



(43) International Publication Date
26 May 2017 (26.05.2017)

(10) International Publication Number
WO 2017/087982 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2016/063198

(22) International Filing Date:
21 November 2016 (21.11.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/257,477 19 November 2015 (19.11.2015) US

(71) Applicants: **THE REGENTS OF UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 12th Floor, Oakland, CA 94607-5200 (US). **SANDIA CORPORATION** [US/US]; P.O.Box 5800, MS-0161, Albuquerque, NM 87185 (US).

(72) Inventors: **SINGER, Steven, W.**; 4812B Lawton Avenue, Oakland, CA 94609 (US). **KOLINKO, Sebastian, D.**; 111 Temescal Circle, Emeryville, CA 94608 (US). **DENZEL, Evelyn**; Neustadt Allee 23, 88471 Laupheim (DE). **HIRAS, Jennifer**; 285 Denison Pkwy E #105, Corning, NY 14830 (US). **GLADDEN, John M.**; 2625 Otis Dr., Alameda, CA 94501 (US).

(74) Agent: **CHIANG, Robin, C.**; Lawrence Berkeley National Laboratory, Innovation & Partnerships Office, One Cyclotron Road, MS 56A-0120, Berkeley, CA 94720-8127 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority

(54) Title: A NOVEL CELLULASE COMPLEX, AND GLYCOSIDASE HYDROLASES THEREOF, AND METHODS OF USING THEREOF

(57) Abstract:



WO 2017/087982 A2

A novel cellulase complex, and glycosidase hydrolases thereof, and methods of using thereof

Inventors: Steven W. Singer, Sebastian D. Kolinko, Evelyn Denzel, Jennifer Hiras, John M. Gladden

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The application claims priority to U.S. Provisional Patent Application Ser. No. 62/257,477, filed November 19, 2015; which is incorporated herein by reference.

STATEMENT OF GOVERNMENTAL SUPPORT

[0002] The invention was made with government support under Contract Nos. DE-AC02-05CH11231 awarded by the U.S. Department of Energy. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention is in the field of saccharification of biomass using ionic liquid.

BACKGROUND OF THE INVENTION

[0004] JTherm is a mixture of enzymes useful as a thermophilic-ionic liquid (IL) tolerant cellulase cocktail. This cocktail is useful in biofuel production based on its application in the production of biodiesel (fatty acid ethyl esters) from switchgrass and its application in pretreatment/saccharification for switchgrass.

SUMMARY OF THE INVENTION

[0005] The present invention provides for a composition comprising a purified or isolated cellulase complex comprising two or more glycosidase hydrolase, or enzymatically active fragment thereof, selected from the group consisting of a GH9 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:1, a GH48 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:2, a GH10 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:3, and a GH6 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:4; and

optionally the composition or the cellulase complex comprising a GH10_2 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:5, and/or an AA10 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:6.

[0006] In some embodiments, the purified or isolated cellulase complex comprises the following glycosidase hydrolases, or enzymatically active fragment thereof: a GH9 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:1, a GH48 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:2, a GH10 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:3, and a GH6 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:4.

[0007] In some embodiments, the purified or isolated cellulase complex comprises the following glycosidase hydrolases, or enzymatically active fragment thereof: a GH9 polypeptide comprising an amino acid sequence at least 90% identical to SEQ ID NO:1, a GH48 polypeptide comprising an amino acid sequence at least 90% identical to SEQ ID NO:2, a GH10 polypeptide comprising an amino acid sequence at least 90% identical to SEQ ID NO:3, and a GH6 polypeptide comprising an amino acid sequence at least 90% identical to SEQ ID NO:4.

[0008] In one embodiment, the purified or isolated cellulase complex comprises a GH9 polypeptide comprising the amino acid sequence of SEQ ID NO:1, a GH48 polypeptide comprising the amino acid sequence of SEQ ID NO:2, a GH10 polypeptide comprising the amino acid sequence of SEQ ID NO:3, and a GH6 polypeptide comprising the amino acid sequence of SEQ ID NO:4.

[0009] In some embodiments, the GH9 polypeptide comprising an amino acid sequence at least 75%, 80%, 85%, 90%, 95%, or 99% identical to SEQ ID NO:1. In some embodiments, the GH9 polypeptide comprises the amino acid sequence of SEQ ID NO:1. In some embodiments, the GH9 polypeptide comprises one or more of the conserved amino residues depicted in FIG. 5. In some embodiments, the GH9 polypeptide comprises one or more of the identical amino residues depicted in FIG. 5, which are in stretches of three or more, or four or more, consecutive identical amino acid residues. In some embodiments, the GH9 polypeptide comprises residues 58 to 452 of SEQ ID NO:1, or residues 43 to 430 of SEQ ID NO:8.

[0010] In some embodiments, the GH48 polypeptide comprising an amino acid sequence at

least 75%, 80%, 85%, 90%, 95%, or 99% identical to SEQ ID NO:2. In some embodiments, the GH48 polypeptide comprises the amino acid sequence of SEQ ID NO:2. In some embodiments, the GH48 polypeptide comprises one or more of the identical amino residues depicted in FIG. 6. In some embodiments, the GH48 polypeptide comprises one or more of the identical amino residues depicted in FIG. 6, which are in stretches of three or more, or four or more, consecutive identical amino acid residues. In some embodiments, the GH48 polypeptide comprises residues 48 to 740 of SEQ ID NO:2, or residues 1128 to 1751 of SEQ ID NO:9.

[0011] In some embodiments, the GH10 polypeptide comprising an amino acid sequence at least 75%, 80%, 85%, 90%, 95%, or 99% identical to SEQ ID NO:3. In some embodiments, the GH10 polypeptide comprises the amino acid sequence of SEQ ID NO:3. In some embodiments, the GH10 polypeptide comprises one or more of the identical amino acid residues when comparing all of the amino acid residues of SEQ ID NO:3 and amino acid residues SEQ ID NO:12 (but excluding amino acid residues 1 to 12). In some embodiments, the GH10 polypeptide comprises SEQ ID NO:3, or SEQ ID NO:12 (but excluding amino acid residues 1 to 12).

[0012] In some embodiments, the GH10_2 polypeptide comprising an amino acid sequence at least 75%, 80%, 85%, 90%, 95%, or 99% identical to SEQ ID NO:5. In some embodiments, the GH10_2 polypeptide comprises the amino acid sequence of SEQ ID NO:5. In some embodiments, the GH10_2 polypeptide comprises one or more of the identical amino residues depicted in FIG. 7. In some embodiments, the GH10_2 polypeptide comprises one or more of the identical amino residues depicted in FIG. 7, which are in stretches of three or more, or four or more, consecutive identical amino acid residues. In some embodiments, the GH10_2 polypeptide comprises residues 48 to 740 of SEQ ID NO:5, or residues 1128 to 1751 of SEQ ID NO:10.

[0013] In some embodiments, the GH6 polypeptide comprising an amino acid sequence at least 75%, 80%, 85%, 90%, 95%, or 99% identical to SEQ ID NO:4. In some embodiments, the GH6 polypeptide comprises the amino acid sequence of SEQ ID NO:4.

[0014] In some embodiments, the AA10 polypeptide comprising an amino acid sequence at least 75%, 80%, 85%, 90%, 95%, or 99% identical to SEQ ID NO:6. In some embodiments, the AA10 polypeptide comprises the amino acid sequence of SEQ ID NO:6. In some

embodiments, the AA10 polypeptide comprises one or more of the identical amino residues depicted in FIG. 8. In some embodiments, the AA10 polypeptide comprises one or more of the identical amino residues depicted in FIG. 8, which are in stretches of ten or more consecutive identical amino acid residues. In some embodiments, the AA10 polypeptide comprises residues 1 to 241 of SEQ ID NO:6, or residues 1 to 241 of SEQ ID NO:11.

[0015] The present invention provides a composition comprising the purified or isolated cellulase complex of the present invention and an ionic liquid (IL).

[0016] The present invention provides for a recombinant or isolated or purified nucleic acid encoding any of the GH polypeptides of a cellulase complex of the present invention. In some embodiments, the recombinant or isolated or purified nucleic acid encodes the GH9 polypeptide, the GH48 polypeptide, the GH10 polypeptide, and the GH6 polypeptide of the cellulase complex of the present invention. In some embodiments, the recombinant or isolated or purified nucleic acid further encoding one or more promoters operatively linked to the nucleotide sequences encoding GH9 polypeptide, the GH48 polypeptide, the GH10 polypeptide, and/or the GH6 polypeptide.

[0017] The present invention provides for a vector comprising the recombinant or isolated or purified nucleic acid of the present invention. In some embodiments, the vector is an expression vector. In some embodiments, the vector is a plasmid.

[0018] The present invention provides for a host cell comprising a nucleic acid of the present invention, wherein the host cell is capable of expressing the GH polypeptides of the present invention. In some embodiments, the nucleic acid is stably integrated into a chromosome of the host cell, or the nucleic acid is capable of stable residence in the host cell.

[0019] The present invention provides for a method for producing a cellulase complex of the present invention comprising: providing a host cell capable of expressing the cellulase complex, culturing the host cell in a culture medium under conditions whereby the cellulase complex is produced, optionally isolating the cellulase complex from the host cell and/or the culture medium, and optionally contacting the cellulase complex and a cellulose, whereby the cellulose is hydrolyzed by the cellulase complex.

[0020] In some embodiments, the providing step comprises: introducing an expression vector capable of expressing the cellulase complex in the host cell into the host cell, and optionally

constructing the expression vector encoding a promoter operatively linked to a nucleic acid encoding the cellulase complex, wherein the constructing step precedes the introducing step.

[0021] In some embodiments, the composition further comprises an ionic liquid (IL). In some embodiments, the composition further comprises a cellulose, wherein the cellulase complex is capable of hydrolyzing the cellulose. In some embodiments, the composition further comprises an IL and a cellulose, wherein the cellulase complex is capable of hydrolyzing the cellulose. In some embodiments of the invention, the composition comprises a pretreatment biomass.

[0022] The present invention provides for a method of hydrolyzing a cellulose, comprising: (a) providing a solution comprising an IL, a cellulose, and a composition of the present invention to the solution, and (b) incubating the solution, such that the cellulose is hydrolyzed by the cellulase complex. In some embodiments of the invention, the solution comprises a pretreatment biomass.

[0023] The present invention provides for a method for converting of the carbohydrates of lignocellulose to sugars with improvements in yield and rate of sugar production has been developed by using the cellulase complex of the invention. In some embodiments of the invention, the cellulase complex is compatible with ionic liquid (IL). In some embodiments, the cellulase complex is introduced to a pretreatment biomass comprising a pretreatment cellulose biomass, pretreatment hemicellulose biomass, pretreatment lingo-cellulose biomass, or a mixture thereof.

[0024] The present invention provides for a method for converting a lignocellulosic biomass to sugars for the production of biofuels using the cellulase complex.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] The foregoing aspects and others will be readily appreciated by the skilled artisan from the following description of illustrative embodiments when read in conjunction with the accompanying drawings.

[0026] FIG. 1 shows the measurement of endoglucanase and endoxylanase activity of fractions from anion-exchange chromatography of 300 L supernatant

[0027] FIG. 2 shows Coomassie and activity staining (zymography) of endoglucanase and

xylanase activity of anion-exchange and affinity digestion fractions of the 300 L supernatant.

[0028] FIG. 3 shows a pie chart depicting relative abundances of metagenome for 300 L culture. Recovered genome in box in figure legend containing GH9, GH48 and GH6/5.

[0029] FIG. 4 shows a reconstructed putative operon containing GH9, GH48 and GH6/5 genes.

[0030] FIG. 5 shows a comparison of the amino acid sequence of the GH9 domain (residues 58 to 418) of SEQ ID NO:1, and the GH9 domain of *Paenibacillus sp. A59* (WP_053783505) (SEQ ID NO:8).

[0031] FIG. 6 shows a comparison of the amino acid sequence of GH48 (SEQ ID NO:2) and the amino acid sequence of *Paenibacillus borealis* GH48 (SEQ ID NO:9).

[0032] FIG. 7 shows a comparison of the amino acid sequence of GH10_2 domain (SEQ ID NO:5) and the amino acid sequence of *Cohnella laeviribosi* GH10 (SEQ ID NO:10).

[0033] FIG. 8 shows a comparison of the amino acid sequence of the AA10 domain (residues 1 to 241) of SEQ ID NO:6, and the AA10 domain of *Caldibacillus cellulovorans* beta-1,4-mannanase precursor (SEQ ID NO:11).

DETAILED DESCRIPTION OF THE INVENTION

[0034] Before the invention is described in detail, it is to be understood that, unless otherwise indicated, this invention is not limited to particular sequences, expression vectors, enzymes, host microorganisms, or processes, as such may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting.

[0035] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limits of that range is also specifically disclosed. Each smaller range between any stated value or intervening value in a stated range and any other stated or intervening value in that stated range is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included or excluded in the range, and each range where either, neither or both limits are included in the smaller ranges is also

encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0036] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0037] As used in the specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to an "IL" includes a single IL compound as well as a plurality of IL compounds, either the same (e.g., the same molecule) or different.

[0038] In this specification and in the claims that follow, reference will be made to a number of terms that shall be defined to have the following meanings:

[0039] The terms "optional" or "optionally" as used herein mean that the subsequently described feature or structure may or may not be present, or that the subsequently described event or circumstance may or may not occur, and that the description includes instances where a particular feature or structure is present and instances where the feature or structure is absent, or instances where the event or circumstance occurs and instances where it does not.

[0040] The terms "expression vector" or "vector" refer to a compound and/or composition that transduces, transforms, or infects a host microorganism, thereby causing the cell to express nucleic acids and/or proteins other than those native to the cell, or in a manner not native to the cell. An "expression vector" contains a sequence of nucleic acids (ordinarily RNA or DNA) to be expressed by the host microorganism. Optionally, the expression vector also comprises materials to aid in achieving entry of the nucleic acid into the host microorganism, such as a virus, liposome, protein coating, or the like. The expression vectors contemplated for use in the present invention include those into which a nucleic acid sequence can be inserted, along with any preferred or required operational elements. Further,

the expression vector must be one that can be transferred into a host microorganism and replicated therein. Particular expression vectors are plasmids, particularly those with restriction sites that have been well documented and that contain the operational elements preferred or required for transcription of the nucleic acid sequence. Such plasmids, as well as other expression vectors, are well known to those of ordinary skill in the art.

[0041] The term "host cell" refers to a living biological cell that can be transformed via insertion of an expression vector. Thus, a host cell as described herein may be a prokaryotic organism (e.g., an organism of the kingdom Eubacteria) or a eukaryotic cell.

[0042] The term "isolated" refers to material that is substantially or essentially free of components that normally accompany it in its native state.

[0043] As used herein, the terms "nucleic acid", "nucleotide" and variations thereof shall be generic to polydeoxyribonucleotides (containing 2-deoxy-D-ribose), to polyribonucleotides (containing D-ribose), to any other type of polynucleotide that is an N-glycoside of a purine or pyrimidine base, and to other polymers containing nonnucleotidic backbones, provided that the polymers contain nucleobases in a configuration that allows for base pairing and base stacking, as found in DNA and RNA.

[0044] The terms "optional" or "optionally" as used herein mean that the subsequently described feature or structure may or may not be present, or that the subsequently described event or circumstance may or may not occur, and that the description includes instances where a particular feature or structure is present and instances where the feature or structure is absent, or instances where the event or circumstance occurs and instances where it does not.

[0045] The terms "expression vector" or "vector" refer to a compound and/or composition that transduces, transforms, or infects a host microorganism, thereby causing the cell to express nucleic acids and/or proteins other than those native to the cell, or in a manner not native to the cell. An "expression vector" contains a sequence of nucleic acids (ordinarily RNA or DNA) to be expressed by the host microorganism. Optionally, the expression vector also comprises materials to aid in achieving entry of the nucleic acid into the host microorganism, such as a virus, liposome, protein coating, or the like. The expression vectors contemplated for use in the present invention include those into which a nucleic acid sequence can be inserted, along with any preferred or required operational elements. Further,

the expression vector must be one that can be transferred into a host microorganism and replicated therein. Preferred expression vectors are plasmids, particularly those with restriction sites that have been well documented and that contain the operational elements preferred or required for transcription of the nucleic acid sequence. Such plasmids, as well as other expression vectors, are well known to those of ordinary skill in the art.

[0046] The term "transduce" as used herein refers to the transfer of a sequence of nucleic acids into a host cell. Only when the sequence of nucleic acids becomes stably replicated by the host cell does the host cell become "transformed." As will be appreciated by those of ordinary skill in the art, "transformation" may take place either by incorporation of the sequence of nucleic acids into the cellular genome, i.e., chromosomal integration, or by extrachromosomal integration. In contrast, an expression vector, e.g., a virus, is "infective" when it transduces a host microorganism, replicates, and (without the benefit of any complementary virus or vector) spreads progeny expression vectors, e.g., viruses, of the same type as the original transducing expression vector to other host cells, wherein the progeny expression vectors possess the same ability to reproduce.

[0047] Standard techniques are used for recombinant DNA, oligonucleotide synthesis, and tissue culture and transformation (e.g., electroporation, lipofection). Enzymatic reactions and purification techniques are performed according to manufacturer's specifications or as commonly accomplished in the art or as described herein. The foregoing techniques and procedures are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification. See e.g., Sambrook et al. *Molecular Cloning: A Laboratory Manual* (3rd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (2001)), which is incorporated herein by reference. The nomenclatures utilized in connection with, and the laboratory procedures and techniques described herein are those well known and commonly used in the art.

[0048] The term "isolated nucleic acid" as used herein shall mean a polynucleotide that has been isolated from its naturally occurring environment. Such polynucleotides may be genomic, cDNA, or synthetic. Isolated polynucleotides preferably are not associated with all or a portion of the polynucleotides they associate with in nature. The isolated polynucleotides may be operably linked to another polynucleotide that it is not linked to in nature.

[0049] The term "isolated protein" referred to herein means a protein that has been isolated from its naturally occurring environment. Such proteins may be derived from genomic DNA, cDNA, recombinant DNA, recombinant RNA, or synthetic origin or some combination thereof, which by virtue of its origin, or source of derivation, the "isolated protein" (1) is not associated with proteins found in nature, (2) is free of other proteins from the same source, (3) is expressed by a cell from a different species, or (4) does not occur in nature.

[0050] The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory or otherwise is naturally-occurring.

[0051] The term "operably linked" as used herein refers to positions of components so described that are in a relationship permitting them to function in their intended manner. For example, a control sequence "operably linked" to a coding sequence is connected in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

[0052] These and other objects, advantages, and features of the invention will become apparent to those persons skilled in the art upon reading the details of the invention as more fully described below.

[0053] The development of the JTherm cellulase enzymatic mixture, which is a cocktail containing supernatant from a thermophilic bacterial consortium maintained with microcrystalline cellulose as its sole carbon substrate as the identified endoglucanase component and cellobiohydrolase and beta-glucosidase components form heterologous proteins expressed in *E. coli* (PLoS One, 7:e37010, 2012; Green Chemistry, 15:2579-2589, 2013). The active component of JTherm was identified as an "endoglucanase". Biochemical techniques are used to enrich the active fraction of JTherm and demonstrate that it contained a protein complex with multiple hydrolases (GH9, GH48, GH6, and GH10) and that this protein complex is capable of hydrolyzing IL pretreated biomass to glucose by addition of a beta-glucosidase, without supplement of the cellobiohydrolase. Metagenomic data show the glycoside hydrolases in the complex map to a low abundance Firmicutes present in the

microbial community from which JTherm was derived. BLAST search of the sequence for xylanase (GH10) indicate that it is 98% identical at the amino acid level to xylanase isolated from “*Caldibacillus cellulovorans*” an unvalidated bacterial isolate (Microbiology, 146:2947-2955, 2000). No characterization of a glycoside hydrolase complex has been reported for “*Caldibacillus cellulovorans*”.

[0054] Four full-length genes are identified using a PCR-based approach in a putative operon that constitutes the complex (FIG. 1). The amino acid sequences of the polypeptides encoded by the genes and the nucleotide sequence of the operon are described herein.

[0055] The amino acid sequence of the GH9 polypeptide of the cellulase complex is as follows:

```
MLRRRALSMLTGAAVVYSAFVPVGSPDSAVVRAAPTSYNYAEALQKAI
YFYDAQRSGLPPDNRVEWRGDSGLNDGADVGVDLTGGWYDAGDHVKFGLPMAYSAAMLA
WAVEYERDAFVQTGQLDYILNNIKWATDYFIKAHSAPNVLWGQVGKGDVDHAWWGPAEVM
QMPRPAYKIDPSCPGSDLAAGTAAAMAAAAVFKPTDPTYASTLIAHAKQLYTFADTYRG
KYSDCITDAQNFYRSWSGYADELTWGA VWLYLATGEQAYLDKAIASVAEWGREGQTPYWG
YKWTQSWDDVHYGAQ LLLARITGDQRFIQSTERNLEYWTDGTDGTGERITYTPGGLAWLD
SWGSLRYAMNASFLAFVYSDWLQSRDPAKAEKYRNFVRQVLYALGDNPRNSSYVVGFR
NPPQRPHHRTAHGSWADSQNPAYHRHILY GALVGGPNQSDAYTDSISDYVGVNEVATDYN
AAFTGNLAKMYLLFGASAGQRPLANFPEPEVREDEFFVEAGVNSSGPNYTEIKALINNRS
GWPARMGDKLSFKYFVDLSEVYAAGYTVNDIKVT TNYNEGAKVSGLLPYDESRRLYYVLV
DFTGTKIYPGGQSAYKKEVQFRLSAPSGTSF WNPNNDFS YQLMSGTSNSSLVKTPYMPVY
DAGVKIFGVEPSSSGSGSSPTPPPTSTPTPTPTPTPTPTLTLTPTPTPTPTPTPTPTPTPT
PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT
NTGSQAVPLTELKVRYWYTKNSTQAEQYFC DWAQIDCSNIRAQFVSLAQPVSGADSYIEL
SFTGGSVPAGGNTGEIQNRIHFTNWMNYNETDDWSYNGTQTTWGPSTRITLYRNGVLVWG
TEPGGGSSTPTPTATPTPTPTSAAPTPTPTAGGSLVVQYRAADTNAGDNQLKPHFRIVNRG
TTSVPLSELSIRYWYTVDGDKPQVFNC DWAQVGC SNLRGSFVKLSTGRTGADYYIEITFT
SGAGSLAAGGSSGDIQVRINKNDWTNYNEANDYSYDPTKTSFADWNRVTLYRNGQLIWGV
EP* (SEQ ID NO:1)
```

[0056] The amino acid sequence of the GH48 polypeptide of the cellulase complex is as follows:

```
MESLAWTLLWKKARIIFLAFALVVSFAFAGFAVSPRSETAYAQTDPQVFKDRFLQLYNQIK
NPANGYFSPEGIPYHSIETLISEAPDYGHMTTSEAFSYWLWLETLYGYFTGDWSKLEQAW
TKMEQFIIPNSTEQPTMGSYNPSSPATYAPEHPYPDRYPTLLNNSVPAGQDPLDAELKAT
YGNNVTYLMHWLLDVNDWYGFGNLLNPSHTATYVNTFQRGEQESVWEAITHPSQDNFRFG
KPNEGFTVTLFVKDNGTPAQQWRYTAASDADARAIQVMYWAKQLGYNNQTYLDKARKMGDY
LRYTLFDKYFQQIGSANDGSPSPGSGKNSAHYLLSWYTAWGGGLGSGGNWAWRIGSSHAH
```

QGYQNPFVAAAYALSAGGLAPRSATAQTDWATSLQRQLEFYTWLQSSEGAIGGGATNSVGGSYQPYPSGRSTFYGMVYDEAPVYRDPSPNSWFGFQAWSVERVAELYIWTSSGNTNTQQFQMVKNIVTKWVDWALDYTFVNQRPVTDAAQGYFLTSSGSRVLGGNNPQIATVSDPGQFYIPSTLEWQGGQPD TWNGYANYTGNPNFHAIKADPGQDVGVTGNYIKLLTFFAAATKAETGNYTALGSQALNVAEQLLNLWNFNDGVGIVRPEQRADYFRYFTKEIYFPSGWSGTYGQGN TIPGPGAVPSDPSKGGNGVYISYAE LRPKIKQDPKWSYLENLYKTSYNPSTGRWENGVP TFTYH RFWAQVDVATAYAEFARLIGGLGASPTPTPSATPTPTPSAGGNLVVQYRAADTNATDNQLKPHFRIVNRGTTSVPLSELTIRYWYTVDGDKPQVFNC DWAQVGCSNVRGSFVKLT TGRTGADYYIEITFTSGAGSLAAGSSSGDIQVRINKNDWTNYNEANDYSCDPTKTSFADWNRVTL YRNGQLVWGVEP* (SEQ ID NO:2)

[0057] The amino acid sequence of the GH10 polypeptide of the cellulase complex is as follows:

MLLVLAIAIGLLLPPIPYLHVASAENVLILQSDVFEDGTTQGWVGRGGVETLTVTSAAAYS
GLSVSGRKTWHGPTLDITSYIQVGKTYQFSAWVKLPSGSSNTRIYMTMQRTMQD TVYYE
QIYFDTASAGNWWQLKAQYKLYEPAVNLQVYFEAPDHATQSFYIDDVRIEQLPDLPKTVE
ENIPSLKDVFAGRFPIGTAFENFELLDEQDRKLILKHFNSVTPGNVLKWDSTEPQEGVF
FTESDKAVAFVQNGMKIRGHTLIWHNQTPNWVIFYDSNGNLVSKEVLYQRMERHIKPVVS
RYKGIIYAWDVVNEVIDPGQPDGLRRSLWYQIAGEEYIEKAFQFAHEADPNALLFINDYN
THESGKSQALYNLVQRLKNKGIPVHGVBHQTHINISWPSISEIENSLVKFSNLGVVQEIT
ELDMSIYNNSSQKYDTLP SDLAQQQATRYRQLFEMFLRRSSLIQNVTFWGKDDANTWLRK
FPVVRNDWPLLFDEQLKAKPAYWAVVGTVPSPPTPTPTSTATPTPTPTPTPTPTPTPT
TPTPTPTSTPTPSASGTLRVEYRVGDSSATDNQMKPQLRIVNTGSQAVPLTELKVRYWYT
KNSTQAEQYFCDWAQIGCSNIRAQFVSLAQPVSGADSYIELSFTGGSVPAGGNTGEIQNR
IHFTNWMNYNETDDWSYNGTQTTWGPSTRITLYRNGVLVWGTEPGGGSPPTPTPTPTSTPT
PTSTPTPTPTPTPTPTATPTPTPTPTPSAGGNLVVQYRAADTNAGDNQLKPHFRIVNRGT
TSVPLSELTIRYWYTVDGDKPQVFNCDAWVGCSNLRGSLVKLTGTGRTGADYYLEITFTS
GAGSLAPGANSQDIQARINKNDWTNYNEANDYSYDPTKTSFADWNRVTLYRNGQLVWGVE
P* (SEQ ID NO:3)

[0058] The amino acid sequence of the GH6 polypeptide of the cellulase complex is as follows:

MTMAWKQRSGLIALILALVAGLLLPWGSLPKAAAEPHVDNPFVGGATAYVNPDYAALVDSS
IARVSDPTLAAKMRTVKTYPTAVWLDRIAIDGGPGRSLVQHLDTALAQQGNTPTITAM
FVIYNMPGRDCAALASNGELPLTQEGQLQRYKTEYIDRIAIFADPKYAGIRIVTVIEPDG
LPNLVTNLSDECAQANSSGIYVEAVRYAINKLSEIPNVYIYLDIAHSGWLGDNNRTGA
VQLYTNVVRGTTKGLSSVDGFVTNVANYTPLEEPYLTDPNLTVGGQPLKSAKFYEWNPYF
DEVYAAAALRSAFISAGWPTSIGMLIDTSRNGWGGPNRPTGASGTTVDAYVNSGRVDRRA
HRGLWCNVSGAGMGMPQVAPAAAYASQGIEAFVWVKPPGESDGASSEIPNDEGKRFRDMC
DPTYTTQYGVLTGALPNAPLAGQWFHDQFVMLVQNAYPAIPTSGGGTPTPSTTVTPTPTP
TPTPTPSATVTPTPTPTPTPTPSATVTPTPTPTPTPTPTVTPTPTSTTSFVARHGQLRVV
GNOLVDONGOPIOLRGISSHGLOWYGHFVNRRSLRWLRDDWGIVFRAALYTAEQGYITN

PSLKEKVKEAVQAAIELGIYVIDWHILSDGDPNTYKEQAKAFFDEMSRLYGSYPNVIYE
IANEPNGVTWEGQVKPYASEVIPVIRANDPDNLIIVGTTTWSQDVHLAADSPLPYSNLAI
ALHFYAGTHGQWLRDRIDYARNKGIAIFVSEWGTSTSTGDGGPYLTESQQWLDLFLNARQI
SWVNWSLSDKAESSAALLPGASATGGWTD AQLSQSGRFVRAQIRSGVLTPPTPTPTPTP
SAAPTPTPTAGGSLVVQYRAADTNAGDNQLKPHFRIVNRGTTSVPLSELSIRYWYTVDGD
KPQVFNCDDWAQVGC SNLRGSFVKLSTGR TGADYYIEITFTSGAGSLAAGGSSGDIQVRIN
KNDWTNYNEANDYSYDPTKTSFADWNRVTLYRNGQLIWGVEP* (SEQ ID NO:4)

[0059] The amino acid sequence of the GH10_2 polypeptide of the cellulase complex is as follows:

MHQIAERRKMIMRNWLQWGMVAALLVFTVSVPVPKEADAGLAKTKFLGNVINNSIPSDFA
 VYWNQVTPENATKWGSVESSRDNMNWSTADMIYNYARSNGFPFKFHTLVWGSQEPGWISG
 LSAAEQQAEEVIEWIQAAGQRYPDADFVDVVNEPLHAKPSYRNAIGGDGSTGWDWVIWSFE
 QARRAFPNSKLLINEYGVENDPNAAQYVQIINLLKSRGLIDGIGIQGHYFNLDTVSVST
 LRTTLGMLAETGLPIYVSELDISGDDATQLARYQEKFPIIWEHPSVQGITLWGYIEGQTW
 RSGTHLITASGVERPALQWLRTYLAGAGSSPTPTPTPTPTPTPTPTPTPTPSANGTLRVE
 YRVGDSSATDNQMKPQLRIVNTGSQAVPLTELKVRYWYTKNSTQAEQYFCDWAQIGCSNI
 RAQFVSLSQPVSGADSYIELSFTGGSIPAGGNTGEIQNRIHFTNWMNYNETDDWSYNGAQ
 TTWGPSTRITLYRNGVLVWGTEPGGSSTPTPTPTPTPTPTPTPSAAPTAPASAGGSLVVQYR
 AADTNATDNQLKPHFRIVNRGTTSVPLSELTIRYWYTVDGDKPQVFNCDDWAQVGCNSVRG
 SFVKLSTGRGTGADYYIEITFTSGAGSLAPGANSGLDIQARINKNDWTNYNEANDYSYDPTK
 TSFADWNRVTLYRNGQLIWGVEP* (SEQ ID NO:5)

[0060] The amino acid sequence of the AA10 polypeptide of the cellulase complex is as follows:

MNRRLLIARLSGMLAMVLIAAVLAYVPKPEPAEAHGGMVFPATRTRYACYVDGKVHNGGDL
 NMINPACLDALAI SGNYQFWNWFGNLISNAGGRHREIIPDGKLCGPTASFDGMNQARTDW
 WTTRLQPGATITVRVNAWAPHPGTWYLYVTRDGDWPTQPLKWSDELEPTPFSQVTNPPINS
 SGPDGAEYSWQVQLPNKQGRHIIYMIWQRSDSPEAFYNCSDAYFGSGPIAYEFGDPREGG
 TMITPPPSGTTPTPTPTPTPTPTPTTLTPTPTPTPTPTPTTLTPTPTPTPTPTSTPTSTP
 TSTPTPTPTPSASGTLRVEYRVGDSSATDNQMKPQLRIVNTGSQAVPLTELKMRYWYTKN
 STQAEQYFCDWAQIGCSNIRAQFVLSQPVSGADSYIELSFTGGSIPAGGNTGEIQNRIH
 FTNWMNynetDDWSYNGAQMTWGPSTRITLYRNGVLVWGTEPGGGSSPPTPTVTPPTPTPT
 STPTPTPTPSAAPTPTPSAGGSLVVQYRAADTNAGDNQLKPHFRIVNRGTTSVPLSELSI
 RYWYTVDGDKQPQVFNCDAQVGCNLRGSFVKLSTGRGTGADYYIEITFTSGAGSLAPGAS
 SGDIQVRINKNDWTNYNEANDYSYDPTKTSFADWNRVTLYRNGQLVWGVEP* (SEQ ID
 NO:6)

[0061] A nucleotide sequence encoding an operon encoding GH9, GH48, GH10,GH6, AA10, and GH10 2 is as follows:

TTGAAGGAGGAACAGACGGTGAATTTCCGACAACTGATGCTGAGACGACGGGCGCTGTCTG
ATGTTGACGGGCGCGGCGGTTCGTTTATTTCGGCTTTCGTGCCGGTTGGAAGTCCGGATTCTG

GCTGTTGTTTCGGGCGGCTCCTACCTCTTACAATTACGCGGAAGCGCTTCAAAAGGCAATT
TATTTCTACGACGCACAACGTTCCGGCAAGCTGCCTCCCGATAACCGCGTCGAATGGCGC
GGCGATTCCGGGCTTAAACGACGGAGCCGACGTCGGCGTCGATTTGACGGGCGGTTGGTAC
GACGCCGGCGACCACGTCAAGTTCGGGTTGCCGATGGCATATTCCGCCGCCATGCTGGCG
TGGGCGGTGTACGAGTACCGCGACGCTTTCTGTCAGACGGGGCAGCTCGATTATATTTTG
AACAACATCAAGTGGGCGACCGACTATTTTCATCAAGGCGCATTCCGCGCCCAACGTGCTT
TGGGGACAGGTTCGGCAAGGGGGACGTCGATCATGCCTGGTGGGGACCGGCAGAAGTGATG
CAGATGCCCGCTCCGGCTACAAAATCGACCCGAGCTGTCCGGGGTCCGATCTCGCTGCC
GGAACGGCCGCCCGCATGGCCGCCGCCGCCGCGGTGTTTAAACCCACTGACCCGACTTAT
GCCTCAACGTTGATCGCTCATGCGAAACAATTGTATACGTTTGCAGATACTTATCGGGG
AAATATTCCGATTGCATCACTGATGCGCAAAATTTCTATCGGTCTGAGCGGTTACGCC
GATGAGCTGACGTGGGTGCCGTCTGGCTTTATCTGCCACCGGCGAGCAGGCCTATCTT
GACAAGGCAATCGCCTCAGTCGCGGAATGGGGGCGCAAGGTCAGACGCCTTATTGGGGT
TACAAATGGACGCAAAGCTGGGACGACGTCCATTACGGCGCTCAGCTGCTGTTGGCAAGA
ATTACGGGCGACCGCGGTTTATCCAGTCGACGGAGCGCAACCTGGAATATTGGACGGAC
GGCAGGACGACACCGGCGAGCGCATCACGTATACGCCGGGGGGCTTGCTTGGCTAGAT
TCTTGGGGTTCACTCCGCTATGCGATGAACGCGTCGTTCTTGGCGTTCGTCTATTCCGAC
TGGCTGCAAAGCCCGCATCCCGCCAAAGCGGAAAAGTACAGGAACCTCGCCGTTCCGAC
GTTCTGTATGCATTGGGCGACAACCCGCGCAATTCTAGCTATGTCGTCGGATTCCGGCGC
AATCCGCCGACGCGCCGCATCACCGGACGGCGCACGGGTCGTGGGCTGACAGCCAGAAC
GTTCCCGCTTACCATCGGCACATTTTATATGGCGCTCTGTGGGCGGTCCGAACAGTCC
GACGCCTATACGATTTCGATCAGCGACTACGTCGGCAATGAGGTTGCGACGGATTACAAC
GCGGCGTTTACGGGAAATCTCGCAAAATGTATCTGCTGTTCCGGCGCCAGCGCCGGACAG
CGGCCGCTTGCCAAATTTCCCGAGCCGGAAGTACGCGAGGACGAGTTTTTCGTCGAAGCC
GGCGTGAACAGCTCCGGACCGAACTACACAGAGATCAAGGCCCTGATCAACAACCGGTCC
GGCTGGCCCGCGCGGATGGGCGACAAGCTTTCGTTCAAGTATTTTCGTCGATTTGTCTGAA
GTTTACGCCCGCGGTTATACGGTCAACGACATTAAGGTGACGACGAACATAACGAAGGC
GCGAAAGTATCCGGTCTGCTTCCGTATGACGAAAGCCGTCGTCTTTACTATGTGCTCGTC
GATTTTACGGGTACGAAGATTTATCCCGCGGTCTAGTCCGCTACAAGAAAGAGGTTTCAG
TTCAGACTGAGCGCTCCGAGCGGGACATCGTTCTGGAATCCGAACAACGATTTCTCGTAC
CAGCTGATGTCCGGCAGTCCAAACAGCAGCCTGGTCAAGACGCCGTATATGCCGGTTTAT
GACGCCGGGTGAAGATTTTCCGGCTGGAGCCGTCGTCCGGAAGCGGGTCGAGCCCGACG
CCGCCACCCACGTCGACGCCGACACCGACACCGACGCCGACACCGACACCGACGCTGACA
CTGACACCGACACCGACGCCGACACCGACGCCGACACCGACGCCGACACCGACGCCGACA
CCGACACCGACGCCGACACCAACGCCGACACCAACGCCGACACCGACGCCGACGCCGACA
CCAACGCCGACGCCGACACCGACACCGACGCCCTAGCGCGAGCGGTACCCTGCGCGTCGAG
TATCGCGTCGGTGACACCGCGCCACCGACAACCAGATGAAGCCGTACCTGCGCATCGTC
AACACCGGCTCGCAAGCCGTGCCGTGACCGAACTGAAGGTGCGCTACTGGTACACGAAG
AACTCGACGCAGGCCGAACAGTACTTCTGCGACTGGGCGCAGATCGACTGCTCGAACATC
CGGGCGCAGTTCTGTGCTGCGTGGCGCAGCCGGTCAGTGGAGCGGACAGCTACATCGAGCTG
AGCTTCACGGGCGGAAGCGTTCCGGCGGGAGGCAACACGGGCGAGATACAGAACCGGATT
CACTTCACGAACCTGGATGAACTACAACGAAACGGACGACTGGTCGTACAACGGGACGCAG
ACGACGTGGGGTCCGTGACGCGGATTACGCTGTATCGTAACGGCGTGCTGGTGTGGGGC
ACCGAGCCGGGCGGCGGATCGTCGACGCCGACACCGACGGCGACACCTACGCCTACGCCG
AGCGCGGCGCCACACCGACGCCGACGGCCGGCGGACGCTGGTTCGTGCAGTATCGCGCG
GCGGACACGAACGCGGGCGACAACACAGCTGAAGCCGCACTTTAGGATTGTGAACCGCGGG
ACGACGAGCGTGCCGCTGTGCGAGCTTTTCGATCCGGTACTGGTACACGGTGGACGGGGAC
AAGCCGCGAGGTGTTCAACTGCGACTGGGCGCAGGTGGGTGTTTCGAACCTGCGGGGCGAGT
TTCGTGAAGCTTTTCGACGGGCCGGACGGGGCGGACTACTACATCGAGATCACGTTTACA
TCGGGCGCGGGCAGCTTGGCGGCTGGGGGAAGCAGTGGGGACATTTCAGGTGCGGATCAAC
AAGAACGACTGGACGAACATAACGAGGCGAACGATTACTCGTATGATCCGACGAAGACG
AGTTTTGCGGATTGGAACCGGGTGACGCTGTATCGCAACGGTCAGCTCATCTGGGGCGTC
GAACCATAGCAAAATTCGGTGGTTTATTCTTTTGAACAGCAAACCCAGGGAACGGAGGA

ATGAGGGTGTGATCGAATTGGCAATCCTATTGCCGACCTTCGTTCCCTGGTTTGCAAAAT
AGAAACGCTACAAAAACAGGAGGAAGAGGAAAAATCATGGAATCACTCGCATGGACGCT
GTTATGGAAGAAAGCAAGAATTATCTTTCTTGCTTTCGCGCTTGTCGTCTCCGCCCTTCGC
GGGCTTCGCTGTGTCTCCTCGTAGCGAAACCGCTTACGCGCAGACGGACCCGAGGTTTT
CAAGGACAGGTTTTTGAGCTGTACAACCAAATCAAAAATCCGGCGAACGGTTACTTTTTC
GCCGGAAGGCATTTCCTTATCACTCCATCGAAACGTTGATTTCCGGAAGCTCCCGACTATGG
GCATATGACGACGTCGGAAGCGTTTCACTTATTGGCTCTGGCTGGAAACGCTATATGGTTA
CTTCACCGGTGACTGGTCGAAACTGGAACAGGCTTGGACAAAAATGGAGCAATTCATTAT
CCCGAACTCGACGGAACAGCCGACGATGGGGTCTTACAACCCATCAAGTCCAGCTACTTA
CGCGCCGGAACATCCGTATCCGGACCGGTATCCAACTTTGCTGAACAATTCCGTGCCGGC
AGGACAGGACCCACTGGATGCGGAACCTCAAAGCGACGTACGGTAATAACGTGACGTATTT
GATGCACTGGCTGCTTGACGTGGACAATTGGTACGGCTTCGGCAACCTGTTGAACCCGTC
GCATACGGCGACCTACGTCAACACGTTCCAGCGCGCGAACAGGAATCGGTCTGGGAGGC
GATCACACATCCGTGCGAGGACAATTTCCGGTTTCGAAAACCGAATGAAGGTTTTGTGAC
GCTGTTTCGTAAGATAACGGAACGCCTGCCAGCAATGGCGTTATACGGCAGCCTCTGA
CGCCGACGCACGCGCCATTCAAGTGATGATTGGGCGAAGCAGCTGGGGTACAACAACCA
GACCTATCTGGATAAGGCGCGCAAGATGGGCGACTATCTGCGCTATACACTGTTTCGACAA
GTATTTCCAACAAATCGGCAGTGCAAACGACGTTCTCCGAGCCCGGGCAGCGGTAAAAA
CTCTGCGCATTACCTTTTGTCTTGGTACACGGCCTGGGGCGGTGGTCTCGGCTCCGGCGG
CAACTGGGCTTGGAGAATCGGATCGAGCCATGCTCATCAGGGTTATCAAAATCCTGTCGC
TGCTTATGCGCTGTCTGCCGGCGGACTGGCGCCGCGTTCCGCAACGGCACAGACCGACTG
GGCGACGTCGTTGCAACGTCAGCTTGAATTCTATACGTGGCTGCAATCGAGCGAAGGCGC
CATCGGCGGCGGGGCGACCAACAGCGTCGGGGGCGAGCTATCAGCCGTATCCTTCCGGTCG
CAGTACGTTCTACGGCATGGTTTACGATGAAGCGCCGGTTTATCGCGATCCGCCCTTCGAA
CTCGTGTTTCGGCTTCCAAGCGTGGTCCGTGCAACGCGTCGCGGAACGTACTATATCTG
GACCAGCAGCGGAAATACCAATACGCAGCAGTTCAGATGGTTAAAAACATCGTCACCAA
ATGGGTGATTTGGGCGCTTGACTATACGTTCTGTGAATCAACGCCCCGGTTACAGACGCTCA
AGGGTATTTTCTGACGAGCAGCGGCAGCCGTGTCTGGGCGGCAACAATCCGCAGATCGC
CACGGTTTTCCGATCCCGGTCACTTCTATATTCGTCGACGCTGGAATGGCAGGGTCAACC
GGACACATGGAACGGATATGCCAATTATACGGGCAATCCCAATTTCCATGCGATTGCGAA
AGACCCCGGCCAAGACGTCGGCGTCACCGGCAACTATATCAAGCTGCTGACGTTCTTTGC
CGCGGCCACGAAAGCGGAGACGGGGAACCTACACCGCTCTCGGAAGCCAGGCGTTGAATGT
CGCCGAACAGTTGCTGAACGTGCTTTGGAATTTCAACGACGGGGTTGGGATTGTCCGTCC
CGAACAACGCGCCGACTACTTCCGCTATTTTACGAAGGAAATTTACTTCCCGAGCGGCTG
GAGCGGCACGTACGACAGGGCAATACCATTCTGGCCGGGCGCGGTTTCTTCCGATCC
GTCGAAAGGCGGAAACGCGGTTTATATCAGCTACGCGGAACGTGCGTCCGAAGATCAAGCA
AGATCCGAAATGGTCGTATCTTGAAATCTGTACAAAACCTTCGTATAATCCGTCCACAGG
TCGCTGGGAAAACGGTGTTCCGACGTTACGTATACCGTTTCTGGGCGCAGGTGATGT
GGCGACGGCGTATGCGGAATTTGCCCGGTTGATCGGCGGTTTGGGCGCTTCGCCGACACC
GACGCCGAGCGGACACCGACGCCGACACCGTCGGCCGGCGGCAACCTGGTCGTGCAGTA
CCGCGCGGCGGACACGAACGCGACGGACAACAGCTGAAGCCGCACTTTAGGATTGTGAA
CCGCGGGACGACGAGCGTGCCGTTGTGCGAGCTGACGATCCGGTACTGGTACACGGTGGA
CGGAGACAAGCCGAGGTGTTCAACTGCGACTGGGCGCAGGTAGGTTGCTCGAATGTGCG
TGGCAGCTTTGTGAAGCTGACGACGGGCGGACGGGGGCGGACTACTACATCGAGATCAC
GTTACGTCGGGCGCGGCAGCTTGGCGGCTGGGGGAAGCAGTGGGGACATTCAAGGTGCG
GATCAACAAGAACGACTGGACGAACATAACGAGCGAACGACTACTCGTGTGATCCGAC
GAAGACGAGTTTTGCGGATTGGAACCGAGTGACGCTGTATCGTAACGGTCAGCTCGTCTG
GGGCGTCGAACCGTAACGCACAACAGTCCATGCCGAGAAGGTTTCGGGGCGGTTCAACTGA
AGACGGGGAACGCCCCCGTTCTTGAAAGTATGTACGAGCAAAACGAGGGAGGAGCCGAG
CATGGGAACGGGCGGTGCCGAGAGTGATCAAAAAATGTTGCTGGTTTTGGCAATTGG
ATTGCTCCTTCCGATTCCATACCTGCATGTCGCTTACGCGGAAAACGTCCTGATTTTGCA
GAGCGATTTTGAGGACGGGACGACGCAAGGTTGGGTCGGTCGCGGGGAGTCGAAACGCT
TACTGTCACTTCCGCGGACGCTACAGCGGAGCTTATGGTTTGTCCGTGAGCGGAAGAAC

GA AACGTGGCATGGTCCGACATTGGACATCACTTCCTATATTCAGGTTGGAAAGACTTA
TCAATTTTCAGCATGGGTTAAATTGCCTTCCGGTTCGTCCAACACACGCATTTATATGAC
GATGCAAAGAACCATGCAGGACACGGTCTACTATGAGCAAATTTATTTGACACGGCTTC
AGCTGGAAATTGGGTTCAATTGAAAGCCCAATACAAGTTGTACGAACCTGCTGTAAACCT
GCAGGTATACTTTGAAGCTCCCGATCATGCTACTCAATCTTTCTATATTGATGACGTCCG
AATTGAACAACCTTCTGATCTTCCGAAGACGGTAGAAGAGAATATTCCGTCCCTGAAAGA
TGTTTTTCGAGGGCGTTTTTCCGATAGGAACGGCGTTTGAAAATTTTGAACCTTCTCGATGA
ACAGGACAGAAAATTGATTTTTAAACATTTCAATAGTGTGACGCCTGGAAACGTGCTGAA
GTGGGACAGCACAGAACCACAAGAAGGAGTCTTTAACTTTACGGAATCGGATAAAGCGGT
TGCTTTTTCGGTTTCAAGCGGAATGAAGATCAGAGGTCATACATTGATTTGGCATAATCA
GACGCCGAATTGGGTGTTTTATGATTCAAACGGAAATTTAGTTTCCAAAGAAGTTCTATA
TCAACGAATGGAAAGACACATTAAACCCGTCGTCAGCCGCTACAAAGGAATCATCTATGC
GTGGGATGTCGTCATGAAGTTATCGATCCCGGACAGCCTGATGGATTGCGTAGAAGCTT
GTGGTATCAGATTGCCGGCAGGAGTATATCGAAAAGCGTTCCAATTTGCGCATGAAGC
TGATCCGAATGCGCTTCTCTTCATCAATGATTATAACACGCATGAATCCGGTAAAAGCCA
AGCATTGTACAATTTGGTACAACGACTGAAAAATAAGGGTATTCTGTTCACGGAGTCGG
ACACCAGACCCATATTAATATTTCTGGCCGTCGATCAGTGAAATCGAAAATTCGCTCGT
CAAGTTCTCGAACCTGGGAGTTGTTTCAAGAAATCACTGAGTTGGATATGAGCATTACAA
CAATTCATCACAGAAGTACGACACATTGCCTTCCGATTTGGCTCAGCAGCAGGCAACCCG
TTACAGACAACCTGTTGAAATGTTCTTGAGAAGGAGCAGTTTGATTCAAACGTTACGTT
CTGGGGCAAAGATGATGCAAATACGTGGTTGCGGAAGTTCCCAGTCGTCCGAAATGACTG
GCCGCTGTTGTTTCGATGAGCAATTAAGGCGAAACCCGCATATTGGGCGGTAGTCGGAAC
TGTTCCGTCACCCACGCCGACACCGACGTCGACGGCAACACCAACGCCGACACCAACGGT
GATACCGACGCCGACACCGACGCCAACGCCGACATCGACGCCGACACCGACGCCGACGTC
GACGCCGACGCCTAGCGCGAGCGGCACCCTGCGCGTCGAGTATCGCGTGGGCGATTCCAG
CGCCACCGACAACAGATGAAACCGCAGCTGCGCATCGTCAACACCGGCTCGCAAGCCGT
GCCGCTGACCGAACTGAAAGTGCCTACTGGTACACGAAGAACTCGACGCAGGCCGAACA
GTACTTCTGCGACTGGGCGCAAATCGGCTGCTCGAACATCCGGGCGCAGTTCTGTGTCGT
GGCGCAGCCAGTCAGCGGAGCGGACAGCTACATCGAGCTGAGCTTCACGGGGGCGAGCGT
TCCGGCGGGAGGCAACACGGGCGAGATCCAGAACCGGATTCACCTTCACGAACCTGGATGAA
CTACAACGAAACGGACGACTGGTTCGTACAACGGGACGCAGACGACGTGGGGGCGGTCGAC
GCGGATTACGCTGTATCGTAACGGCGTGCTGGTGTGGGGCACCGAGCCGGGCGGCGGATC
GCCGACACCGACACCAACGCCGACGTCAACGCCGACGCCGACGTCAACGCCGACACCGAC
GCCGACGCCGACACCGACACCGACGCCGACACCGACACCCACACCCACGCCGACGCCGTC
GGCCGGCGGCAACCTGGTTCGTGACGTACCGCGCGGCGGACACGAACCGGGCGGCAACCA
GCTGAAGCCGCATTTTCGGATTGTGAACCGCGGGACGACGAGCGTACCGTTGTGCGAGCT
TACGATTCCGTACTGGTACACGGTGGACGGCGACAAGCCGAGGTGTTCAACTGTGACTG
GGCGTGGGTGCGGATGTTGAAACCTGCGCGGCGAGTCTGGTGAAGTTGACGACGGGCGGAC
GGGGGCGGACTACTACCTTGAGATCACGTTACATCGGGCGCGGGCAGCCTGGCGCCTGG
GGCGAACAGCGGAGACATTCAGGCGCGGATCAACAAGAAGCAGCTGGACGAACCTACAACGA
GGCGAACGACTACTCGTATGATCCGACGAAGACGAGTTTTCGGATTGGAACCGGGTGAC
GCTGTATCGGAATGGTCAGCTCGTCTGGGGCGTCGAGCCGTAAGGGTATACCTAAGAGCG
GCGTGGCGGAGTCGATAAGCGGTGATGATTCCGCCTCGCTCGAGGACCGGTGACTGCCA
CAGAAGGCTTTGTGAAGGAGGTGATGGACGGAAGATCCGAAAAAGAAAGAAATATGAAGGT
TTTGTGGGTGGTTTTTGGTAAAAAAGAATCCATGAGGAACCAACGAAAGAGGGGAGTGA
CACAGGCATGACGATGGCGTGGAAACAGCGCAGCGGATTGATCGCGTTGATTTTGGCATT
GGTAGCGGGTTTGTCTGCTGCCATGGGGATCGCTGCCGAAAGCGGCGGCGGAGCCGCATGT
GGACAATCCGTTTGTAGGAGCGACGGCTTACGTCAATCCGGACTATGCGGCGCTGGTCCGA
TTCGTGATCGCGAGGGTGAGCGATCCAACGCTGGCGGCGAAGATGCGTACGGTCAAGAC
GTATCCGACGGCGGTGTGGTTGGATCGGATCGCGGCGATTGACGGAGGGCCGGGAAGACG
GAGCTTGGTGCAGCATTGGATACGGCGTTGGCGCAGAAGCAAGGAATACGCCGATTAC
GGCGATGTTTGTGATTTACAATATGCCGGGTGCGGACTGCGCGGCGCTGGCGTCCGAACGG
GGAGCTGCCGCTGACGCAGGAAGGGCTGCAGAGGTACAAGACGGAGTATATTGACCGAAT

TGCGGCAATTTTTGCAGATCCGAAGTATGCGGGAATTCGGATCGTGACGGTGATTGAACC
GGACGGCTTGCCGAACCTGGTGACGAACCTGAGCGATCCGGAATGCGCGCAGGCGAATTC
AAGCGGAATTTATGTAGAGGCAGTACGATATGCGATCAACAAGTTGAGCGAAATTCGGAA
CGTGTATATTTACCTGGACATCGCGCATTTCGGGATGGCTGGGCTGGGACAACAACCGGAC
CGGCGCGGTGCAGCTGTATACGAACGTGGTGCGAGGGACGACGAAAGGGCTTTTCGAGCGT
GGACGGGTTTGTGACGAACGTGGCGAACTATACGCCGCTCGAGGAGCCGTATTTGACGGA
TCCAAACCTGACGGTGGGAGGTGAGCCGCTTAAGTCAGCGAAGTTTTATGAGTGGAACCC
GTATTTTGATGAAGTAGATTATGCGGCAGCGTTGCGGTTCGGCGTTTATCAGTGCAGGGTG
GCCGACGAGCATCGGGATGTTGATCGACACGAGCCGCAACGGCTGGGGCGGGCCGAACCG
GCCGACGGGAGCGAGCGGGACGACGGTGGACGCGTATGTGAATTCGGGGCGTGTGGACCG
TCGGGCGCATCGCGGGCTGTGGTGTAACGTGACGGAGCGGGGATGGGAATGCCGCCGCA
GGTGGCGCCGGCGGCGTATGCGTCGCAAGGGATCGAGGCATTCTGTATGGGTGAAGCCGCC
CGGGGAGTCGGACGGAGCGAGTTCGGAGATACCGAACGACGAAGGCAAGCGGTTTGACCG
GATGTGCGATCCGACGTATACGACGCAATACGGGGTGTGACGGGGCGTTCGCCAACGC
GCCGTTGGCGGGGCAATGGTTCCATGATCAGTTTGTGATGTTGGTGCAGAATGCGTATCC
GGCGATTCCGACGAGCGGCGGTGGGACACCGACGCCGAGTACGACGGTGACGCCGACACC
GACACCGACGCCGACACCGACGCCGAGTGCAGCGGTGACGCCGACACCGACACCGACGCC
GACACCGACGCCGAGTGCAGCGGTGACGCCGACACCGACACCGACGCCGACACCGACGCC
GACCGGTGACGCCGACGCCGACATCGTCGACAAGTTTTGTGGCCAGGCACGGGCAATTGAG
AGTCGTGGGGAATCAATTGGTCGACCAAAATGGACAACCCATCCAATAAGAGGCATTAG
TTCTCATGGGTTACAATGGTATGGGCATTTCTGCAATCGAGACAGCCTCCGATGGCTCCG
AGATGATTGGGGAATAACAGTTTTCCGAGCAGCTCTGTATACTGCCGAACAAGGATATAT
CACGAATCCGTCTTTAAAGAAAAAGTGAAGGAAGCTGTACAAGCCGCAATTGAACTCGG
TATTTATGTGATCATCGACTGGCACATTTTGTCTGATGGCGATCCGAACACGTACAAGGA
GCAAGCGAAAGCGTTTTTCGATGAAATGTCGCGATTGTACGGCAGTTATCCGAACGTGAT
TTATGAGATCGCCAACGAACCGAATGGTGTGACATGGGAAGGACAGGTTAAGCCGTACGC
TTCGAGGTGATCCCGGTGATCCGTGCTAATGACCCGATAATCTCATTATTGTCCGAAC
AACAACGTGGAGTCAGGATGTCCATCTTGACGAGATAGCCCGCTACCTTACAGCAACCT
GGCGTACGCTCTGCATTTCTATGCCGGTACGCATGGTCAATGGTTGAGAGACCGGATCGA
CTATGCGAGGAATAAAGGCATCGCGATTTTCTGTAGTGAATGGGGGACAAGCACTTCGAC
AGGCGATGGAGGCCCTTATCTACGGAGTCGCAACAATGGTTGGATTTCTTAATGCTCG
GCAGATCAGTTGGGTGAACTGGTCGTTGAGCGACAAGGCCGAGTCATCCGCAGCATTGTT
GCCTGGCGCAAGCGCAACAGGTGGTTGGACGGACGCACAATTGTCTCAGTCGGGGCGTTT
TGTTTCGCGCTCAGATTTCGACGCGGTGTATTGACGCCGACACCGACGCCGACACCTACGCC
TACGCCGAGTGCGGCGCCACACCGACGCCGACGCCGCGGCGGAGCCTGGTCTGTCAGTA
TCGCGCGGCGGACACGAACCGCGGCGACAACCAGCTGAAGCCGCATTTTCGATTGTGAA
CCGCGGGACGACGAGCGTGCCGCTGTGCGAGCTTTCGATCCGGTACTGGTACACGGTGGA
CGGAGACAAGCCGCAGGTGTTCAACTGCGACTGGGCGCAGGTGGGTTGTTGAACTTGCG
GGGAGTTTTCGTGAAGCTTTCGACGGGCGGACGGGGGCGGACTACTACATTGAGATCAC
GTTACGTCGCGGCGGGCAGCTTGGCGGCTGGGGGAAGCAGCGGGGACATTACGGTGCG
GATCAACAAGAACGACTGGACGAACATAATGAGGCGAAGGACTACTCGTATGATCCGAC
GAAGACGAGTTTTGCGGATTGGAACCGGGTGACGCTGTATCGCAACGGTCAGCTCATCTG
GGGCGTCGAGCCTTGATTGCCAACCGACGGTATGGACCTGGCGGACGGTAAGTCCGTTTCG
CCAGGTTTCTATAAAAAACAACATCTCCGCTCGAAAAACCTTGAAGGAAGGGAGAGGATT
TTTTATGAATCGACGCCCTATCGCCCGCCTCAGCGGCATGTTGGCGATGGTTCTCATCGC
CGCAGTGTTGGCGTACGTTCCGAAGCCTGAACCGGCCGAGGCGCACGGAGGTATGGTGTT
TCCAGCCACGCGAACGTATGCCTGTTATGTTGACGGCAAGGTTTACGGCAATGGCGGAGA
CTTGAAACATGATCAATCCGGCGTGTCTTGATGCCCTTGGCGATCTCGGGCAACTATCAGTT
CTGGAACGTGGTTTCGAAAATCTGATCAGTAATGCCGAGGACGCCATAGGGAAATCATTCC
TGACGGCAAACGTGTCGGACCAACGGCCAGTTTTGATGGTATGAACCAGGCGCGTACAGA
CTGGTGGACGACTCGTCTGACGCCGGGCGCAACGATTACGGTGCGAGTCAACGCATGGGC
GCCGCATCCCGGCACGTGGTATTTGTATGTAACCGGGACGGATGGGATCCGACACAACC
GCTGAAATGGTCGGATCTGGAACCGACGCCCTTCAGCCAGGTGACTAATCCGCCGATCAA

CTCGAGCGGACCGGACGGGGCCGAGTACAGCTGGCAGGTGCAGCTGCCGAACAAGCAAGG
GCGACACATCATTTTATATGATATGGCAGAGATCCGACAGTCCGGAGGCATTTTACAACTG
TTCGGATGCGTATTTTCGGATCGGGGGCCGATTGCTTATGAATTTGGTGACCCGCGGGAAGG
AGGAACGATGATTACGCCGCCCGCTCGGGCACGACGCCGACACCGACGCCGACACCGAC
GCCGACACCGACGCCGACACTGACGCCGACACCGACGCCGACACCGACGCCGACACCGAC
GCCGACACTGACGCCGACACCGACGCCGACACCGACGCCGACATCGACGCCGACGTCGAC
ACCGACGTCGACGCCGACACCGACACCGACGCCCTAGCGCGAGCGGCACCCTGCGTGTCGA
GTATCGCGTGCGGCGATTCCAGCGCCACCGACAACCAGATGAAACCGCAGCTGCGCATCGT
CAACACCGGCTCGCAAGCCGTGCCGCTGACCGAGCTGAAGATGCGCTACTGGTACACGAA
GAACTCGACGCGAGGCCGAACAATACTTCTGCGACTGGGCGCAGATCGGCTGCTCGAACAT
CCGGGCGCAGTTCTGTGCTGTGCGAGCCGGTCAGCGGGGCGGACAGCTACATCGAGCT
GAGCTTTACGGGCGGAAGCATTCCGGCGGGAGGCAACACGGGCGAGATTGAGAACCGGAT
TCACTTCACGAACTGGATGAACTACAACGAAACGGACGACTGGTTCGTACAACGGGGCGCA
GATGACGTGGGGGCCGTCGACGCGGATTACGCTTTATCGCAACGGCGTGCTGGTGTGGGG
CACGGAGCCGGGCGCGGATCGTCGCCGCCGACGCCGACGGTGACACCGACACCTACACC
GACATCGACGCCGACACCTACGCCCTACGCCGAGTGCGGCGCCACACCGACGCCGTGCGC
CGGCGGCAGCCTAGTCGTGCAGTATCGCGCGGCGGACACGAACGCGGGCGACAACCAGCT
GAAGCCGCATTTTCGGATTGTGAACCGCGGGACGACGAGCGTGCCGCTGTTCGGAGCTTTC
GATCCGGTACTGGTACACGGTGGACGGGGACAAGCCGACGGTGTTCAACTGCGACTGGGC
GCAGGTGGGTTGTTTCGAACTTGCGGGGCGAGCTTCGTGAAGCTTTCGACGGGCGGACGGG
GGCGGACTACTACATCGAGATCACGTTTACGTCGGGCGCGGGCAGTCTGGCGCCTGGGGC
GAGCAGCGGAGACATTGAGGTGCGGATCAACAAGAACGACTGGACGAACTACAACGAGGC
GAACGACTACTCGTATGACCCGACGAAGACGAGTTTTCGGATTGGAACCGGGTGACGCT
GTATCGGAATGGTCAGCTCGTCTGGGGCGTTGAACCATAATAACGGCAAGCACAACTCGG
CCAGGTCTGTTTCTCCAAAGCCCTTCTTTTCGGAAGTATCGAAAGAAGGGCTTTCCTTCTAA
ACTTTTTTCGGGTGACATCTAAAGTTTATCCCGTACTCGAAGGATCGAGAGAAACGATAG
AATAGGCAATAAGTTACTGTAAATCTTGATGAACAGAAAGGAGATGATTACAAAAGGAC
GGATTTCATCGTTTTTTTTTCGAAAAGGTCGGCAGTAGGTCTGTTTGAGACGGGATCACACA
TGCATCAAATTGCAGAGAGGAGGAAAATGATAATGCGAAACTGGCTCCAATGGGGCATGG
TTGCGGCTTTGCTCGTTTTTACGGTATCGGTCTGTCGCCCCGAAAGAAGCCGATGCAGGGC
TAGCCAAGACAAAATTCTTGGGGAATGTCATCAACAATAGCATCCCTTCTGATTTTGCTG
TTTACTGGAATCAGGTTACCCCTGAAAACGCTACCAAGTGGGGTTTCGGTCAATCCAGCC
GCGACAACATGAACTGGTCGACGCCGATATGATTTACAACCTACGCTCGCAGTAACGGTT
TTCCGTTCAAATTCCACACACTGGTCTGGGGGAGTCAGGAGCCCGGCTGGATCAGCGGGC
TTTCGGCTGCAGAACACAGGCCGAAGTGATCGAATGGATCCAAGCGGCCGGTACGCGTT
ATCCCGACGCAGACTTCGTGACGTCGTCAACGAACCGCTGCACGCCAAACCTTCTTACC
GCAATGCCATCGGCGGAGACGGCTCGACAGGTTGGGACTGGGTCATCTGGTCTGTCGAAC
AGGCGCGCCGCGCATTCCCAATTCCAATTGCTGATTAACGAGTACGGCGTCGAGAACG
ACCCGAATGCGGCGAGCCAATATGTCCAAATCATCAATCTGTTAAAAAGCCGCGGCTTGA
TCGACGGCATCGGCATTCAAGGTCATTATTTCAATCTTGACACGGTTTCAGTCAGTACGC
TGCGAACACGCTCGGTATGCTTGCTGAAACAGGTTTGCTATTTATGTGTCAGAACTGG
ATATTTTCGGGTGATGACGCCACGCAATTGGCTAGATATCAAGAAAAGTTCCCAATTCTAT
GGGAACATCCTTCTGTCCAAGGGATTACGCTGTGGGGCTATATTGAAGGTCAAACCTGGA
GATCCGGCACGCATTTGATTACGGCTTCGGGCGTGGAACGACCTGCGTTGCAATGGTTGC
GGACGTATTTGGCAGGAGCCGGATCCTCGCCGACACCAACGCCGACGCCACACCGACCG
TGACACCAACGGTGACGCCGACACCGACGCCCTAGCGCGAACGGCACCCCTGCGCGTCGAGT
ATCGCGTGCGGCGACTCTAGCGCCACCGACAACCAGATGAAACCGCAGCTGCGCATCGTCA
ACACCGGCTCCCAAGCCGTGCCTCTGACCGAGCTGAAGGTGCGCTACTGGTACACGAAGA
ACTCGACGCGAGGCCGAACAGTACTTCTGCGACTGGGCGCAGATCGGCTGCTCGAACATCC
GGGCGCAGTTCTGTGCTGTGCGAGCCGGTCAGCGGGGCGGACAGCTACATCGAGCTGA
GCTTCACGGGCGGAAGCATTCCGGCGGGAGGCAACACGGGCGAGATACAGAACCGGATTC
ACTTCACGAACTGGATGAACTACAACGAAACGGACGACTGGTTCGTACAACGGGGCGCAGA
CGACGTGGGGGCCGTCGACGCGGATTACGCTGTATCGCAACGGCGTGCTGGTATGGGGCA

CGGAGCCGGGCGGATCGTTCGACGCCGACACCGACACCGACGCCGACCCCTACGCCTACGC
 CGAGCGCGGCCGCCACACCGGCGCCGTCGGCCGGCGGCAGCCTGGTTCGTGCAGTATCGCG
 CGGCGGACACGAACGCGACGGACAACCAGCTGAAGCCGATTTTCGGATTGTGAACCGCG
 GGACGACGAGCGTGGCGCTGTTCGGAGCTAACGATTCCGGTACTGGTACACGGTAGACGGAG
 ACAAGCCGCGAGGTGTTCAACTGCGACTGGGCGCAGGTGGGCTGCTCAAACGTGCGGGGCA
 GCTTCGTGAAGCTTTCGACGGGCGGACGGGGCGGACTACTATATTGAGATCACGTTCA
 CGTCAGGCGCGGGGAGCCTGGCGCCTGGGGCGAACAGCGGAGACATTACAGGCGCGGATCA
 ACAAGAACGACTGGACGAACACTACAACGAGGCGAACGACTACTCGTATGATCCGACGAAGA
 CGAGTTTTGCGGATTGGAACCGGGTGACGCTGTATCGGAACGGTCAGCTCATCTGGGGCG
 TCGAACCTGA (SEQ ID NO:7)

[0062] The amino acid sequence of the *Paenibacillus sp. A59* (WP_053783505) GH9 is as follows:

MKGSWRRRVILALSTGLLAGSTSIQAWNGKADAAAGNHNYAEALQKAIYFYETQSRGKLPEDNRVEWRG
 DSGLNDGADVGVDLTGWDYDAGDHVKFGVPMAYSATMLAWSVVEYREGYEQAGQLEEIKDNLKWATDYFV
 RAHTKPNEIHWGQVGAGNTDHAWWGPAEVMQMNRPAYKIDASCPGSELAGETAAALASSSIVFRSDPAYA
 NKLLQHAKELYSFADTYRGKYSDCITDAQSFYNSWTGYDELAWAATWLYMATNDSAYLSKAIATANLWQ
 ADGQSGNWAYTWTQGWDDKHYGAQILLARITSSSLNMPAARFIQSTERNLDYWSVGTNGQRIKYTPGGLA
 WLDTWGSLRYAANASFIAFVYSDWVSDPVKKARYQDFAVSQMNYILGDNPRQSSYVVGYGQNPCKPHHR
 TSHSSWTNNENVPSEHRHTLYGAMVGGPDASDAYTDSIGDYVSNEVATDYNAGFTGALAKMNLFGQNNQ
 PIANFPAPEVKSDEFFVEASVKASGSNYTEIKAQLNNRSGWPARMGDKLSFRYFVDLSEVYAAGYTVSDV
 HVTTAYAEGAIVSQPDVDAVKRIYAVTADFTGTKIYPGGEGHYRKEVQFRITGPEGAWNANNDHSFQGL
 GTGNVAKSAYLPVYDAGIRIYQEPGITPVVTPIAPSGVQAVSGNAQVILNWIASPAGESYTVKRAEVNG
 GPYTSVATNVLGLTYTNTGLTNGKTYYYVVTAVNSVGESPSAQATATPQAGTSLPGALTLSGTAGNTQS
 ILTWTAAATGAVSYKVQRAAGGSAYADVATGLAVLNLYTDATAANGTAYSYRIAAVNASGQTLNIVTLTPS
 APPATTGTLEVQYRNGSGASGNAVTPQFNLKNTGTQPIDLSTVKLRYYFTKDGTDLTFWCDYAIQIGSA
 NIEGKFVTLNPAKGTADTVLEISFQSGAGSLAAGAEAGVIQGRFSKNNWSNFDQSNDSYDATKTAFTTW
 NQVIGYQGGTKVWGIEP (SEQ ID NO:8)

[0063] The amino acid sequence of the *Paenibacillus borealis* GH48 (AIQ60376) is as follows:

MKLSSIKKPF SIVMAAILIISLTSGIFNFRPGTAKAASVEKTRFLQLYAQLKDPASGYFSAEGIPYHSVE
 TLLSEAPNYGHMTTSEAYSWMWLEVLYGYNTGDWSKLEAAWDNMEKYIIPINEGDGVQEQT MNYNPN
 SPATYASELAQPDQYP SALNGKYAPGKDPLDAELKAAYGNNQTYLMHWLVDVDNWYGFGNLLNPTHAA
 VNTFQRGVQESVWEAVAHPSQDDKSFGKTNEGFMSLFTKENSVP SAQWRYTNATDADARAVQAMYWAKDL
 GYTNTVYLNKAKKMGDFLRYGMYDKYFQKVGSAADGTPEAGTGKDSQYLLAWYTAWGGGLGTTGNWAWR
 IGASHAHQGYQNVVAAALSTADGGLIPASATAGEDWGKSLTRQLEFYNNWLSSEGAIAAGGATNSYGGSY
 SAYPSGTSTFYGMAYDEAPVYHDPSPNNWFGMQAWSLIERVAELYIILASSGDTTSANFKMAKRVIENWID
 WSADYAFAGSRPVTDAAGYYLDLQGNRILGGDDPQIATVSAPGEFWIPGNVEWQQQPD TWSGFSSFSGNS
 GLKAVTKDPGQDTGVLGSYIKALTFFAAGNKAEHGSYALGGTASQLAKSLLD TAWGYNDGVGITTLEKR
 ADYFRFFTKEVYFPAGWTGTFGQGNTPGSSTVPSDPAKGGNGVYASYTDVLPDIKNDPKWSYLEGKYNS

SYNKTTKTWDNGAPEFTYHREWSQVDIATAYAEDRLINNGSGP IPTATPTTPTATPTVTP TATPTATP
TVTP TATPTVTP TATP IATP TATPTATPTATPTATPAAANLVQYRTDTNATDQQFRPQLRIVNNGTTA
VDLSKVKLRYYYYTIDGEKAQQFNVDYATLGGSNVLGSFVKLEPAVAGADYYVEISFSTGAGSLAPGANTG
EIQLRINKTDWSNYNKADDYSYDSTKTAYTDWNRVTLYLNGVRVWGVQPQ (SEQ ID NO:9)

[0064] The amino acid sequence of the *Cohnella laeviribosi* GH10 is as follows:

MSKAKAMKICASLLLLGSVFSFIATSDADAGLARKSKFLGNVIASSVPSNFATYWNQVTPENSTKWGSVEA
TRNVMNWSAADLAYNYAKSNGFPFKFHTLVWGSQQPGWISGLSQAEQKAEVLQWIQAAGQRYPNADFVDV
VNEPLHAKPSYRNAIGGDGATGWDWVIWSFQEARKAFFNAKLLINEYGIISDPNAANQYVQIINLLKSRG
LIDGIGIQCHYFNMDSVSVSTMNSVLNLTAAATGLPIYVSELDMTGDDSTQLARYQQKFPVLWEHSAVKGV
TLWGYIEGQTWASNTHLVRNNGTERPALQWLRTYLSLTH (SEQ ID NO:10)

[0065] The amino acid sequence of the *Caldibacillus cellulovorans* beta-1,4-mannanase precursor is as follows:

[illegible]

[0066] The amino acid sequence of the *Caldibacillus cellulovorans* beta-1,4-xylanase XynA precursor is as follows:

MGTGRAGEWIKKMLLVLAIGLLIPIPHYVASAENVLILQSDFFEDGTTQGWVGRGGVETLTVTSAAAYS
 AYGLSVSGRTETWHGPTLIDITSYIQVGKTYQFSAWVKLPSGSSNTRISMTMQRTMQDVTYYEQIYFDAL
 SGNWQLKAQYKLYEPAVNLOVYFEAPDHATQSFYIDDVRIEQLPDLPKTVEENIPSLKDVFAGRFPIGT
 AFENFELLDEQDRKLLIKHFNSVTPGNVLKWDSTEPQEGVFNFTESDKAVAFVQNGMKIRGHTLIWHNQ
 TPNWVFYDSNGNLVSKEVLYQRMERHIKTVVSRYKGIYAWDVVNEVIDPGQPDGLRRSLWYQIAGEEYI
 EKAQFAHEADPNALLFINDYNTHESGKSQALYNLVQRLKSKGIPVHGVGHQTHINITWPSISEIENSLV
 KFSNLGVVQEITELDMSIYNNSSQKYDTLPSDLAQQQATRYRQLFEMFLRRSSLIQNVTFWGKDDANTWL
 RKFPVVRNDWPLLFDEQLKAKPAYWAVVGTVPSPTPTPTSTATPTPTPTVIPTPTPTPTPTSTPTPTPTPT

SASGTLRVEYRVGDSSATDNQMKPQLRIVNTGSQAVPLTELKVRWYTKNSTQAEQYFCDWAQIGCSNIR
 AQFVSLAQPVSGADSYIELSFTGGSVPAGGNTGEIQNRHFTNWMNYNETDDWSYNGTQTTWGPSTRITL
 YRNGVLVWGTEPGGGSSPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTATPTPTPTPTPSAG
 GNLVVQYRAADTNAGDNQLKPHFRIVNRGTTSVPLSELTIYRYWYTVDGDKPQVFNCDAWVGCSNLRGSL
 VKLTTGRTGADYYLEITFTSGAGSLAPGANSQDIQARINKNDWTNYNEANDYSYDPTKTSFADWNRVTLY
 RNgQLVWGVEP (SEQ ID NO:12)

Ionic liquid (IL)

[0067] The suitable IL that can be used in the present invention can be any IL that does not impede the enzymatic activity of any of the GH polypeptide or the cellulase complex. In some embodiments of the invention, the IL is suitable for pretreatment of biomass and for the hydrolysis of cellulose by any of the GH polypeptide or the cellulase complex. Suitable IL are taught in ChemFiles (2006) 6(9) (which are commercially available from Sigma-Aldrich; Milwaukee, WI). Such suitable IL include, 1-alkyl-3-alkylimidazolium alkanate, 1-alkyl-3-alkylimidazolium alkylsulfate, 1-alkyl-3-alkylimidazolium methylsulfonate, 1-alkyl-3-alkylimidazolium hydrogensulfate, 1-alkyl-3-alkylimidazolium thiocyanate, and 1-alkyl-3-alkylimidazolium halide, wherein an “alkyl” is an alkyl group comprising from 1 to 10 carbon atoms, and an “alkanate” is an alkanate comprising from 1 to 10 carbon atoms. In some embodiments, the “alkyl” is an alkyl group comprising from 1 to 4 carbon atoms. In some embodiments, the “alkyl” is a methyl group, ethyl group or butyl group. In some embodiments, the “alkanate” is an alkanate comprising from 1 to 4 carbon atoms. In some embodiments, the “alkanate” is an acetate. In some embodiments, the halide is chloride.

[0068] Such suitable IL include, but are limited to, 1-ethyl-3-methylimidazolium acetate (EMIN Acetate), 1-ethyl-3-methylimidazolium chloride (EMIN Cl), 1-ethyl-3-methylimidazolium hydrogensulfate (EMIM HOSO₃), 1-ethyl-3-methylimidazolium methylsulfate (EMIM MeOSO₃), 1-ethyl-3-methylimidazolium ethylsulfate (EMIM EtOSO₃), 1-ethyl-3-methylimidazolium methanesulfonate (EMIM MeSO₃), 1-ethyl-3-methylimidazolium tetrachloroaluminate (EMIM AlCl₄), 1-ethyl-3-methylimidazolium thiocyanate (EMIM SCN), 1-butyl-3-methylimidazolium acetate (BMIM Acetate), 1-butyl-3-methylimidazolium chloride (BMIM Cl), 1-butyl-3-methylimidazolium hydrogensulfate (BMIM HOSO₃), 1-butyl-3-methylimidazolium methanesulfonate (BMIM MeSO₃), 1-butyl-3-methylimidazolium methylsulfate (BMIM MeOSO₃), 1-butyl-3-methylimidazolium tetrachloroaluminate (BMIM AlCl₄), 1-butyl-3-methylimidazolium thiocyanate (BMIM

SCN), 1-ethyl-2,3-dimethylimidazolium ethylsulfate (EDIM EtOSO₃), Tris(2-hydroxyethyl)methylammonium methylsulfate (MTEOA MeOSO₃), 1-methylimidazolium chloride (MIM Cl), 1-methylimidazolium hydrogensulfate (MIM HOSO₃), 1,2,4-trimethylpyrazolium methylsulfate, tributylmethylammonium methylsulfate, choline acetate, choline salicylate, and the like. The ionic liquid can comprises one or a mixture of the compounds. Further ILs are taught in U.S. Patent No. 6,177,575; herein incorporated by reference.

[0069] The ionic liquid (IL) is of a concentration of more than 0% to 20% of the composition or solution. In some embodiments, the concentration of IL is equal to or more than 1%, equal to or more than 2%, equal to or more than 3%, equal to or more than 5%, equal to or more than 10%, equal to or more than 15%. In some embodiments, the upper range of the concentration of IL is equal to or less than 20%. In some embodiments of the invention, the IL has a concentration from more than 0% to about 20%. In some embodiments of the invention, the IL has a concentration from about 5% to about 20%.

[0070] The solution containing IL can further comprise NaCl, such as up to 10 mM of NaCl. The solution can further comprise a suitable buffer.

Host cells

[0071] The host cells of the present invention are genetically modified in that heterologous nucleic acid have been introduced into the host cells, and as such the genetically modified host cells do not occur in nature. The suitable host cell is one capable of expressing from the nucleic acid of the present invention.

[0072] Any prokaryotic or eukaryotic host cell may be used in the present method so long as it remains viable after being transformed with a sequence of nucleic acids. Generally, although not necessarily, the host microorganism is bacterial. Examples of bacterial host cells include, without limitation, those species assigned to the Escherichia, Enterobacter, Azotobacter, Erwinia, Bacillus, Pseudomonas, Klebsiella, Proteus, Salmonella, Serratia, Shigella, Rhizobia, Vitreoscilla, and Paracoccus taxonomical classes. Preferably, the host cell is not adversely affected by the transduction of the necessary nucleic acid sequences, the subsequent expression of the proteins (i.e., enzymes), or the resulting intermediates required for carrying out the steps associated with the mevalonate pathway. For example, it is preferred that minimal "cross-talk" (i.e., interference) occur between the host cell's own

metabolic processes and those processes involved with the mevalonate pathway.

[0073] Suitable eukaryotic cells include, but are not limited to, fungal, insect or mammalian cells. Suitable fungal cells are yeast cells, such as yeast cells of the *Saccharomyces* genus.

[0074] The present invention provides for an isolated GH polypeptide produced from the method of the present invention. Isolating the GH polypeptide involves the separating at least part or all of the host cells, and parts thereof, from which the GH polypeptide was produced, from the isolated GH polypeptide. The isolated GH polypeptide may be free or essentially free of impurities formed from at least part or all of the host cells, and parts thereof. The isolated GH polypeptide is essentially free of these impurities when the amount and properties of the impurities do not interfere in the use of the GH polypeptide as an enzyme. In some embodiments, these host cells are specifically cells that do not in nature produce the GH polypeptide. The impurities are no more than 5%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% by weight of a composition comprising one or more of the GH polypeptide.

[0075] It is to be understood that, while the invention has been described in conjunction with the preferred specific embodiments thereof, the foregoing description is intended to illustrate and not limit the scope of the invention. Other aspects, advantages, and modifications within the scope of the invention will be apparent to those skilled in the art to which the invention pertains.

[0076] All patents, patent applications, and publications mentioned herein are hereby incorporated by reference in their entireties.

[0077] The invention having been described, the following examples are offered to illustrate the subject invention by way of illustration, not by way of limitation.

EXAMPLE 1

[0078] This invention describes the identification of a new type of cellulase complex previously unobserved in natural systems. The active component of a cellulolytic microbial community was partially purified by affinity digestion and column chromatography and

visualized by native polyacrylamide gels (Native PAGE), using both Coomassie staining and endoglucanase activity staining, which revealed a 350 kDa protein complex containing three cellulases, a glycoside hydrolase (GH) family 48 protein, a GH9 protein and GH6/5 protein. Reconstruction of the genes for these proteins indicated that each of these proteins contained at least cellulose-binding domain (CBM3) that may be involved in complex formation. The three proteins were located in a putative operon that also contained multiple GH10 proteins and an AA10 polysaccharide monooxygenase. The gene sequences of the complex lacked a scaffoldin domain and cohesion-dockerin pairs characteristic of anaerobic cellulosomes. The complex can be produced by heterologous expression of individual glycoside hydrolases in *E. coli* and used for the hydrolysis of polysaccharides in biomass. The catalytic domains in the GH9, GH48, and GH6/5 proteins can be replaced by heterologous expression with new catalytic domains to change or improve the hydrolysis of biomass and/or add new functions to the complex.

MATERIALS AND METHODS

[0079] Cultivation of Adapted Consortia

[0080] Sample collection and enrichment procedures were described previously (1, 2). Briefly, compost sample are derived from Jepson Prairie (JP) Organics located in Vacaville, California. The compost-derived microbial consortium was adapted to switchgrass (SG) and microcrystalline cellulose (Avicel) as the sole carbon source in liquid culture at 60°C (JP-148, 1% Avicel, and 1% SG) and passaged every 14 days. Characterizations were performed on a 300 L scale up of a passage carried out at the Advanced Biofuels Process Demonstration Unit-Lawrence Berkeley National Lab. The supernatant was separated from residual biomass by centrifugation, lyophilized, and stored at -20°C for subsequent analysis.

[0081] Cellulase enrichment by column chromatography

[0082] The lyophilized supernatant was resuspended in an appropriate volume of H₂O and passed through a 0.2 µm filter. The supernatant was desalted through PD-10 desalting columns (GE Healthcare, Little Chalfont, UK) or by dialysis against running buffer (20 mM Tris, pH 8.0) followed by gradual fractionization by anion-exchange chromatography (AIEX) with a 30 ml gradient with elution buffer (20 mM Tris, 2 M NaCl, pH 8.0) from HiTrap Q HP columns on an ÄKTA Protein Purification System (GE Healthcare, Little Chalfont, UK).

[0083] Furthermore, glycosidases in the supernatant protein concentrate were enriched solely and in addition to AIEX fractionation via binding to amorphous cellulose and subsequent affinity digestion as described previously (3). Briefly, acid-swollen Avicel PH-105 was added to the supernatant. After binding, the amorphous cellulose was separated by centrifugation and resuspended in 10 ml of reaction buffer containing 25 mM 2-(N-morpholino)ethanesulfonic acid (MES), pH 6.5. Affinity digestion was performed with dialysis membranes (SnakeSkin and Slide-A-Lyzer (Fisher Scientific, Pittsburgh, PA, USA) with a 3.5- to 10-kDa cutoff) against reaction buffer at 55- to 60°C for up to 48 h with frequent changes of the dialysis membranes and buffer to prevent possible membrane ruptures and product inhibition. The reaction was considered complete after no visible changes of the substrate were observable.

[0084] Glycoside hydrolase activity assays

[0085] The glycoside hydrolase activities were measured using the DNS (3,5-dinitrosalicylic acid) reducing sugar assay (endoglucanase, endoxylanase) and the p-nitrophenol (pNP) assay (β -D-glucosidase, cellobiohydrolase, β -D-xylosidase, and α -L-arabinofuranosidase) (5, 29). Heat-killed samples generated by heating the supernatant to 95°C for 16 h were used as blanks. Activity units for all assays were calculated as μmol of sugar liberated $\text{min}^{-1} \text{ml}^{-1}$ and reported as U/ml.

[0086] Zymography

[0087] SDS-PAGE and Native-PAGE (4) were performed with 8 to 16% Protean TGX protein gels (Biorad) with the Tris-glycin-SDS-buffer and Tris-glycin-buffer, respectively. Protein bands were stained with Coomassie blue dye. Protein bands with endoglucanase or endoxylanase activity were visualized using a modification of the zymogram technique as described previously (5). Gels were incubated in 2% carboxymethyl cellulose (CMC) or 2% xylan from Birchwood

[0088] Protein Identification

[0089] Proteins were excised from native gels and analyzed as previously described (6).

[0090] Reconstruction of Glycoside Hydrolase Genes

[0091] Small discontinuously contigs encoding partial glycoside hydrolases were identified

by combining proteomics with metagenomics. PCR primers (Table 1) were deduced using the CLC Main Workbench (Qiagen, Hilden, Germany) and PCR products were cloned into pJET1.2/blunt Cloning Vector (Fermentas, Waltham, MA, USA) and sequenced with an ABI system according to the instructions of the manufacturer. Assembly of gene sequences into an operon and annotation of genes was performed with the CLC Main Workbench and checked for chimera using the Bellerophon algorithm (7).

RESULTS

[0092] Identification of cellulases produced a thermophilic cellulolytic consortium

[0093] Supernatant was obtained from a 300 L culture of a community grown in microcrystalline cellulose. The supernatant from this community high endo-glucanase and – xylanase activity. The supernatant was fractionated on an anion exchange column with an NaCl gradient. The majority of the glycoside hydrolase activity eluted in the salt gradient at <50 mM NaCl (FIG. 1).

[0094] In a second enrichment strategy, the cellulose-binding proteins were isolated by incubated the supernatant with amorphous cellulose; this step is referred to as affinity digestion (AD). In a subsequent step, the cellulose fraction was heated to 55 °C, causing the cellulose to be digested. The supernatant after affinity digestion was separated using an anion-exchange column, producing a fraction enriched in complex identified by native gel that had an observed molecular weight of 350 kDa that stained positive for endoglucanase activity. Native PAGE in combination with zymograms showed a 100 kDa element in the anion-exchange fraction with high xylanase activity, which was missing in the AD fractions (FIG. 2).

[0095] The 350 kDa band was excised and trypsin digested peptides analyzed against a database containing metagenomic data obtained from DNA isolated from the 300 L culture. Liquid chromatography–mass spectrometry (LC-MS) analysis of this complex identified unambiguously that the band consisted of three glycoside hydrolase (GH) proteins from low abundance member of the Firmicutes present at 0.6% abundance (FIG. 3). Partial genes were recovered from the metagenome that indicated that the three proteins were a GH9, GH48 and GH6/5.

[0096] PCR-based extensions of the original genes indicated that the three genes coding for

these cellulases were arranged in a putative operon with two GH10 proteins and on AA10 protein (FIG. 4).

[0097] A striking feature of the proteins in this operon was that each glycoside hydrolase contained at least one cellulose-binding domain (CBM), all in the CBM3 family. The GH9 contained three CBM3 domains and the GH48 and GH6/5 proteins contained one CBM3 each. No cohesion or dockerin domains were detected, leading to the hypothesis that the proteins in the complex were bound through the action of the multiple CBM domains. This structure represents a new motif in cellulase complex formation.

[0098] References cited:

1. C. Li, B. Knierim, C. Manisseri, R. Arora, H. V. Scheller, M. Auer et al., Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification, *Bioresource technology*. 101, 4900–4906 (2010).
2. J. M. Gladden, M. Allgaier, C. S. Miller, T. C. Hazen, J. S. VanderGheynst, P. Hugenholtz et al., Glycoside hydrolase activities of thermophilic bacterial consortia adapted to switchgrass, *Applied and environmental microbiology*. 77, 5804–5812 (2011).
3. E. Morag, E. A. Bayer, R. Lamed, Affinity digestion for the near-total recovery of purified cellulosome from *Clostridium thermocellum*, *Enzyme and Microbial Technology*. 14, 289–292 (1992).
4. U. K. Laemmli, Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4, *Nature*. 227, 680–685 (1970).
5. Wolfgang H. Schwarz, Karin Bronnenmeier, Folke Gräbnitz, Walter L. Staudenbauer, Activity staining of cellulases in polyacrylamide gels containing mixed linkage β -glucans, *Analytical Biochemistry*. 164, 72–77 (1987).

6. Park JI, Steen EJ, Burd H, Evans SS, Redding-Johnson AM, Batth T, Benke PI, D'Haeseleer P, Sun N, Sale KL, et al: A Thermophilic Ionic Liquid-Tolerant Cellulase Cocktail for the Production of Cellulosic Biofuels. PLoS ONE 7, e37010 (2012).
7. T. Huber, G. Faulkner, P. Hugenholtz, Bellerophon: a program to detect chimeric sequences in multiple sequence alignments, Bioinformatics (Oxford, England). 20, 2317–2319 (2004).

[0099] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

What is claimed is:

1. A composition comprising a purified or isolated cellulase complex comprising two or more glycosidase hydrolase, or enzymatically active fragment thereof, selected from the group consisting of a GH9 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:1, a GH48 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:2, a GH10 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:3, and a GH6 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:4; and optionally the composition or the cellulase complex comprising a GH10_2 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:5, and/or an AA10 polypeptide comprising an amino acid sequence at least 70% identical to SEQ ID NO:6..
2. A composition comprising the composition of claim 1, and an ionic liquid (IL).
3. A recombinant or isolated or purified nucleic acid encoding the cellulase complex, or the GH9 polypeptide, the GH48 polypeptide, the GH10 polypeptide, and/or the GH9 polypeptide of claim 1.
4. A vector comprising the recombinant or isolated or purified nucleic acid of claim 3.
5. A host cell comprising the nucleic acid of claim 3 or the vector of claim 4.
6. A method for producing a cellulase complex of the present invention comprising: providing a host cell of claim 5, culturing the host cell in a culture medium under conditions whereby the cellulase complex is produced, optionally isolating the cellulase complex from the host cell and/or the culture medium, and optionally contacting the cellulase complex and a cellulose, whereby the cellulose is hydrolyzed by the cellulase complex.
7. The method of claim 6, wherein the providing step comprises: introducing an expression vector capable of expressing the cellulase complex in the host cell into the host cell, and optionally constructing the expression vector encoding a promoter operatively linked to a nucleic acid encoding the cellulase complex, wherein the constructing step precedes the introducing step.

8. A method of hydrolyzing a cellulose, comprising: (a) providing a solution comprising an IL, a cellulose, and the composition of claim 1, and (b) incubating the solution, such that the cellulose is hydrolyzed by the cellulase complex.

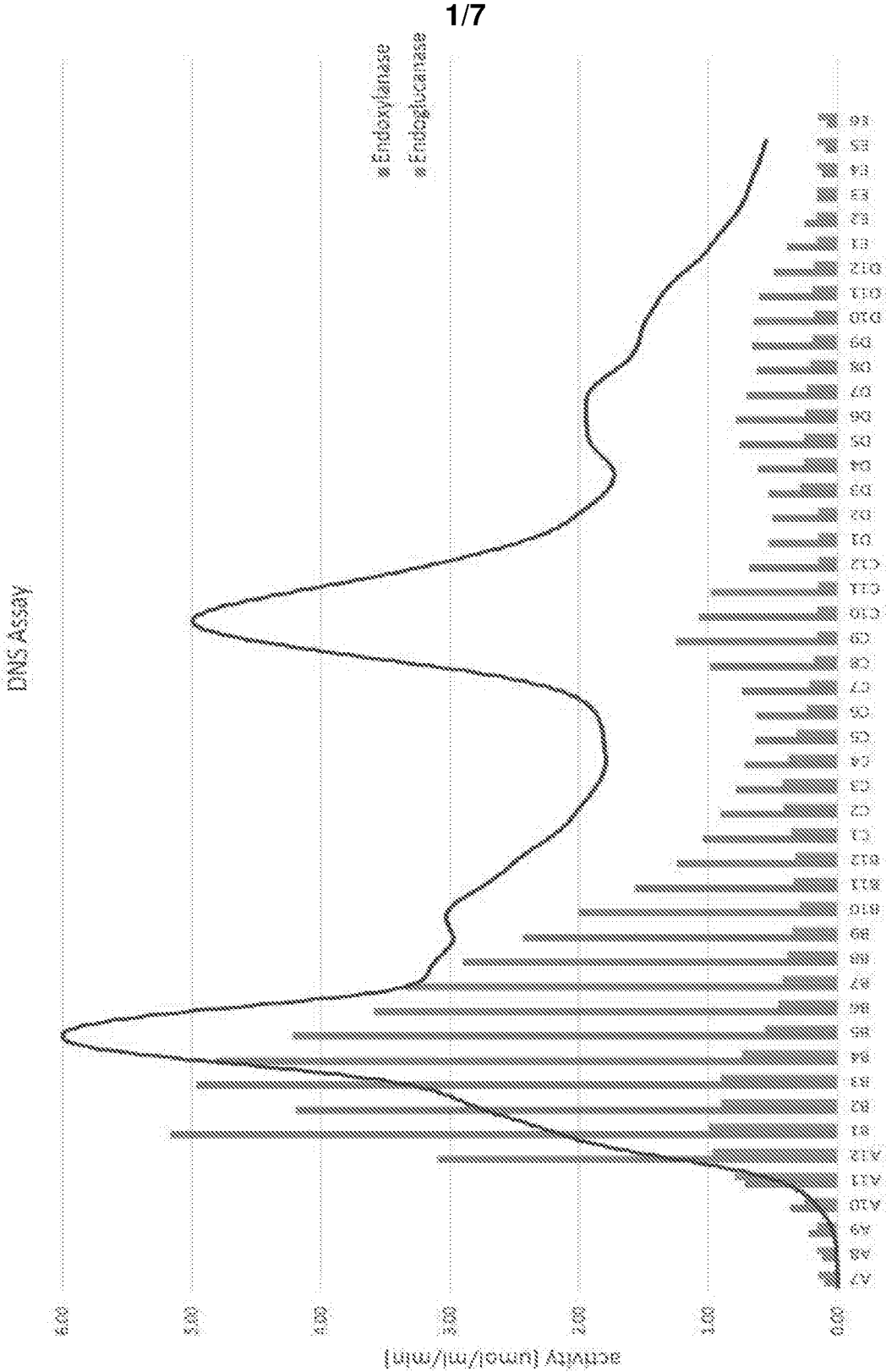


Figure 1

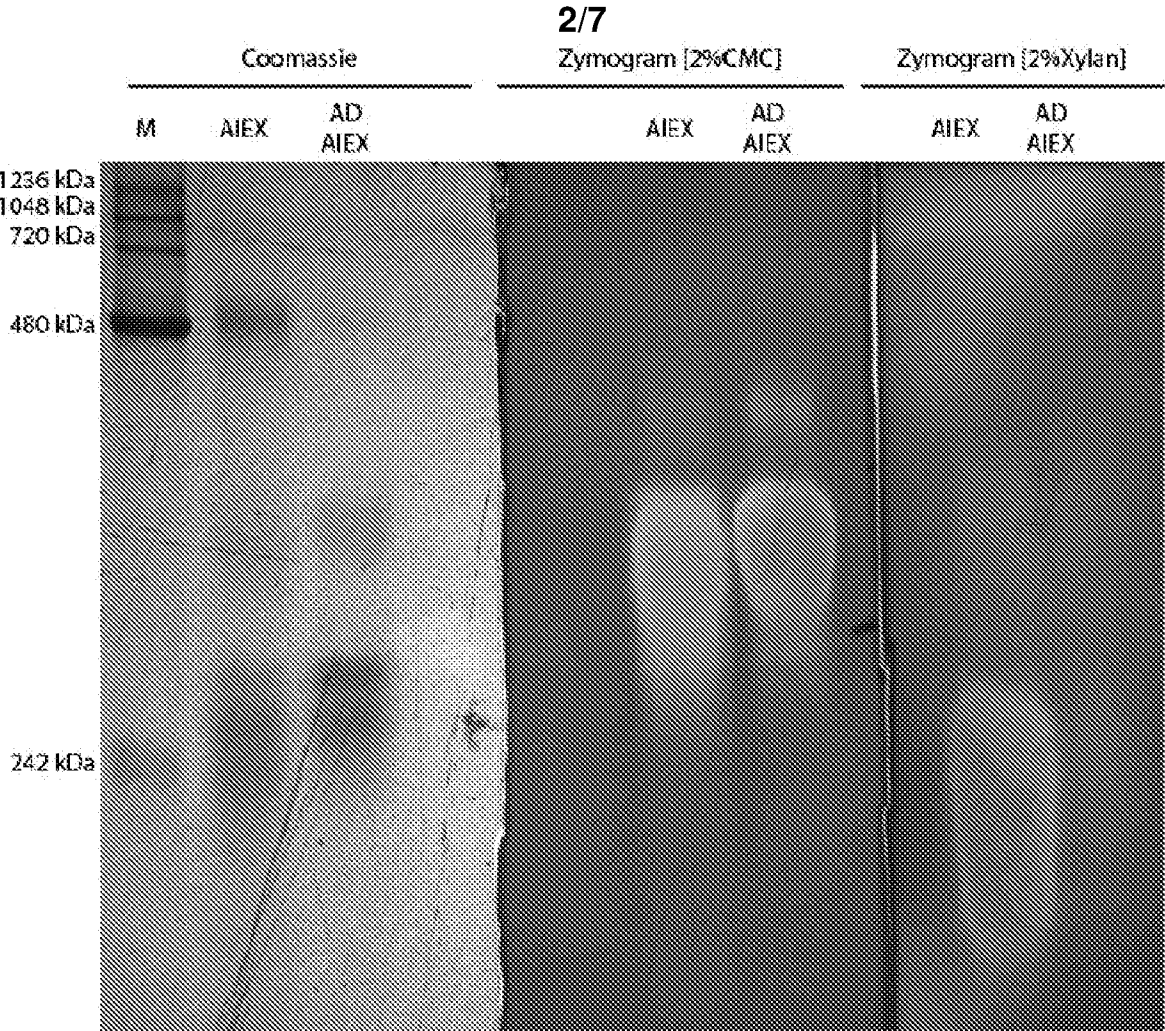


FIG. 2

3/7

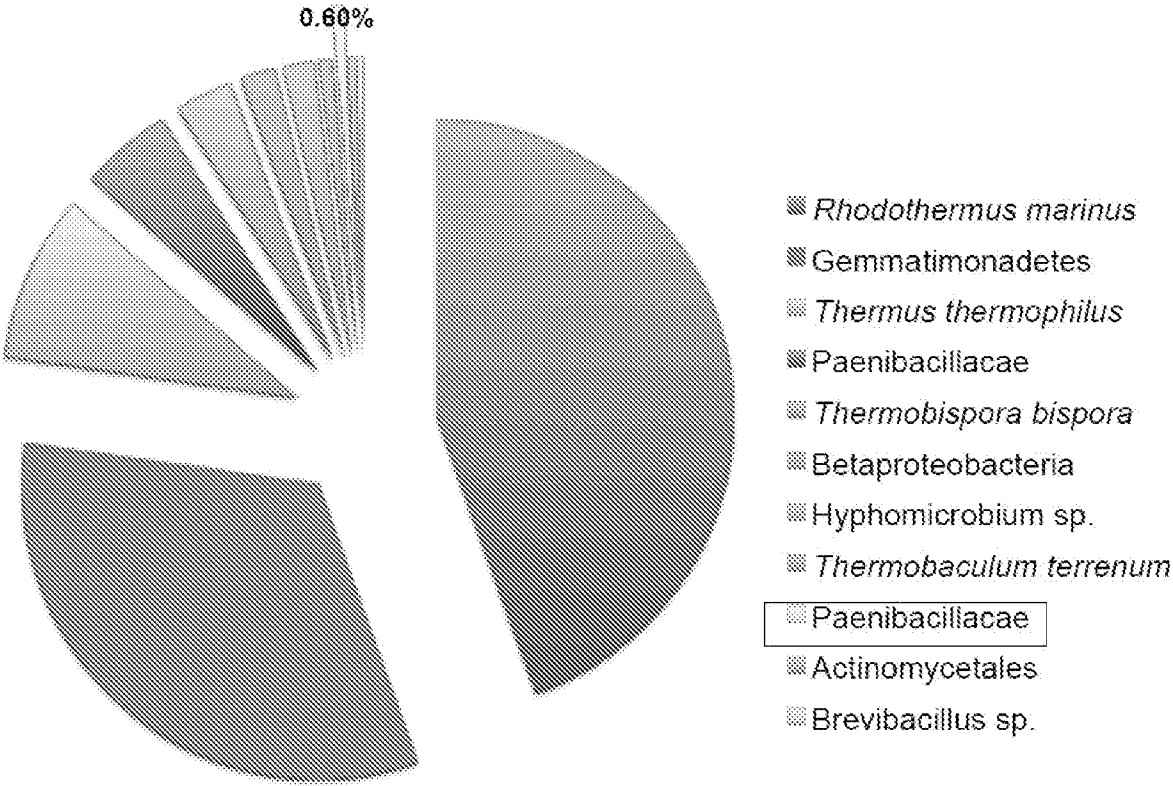


FIG. 3

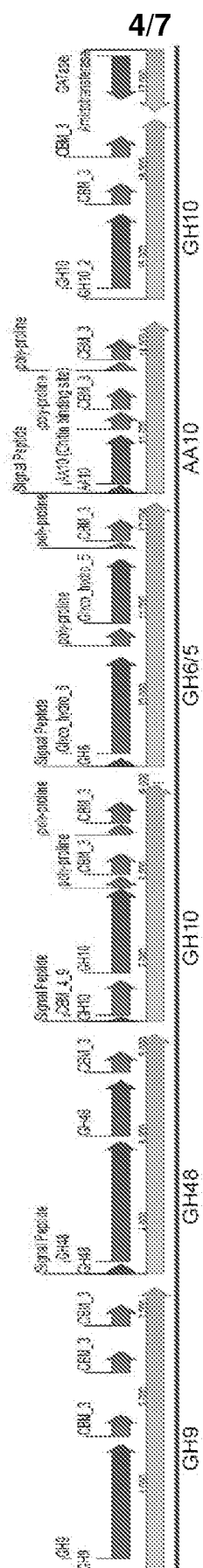


Figure 4

5/7

58 KAIYFYDAQRSGLPPDNRVEWRGDSGLNDGADVGVDLTGGWYDAGDHVKFGLPMAYSAA
 KAI FY+ Q SGKLP R WRGDS L DG D G+DLTGGW+DAGDHVKF LPM+Y+
 43 KAIMFYEFQMSGKLPSWVRNNWRGDSALKDGQDNGLDLTGGWFDAGDHVKFNLPMSTGT

 118 MLAWAVYEYRDAFVQTGQLDYILNNIKWATDYFIKAHSAPNVLWGQVGKGDVDHAWWGPA
 ML+WAVYEY+DAFV++GQL++ILN I+W DYF+K H + V + QVG G DHAWWGPA
 103 MLSWAVYEYKDAFVKSGQLEHILNQIEWVNDYFVKCHPSKYVYYYQVGDGSKDHAWWGPA

 178 EVMQMPRPAYKIDPSCPGSDLAAGTAAAMAAAAAVFKPTDPTYASTLIAHAKQLYTFADT
 EVMQM RP++K+ S PGS + A TAA++AAA+ V K +PT A+T + HAK LY FA+
 163 EVMQMERPSFKVTQSSPGSTVVAETAASLAAASIVLKDRNPTKAATYLQHAKDLYEFAEV

 238 YRGKYSDCITDAQNFYRSWSGYADELTWGA VWLYLATGEQAYLDKAIASVAEWGREGQTP
 K T A +Y SWSG+ DEL+W AVWLYLAT + YL KA + V W + +
 223 --TKSDSGYTAANGYNSWSGIFYDELSSAAVWLYLATNDSTYLTKAESYVQNWPKISGSN

 298 YWGYKWTQSWDDVHYGAQLLLARITGDQRFIQSTERNLEYWTDGTDDTGERITYTPGGLA
 YKW WDDVH GA LLLA+ITG + Q E +L+YWT G + GERI YTP GLA
 281 TIDYKWAHCWDDVHNGAALLAKITGKDTYKQIEMHLDYWTTGYN--GERIKYTPKGGLA

 358 WLD SWGSLRYAMNASFLAFVYSDWLQSRDPAKAEKYRNFAVRQVLYALGDNPRNSSYVVG
 WLD WGSRLRYA +FLAFVYSDW K E YR F Q+ YALG R S+VVG
 339 WLDQWGSRLRYATTTAFLAFVYSDWAGC-PSTKKETYRKFGESQIDYALGSAGR--SFVVG

 418 FGRNPPQRPHHRTAHGSWADSQNPAYHRHILYGA 452
 FG NPP+RPHHRTAH SWADSQ++P+YHRH LYGA
 396 FGTPNPPKRPHHRTAHSSWADSQSIPSYHRHTLYGA 430

FIG. 5

6/7

48 FKDRFLQLYNQIKNPANGYFSPEGIPYHSIETLISEAPDYGHMTTSEAFSYWLWLETLYG
 + RF+ L+N+I +PANGYF+ +GIPYHS+ETLI EAPDYGH+TTSEAFSY++WLE +YG
 1128YGQRFMWLWNKIHDHPANGYFNQDGIPYHSVETLICEAPDYGHLLTSEAFSYVWLEAVYG

 108 YFTGDWSKLEQAWTKMEQFIIPNSTEQPTMGSYNPSSPATYAPEHPYPDRYPTLLNNSVP
 TGDWSK + AW +E+++IP++ +QP + SY+P+ PATYA E PD+YP+ L +VP
 1188KLTGDWSKFKTAWDTLEKYMIPSAEDQP-IRSYDPNKPATYAGEWETPDKYPSPLEFNVP

 168 AGQDPLDAELKATYGNNVTYLMHWLLDVDNWWYGFNLLNPSHTATYVNTFQRGEQESVWE
 G+DPL EL +TYG+ + Y MHWL+DVDNWWYG+G + A+++NTFQRG +ESVWE
 1247VGKDPLHNELVSTYGSTLMYGMHWLMDVDNWWYGYGKRGDGVSRSASFINTFQRGPEESVWE

 228 AITHPSQDNFRFGKPNEGFVTLFVKDNGTPAQQWRYTAASDADARAIQVMYWAK-----Q
 + HPS + F++G PN GF+ LF+KD ++QWRYT A DADARAIQ YWAK Q
 1307TVPHPSWEEFKWGGPN-GFLDLFIKDQNY-SKQWRYTDAPDADARAIQATYWAKVWAKEQ

 283 LGYNN-QTYLDKARKMGDYLRYTLFDKYFQQIGSANDGSPSPGSGKNSAHYLLSWYTAWG
 +N +Y+ KA KMGDYLRY +FDKYF+ +G D + + G+G +SAHYLLSWY AWG
 1365GKFNEISSYVAKAAKMGDYLRYAMFDKYFKPLG-CQDKNAAGGTGYDSAHYLLSWYYAWG

 342 GGLGSGGNWAWRIGSSHAHQGYQNPVAAAYALS-AGGLAPRSATAQTDWATSLQRQLEFYT
 G L G W+W+IG SHAH GYQNP+AA+AL+ + P+S T DWA SL+RQ+EFY
 1424GALD--GAWSWKIGCSHAHFGYQNPMAAWALANDSDMKPKSPTGANDWAKSLKRQIEFYR

 401 WLQSSEGAIGGGATNSVGGSYQPYPSSGRSTFYGMVYDEAPVYRDPPSNSWFGFQAWSVER
 WLQS+EGAI GGATNS G Y+ YP+G +TFYGM Y+ PVYRDP SN+WFGFQAWS++R
 1482WLQSAEGAIAGGATNSWNGRYEKYPAGTATFYGMAYEPNPVYRDPGSNTWFGFQAWSMQR

 461 VAELYIWTSSGNTNTQQFQMVKNIVTKWVDWALDYTFVNQRPVTDAAQGYFLTSSGSRVL
 VAE YY+ +G+ + ++ KWV W
 1542VAEYYYV---TGDKDAGA-----LLEKWVSWI-----

 521 GGNNPQIATVSDPGQFYIPSTLEWQGQPDWNGYANYTGNNPNFHAIKDPGQDVGVTGNY
 + ++ G F IPSTL+W GQPDW G YTGNNPN H D G D+G+T +
 1566---KSVVKLNSDGTFAIPSTLDWSGQPDWKG--TYTGNNPNLHVKVVDYGTDLGITASL

 581 IKLLTFFAAATKAETGNYTALGSQALNVAEQLLNVLWN-FNDGVGIVRPEQRADYFRYFT
 L +++A TK Y +A N+A++LL+ +W + D G+ PE+RADY R+F
 1620ANALLYYSAGTK---KYGVFDEEAKNLAKELLD RMWKLYRDNKGLSAPEKRADYKRFFE

 640 KEIYFPGSGWSTYGGNTIPGPGAVPSDPSKGGNGVYISYAE LRPKIKQDPKWSYLENLY
 +E+Y P+GW+G G+ I K G + + ++R K QDP W LE Y
 1676QEVYIPAGWTGKMPNGDVI-----KSG---VKFIDIRSKYKQDPDWPKLEAAY

 700 KTSYNPSTGRWENGVPFTFYHRFWAQVDVATAYAEFARLIG 740
 K+ VP F YHRFWAQ D+A A A + L G
 1721KSG-----QVPEFRYHRFWAQCDIAIANATYEILFG 1751

FIG. 6

7/7

20 MVAALLVF--TVSVVPPKEADAGLAKTKFLGNVINNSIPSDFAVYWNQVTPENATKWGSV
 + A+LL+ S + +ADAGLA++KFLGNVI +S+PS+FA YWNQVTPEN+TKWGSV
 9 ICASLLLLGSVFSFIATSDADAGLARSKFLGNVIASSVPSNFATYWNQVTPENSTKWGSV

78 ESSRDNMNWSTRADMIYNYARSNGFPFKFHTLVWGSQEPGWISGLSAAEQQAEEVIEWIQAA
 E++R+ MNWS AD+ YNYA+SNGFPFKFHTLVWGSQ+PGWISGLS AEQ+AEV++WIIQAA
 69 EATRNVMNWSAADLAYNYAKSNGFPFKFHTLVWGSQQPGWISGLSQAEQKAEVLQWIIQAA

138 GQRYPDADFVDVNEPLHAKPSYRNAIGGDGSTGWDWVIWSFEQARRAFPNSKLLINEYG
 GQRYP+ADFVDVNEPLHAKPSYRNAIGGDG+TGWDWVIWSF++AR+AFP+KLLINEYG
 129 GQRYPNADFVDVNEPLHAKPSYRNAIGGDGATGWDWVIWSFQEARAKAFNAKLLINEYG

198 VENDPNAASQYVQIINLLKSRGLIDGIGIQGHYFNLDTVSVSTLRITLGLAETGLPIYV
 + +DPNAA+QYVQIINLLKSRGLIDGIGIQ HYFN+D+VSVST+ + L LA TGLPIYV
 189 IISDPNAANQYVQIINLLKSRGLIDGIGIQCHYFNMDSVSVSTMNSVLNTLAATGLPIYV

258 SELDISGDDATQLARYQEKFPILWEHPSVQGITLWGYIEGQWRSGLITASGVERPAL
 SELD++GDD+TQLARYQ+KFP+LWEH +V+G+TLWGYIEGQW S THL+ ++G ERPAL
 249 SELDMTGDDSTQLARYQQKFPVLWEHSAVKGVTLWGYIEGQTWASNTHLVRNNGTERPAL

318 QWLRTYLA 325
 QWLRTYL+
 309 QWLRTYLS 316

FIG. 7

1 MNRRLIARLSGMLAMVLIAAVLAYVPKPEPAEAHGGMVFPATRTYACYVDGKVHGNNGDDL
 MNRRLIARLSGMLAMVLIAA+LAYVPKPEPAEAHGGMVFPATRTYACYVDGKVHGNNGDDL
 1 MNRRLIARLSGMLAMVLIAAMLAYVPKPEPAEAHGGMVFPATRTYACYVDGKVHGNNGDDL

61 NMNIPACLDALAISGNYQFNNWFGNLISNAGGRHREIIPDGKLCGPTASFDGMNQARTDW
 NMNIPACLDALAISGNYQFNNWFGNLISNAGGRHREIIPDGKLCGPTASFDGMNQARTDW
 61 NMNIPACLDALAISGNYQFNNWFGNLISNAGGRHREIIPDGKLCGPTASFDGMNQARTDW

121 WTTRLQPGATITVRVNAWAPHPGTWYLYVTRDGWDPTQPLKWSLEPTPFSQVTNPPINS
 WTTRLQPGATITVRVNAWAPHPGTWYLYVTRDGWDPTQPLKWSLEPTPFSQVTNPPINS
 121 WTTRLQPGATITVRVNAWAPHPGTWYLYVTRDGWDPTQPLKWSLEPTPFSQVTNPPINS

181 SGPDGAEYSWQVQLPNKQGRHIIYMIWQRSDSPEAFYNCSDAYFGSGPIAYEFGDPREGG
 SGPDGAEYSWQVQLPNKQGRHIIYMIWQRSDSPEAFYNCSD YFGSGPIAYEFGDPREGG
 181 SGPDGAEYSWQVQLPNKQGRHIIYMIWQRSDSPEAFYNCSDVYFGSGPIAYEFGDPREGG

241 T
 T
 241 T

FIG. 8

PATENT COOPERATION TREATY

PCT

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT (PCT Article 17(2)(a), Rules 13ter.1(c) and (d) and 39)

Applicant's or agent's file reference 2015-010-02PCT	IMPORTANT DECLARATION	Date of mailing (day/month/year) 06 APR 2017
International application No. PCT/US16/63198	International filing date (day/month/year) 21 November 2016	(Earliest) Priority Date (day/month/year) 19 November 2015
International Patent Classification (IPC) or both national classification and IPC IPC: C12Q 1/68; CPC: C12Q 2600/158		
Applicant THE REGENTS OF THE UNIVERSITY OF CALIFORNIA		

This International Searching Authority hereby declares, according to Article 17(2)(a), that no international search report will be established on the international application for the reasons indicated below.

1. ☐ The subject matter of the international application relates to:
 - a. ☐ scientific theories
 - b. ☐ mathematical theories
 - c. ☐ plant varieties
 - d. ☐ animal varieties
 - e. ☐ essentially biological processes for the production of plants and animals, other than microbiological processes and the products of such processes
 - f. ☐ schemes, rules or methods of doing business
 - g. ☐ schemes, rules or methods of performing purely mental acts
 - h. ☐ schemes, rules or methods of playing games
 - i. ☐ methods for treatment of the human body by surgery or therapy
 - j. ☐ methods for treatment of the animal body by surgery or therapy
 - k. ☐ diagnostic methods practised on the human or animal body
 - l. ☐ mere presentations of information
 - m. ☐ computer programs for which this International Searching Authority is not equipped to search prior art
2. ☐ The failure of the following parts of the international application to comply with prescribed requirements prevents a meaningful search from being carried out:

☐ the description
 ☐ the claims
 ☐ the drawings
3. ☒ A meaningful search could not be carried out without the sequence listing; the applicant did not, within the prescribed time limit:

☒ furnish a sequence listing in the form of an Annex C/ST.25 text file, and such listing was not available to the International Searching Authority in a form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ furnish a sequence listing on paper or in the form of an image file complying with the standard provided for in Annex C of the Administrative Instructions, and such listing was not available to the International Searching Authority in a form and manner acceptable to it; or the sequence listing furnished did not comply with the standard provided for in Annex C of the Administrative Instructions.

☐ pay the required late furnishing fee for the furnishing of a sequence listing in response to an invitation under Rule 13ter.1(a) or (b).
4. Further comments:

Applicant failed to submit a valid electronic seq. listing in response to the ISA/225.

Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774
---	---