagcgtctgcccattctgacgctg
4981      cccattgttttgtgtcgcatctgtcttcggatgtcggtcaagagtcgggatttttcc
5041      cctgtgcttccagcctaattctgagcgggagctgggttcgagtgagggttgccttgt
5101      tgggtggagcagcaaccagccaattcactccccgcatttctcgggccgcccaggcatccc
5161      cggcatgcgtttggcggttaactactccgtactggggtagggtgaaattggttctccgctc
5221      gcaggaggctcgtgctcggtcaggggagaacaaagtccaactgctccttctggcaacaa
5281      tgagaggggggttctattgccaacgttgacgaaaggagcagccacaaaacccaaaagcag
5341      gttaccttactgtacctgagcttgaacgtcgcgtagcattggagctctcgtctaccggcg
5401      gcgtcacactccattggcaggtcaaggcagtcagtggcagcgacccaacaacgtcaatgc
5461      ttgttaccgccagaattacccggggtgcgaactgcagggggccgcgcgcatgttgatca
5521      ccggttgattacttctcgcccgcaaccgggagatgagaagcagaactttgttctctttt
5581      caaaaaggacctgacttgcggggaacgcactgccggcagtgagggtggatgcacgctagt
5641      atatgtttccgcccacccagtcgcccgtcgcgccggtgaggtcagtttggttccc
5701      gtgccgccgacaaacgagcgggtgcataattacatttctgctccatgtaccgtgcacccctcc
5761      ccgttcgcgaccgtagta

```

FIGURE 13

SEQ ID NOs: 11 and 12 . bgl1 gene encoding BGL

```

1      cccggcctccagttccaggagcttggtctgcccacatactgtgtacactaggaattctct
61     tatgcggggtgtgcgcggggaaatgttggggaactcgagttgggtcatgtggacaagacc
121    aatgggagctgacatcatttgtgcgacccgttaaaccggaagctacaacaacattctggat
181    tctacactagtggaagaggtaagtaattgacgacaagcaagaagcattgccatgttctgc
241    gaaggatgcgggtgtttttgcatgagcaggaagctgtggctttttagtgtcctttgtgc
301    tcgcccggcgcgagaaactaccgaaacgcaggggactgcgtgcctctggggtcgaatg
361    ccgatccccatcttcacattcccaccatcgtgttctgttaacgaagccggagcggcgggga
421    actcgaagctccactacgtatggatacttgggaccgtacggagtgtgttggtacggatgc
481    ctgcacaagtgttgtgtcttctacgaagacgccaacccacataatacacaaaagctgttg
541    taagtcgagttacctcaggcacgttcgggcaactcgggcaacctgacgagatttccccgc
601    cattccgccaagaggccggcgctgcctgattaggcagctcttggacaataactatgta
661    gaatggaagctccatccatagtcagctccattggcgggtcccagtgatctcgatggctgga
721    tggctgctctgtacggtacatacatagtaagttctcgcttgagagcccaattcgctgca
781    atagcatctttccccgcagtgcgccggccgctgggtcccgtccacaatgacctgct
841    tctggagcttctcgacgaacagatcgcccggttcttctccacaccaatccgaaccagtc
901    gggagcatggctgcggatgcgacgcagccttcttcgcgctgtacaaacagctccgggaa
961    cgtcgactggtatgtacggactacagtaagtacactacgagtgcacatactgacgaatac
1021   cggcctcagaggaacctggcaggaccctacccacacgaaaccacagcgagaaagcgcaa
1081   tggatcagtaactactgcgaagtaaccgtggtcccgggcaaaggatctgagggccgatcg
1141   ctcgtggggctgcgaggcgaggagagcaaaacagccagtcctcccgcgaacctggaaaa
1201   tcacttataaacacacgtcacccggcgccgggtgcgcgccatgtgtcacctccaggtcc
1261   tcccgggcgatgatctctgcgggtgcatcaatcatctcggttcgcccgcagctgcttctt
1321   tctgtgcagtgaacgctctcaaactgcaacgagctgtccgacatgaaggctgctgcgct
1381   ttctgcctcttcggcagtagcccttgccgttgaggcgccattgaatcgagaaaggtag
1441   gacgggctttcgtcaaagactcgctccccgatcaacttccccttcatccagaccacccc
1501   aaccctcccagtcctgcttcgagcacgatctcttcgggcagcaccaccacacatccact
1561   cagattagcggcgacaccggtgactgttgcaatccgcaatcgacATGCAACTTCCAGCCG
                                     M Q L P A
1621   CAGCCCAATGGCTGCTCAGCTTCCCGGAAAGCCTCACTTGCTGACAATCATCGTCAGG
      A A Q W L L T L P A K A S L A D N H R Q
1681   TTCACCAGAAGCCCCTCGCGAGATCTGAACCTTTTTACCCGTCGCCATGGATGAATCCCA
      V H Q K P L A R S E P F Y P S P W M N P
1741   ACGCCGACGGCTGGGCGGAGGCCTATGCCAGGCCAAGTCCTTTGTCTCCCAAATGACTC
      N A D G W A E A Y A Q A K S F V S Q M T

```

FIGURE 13 (CONT'D)

..SEQ ID NOs: 11 and 12 ..bgl1. gene encoding BGL

1801 TGCTAGAGAAGGTCAACTTGACCACGGGAGTCGGgtaagttttgtcattttgtccaggtta
 L L E K V N L T T G V G

C1 Bgl1 236 for

1861 acatgcaaattggttctgctaacaataacttaccgtagCTGGGGGCTGAGCAGTGGCTCG
 W G A E Q C V

1921 GCCAAGTGGGCGCGATCCCTCGCCTTGGACTTCGCAGTCTGTGCATGCATGACTCCCTC
 G Q V G A I P R L G L R S L C M H D S P

1981 TCGGCATCCGAGGAGCCGACTACAACTCAGCGTTCCCTCTGGCCAGACCGTTGCTGCTA
 L G I R G A D Y N S A F P S G Q T V A A

2041 CCTGGGATCGCGGTCTGATGTACCGTCGCGGCTACGCAATGGGCCAGGAGGCCAAAGGCA
 T W D R G L M Y R R G Y A M G Q E A K G

2101 AGGGCATCAATGTCCTTCTCGGACCAGTCGCGGGCCCCCTTGGCCGCATGCCCGAGGGCG
 K G I N V L L G P V A G P L G R M P E G

2161 GTCGTAAGTGGGAAGGCTTCGCTCCGGATCCCGTCCTTACCGGCATCGGCATGTCCGAGA
 G R N W E G F A P D P V L T G I G M S E

C1Bgl1 682 rev

2221 CGATCAAGGGCATTTCAGGATGCTGGCGTCATCGCTTGTGCGAAGCACTTTATTGGAAACG
 T I K G I Q D A G V I A C A K H F I G N

2281 AGCAGGgtgagtagtcaaagacgggcccgtctcggaccgcgggcttcaagctgctgactct
 E Q

2341 gctgcagAGCACTTCAGACAGGTGCCAGAAGCCCAGGGATACGGTTACAACATCAGCGAA
 E H F R Q V P E A Q G Y G Y N I S E

2401 ACCCTCTCCTCCAACATTGACGACAAGACCATGCACGAGCTCTACCTTTGGCCGTTTGCC
 T L S S N I D D K T M H E L Y L W P F A

2461 GATGCCGTCCGGGCCGGCGTCCGCTCTGTTCATGTGCTCGTACCAGCAGGTCAACAACCTCG
 D A V R A G V G S V M C S Y Q Q V N N S

2521 TACGCCTGCCAGAACTCGAAGCTGCTGAACGACCTCCTCAAGAACGAGCTTGGGTTTCAG
 Y A C Q N S K L L N D L L K N E L G F Q

2581 GGCTTCGTCATGAGCGACTGGCAGGCACAGCACACTGGCGCAGCAAGCGCCGTGGCTGGT
 G F V M S D W Q A Q H T G A A S A V A G

2641 CTCGATATGTCCATGCCGGGCGACACCCAGTTCAACACTGGCGTCAGTTTCTGGGGCGCC
 L D M S M P G D T Q F N T G V S F W G A

2701 AATCTCACCCCTCGCCGTCCTCAACGGCACAGTCCCTGCCTACCGTCTCGACGACATGGCC
 N L T L A V L N G T V P A Y R L D D M A

FIGURE 13 (CONT'D)

SEQ ID NOs: 11 and 12 *bgl1* gene encoding BGL

```

2761      ATGCGCATCATGGCCGCCCTCTTCAAGGTCACCAAGACCACCCACCTGGAACCCATCAAC
          M R I M A A L F K V T K T T H L E P I N
2821      TTCTCCTTCTGGACCGACGACACTTATGGCCCGATCCACTGGGCCGCCAAGCATGGCTAC
          F S F W T D D T Y G P I H W A A K H G Y
2881      CAGAAGATTAATTCCCACGTTGACGTCCGCGCCGACCACGGCAACCTCATCCGGGAGATT
          Q K I N S H V D V R A D H G N L I R E I
2941      GCCGCCAAGGGTACGGTGCTGCTGAAGAATAACGGCTCTCTACCCCTGAACAAGCCAAAG
          A A K G T V L L K N T G S L P L N K P K
3001      TTCGTGGCCGTCAATCGGCGAGGATGCTGGGTGAGCCCCAACGGGCCAACGGCTGCAGC
          F V A V I G E D A G S S P N G P N G C S
3061      GACCGCGGCTGTAACGAAGGCACGCTCGCCATGGGCTGGGGATCCGGCACAGCCAACTAT
          D R G C N E G T L A M G W G S G T A N Y
3121      CCGTACCTCGTTTTCCCCGACGCCGCGCTCCAGGCCCGGGCCATCCAGGACGGCACGAGG
          P Y L V S P D A A L Q A R A I Q D G T R
3181      TACGAGAGCGTCCTGTCCAACACGCGGAGGAAAAGACAAAGGCTCTGGTCTCGCAGGCC
          Y E S V L S N Y A E E K T K A L V S Q A
3241      AATGCAACCGCCATCGTCTTCGTCAATGCCGACTCAGGCGAGGGCTACATCAACGTGGAC
          N A T A I V F V N A D S G E G Y I N V D
3301      GGTAACGAGGGCGACCGTAAGAACCTGACTCTCTGGAACAACGGTGATACTCTGGTCAAG
          G N E G D R K N L T L W N N G D T L V K
3361      AACGTCTCGAGCTGGTGCAGCAACACCATCGTTCGTATCCACTCGGTCGGCCCGGTCTCTC
          N V S S W C S N T I V V I H S V G P V L
3421      CTGACCGATTGGTACGACAACCCCAACATCACGGCCATTCTCTGGGCTGGTCTTCCGGGC
          L T D W Y D N P N I T A I L W A G L P G
3481      CAGGAGTCGGGCAACTCCATCACCGACGTGCTTTACGGCAAGGTCAACCCCGCCGCCCGC
          Q E S G N S I T D V L Y G K V N P A A R
3541      TCGCCCTTCACTTGGGGCAAGACCCGCGAAAGCTATGGCGCGACGTCTGTACAAGCCG
          S P F T W G K T R E S Y G A D V L Y K P
3601      AATAATGGCAATGGTGCGCCCAACAGGACTTCACCGAGGGCGTCTTCATCGACTACCGC
          N N G N G A P Q Q D F T E G V F I D Y R
3661      TACTTCGACAAGGTTGACGATGACTCGGTCTATCTACGAGTTCGGCCACGGCCTGAGCTAC
          Y F D K V D D D S V I Y E F G H G L S Y
3721      ACCACCTTCGAGTACAGCAACATCCGCGTCGTCAAGTCCAACGTGAGCGAGTACCGGCC
          T T F E Y S N I R V V K S N V S E Y R P
3781      ACGACGGGCACCACGGCCCAGGCCCGACGTTTGGCAACTTCTCCACCGACCTCGAGGAC

```

FIGURE 13 (CONT'D)

SEQ ID NOs: 11 and 12 --bgl1-gene encoding BGL

```

      T T G T T A Q A P T F G N F S T D L E D
3841 TATCTCTTCCCAAGGACGAGTTCCCTACATCTACCAGTACATCTACCCGTACCTCAAC
      Y L F P K D E F P Y I Y Q Y I Y P Y L N
3901 ACGACCGACCCCGGAGGGCTCGGCCGATCCCCACTACGGCCAGACCGCCGAGGAGTTC
      T T D P R R A S A D P H Y G Q T A E E F
3961 CTCCCGCCCCACGCCACCGATGACGACCCCGAGCCGCTCCTCCGGTCTCGGGCGGAAAC
      L P P H A T D D D P Q P L L R S S G G N
4021 TCCCCGGCGGCAACCGCCAGCTGTACGACATTGTCTACACAATCAGGCCGACATCAAG
      S P G G N R Q L Y D I V Y T I T A D I T
4081 AATACGGGCTCCGTTGTAGGCGAGGAGGTACCGCAGCTCTACGTCTCGCTGGGCGGTCCC
      N T G S V V G E E V P Q L Y V S L G G P
4141 GAGGATCCCAAGGTGCAGCTGCGCGACTTTGACAGGATGCGGATCGAACCCGCGGAGACG
      E D P K V Q L R D F D R M R I E P G E T
4201 AGGCAGTTCACCGCCCGCTGACGCGCAGAGATCTGAGCAACTGGGACGTCACGGTGCAG
      R Q F T G R L T R R D L S N W D V T V Q
4261 GACTGGGTCTACGAGGTATCCCAAGACGGCATATGTTGGGAGGAGCAGCCGGAAGTTG
      D W V I S R Y P K T A Y V G R S S R K L
4321 GATCTCAAGATTGAGCTTCCTTGAATGAGTTTCATCAGGGGCTGCAGAGGGATGGTAACA
      D L K I E L P *
4381 CGTTCTTAATCAGAAGTATGATGGAGAAAAGCACTTGGCAAGTTCGGTGAGCAAAAAGA
4441 AGGCACCTTATTAAGTGTAGGCGGTGTTCTATGTTTAAAGGTGCTATGTTTACATATAA
4501 TTAGTATATAATGATTTAATAATTATGTTTAGCAGTTGCTAATGTGTAATTTTCGGCGT
4561 GTGATGACTGCTACAACACTGGTCTGTCTTCTAGTCGCCATTGTTAATTATGAAGGTTA
4621 TTGTCTACAATTTCTAATACCTTATGGATGATTGCCCAGCTGGTTTCAAACTCGTTACGC
4681 GCAATGGTACGATTGAGGTATTATTTCATTGTAAGTACCTCCGTACAGCGTCCCAACTA
4741 TTTCCATTACGAGATGCCTCGCTTTTCGGTGCTTTTCGGAACAGGGCTGGCAGCGGATCA
4801 TGGCGCGATCAAAACATGGCGAGCAGCTGTCCAGGACGGAGGACAGGTTGGGGACTGATG
4861 CCTCCCGGACGCATTAAGGTCAGAAGATAGACACGTTTACACAGCGTTGAGACCGACAA
4921 GCCACATTAGGCAGCGCCGGTTGCACCACCGCCGTCACGGGCAACGGTTCAATCAATCGA
4981 CAACAGTGGAAGACAAAGTACTGAAGATCAGGTATTAAATAGTGTGAGAGAGAAAACAGACG
5041 GTGGAAGTGGGTGCTAATATTTCTCTTGATTTCGGTGTCCATGGTAGTACAGAACACAA
5101 GAAAAAGAAGGAGGAGTGAGCGGAGAAGGAGGAGGGGAAGCCAGAAAAAGAACATGAA
5161 AAAGCATACACATTGGAGTCGGTCAGTCGGTTGATTGGTTTGGTAGAGAGCGAAAAAGCA
5221 AGCGTCACCTGTAGGATTTCGAACCTACGCTCCCGAAGGAACTGCCTAAGAACGCTAAGCA

```

FIGURE 13 (CONT'D)

SEQ ID NOs: 11 and 12 bgl1 gene encoding BGL

```

5281      AGGTTAGCAGGGCAGCGCGTTAACCACTCCGCCAAAGTGACTGTCGTTGATCATGGTCGA
5341      ATTCAGTAGCTTATAGGAGTTCAACCAGATCACAAATGCATAGGTGCTCGTAGAACGGT
5401      CTAAGTATGAGTTGATTATAAGCAACCGAATGGCTCTCAGCGGCAACACCGTAGCTGAAG
5461      TAACAAAACGCACCTTTGGTTACTTTCTGACTATAAAAATGGGATATTTGGAAATGACCA
5521      CCCGATAAGGTGTCAAATCTAAATGACTGTCTGGGTGTGAAGATGTTACTGTGGTTCCA
5581      CCACGAACCAGTTTTAGTATCCGCATGCTTCAGTCTCTGCGCCTCGACAGGCGGAGGGTG
5641      TGTGTTAGATCAGAATCGATGTGACGCTGTGACCGGAGGCTCTCGAGCCTAGGTGCGGT
5701      AGTTCTGTTCAAAAAGAAGTGTGTGGCCGGGTTTGGGCGCCCTTATAGCCTACCATCCTG
5761      GCTGTGGTTCCCGAGCGGGAGCCGGTTCTCCGTTTTGGTTCCGATAAAGTGTCAATCTG
5821      CCTCCCGGTTTCGCATCTAATTTCTGACTTCGTTCCGGACCTCTGGAGACGTAGGGATAG
5881      GTATGGGATATGCCCGGCATTTCTGTAATGTCCATAGTCTCTTTCCGGACGAGGCGGCAA
5941      GCTCTCAGAGCTATCTAAGCTTAACCAACCCCTGATCCTTAACCCCTCCCAGACCACACT
6001      CCTGGGAGAATAAACCGGGCTCCAAGATCGAAATCGAAATCAGTGCGCGAACTTGAAATC

```

FIGURE 14

SEQ ID NOs: 13 and 14 eg5 gene encoding EGV

atg	cat	ctc	tcc	gcc	acc	acc	ggg	ttc	ctc	gcc	ctc	ccg	gcc	ctg	gcc	48
Met	His	Leu	Ser	Ala	Thr	Thr	Gly	Phe	Leu	Ala	Leu	Pro	Ala	Leu	Ala	
1				5					10					15		
ctg	gcc	cag	ctc	tcg	ggc	agc	ggc	cag	acg	acc	cgg	tac	tgg	gac	tgc	96
Leu	Ala	Gln	Leu	Ser	Gly	Ser	Gly	Gln	Thr	Thr	Arg	Tyr	Trp	Asp	Cys	
				20				25						30		
tgc	aag	ccg	agc	tgc	gcc	tgg	ccc	ggc	aag	ggc	ccc	tcg	tct	ccg	gtg	144
Cys	Lys	Pro	Ser	Cys	Ala	Trp	Pro	Gly	Lys	Gly	Pro	Ser	Ser	Pro	Val	
		35						40						45		
cag	gcc	tgc	gac	aag	aac	gac	aac	ccg	ctc	aac	gac	ggc	ggc	tcc	acc	192
Gln	Ala	Cys	Asp	Lys	Asn	Asp	Asn	Pro	Leu	Asn	Asp	Gly	Gly	Ser	Thr	
		50					55							60		
cgg	tcc	ggc	tgc	gac	gcg	ggc	ggc	agc	gcc	tac	atg	tgc	tcc	tcc	cag	240
Arg	Ser	Gly	Cys	Asp	Ala	Gly	Gly	Ser	Ala	Tyr	Met	Cys	Ser	Ser	Gln	
65						70				75					80	
agc	ccc	tgg	gcc	gtc	agc	gac	gag	ctg	tcg	tac	ggc	tgg	gcg	gcc	gtc	288
Ser	Pro	Trp	Ala	Val	Ser	Asp	Glu	Leu	Ser	Tyr	Gly	Trp	Ala	Ala	Val	
				85					90					95		
aag	ctc	gcc	ggc	agc	tcc	gag	tcg	cag	tgg	tgc	tgc	gcc	tgc	tac	gag	336
Lys	Leu	Ala	Gly	Ser	Ser	Glu	Ser	Gln	Trp	Cys	Cys	Ala	Cys	Tyr	Glu	
				100					105					110		
ctg	acc	ttc	acc	agc	ggg	ccg	gtc	gcg	ggc	aag	aag	atg	att	gtg	cag	384
Leu	Thr	Phe	Thr	Ser	Gly	Pro	Val	Ala	Gly	Lys	Lys	Met	Ile	Val	Gln	
				115				120						125		

FIGURE 14 (CONT'D)

SEQ ID NOs: 13 and 14 eg5 gene encoding EGV

gcg acc aac acc ggt ggc gac ctg ggc gac aac cac ttt gac ctg gcc	432
Ala Thr Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Ala	
130 135 140	
gtgagttgcc tccccctctc cccggaccgc tcagattaga tgagattaga ctttgcctgt	492
aaatcggtcc aagattccct tgactgacca acaaacaatca tacgggcag atc ccc ggt	550
Ile Pro Gly	
145	
ggc ggt gtc ggt att ttc aac g gtaagctggt gcccccggac ccctcccgg	602
Gly Gly Val Gly Ile Phe Asn	
150	
acccctcccc cttttcctcc agcgagccga gttgggacgc ccgagatcga gaactcacac	662
aacttctctc tcgacag cc tgc acc gac cag tac ggc gct ccc ccg aac	711
Ala Cys Thr Asp Gln Tyr Gly Ala Pro Pro Asn	
160 165	
ggc tgg ggc gac cgc tac ggc ggc atc cat tcc aag gaa gag tgc gaa	759
Gly Trp Gly Asp Arg Tyr Gly Gly Ile His Ser Lys Glu Glu Cys Glu	
170 175 180	
tcc ttc ccg gag gcc ctc aag ccc ggc tgc aac tgg cgc ttc gac tg	806
Ser Phe Pro Glu Ala Leu Lys Pro Gly Cys Asn Trp Arg Phe Asp Trp	
185 190 195	
gtacgttgc ttagacatacc ggaacccaat tctccaacc ccccccttt tctccccaa	866
ctccgggggt agtcggaatg tcgcgactga ccctatttca g g ttc caa aac gcc	920
Phe Gln Asn Ala	
200	

FIGURE 14 (CONT'D)

.. SEQ ID NOs: 13 and 14 .. eg5 gene encoding EGV

gac aac ccg tcg gtc acc ttc cag gag gtg gcc tgc ccg tcg gag ctc	968
Asp Asn Pro Ser Val Thr Phe Gln Glu Val Ala Cys Pro Ser Glu Leu	
205 210 215	
acg tcc aag agc ggc tgc tcc cgt taa	995
Thr Ser Lys Ser Gly Cys Ser Arg	
220 225	

FIGURE 15

SEQ ID NOs: 15 and 16 eg7 encoding EG VI

```

1      GCGCTTCCGGCCTGGGCGAGTAAATGACGGAAGCCgggccccgtccgactgcggttgtc
61     ccaactcggaagcaggcatcggtttttgggcgggaggaagcgttgcaacacgcactatcg
121    ccaaggtggactcggcgcaatctggagggttcggcccgaggagacggaatccgggctgaa
181    tctgcgcaaaggctgaccctgcgatgggtgggaaatgtaaatatgtgaagttataggcat
241    ataggactcagcgatgacatggaaattgcagaggcatgtgggatttcagcgtttggcatg
301    cattgggtcggtatctctcgccctgtctgatgtgatcccgccggagggtgttcggtctctgg
361    ggaagggacccccctggccccccacctgccccgcacatgcctcgccacgactcccgcg
421    cgccgagggaagaacttcgggtctttgtgacgggagattccactgagtgaacattggccaa
481    ccaagcacacaattactccgtacatacacagtacttctgactccgtaaagtaaacggtgt
541    gtttcaaagatcggtaatccgtaacagggtactccgtatctaaggtaaatattaccctgtgc
601    acggagcagaacctgaacttcttccccctcttactcgagtagtcaccctactccaacca
661    gcggcttttcaactcgcaaagtcttgtttataacagtgcataatacctgcatttcgtatct
721    cgctagtgtaaagacgaccacacgcggacaagaagaaaaatccaattgcccgatggct
781    cttagtttgaggacagcagogaaggactacactgcgccgtagtaccaggccaagaaacg
841    cgaatcgatatattaacggcacaatcaaatggattatatgccatttcgctccgggttgcg
901    tgcctcgccgaagtctgggtgccgatcgattgcgaacccccggaatcgccgggatgattcct
961    acagccgcgaaaggggggggggggggggggggtctggacgggacgtgcataaacttcgaa
1021   tttctagaatattgcggattgggttcccttcagccctgcgagcgcgcccccttctggaac
1081   cgcacccttcaccggttcacacacacagaggacatgggtggaaatgtgtacctgacgggtg
1141   cccctttgggacagtggagaggcggatgttcggataaaccatccggagcgcagtgctgcac
1201   caagatcttggcttaccatcgacaccaacatgcggactcgtccctcagtcatggagcctt
1261   ggctcgccgagcctccgttcgaagcggctatcccgctcctgccagcggaggatctcgtaac
1321   gcttcgcgaaactgtgaatgtcctgggtataagagcatggcgcgaccttgtctcgtcagg
1381   aacggggaggaggagggttgggttagggtcgcggtcgtttggagattgctgagctctgag
1441   ccttcgggtccttggatccctgcgggtccccgggtctcctctctctctctctctctctc
1501   tctctctctctcttcccacgctcgttcgacagacgcctccccctctcttcgctctctctccc
1561   tcgcacgtagcacactaatagtgcaccATGCGCGTCTCTAGTTTGGTCGCGGCCCTTGCT
                                     M R V S S L V A A L A

1621   ACCGGTGGTCTTGTCGCCGCCACGCCTAAGCCCAAGGGGTGTCGCCCCCTGGGGCCGTG
       T G G L V A A T P K P K G S S P P G A V

1681   GACGCGAACCCTTTCAAGGGCAAGACGCAGTTGTC AACC CGGCATGGGCGGCCAAGCTG
       D A N P F K G K T Q F V N P A W A A K L

```

FIGURE 15 (CONT'D)

SEQ ID NOs: 15 and 16 eg7 encoding EG VI

1741	GAACAGACCAAAAAGGCGTTCCTGGCCAGGAACGACACCGTCAATGCCGCCAAGACGGAG E Q T K K A F L A R N D T V N A A K T E
1801	AAGGTCCAGCAGACCAGCTCGTTCGTCTGGGTCTCGAGGATCGCCGAGCTCTCCAACATC K V Q Q T S S F V W V S R I A E L S N I
1861	GACGACGCCATCGCGGCTGCCCGCAAGGCGCAGAAGAAGACGGGCAGGAGGCAGATCGTC D D A I A A A R K A Q K K T G R R Q I V
1921	GGCCTGGTGCTCTACAACCTTCCGGACCGGACTGCAGCGCGGGCAGAGCGCGGGCGAG G L V L Y N L P D R D C S A G E S A G E
1981	CTCAGCAGCGACAAGAACGGGCTCGAGATCTACAAGACTGAGTTCGTCAAGCCCTTCGCC L S S D K N G L E I Y K T E F V K P F A
2041	GACAAGGTGGCGGCCGCAAAGGACCTCGACTTCGCCATCGTCTCTGGAGCCCGACTCGCTG D K V A A A K D L D F A I V L E P D S L
2101	GCCAACCTGGTCACCAACCTGGGCATCGAGTTCTGCGCCAACGCCGCCCGCTCTACCGC A N L V T N L G I E F C A N A A P V Y R
2161	GAGGGCATCGCCTATGCCATCTCCAGCCTTCAGCAGCCAAACGTGCACTTGTACATCGAT E G I A Y A I S S L Q Q P N V H L Y I D
2221	GCTGCCCACGGCGGCTGGCTCGGCTGGGACGACAACCTGCCGCTGGCCGCCAAGGAGTTT A A H G G W L G W D D N L P L A A K E F
2281	GCCGAGGTGGTCAAGCTTGCCGGCGAGGGCAAGAAGATCCGCGGCTTCGTCAACCAACGTG A E V V K L A G E G K K I R G F V T N V
2341	TCCAACCTACAACCCCTTCCACGCCGTCGTGCGCGAGAACTTTACCGAGTGGAGCAACTCG S N Y N P F H A V V R E N F T E W S N S
2401	TGGGACGAGTCTCACTACGCCTCCTCGCTCACACCGTTCTCGAGAAAGAGGGGCTGCCG W D E S H Y A S S L T P F L E K E G L P

FIGURE 15 (CONT'D)

SEQ ID NOs: 15 and 16 eg7 encoding EG VI

2461 GCACGCTTCATCGTCGACCAGGGTCGCGTTGCCCTCCCGGGAGCCCGCAAGGAGTGgtga
 A R F I V D Q G R V A L P G A R K E W

2521 gtttcgaccagattgaccctcgacccatgcgaccgagattgctgacgattgaattgogtg

2581 tcccgtccccagGGGTGAATGGTGCAACGTGGCACCCGCGGATTTGGCCCCGCGCCCA
 G E W C N V A P A G F G P A P

2641 CGACCAGGGTCAACAACACCGTCGTCGATGCTCTCGTCTGGGTCAAGCCTGGCGGCGAGA
 T T R V N N T V V D A L V W V K P G G E

2701 GCGACGGCGAGTGTGGCTTGGCTGGCGCCCCCAAGGCCGGCCAGTGGTTCGACGAGTACG
 S D G E C G L A G A P K A G Q W F D E Y

2761 CCCAGATGCTGGTCGAGAATGCCACCCGCTCTGTCTGTCCACAAGTGGTAGataaaattttg
 A Q M L V E N A H P S V V H K W *

2821 gagtccgagaaggggtcccagatagacttttggttttaaaacaaatgcaaggtgtcgacag
 2881 atactggcttaacattaaccaagcaccatgaacatgacttgtcaacatattgatacattc
 2941 cgctgctttcccatagctgctctcaggtctcagggatcaaattggataggtcggttaatgca
 3001 aaacgatccattggatatccagaagagagaaaaaaaaggacatgcatgccttgtctgt
 3061 catcatgaggaaacaaaggaaaaacaaacgatcgtcgtgttccaacaagctttccaagac
 3121 cacaagacccatccaccaacacaaccaaacgacaagcaatacgaatggaccgctgtgttc
 3181 catctctcaagagctgactaaacgaacagtcgttgaaatcatcctacatgagtacgccgc
 3241 accacctgttatcgtgtaaaccaaatacgctgttaaagtgcacatctcttaggtatgat
 3301 cgtaagttccggtcacgggtcacggatcagggatggttctcaattcgtgtgtcgcgtagcc
 3361 gccgccgtatctggacaagacttcttgtattgctccgaacccgcttttgccgccctaata
 3421 atctgtagccttcttacctgggtggtgccttgaaagacgcggcaggcaacacttcgcaggt
 3481 ctgtggcgaccagcaccaggctgtggtgatccccggaaccggtcgtcgacttgctcgc
 3541 ggtgtcctcggtggtggggtgaggggtgatgagggcttgaggggtgttgttcgcgccgc
 3601 aacatccggctccggctccggaccgtccacagacattggacctgcgagcatgaactcgtgc
 3661 ctccagccagaccaaagccatgccatcatcgctctgcccagcgtgttgagcgggaggtc
 3721 gatgttctcagccagaactcggggctgtacggccatgacctgggctgttcggtctggcc
 3781 gtcttgccggcggtttctccctgccagcttgttggtgcgcggtgcctgcgagattcgacttc

FIGURE 15 (CONT'D)

SEQ ID NOs: 15 and 16 eg7 encoding EG VI

```
3841      gacctgggcgtggcagaggggtgacgagggacggttgacgccttgatctccttgcctcccat
3901      gtocctccaccggtacagggcgacgggtgccatagcggtccacagcctgcacgagaacct
3961      cagggcgtcgtcaatgagttctgtcaacttgctctccagcctctctatgccgcgagcatc
4021      ctgatcctggagcagaaaccgtgccgagcctccgaggaaacgctccttcagcttccgcgc
4081      gtagtttaggcgtgattcaacaaacgtccggcgggactcgttggtgcccgcagcagcgac
4141      gtccttgatgctgaagccgcctcgggcgaacaggcgcatcatctggggccc
```

FIGURE 16

SEQ ID NOs: 17 and 18 xyl2 gene encoding Xyl II

```

1      cgcgcccccgctctttgaacgcttgagaagcgacgggtgaagaaccatcaactccgattcc
61     gctcctcatcctcccaagcgcgattgaaatagccacagcggtatgtacggattactc
121    tgctccgctttgcacatccatacacagcgctatttttaaagttcaggacggccaagcccg
181    gttcttggaacggacgacccggattccgaaagctccagcgctcaatgcggtcagtcgtgg
241    cgctgatcctgctgatctgctgatctcataaaccgcaacttcaacttttcactttgaag
301    cgtatacacgcagcgctctttcaccggcgcttatactcgcaaattaaccgctaataat
361    cctcgacttggataaatgtgtagccgacacggaggaggggggttgggggggggttggggg
421    gagacatgatggctgtcccaacggatattatttttgttgtttgtataattactcgcg
481    caacattctcaaagggcgctgcctcgcgcgaggaaagcccatgacagagaattggacag
541    ctccaagctcgcatatactctaacaacggcgctgactcggcaatgaaggcctgccgctcg
601    agtgatagggcggaagtaaaacggacgcttacatgcggcacttagccggctgatgccggaga
661    atacgggattcaacgatacaatcacacgatgcgacacacctcggcgacttggcgctctat
721    ggaagaaggtcgggttaaagctggcgtagattttgcgctcttggtttcttaaccgggtt
781    atttctatttctcatatgcgcgagcgaatgcggggtgcagagcgcccgaggatcgatgg
841    tcctatcagacaagagcctggccccggaacctgggataatagaagccaaattaagccatg
901    ggagtatcgtccgggggtaggaaacgcacggggcaactagaggaggaagaatttggtataa
961    agggaggacggcggaacaggcttgatggacatgaatcagaagacgacactgggcaactaa
1021   acagcttcagcagaggttttgccttgcataggccctcgatatcATGGTCTCGTTCACT
                                           M V S F T
1081   CTCCTCCTCACGGTCATCGCCGCTGCGGTGACGACGGCCAGCCCTCTCGAGGTGGTCAAG
361    L L L T V I A A A V T T A S P L E V V K

1141   CGCGGCATCCAGCCGGGCACGGGCACCCACGAGGGGTACTTCTACTCGTTCTGGACCGAC
381    R G I Q P G T G T H E G Y F Y S F W T D
1201   GGCCGTGGCTCGGTGCACTTCAACCCCGGGCCCCGCGGCTCGTACAGCGTCACCTGGAAC
401    G R G S V D F N P G P R G S Y S V T W N
1261   AACGTCAACAACTGGGTGCGCGCAAGGGCTGGAACCCGGGCCCCGCCGCAAGATTGCG
421    N V N N W V G G K G W N P G P P R K I A
1321   TACAACGGCACCTGGAACAACTACAACGTGAACAGCTgtgcgttgctcctcctcttctcc
441    Y N G T W N N Y N V N S
1381   Ctttcgcttggttttccttgatgattgggatccatttttaaagagaaggaaaaaaaaaaca
C1 xyl10 423 for
1441   aaggaaaatagaagataactaacgccaagctctggcagACCTCGCCCTGTACGGCTGGAC
Y L A L Y G W T

```

FIGURE 16 (CONT'D)

SEQ ID NOs: 17 and 18 xyl2 gene encoding Xyl II

1501 TCGCAACCCGCTGGTCGAGTATTACATCGTGGAGGCATACGGCACGTACAACCCCTCGTC
 501 R N P L V E Y Y I V E A Y G T Y N P S S
 1561 GGGCACGGCGCGGCTGGGCACCATCGAGGACGACGGCGGCGTGACGACATCTACAAGAC
 521 G T A R L G T I E D D G G V Y D I Y K T
 1621 GACGCGGTACAACCAGCCGTCATCGAGGGGACCTCCACCTTCGACCAGTACTGGTCCGT
 541 T R Y N Q P S I E G T S T F D Q Y W S V
 1681 CCGCCGCCAGAAGCGCGTCGGCGGCACTATCGACACGGGCAAGCACTTTGACGAGTGGAA
 561 R R Q K R V G G T I D T G K H F D E W K
 C1 Xyl110 722 rev
 1741 GCGCCAGGGCAACCTCCAGCTCGGCACCTGGAACTACATGATCATGGCCACCGAGGGCTA
 581 R Q G N L Q L G T W N Y M I M A T E G Y
 1801 CCAGAGCTCTGGTTCCGGCCACTATCGAGGTCCGGGAGGCCTAAagaagccaggcgcccttt
 601 Q S S G S A T I E V R E A *
 1861 cttttgttttgaggaggggtagagggggggggagggaacgaaaagtagcaggggt
 1921 ggtttttatgccggcagccgtggggcattcgagtgcaacctgtatctctctctcccaag
 1981 tctccgggctccttctcagagaacttcaatatgtctggggacaaaccaccttgtaaata
 2041 caacggtaattatctaagtttgagtgccctatcgatgcttctgaaaatttctgctcct
 2101 tgatacaagtcggtttgagccgagccaatgagactgtgtcgattgatagaggccctgaag
 2161 gatcaagcgcatgcaacaattaagcatgactacgtgcctagctgcagataaatggaagc
 2221 cactcaccaagggtcaacccgcataactggcacgtaagaaccttccgtgtacaaggcccaa
 2281 ccgactcacatatctatctgcttgggttttgggatgagggtttttaccacaaaaacaaat
 2341 ttgatacaatgctctgctgtgcccgggttgcctgagaccaagccgtaatacagcgggcaggg
 2401 aatcgagtaggtcacgcctgtgcttgggtctagaacaaactaatatataaaagccttggtg
 2461 ctcggcacacatacagaactcgacctgaggcatgttcttggaaggcggctagccagtcac
 2521 gtctggcaccaggccttggtctcgctcagggataccgagggcgaggaggatgaggaagacc
 2581 tctttctcgccctcagatctcttaggggacgaagaagacaacgcgggagccacacaataat
 2641 taggtctcatatcagacgtttcggcctggccgagctaataatgtctaattatgcccatcag
 2701 ccgtatgtcgaggcaggttgaccgatacgctcgccgcccgcctcattcatctccgact
 2761 gggcaccaatgtcgccatctcgccgtcaaggtggtgcaagatacctattatgcaagcaga
 2821 ggatcagatggcgggcccatacgagcggctgctccggcttgcgagaaagccgcttcgcag
 2881 caaggatcgctggcaggccgccattttcggttgggtattctttgtcttgtttgcttcgta
 2941 attatgtcctggctggcattgtgggaagggggcgaacctcttgatttccgatgggggtcga

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 6015707 A [0007]
- US 6573086 B [0009]
- US 4940838 A [0066]
- US 5464763 A [0066]

Non-patent literature cited in the description

- BUNGAY H.R. Energy: the biomass options. Wiley, 1981 [0002]
- OLSSON L ; HAHN-HÄGERDAL B. Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme Microb Technol*, 1996, vol. 18, 312-31 [0002]
- ZALDIVAR J ; NIELSEN J ; OLSSON L. Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl Microbiol Biotechnol*, 2001, vol. 56, 17-34 [0002]
- GALBE M ; ZACCHI G. A review of the production of ethanol from softwood. *Appl Microbiol Biotechnol*, 2002, vol. 59, 618-28 [0002]
- The Rise Of Industrial Biotech. *Forbes*, 24 July 2006 [0002]
- BUKHTOJAROV et al. *Biochemistry (Moscow)*, vol. 69 (5), 666-677 [0007]
- SAMBROOK et al. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Labs Press, 1989 [0039] [0060]
- STEMMER. *PNAS*, 1994, vol. 91, 10747 [0048]
- Procedures for Introducing Foreign DNA into Plants. MIKI et al. *Methods in Plant Molecular Biology and Biotechnology*. CRC Press, Inc, 1993, 67-88 [0065]
- Vectors for Plant Transformation. GRUBER et al. *Methods in Plant Molecular Biology and Biotechnology*. CRC Press, Inc, 1993, 89-119 [0065]
- HORSCH et al. *Science*, 1985, vol. 227, 1229 [0066]
- KADO, C.I. *Crit. Rev. Plant. Sci.*, 1991, vol. 10, 1 [0066]
- MOLONEY et al. *Plant Cell Reports*, 1989, vol. 8, 238 [0066]
- SANFORD et al. *Part. Sci. Technol.*, 1987, vol. 5, 27 [0067]
- SANFORD, J.C. *Trends Biotech.*, 1988, vol. 6, 299 [0067]
- SANFORD, J.C. *Physiol. Plant*, 1990, vol. 79, 206 [0067]
- KLEIN et al. *Biotechnology*, 1992, vol. 10, 268 [0067]
- ZHANG et al. *Bio Technology*, 1991, vol. 9, 996 [0068]
- DESHAYES et al. *EMBO J.*, 1985, vol. 4, 2731 [0068]
- CHRISTOU et al. *Proc Natl. Acad. Sci. USA*, 1987, vol. 84, 3962 [0068]
- HAIN et al. *Mol. Gen. Genet.*, 1985, vol. 199, 161 [0068]
- DRAPER et al. *Plant Cell Physiol.*, 1982, vol. 23, 451 [0068]
- DONN et al. *Abstracts of VIIth International Congress on Plant Cell and Tissue Culture IAPTC*, 1990, vol. A2-38, 53 [0068]
- D'HALLUIN et al. *Plant Cell*, 1992, vol. 4, 1495-1505 [0068]
- SPENCER et al. *Plant Mol. Biol.*, 1994, vol. 24, 51-61 [0068]
- GUSAKOV AV ; SINITSYN AP ; SALANOVICH TN ; BUKHTOJAROV FE ; MARKOV AV ; USTINOV BB ; VAN ZEIJL C ; PUNT P ; BURLINGAME R. Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*. *Enzyme Microb Technol*, 2005, vol. 36, 57-69 [0088] [0092]
- BUKHTOJAROV FE ; USTINOV BB ; SALANOVICH TN ; ANTONOV AI ; GUSAKOV AV ; OKUNEV ON ; SINITSYN AP. Cellulase complex of the fungus *Chrysosporium lucknowense*: isolation and characterization of endoglucanases and cellobiohydrolases. *Biochemistry (Moscow)*, 2004, vol. 69, 542-51 [0088]
- SMITH BE. Protein sequencing protocols. Humana Press, 1997 [0090]
- Proteome research: mass spectrometry. Springer-Verlag, 2001 [0090]
- Methods of investigation and properties of cellulolytic enzymes. SINITSYN AP ; CHEMOGLAZOV VM ; GUSAKOV AV. *Biotechnology Series*. VINITI Press, 1990, vol. 25 [0091]
- SOMOGYI M. Notes on sugar determination. *J Biol Chem*, 1952, vol. 195, 19-23 [0091]
- GHOSE TK. Measurement of cellulase activities. *Pure Appl Chem*, 1987, vol. 59, 257-68 [0091]

- Methods of investigation and properties of cellulolytic enzymes. **SINITSYN AP ; CHERNOGLAZOV VM ; GUSAKOV AV.** Biotechnology Series. VINITI Press, 1990, vol. 25 **[0093]**

Szabadalmi igénypontok

1. Enzim-készítmény cellulóz hidrolízisére, amely készítmény olyan béta-glukozidázt (BGL) tartalmaz, amely a SEQ ID NO: 11 szerinti szekvenciájú *bglI* gén expresszáltatásával állítható elő.

2. Eljárás fermentálható cukrok előállítására lignocellulóz anyagból, amely eljárás magában foglalja

(a) cellulóz hidrolízisére szolgáló enzim-készítmény biztosítása, amely készítmény olyan béta-glukozidázt (BGL) tartalmaz, amely a SEQ ID NO: 11 szerinti szekvenciájú *bglI* gén expresszáltatásával állítható elő;

(b) a készítmény alkalmazását lignocellulóz anyagon, fermentálható cukrokat hozva létre.

3. Eljárás fermentációs termék vagy fermentációs termék előállítására szolgáló kiindulási anyag előállítására fermentálható cukorból, amely eljárás magában foglalja

(a) cellulóz hidrolízisére szolgáló enzim-készítmény biztosítása, amely készítmény olyan béta-glukozidázt (BGL) tartalmaz, amely a SEQ ID NO: 11 szerinti szekvenciájú *bglI* gén expresszáltatásával állítható elő;

(b) a készítmény alkalmazását lignocellulóz anyagon, fermentálható cukrokat hozva létre; és

(c) a fermentálható cukor fermentálását, fermentációs terméket hozva létre.

4. A 3. igénypont szerinti eljárás, ahol a fermentációs termék bio-üzemanyag, vagy a következők által alkotott csoportból van kiválasztva: tejsav, műanyagok, speciális vegyszerek, szerves savak, oldószerek, takarmányadalekkek, gyógyszerek, vitaminok, aminosavak, ipari enzimek, és kémiai alapanyagok.

5. Eljárás energia előállítására fermentálható cukorból, amely eljárás magában foglalja

(a) cellulóz hidrolízisére szolgáló enzim-készítmény biztosítása, amely készítmény olyan béta-glukozidázt (BGL) tartalmaz, amely a SEQ ID NO: 11 szerinti szekvenciájú *bglI* gén expresszáltatásával állítható elő;

(b) a készítmény alkalmazását lignocellulóz anyagon, fermentálható cukrot hozva létre;

(c) a fermentálható cukor fermentálását, éghető fermentációs terméket hozva létre;

(d) az éghető fermentációs termék elégetését, energiát állítva elő.

6. A 2., 3. vagy 5. igénypont szerinti eljárás, ahol a fermentálható cukor a következők által alkotott csoportból van kiválasztva: glükóz, xilóz, arabinóz, galaktóz, mannóz, ramnóz, szacharóz és fruktóz.

7. A 2., 3. vagy 5. igénypont szerinti eljárás, ahol a lignocellulóz anyag a következők által alkotott csoportból van kiválasztva: kertészeti nyesedékek, bozót, malom hulladék, városi fahulladék, kommunális hulladék, fakitermelési hulladék, erdőritkítási hulladék, rövid-tenyészidejű fás növények, ipari hulladék, búza szalma, zab szalma, rizs szalma, árpa szalma, rozs szalma, len szalma, szójahej, kukoricasíker-takarmány, zabkorpa, cukornád, kukorica takarmány, kukoricaszár, kukorica csövek, kukoricacsuhé, prérfű, gamagrass fű, ecsetpázsit, cukorrépa-pép, citruspép, vetőmag héj, cellulóz állati hulladékok, gyp nyesedék, gyapot, hínár, fák, cserjék, fűvek, búza, cukornádrost, kukorica, kukoricamag, magrost, magok nedves vagy száraz őrléséből származó termékek és melléktermékek, kommunális szilárd hulladék, papírhulladék, udvari hulladék, lágyszárú anyag, mezőgazdasági hulladékok, erdészeti hulladékok, cellulóz, papírgyári maradványok, ágak, bokrok, vesszők, energianövények, erdők, gyümölcsök, virágok, magok, fűvek, lágyszárú növények, levelek, kéreg, tűk,

fatörzsek, gyökerek, csemeték, cserjék, vesszős köles, fák, zöldségek, gyümölcs héjak, indák, közepes búza, kemény- és puhafák, mezőgazdasági folyamatokból származó szerves hulladékanyagok, erdészeti fahulladék, és ezek kombinációi.

8. Az 5. igénypont szerinti eljárás, ahol az éghető fermentációs termék alkohol.

9. A 2., 3. vagy 5. igénypont szerinti eljárás, ahol a lignocellulóz anyag a előkezelésnek van alávetve az enzimekkel való érintkeztetést megelőzően, ahol adott esetben az előkezelés magában foglalja a lignocellulóz biomassza kitételét sav, bázis, oldószer, hő, peroxid, ózon, mechanikai aprítás, darálás, őrlés, gyors nyomáscsökkentés, vagy ezek kombinációjának, ahol adott esetben az oldószer aceton/etanol keverék.

10. Fehérje, amely a SEQ ID NO: 11 szerinti szekvenciájú *bgII* gén expresszáltatásával állítható elő.

11. A SEQ ID NO: 11 szerinti nukleinsav-szekvenciájú *bgII* gén.