



(43) International Publication Date
18 October 2012 (18.10.2012)

(51) International Patent Classification:

C12N 9/24 (2006.01) C12N 11/00 (2006.01)
C07K 14/195 (2006.01) C12N 9/42 (2006.01)
C12N 15/70 (2006.01)

(21) International Application Number:

PCT/US2012/033138

(22) International Filing Date:

11 April 2012 (11.04.2012)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/474,633 12 April 2011 (12.04.2011) US

(71) Applicant (for all designated States except US):

ALLOPARTIS BIOTECHNOLOGIES, INC. [US/US]; 409 Illinois Street, San Francisco, California 94158 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only):

COHEN, Richard N. [US/US]; 536 63rd Street, Oakland, California 94609 (US). KOPPEL, Nitzan [US/US]; 501 Hope Terrace #1, Sunnyvale, California 94087 (US). TORIELLO, Nicholas

M. [US/US]; 650 Glencrag Way, Woodside, California 94062 (US). BLAZEJ, Robert G. [US/US]; 29 Coleridge Street, San Francisco, California 94110 (US). EMRICH, Charles A. [US/US]; 1335B South Van Ness, San Francisco, California 94110 (US).

(74) Agents: HALIDAY, Emily, M. et al.; WEAVER AUSTIN VILLENEUVE & SAMPSON LLP P.O. Box 70250, Oakland, California 94612-0250 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ,

[Continued on next page]

(54) Title: IMPROVED CELLULOSE VARIANTS

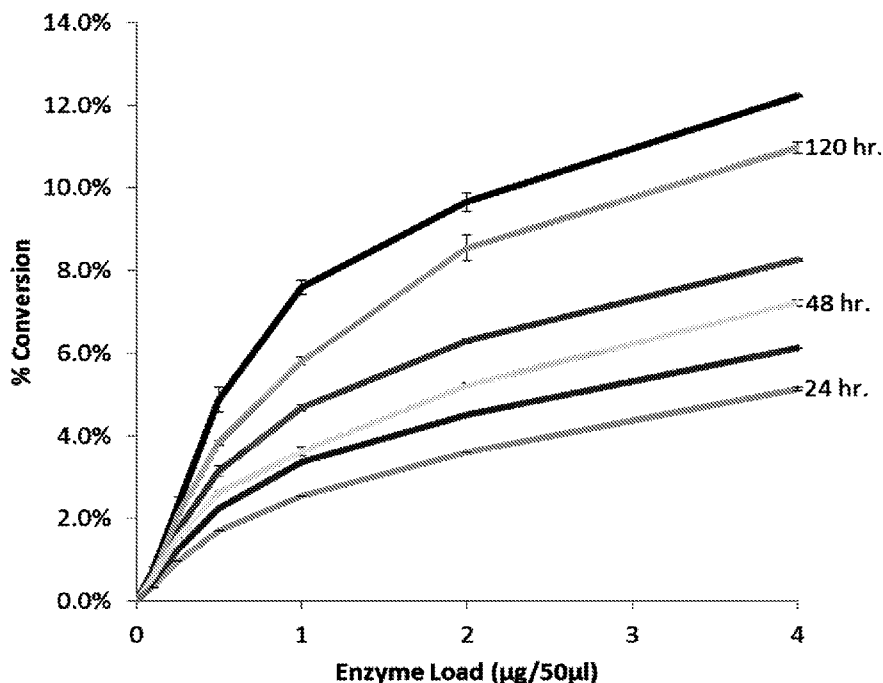


Fig. 3

(57) Abstract: This invention provides novel variant cellulolytic enzymes having improved activity and/or stability. In certain embodiments the variant cellulolytic enzymes comprise a glycoside hydrolase with or comprising a substitution at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A enzyme. In certain embodiments the glycoside hydrolase is a variant of a family 9 glycoside hydrolase. In certain embodiments the glycoside hydrolase is a variant of a theme B family 9 glycoside hydrolase.

WO 2012/142171 A2

UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

IMPROVED CELLULASE VARIANTS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of and priority to USSN 61/474,633, filed on April 12, 2011, which is incorporated herein by reference in its entirety for all purposes.

5 STATEMENT OF GOVERNMENTAL SUPPORT

[0002] This work was supported in part by Grant No DE-SC0002302 from the Department of Energy. The Government has certain rights in this invention.

FIELD OF THE INVENTION

[0003] The present invention relates to cellulolytic enzymes useful in the processing
10 of plant biomass and other cellulosic materials.

BACKGROUND OF THE INVENTION

[0004] Cellulases or cellulolytic enzymes are enzymes involved in hydrolysis of cellulose. Typically there are three major types of cellulase enzymes involved in the hydrolysis of native cellulose, namely cellobiohydrolase (1,4-beta-D-glucan
15 cellobiohydrolase, EC 3.2.1.91), endo-beta-1,4-glucanase (endo-1,4-beta-D-glucan 4-glucanohydrolase, EC 3.2.1.4) and beta-glucosidase (EC 3.2.1.21).

[0005] The endo-beta-1,4-glucanases (EC No. 3.2.1.4) constitute a group of hydrolases of particular interest in various industrial applications. Endoglucanases catalyse endo hydrolysis of 1,4-beta-D-glycosidic linkages in cellulose, cellulose derivatives (such
20 as carboxy methyl cellulose and hydroxy ethyl cellulose), lichenin, beta-1,4 bonds in mixed beta-1,3 glucans such as cereal beta-D-glucans or xyloglucans, and other plant material containing cellulosic components

[0006] Cellulases are synthesized by a large number of microorganisms that include, for example, fungi, actinomycetes, myxobacteria and true bacteria, and certain plants.
25 Endoglucanases of a wide variety of specificities have been identified.

[0007] One industrial use of cellulolytic enzymes is for treatment of cellulosic textiles or fabrics, *e.g.*, as ingredients in detergent compositions or fabric softener compositions, for bio-polishing of new fabrics (garment finishing), and for obtaining a "stone-washed" look of cellulose-containing fabrics, especially denim. Another important

industrial use of cellulolytic enzymes is the use for treatment of paper pulp, *e.g.*, for improving the drainage or for deinking of recycled paper.

[0008] Bioconversion of renewable lignocellulosic biomass to a fermentable sugar that is subsequently fermented to produce alcohol (*e.g.*, ethanol) or other fuels as an
5 alternative to petroleum-based fuels is also an important use of cellulosic enzymes.

Fermentable sugars are also used to produce plastics, polymers and other bio-based products and this industry is expected to grow substantially increasing the demand for abundant low cost fermentable sugars that can be used as a feed stock in lieu of petroleum based
10 feedstocks. Cellulosic biomass is the most abundant renewable natural resource. Generated at a rate of ~100 billion dry tons/year by the biosphere, cellulosic biomass has the potential to replace the world's demand for diminishing fossil fuels.

[0009] The major polysaccharides comprising different lignocellulosic residues, that may be considered as a potential renewable feedstock, include, for example, cellulose and hemicelluloses (xylans). The enzymatic hydrolysis of these polysaccharides to soluble
15 sugars, for example glucose, xylose, arabinose, galactose, mannose, and other hexoses and pentoses can occur 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 cellooligosaccharides (cellobiose as a main product), while β -glucosidases (BGL) convert the oligosaccharides to glucose. Xylanases together with other accessory
20 enzymes (non-limiting examples of which include α -L-arabinofuranosidases, feruloyl and acetylxylan esterases, glucuronidases, and (β -xylosidases) catalyze the hydrolysis of hemicelluloses.

[0010] Regardless of the type of cellulosic feedstock, the cost and hydrolytic efficiency of enzymes are major factors that restrict the commercialization of the biomass
25 bioconversion processes. According to Zhang, Y. H. P., "[o]ne of the most important and difficult technological challenges is to overcome the recalcitrance of natural lignocellulosic materials, which must be enzymatically hydrolyzed to produce fermentable sugars" (*see, Zhang et al. (2006) Biotechnol. Adv. 24: 452-481*). A major limitation for the conversion of biomass to biofuel and renewable chemicals is the high cost and large quantities of
30 enzymes required for hydrolysis.

SUMMARY

[0011] In various embodiments novel variant cellulolytic enzymes are provided having improved activity and/or stability. In certain embodiments the variant cellulolytic enzymes comprise a glycoside hydrolase consisting of or comprising a substitution at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A enzyme. In certain embodiments the enzyme comprises no more than 5, additional variations, or no more than 4 additional variations, or no more than three additional variations, or no more than two additional variations, or no more than one additional variation at positions other than those corresponding to F64, A226, and/or E246 in *Thermobifida fusca* Cel9A. In certain embodiments the glycoside hydrolase is a variant of a family 9 glycoside hydrolase. In certain embodiments the glycoside hydrolase is a variant of a theme B glycoside hydrolase. In certain embodiments the variant comprises a substitution at a position corresponding to F64 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at a position corresponding to A226 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at a position corresponding to E246 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to F64 and A226 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to F64 and E246 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to A226 and E246 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to F64, A226, and E246 in *Thermobifida fusca* Cel9A. In certain embodiments the substitution at a position corresponding position F64 in *Thermobifida fusca* Cel9A is a substitution to I, V, L, or M. In certain embodiments the substitution at a position corresponding position F64 in *Thermobifida fusca* Cel9A is a substitution to I. In certain embodiments the substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V, I, L, M, or T. In certain embodiments the substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V, I, L, or M. In certain embodiments the substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V. In certain embodiments the substitution at a position corresponding position E246 in *Thermobifida fusca* Cel9A is a substitution to G, A, N, or S. In certain embodiments the substitution at a position corresponding position E246 in *Thermobifida fusca* Cel9A is a substitution to G. In certain embodiments the enzyme is a variant of a glycoside hydrolase

from an organism selected from the group consisting of archaea, bacteria, and eukaryota. In certain embodiments the enzyme is a variant of a glycoside hydrolase from a gram negative bacterium. In certain embodiments the enzyme is a variant of a glycoside hydrolase from a gram positive bacterium. In certain embodiments the enzyme is a variant of glycoside
5 hydrolase from a bacterial family selected from the group consisting of *Thermofidia*, *Micromonospora*, *Cellulomonas*, *Listeria*, *Pseudomonas*, *Ruminococcus*, *Saccharophagus*, *Streptomyces*, *Vibrio*, *Xanthomonas*, and *Clostridium*. In certain embodiments the enzyme is a variant of *Thermobifida fusca* Cel9A. In certain embodiments the enzyme is a variant of *Clostridium phytofermentans* Cphy3367.

10 [0012] Variant cellulolytic enzymes, as contemplated herein, do not include any wild-type and/or naturally occurring enzymes. In certain embodiments the enzyme has at least 50%, or at least 60%, or at least 65%, or at least 70%, or at least 75%, or at least 80%, or at least 85%, or at least 86%, or at least 87%, or at least 88%, or at least 89%, or at least
15 90%, , or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity with F1, and/or Cel9A, and/or Cphy3367, and/or G03, and/or G11. In certain embodiments the enzyme is attached to or operably linked to a signal peptide. In certain embodiments the cellulolytic enzyme is present and/or displayed on the surface of an organism selected from the group consisting of a phage, a fungus, an alga, and a bacterium. In certain embodiments the
20 cellulolytic enzyme is present and/or displayed on a substrate or the surface of a particle (e.g., a particle comprising a material selected from the group consisting of a plastic, a glass, a mineral, a synthetic polymer, a biological polymer, and a metal). In certain embodiments the substrate or particle comprises a surface of a microfluidic channel or chamber, a surface of a bioreactor, or a surface of a bioreactor. In certain embodiments the
25 cellulolytic enzyme is a component of a cellulosome or a minicellulosome. In certain embodiments the cellulosome is on the surface of a yeast, bacteria, or non-yeast fungus. In certain embodiments cellulosome comprises one or more additional enzymes selected from the group consisting of an endocellulase, an exocellulase, a beta-glucosidase (cellobiase), an oxidative cellulose, a xylanase, a hemicellulase, a lichenase, a chitenase, a xylanase, a
30 cellulose phosphorylase, and a cellulose disrupting protein.

[0013] In various embodiments nucleic acid encoding the variant enzymes, vectors comprising the nucleic acids and host cells transfected to produce the variant enzymes are also provided. In certain embodiments the nucleic acid comprises codons optimized for expression in a host cell (e.g., optimized for expression in a host selected from the group

consisting of a bacterium, a yeast, a fungus, an alga, an insect, and a mammalian cell). In certain embodiments the nucleic acid comprises codons optimized for expression in *E. coli*. In certain embodiments the host cell is selected from the group consisting of an archiobacterium, a bacterium, a yeast cell, a fungal cell, an algal cell, a plant cell, an insect cell, and a mammalian cell.

[0014] In various embodiments a method of producing a cellulase is provided where the method comprises a) culturing a host cell comprising a nucleic acid that encodes a cellulase as described herein in a suitable culture medium under suitable conditions to produce cellulase; and obtaining the produced cellulase.

10 [0015] A method of producing an enzyme variant is provided that comprises introducing a substitution in the amino acid sequence of an endoglucanase at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A.

[0016] Also provided are methods of degrading cellulosic biomass into fermentable sugars. The methods typically involve contacting the cellulosic biomass with a cellulolytic enzyme variant as described herein under conditions in which the enzyme partially or fully degrades cellulose in said cellulosic biomass to form one or more fermentable sugars. In certain embodiments the cellulosic biomass comprises one or more materials selected from the group consisting of an agricultural plant waste (e.g., corn stover, cereal straw, sugarcane bagasse), a plant waste from an industrial processes (e.g., sawdust, paper pulp), an non-food energy crop (e.g., switchgrass). In certain embodiments the cellulosic biomass comprises one or more materials selected from the group consisting of grasses, rice hulls, bagasse, jute, hemp, flax, bamboo, sisal, abaca, straw, corn cobs, corn stover, alfalfa, hay, coconut hair, seaweed, and algae.

25 [0017] In certain embodiments a cellulolytic compositions is provided where the composition comprises a plurality of cellulolytic enzymes at least one of which is an enzyme variant as described herein. In certain embodiments the composition comprise a second enzyme selected from the group consisting of an endocellulase, an exocellulase, a beta-glucosidase (cellobiase), an oxidative cellulose, a xylanase, a hemicellulase, a lichenase, a chitenase, a xylanase, and a cellulose phosphorylase.

[0018] Also provided is a detergent composition comprising a cellulolytic enzyme variant as described herein, and a surfactant. In certain embodiments the detergent comprises a laundry detergent, a dish detergent, or an industrial detergent.

[0019] Uses of the variant cellulolytic enzymes described herein are also provided. Such uses include, but are not limited to methods of processing textiles, or paper, and methods of processing cellulosic stock materials (*e.g.*, lignocellulolytic materials) to produce sugars, alcohols, and/or other compounds, and the like. In certain embodiments
5 such uses include, but are not limited to the treatment of a cellulose containing textile, stone washing or dyeing of a fabric, the treatment of wood pulp or other lignocellulosic material, and the reduction of biomass to glucose.

DEFINITIONS

[0020] The terms "polypeptide", "peptide" and "protein" are used interchangeably
10 herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. The term also includes variants on the traditional peptide linkage joining the amino acids making up the polypeptide. Preferred "peptides", "polypeptides", and
15 "proteins" are chains of amino acids whose alpha carbons are linked through peptide bonds. The terminal amino acid at one end of the chain (amino terminal) therefore has a free amino group, while the terminal amino acid at the other end of the chain (carboxy terminal) has a free carboxyl group. As used herein, the term "amino terminus" (abbreviated N-terminus) refers to the free α -amino group on an amino acid at the amino terminal of a peptide or to
20 the α -amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the peptide. Similarly, the term "carboxy terminus" refers to the free carboxyl group on the carboxy terminus of a peptide or the carboxyl group of an amino acid at any other location within the peptide. Peptides also include essentially any
25 polyamino acid including, but not limited to peptide mimetics such as amino acids joined by an ether as opposed to an amide bond.

[0021] A "variant enzyme" refers to an enzyme that is derived from a precursor enzyme (*e.g.*, the native enzyme, or a previously known variant) by addition of one or more amino acids to either or both the C- and N-terminal end, substitution of one or more amino acids at one or a number of different sites in the amino acid sequence, deletion of one or
30 more amino acids at either or both ends of the protein or at one or more sites in the amino acid sequence, or insertion of one or more amino acids at one or more sites in the amino acid sequence. In one embodiment, the preparation of an enzyme variant can be achieved by preparing a nucleic acid sequence (*e.g.*, by modifying a DNA sequence that encodes for

the native protein), transformation of that nucleic acid sequence into a suitable host, and expression of the modified sequence to form the variant. In another illustrative embodiment, the enzyme variant may also be prepared by expressing the DNA sequence in an *in vitro* expression system without the need for a host. In certain embodiments the
5 variant enzyme can be chemically synthesized.

[0022] The term "family 9 glycoside hydrolase" refers to a family of cellulolytic enzymes characterized by a catalytic domain structure having an (a/a)₆ barrel fold that contains an open active site cleft that contains at least six sugar binding subsites -4 to +2. In processive endoglucanases the catalytic domain is joined to a family 3c carbohydrate-
10 binding module that is aligned with the active site cleft. Sequence-based algorithmic methods are used to assign enzymes to various families. The glycoside hydrolases have been classified into more than 100 families. Each family (GH family) contains proteins that are related by sequence, and by corollary, fold (*see* Henrissat (1991) *Biochem J.* 280:309-316).

15 [0023] A theme B subfamily of said "family 9 glycoside hydrolase " refers to enzymes in which the catalytic module is fused to a family 3 carbohydrate-binding module (*see* Gilad *et al.* (2003) *J Bacteriol* 185:391–398)

[0024] The term "cellulosic biomass" refers to plant, algal, or other biomass that contains cellulose.

20 [0025] Lignocellulosic biomass refers to plant biomass that typically contains cellulose, hemicellulose, and lignin. The carbohydrate polymers (cellulose and hemicelluloses) are often tightly bound to the lignin. Lignocellulosic biomass can be grouped into four main categories: (1) agricultural residues (including corn stover and sugarcane bagasse), (2) dedicated energy crops, (3) wood residues (including sawmill and
25 paper mill discards), and (4) municipal paper waste. Illustrative lignocellulosic biomass sources include, but are not limited to grasses, rice hulls, bagasse, jute, hemp, flax, bamboo, sisal, abaca, straw, corn cobs, corn stover, alfalfa, hay, coconut hair, seaweed, algae,

[0026] A cellulase is an enzyme that breaks down cellulose, especially in the wall structures, and a "cellulosome" is an array, cluster, or sequence of enzymes or cellulases
30 that degrades cellulose. In various embodiments cellulosomes comprise catalytic subunits such as glycoside hydrolases, polysaccharide lyases and carboxyl esterases bound together by scaffoldins consisting of cohesins connected to other functional units such as the enzymes and carbohydrate binding modules via dockerins.

[0027] The terms "percent identity," "% identity," "percent identical," and "% identical" are used interchangeably herein to refer to the percent amino acid sequence identity that is obtained using standard alignment algorithms. For example, the percent sequence identity can be obtained by optimally aligning two sequences, counting the number of identical matches in the alignment and dividing such number of identical matches by the length of the reference sequence.

[0028] In various embodiments two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (*e.g.*, BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art (*see, e.g.*, Dayhoff *et al.* (1978), "A model of evolutionary change in proteins"; "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (Ed. M. O. Dayhoff), pp. 345-352, Natl. Biomed. Res. Round., Washington, D.C.; and Henikoff *et al.* (1992) *Proc. Natl. Acad. Sci. USA*, 89: 10915-10919, both of which are incorporated herein by reference). The BLOSUM62 matrix is often used as a default scoring substitution matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acid position of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, *e.g.*, gapped BLAST 2.0, described in Altschul *et al.* (1997) *Nucleic Acids Res.*, 25: 3389-3402, and made available to the public at the National Center for Biotechnology Information Website. Optimal alignments, including multiple alignments can be prepared using readily available programs such as PSI-BLAST, which is described by Altschul *et al.* (1997) *Nucleic Acids Res.*, 25: 3389-3402. In certain embodiments preferred parameters for BLASTp are: Expectation value: 10 (default); Filter: seg (default); Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Max. alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOSUM62.

[0029] One skilled in the art may also use the ALIGN program incorporating the non-linear algorithm of Myers and Miller (1988) *Comput. Appl. Biosci.* 4: 11-17). For

amino acid sequence comparison using the ALIGN program illustrative parameters include a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4.

[0030] The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, preferably more than about 35 residues, and typically the full length of the mature protein. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. Pearson, *Methods Enzymol.* 183:63-98 (1990) (herein incorporated by reference). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

[0031] Amino acid residues are said to "correspond to each other" or to be "corresponding amino acids" when they occupy the same position in optimally aligned sequences and/or when they occupy the same position in the three dimensional conformation of the subject molecule(s), and/or occupy the same position in a local alignment of conserved domains (the entire sequence may not align, but local domains often do).

[0032] An amino acid "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. Owing to deletions, insertions, truncations, fusions, and the like that must be taken into account when determining an optimal alignment. The amino acid residue number in a test sequence determined by simply counting from the N-terminal may not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not

correspond to any amino acid in the corresponding sequence. The identity of residues in a "test" (variant) sequence corresponding to particular residues in a "parent" sequence can be determined by alignment (*e.g.*, of the entire protein or particular domains) and/or inspection of protein conformation, and/or particular chemistries at designated positions.

5 [0033] An "expression vector" refers to a nucleic acid construct comprising a nucleic acid sequence (*e.g.*, DNA sequence) that is operably linked to a suitable control sequence capable of effecting the expression of the nucleic acid in a suitable host. Such control sequences may include a promoter to effect transcription, an optional operator sequence to control transcription, a sequence encoding suitable ribosome-binding sites on
10 the mRNA, and sequences that control termination of transcription and translation. Different cell types are typically used with different expression vectors. For example, an illustrative promoter for vectors used in *Bacillus spp.* is the AprE promoter; an illustrative promoter used in *E. coli* is the Lac promoter, an illustrative promoter used in *Saccharomyces spp.* is PGK1, an illustrative promoter used in *Aspergillus spp.* *glaA*, and an illustrative promoter
15 for *Trichoderma spp.* is *cbhI*. In certain embodiments the vector may be a plasmid, a phage particle, or simply a potential genomic insert. Once transformed into a suitable host, the vector may replicate and function independently of the host genome, or may, under suitable conditions, integrate into the genome itself. In the present specification, plasmid and vector are sometimes used interchangeably. However, the invention is intended to include other
20 forms of expression vectors that serve equivalent functions and which are, or become, known in the art. Thus, a wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences described herein. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic DNA sequences such as various known derivatives of SV40 and known bacterial
25 plasmids, *e.g.*, plasmids from *E. coli* including *col E1*, *pCR1*, *pBR322*, *pMb9*, *pUC 19* and their derivatives, wider host range plasmids, *e.g.*, *RP4*, phage DNAs *e.g.*, the numerous derivatives of phage *lamda*, *e.g.*, *NM989*, and other DNA phages, *e.g.*, *M13* and filamentous single stranded DNA phages, yeast plasmids such as the *2 mu* plasmid or derivatives thereof, vectors useful in eukaryotic cells, such as vectors useful in animal cells
30 and vectors derived from combinations of plasmids and phage DNAs, such as plasmids which have been modified to employ phage DNA or other expression control sequences. Expression techniques using the expression vectors of the present invention are known in the art and are described generally in, for example, Sambrook.

[0034] "Host strain" or "host cell" means a suitable host for expressing nucleic acids comprising an expression vector as described herein. Illustrative host cells include prokaryotic or eukaryotic hosts, including any transformable microorganism in which expression can be achieved. Certain typical host strains include, but are not limited to, 5 *Bacillus subtilis*, *Escherichia coli*, *Trichoderma reesei*, *Saccharomyces cerevisiae*, *Aspergillus niger*, and the like.

[0035] The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) exists in a purity 10 not found in nature, where purity can be adjudged with respect to the presence of other cellular material (*e.g.*, is free of other proteins from the same species) (3) is expressed by a cell from a different species, or (4) does not occur in nature (*e.g.*, it is a fragment of a polypeptide found in nature or it includes amino acid analogs or derivatives not found in nature or linkages other than standard peptide bonds). Thus, a polypeptide that is chemically 15 synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art. As thus defined, "isolated" does not necessarily require that the protein, polypeptide, peptide or 20 oligopeptide so described has been physically removed from its native environment.

[0036] The term "polypeptide fragment" as used herein refers to a polypeptide that has a deletion, *e.g.*, an amino-terminal and/or carboxy-terminal deletion and/or internal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is 25 identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long. In certain embodiments the polypeptide 30 fragment is at least 100, or at least 200, or at least 300, or at least 400, or at least 440, or at least 450, or at least 500 amino acid in length. In certain embodiments the fragment comprises or substantially comprises the catalytic domain of an enzyme described herein.

[0037] When "homologous" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A conservative amino acid substitution is typically one that will not substantially change the functional properties of the protein. In various embodiments a
5 "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (e.g., charge or hydrophobicity). In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative
10 nature of the substitution. Means for making this adjustment are well known to those of skill in the art (see, e.g., Pearson (1994) *Meth. Mol. Biol.* 24: 307-331 and 25:365-389).

[0038] The following six illustrative groups each contain amino acids that are often conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5)
15 Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W). These "conservative" substitutions are illustrative and not limiting. Other conservative substitutions are known. In addition numerous methods are available for the prediction of conservative substitutions. For example, SIFT (Sorting Intolerant From Tolerant) is a program that predicts whether an amino acid substitution
20 affects protein function. SIFT has been demonstrated to distinguish between functionally neutral and deleterious amino acid changes in mutagenesis studies and on human polymorphisms (see, e.g., Ng and Henikoff (2001) *Genome Res.*, 11: 863-875 and Ng and Henikoff (2003) *Nucl. Acids Res.*, 31(13): 3812-3814). SIFT is available at <http://blocks.fhcrc.org/sift/SIFT.html>. Other illustrative methods of identifying
25 conservative mutations (or non-conservative mutations) include, but are not limited to PolyPhen (see, e.g., Stitzel et al. (2004) *Nucleic Acids Res.* 32: D520-522; Stitzel et al. (2003) *J. Mol. Biol.* 327:1021-1030; and "<http://www.bork.emblheidelberg.de/PolyPhen>"), SNPs3D (see, e.g., Yue et al. (2005) *J. Mol. Biol.* 353:459-473; Yue and Moulton (2005) *J. Mol. Biol.* 356:1263-1274; and "<http://www.snps3d.org/>"), PANTHER PSEC (see, e.g.,
30 Thomas et al. (2003) *Genome Res.* 13: 2129-2141; and "<https://panther.appliedbiosystems.com/methods/csnpscoreForm.jsp>"), PMUT (see, e.g., Ferrer-Costa et al. (2005) *Bioinformatics*, 21: 3176-3178; Ferrer-Costa et al. (2002) *J. Mol. Biol.* 315: 771-786; Ferrer-Costa et al. (2004) *Proteins* 57:811-819; and "<http://mmb2.pcb.ub.es:8080/PMut>"), TopoSNP (see, e.g., Stitzel et al. (2003) *supra*;

Stitzel *et al.*(2004) supra.; and"<http://gila.bioengr.uic.edu/snp/toposnp>), and the like. A review is provided by Ng and Henikoff (2006) *Annu. Rev. Human Genet.* 7: 61-80.

[0039] The term "purified" does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified product preparation, is one in which the product is more concentrated than the product is in its environment within a cell. For example, a purified wax is one that is substantially separated from cellular components (nucleic acids, lipids, carbohydrates, and other peptides) that can accompany it. In another example, a purified wax preparation is one in which the wax is substantially free from contaminants, such as those that might be present following fermentation.

10 [0040] A "recombinant" nucleic acid molecule or protein is one that has a sequence that is not naturally occurring, has a sequence that is made by an artificial combination of two otherwise separated segments of sequence, or both. This artificial combination can be achieved, for example, by chemical synthesis or by the artificial manipulation of isolated segments of nucleic acid molecules or proteins, such as genetic engineering techniques.

15 Recombinant is also used to describe nucleic acid molecules that have been artificially manipulated, but contain the same regulatory sequences and coding regions that are found in the organism from which the nucleic acid was isolated.

[0041] The standard amino acid abbreviations shown below in Table 1 are used herein.

20 Table 1. Amino acid abbreviations.

Abbreviation	1 letter abbreviation	Amino acid name
Ala	A	Alanine
Arg	R	Arginine
Asn	N	Asparagine
Asp	D	Aspartic acid
Cys	C	Cysteine
Gln	Q	Glutamine
Glu	E	Glutamic acid
Gly	G	Glycine
His	H	Histidine
Ile	I	Isoleucine
Leu	L	Leucine
Lys	K	Lysine
Met	M	Methionine
Phe	F	Phenylalanine
Pro	P	Proline
Pyl	O	Pyrrolysine
Ser	S	Serine

Sec	U	Selenocysteine
Thr	T	Threonine
Trp	W	Tryptophan
Tyr	Y	Tyrosine
Val	V	Valine
Asx	B	Aspartic acid or Asparagine
Glx	Z	Gultamic acid or Glutamine
Xaa	X	Any amino acid
Xle	J	Leucine or Isoleucine

[0042] When a Markush Group or a list of particular compounds is described in the specification and/or claims it is intended that in various additional or alternative embodiments any subset of that Markush group or list is contemplated. Thus, for example, a Markush group or list consisting of elements A, B, and C also comprises a disclosure of a Markush Group or list consisting of A, and B, a Markush Group or list consisting of B, and C, and a Markush Group or list consisting of A and C as well as elements A, B, and C individually. Similarly, when elements in one list or Markush group are described as being in combination with elements in another list or Markush group the combination of any one element of one group with any one element of another group is contemplated. Thus where a list or Markush Group consisting of A, B and C is disclosed in combination with another list or Markush Group consisting of D, E, and F, the description is to be recognized as contemplating the combinations A/D, A/E, A/F, B/D, B/E, B/F, C/D, C/E, C/F and/or any combination of A, B, or C with any subgroup or member of D, E, and F, and/or any combination of D, E, or F, with any subgroup or member of A, B, and C.

15

BRIEF DESCRIPTION OF THE DRAWINGS

[0043] Figure 1A shows the amino acid sequence of the novel *Thermobifida fusca* Cel9A variant (designated "F1", SEQ ID NO:1). Figure 1B shows the DNA sequence of F1 (SEQ ID NO:2). The sequence shown is a synthetic DNA sequence that has been codon optimized for expression in *E. coli*. With the exception of the novel changes at positions F64, A226, and E246, this DNA sequence encodes the wild-type *Thermobifida fusca* Cel9A amino acid sequence, however, the nucleotide sequence is dissimilar from wild-type *Thermobifida fusca* Cel9A DNA. Note that nucleic acid sequences and amino acid position numbering are for the mature enzyme. They exclude the signal peptide, e.g., MSVTEPPPPRRRG RHSRARRFLTSLGATAALTAGMLGVPLATGTAHA (SEQ ID NO:3), although it will be recognized that in certain embodiments, protein(s) comprising the variant cellulolytic peptides described herein attached to a signal peptide are contemplated.

25

[0044] Figure 2 shows an alignment of wild-type *Thermobifida fusca* Cel9A (SEQ ID NO:4) and F1 (SEQ ID NO:1) amino acid sequences. As shown herein, F1 contains the following amino acid changes: F64I, A226V, and E246G.

[0045] Figure 3 shows data illustrating the improved activity of F1 as compared to wild-type. F1 exhibits 20-30% increased specific activity. Activity of wild-type Cel9A and F1 at various enzyme loadings for 24, 48, and 120 hour digestions. Key: F1 activity dark (upper) line, wild-type activity (lower) light line.

[0046] Figure 4 shows an alignment between wild-type *Thermobifida fusca* Cel9A (SEQ ID NO:4) and *Clostridium phytofermentans* Cphy3367 (SEQ ID NO:5).

[0047] Figure 5 shows structures of wild-type *Thermobifida fusca* Cel9A and *Clostridium phytofermentans* Cphy3367 illustrating 3D similarity. It is believed that the novel F1 variants if transferred, singly or in combination, to homologous positions in Cphy3367 will confer increased activity. Cel9A and Cphy3367 share 66% similarity. At homologous positions, two of the F1 variant positions are identical in Cphy3367 (A226, E246) and the third F1 variant (I64) shares close similarity to wild-type amino acid found in the homologous position in Cphy3367 (L).

[0048] Figure 6 shows additional examples of novel variants cellulolytic peptides exhibiting increased cellulolytic activity (as compared to wild-type). The peptides are designated G03 (SEQ ID NO:6) and G11 (SEQ ID NO:7).

20

DETAILED DESCRIPTION

[0049] In various embodiments novel variant cellulolytic enzymes (cellulases) having improved cellulolytic activity and/or stability are provided herein as well as methods of use of such enzymes. The variant enzymes provided herein have widespread utility in biomass conversion into renewable chemical and biofuel production, textile processing, waste water treatment, animal feed treatment, paper and pulp processing, detergent formulation and the like.

[0050] In addition to the enzymes themselves, in certain embodiments, nucleic acids encoding the enzymes, vectors incorporating the nucleic acids, and host cells that express the enzymes are provided.

[0051] In certain embodiments the enzymes are displayed on the surface of a cell (e.g., a bacterial cell, a phage, a yeast cell) or on the surface of a particle or component of a bioreactor. In certain embodiments the enzymes are provided as a component of a

cellulosome or minicellulosome (*e.g.*, displayed on a bacteria such as Clostridium, on a yeast, and the like).

[0052] Cellulases are known in the art as enzymes that hydrolyze cellulose (beta-1,4-glucan or beta D-glucosidic linkages) resulting in the formation of glucose, cellobiose, cellooligosaccharides, and the like. Cellulases have been traditionally divided into three major classes: endoglucanases (EC 3.2.1.4) ("EG"), exoglucanases or cellobiohydrolases (EC 3.2.1.91) ("CBH") and beta-glucosidases (EC 3.2.1.21) ("BG") (*see, e.g.*, Knowles *et al.* (1987) *TibTech*, 5: 255-261).

[0053] It is believed that endoglucanase-type cellulases hydrolyze internal beta-1,4-glucosidic bonds in regions of low crystallinity of the cellulose and exo-cellobiohydrolase-type cellulases hydrolyze cellobiose from the reducing or non-reducing end of cellulose. It follows that the action of endoglucanase components can greatly facilitate the action of exo-cellobiohydrolases by creating new chain ends which are recognized by exo-cellobiohydrolase components. Further, beta-glucosidase-type cellulases have been shown to catalyze the hydrolysis of alkyl and/or aryl β -D-glucosides such as methyl β -D-glucoside and p-nitrophenyl glucoside as well as glycosides containing only carbohydrate residues, such as cellobiose. This yields glucose as the sole product for the microorganism and reduces or eliminates cellobiose which inhibits cellobiohydrolases and endoglucanases.

[0054] Cellulases also find a number of uses in detergent compositions including to enhance cleaning ability, as a softening agent and to improve the feel of cotton fabrics (Hemmpel (1991) *ITB Dyeing/Printing/Finishing* 3: 5-14; Tyndall (1992) *Textile Chemist and Colorist* 24: 23-26; Kumar *et al.* (1997) *Textile Chemist and Colorist*, 29: 37-42). Without being bound to a particular theory, softening and color restoration properties of cellulase have been attributed to the endoglucanase components in cellulase compositions, as exemplified by U.S. Patent Nos. 5,648,263, 5,691,178, and 5,776,757, which disclose that detergent compositions containing a cellulase composition enriched in a specified endoglucanase component impart color restoration and improved softening to treated garments as compared to cellulase compositions not enriched in such a component.

[0055] Cellulase compositions have also been shown to degrade cotton-containing fabrics, resulting in reduced strength loss in the fabric (U.S. Patent No: 4,822,516), contributing to reluctance to use cellulase compositions in commercial detergent applications. Cellulase compositions comprising endoglucanase components have been

suggested to exhibit reduced strength loss for cotton-containing fabrics as compared to compositions comprising a complete cellulase system.

[0056] Cellulases have also been shown to be useful in degradation of cellulase biomass to ethanol (wherein the cellulase degrades cellulose to glucose and yeast or other
5 microbes further ferment the glucose into ethanol), in the treatment of mechanical pulp (Pere *et al.* (1996) *In Proc. Tappi Pulping Conf., Nashville, Tenn., 27-31*, pp. 693-696), for use as a feed additive (WO 1991/004673) and in grain wet milling.

Variant enzymes.

[0057] In certain embodiments a prototypic variant enzyme is provided designated
10 F1. This enzyme is a variant of *Thermobifida fusca* Cel9A. The amino acid sequence is shown in Figure 1 and a nucleic acid sequence (codon optimized for expression in *E. coli*) is shown in Figure 2. The reference/"parent" *Thermobifida fusca* Cel9A wild-type sequence (SEQ ID NO:4) is described in US Patent 7,314,974 (which is incorporated here by reference for the sequences described therein), and is listed also as database entry EMBL-
15 Bank CDS AAB42155. Cel9A is also known as CelD, E-4, E4, Cel9D, and YP_290232. It was a surprising discovery that alterations of one or more residues corresponding to positions F64, A226, and E246 result in an enzyme with greater activity and/or stability.

[0058] In view of the conserved nature of glycoside hydrolases, and the highly conserved nature of family 9 glycoside hydrolases, and the very highly conserved nature of
20 theme B family 9 glycoside hydrolases, it is believed that variation of residues in other glycoside hydrolases at one or more positions corresponding to s F64, A226, and/or E246 in *Thermobifida fusca* Cel9A will produce cellulolytic enzymes having increase activity and/or stability.

[0059] Thus, for example, without being bound to a particular theory, in view of the
25 strong alignment between wild-type *Thermobifida fusca* Cel9A and *Clostridium phytofermentans* Cphy3367 (*see*, Figure 4) and structures showing 3D similarity (*see*, Figure 5), it is believed that variants if transferred, signally or in combination, to homologous positions in Cphy3367 will confer increased activity. In this regard, it is noted that Cel9A and Cphy3367 share 66% similarity. At homologous positions, two of the F1
30 variant positions are identical in Cphy3367 (A226, E246) and the third F1 variant (I64) shares close similarity to wild-type amino acid found in the homologous position in Cphy3367 (L).

[0060] Thus, in certain embodiments variant cellulolytic enzymes comprising a glycoside hydrolase with or comprising a substitution at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A are contemplated. Variant cellulolytic enzymes, as contemplated herein, do not include any wild-type and/or naturally occurring enzymes. In certain embodiments the enzyme has at least 50%, or at least 60%, or at least 65%, or at least 70%, or at least 75%, or at least 80%, or at least 85%, or at least 86%, or at least 87%, or at least 88%, or at least 89%, or at least 90%, or at least 91%, or at least 92%, or at least 93%, or at least 94%, or at least 95%, or at least 96%, or at least 97%, or at least 98%, or at least 99% sequence identity with F1, and/or Cel9A, and/or Cphy3367, and/or G03, and/or G11. In certain embodiments the sequence identity is over a region of at least 30 contiguous amino acids, or at least 50 contiguous amino acids, or at least 100 contiguous amino acids, or at least 150 contiguous amino acids, or at least 200 contiguous amino acids, or at least 250 contiguous amino acids, or at least 300 contiguous amino acids, or at least 350 contiguous amino acids, or at least 400 contiguous amino acids, or at least 450 contiguous amino acids, or at least 500 contiguous amino acids, or at least 550 contiguous amino acids, or at least 600 contiguous amino acids, or over the full length of the enzyme(s).

[0061] In various embodiments the enzyme comprises no more than 5 additional variations (substitutions, and/or insertions, and/or deletions) at positions other than those corresponding to F64, A226, and/or E246 in *Thermobifida fusca* Cel9A. In various embodiments the variant is a variant of a family 9 glycoside hydrolase. In various embodiments the variant is a variant of a theme B subfamily of said family 9 glycoside hydrolase.

[0062] In certain embodiments the variant comprises a substitution at a position corresponding to F64 in *Thermobifida fusca* Cel9A. In various embodiments the variant comprises a substitution at a position corresponding to A226 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at a position corresponding to E246 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to F64 and A226 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to F64 and E246 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to A226 and E246 in *Thermobifida fusca* Cel9A. In certain embodiments the variant comprises a substitution at positions corresponding to F64, A226, and E246 in *Thermobifida fusca* Cel9A.

[0063] In certain embodiments the substitution at a position corresponding position F64 in *Thermobifida fusca* Cel9A is a substitution to I, V, L, or M. In certain embodiments the substitution at a position corresponding position F64 in *Thermobifida fusca* Cel9A is a substitution to I. In certain embodiments the substitution at a position corresponding
5 position A226 in *Thermobifida fusca* Cel9A is a substitution to V, I, L, M, or T. In certain embodiments the substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V, I, L, or M. In certain embodiments the substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V. In certain embodiments the substitution at a position corresponding position E246 in
10 *Thermobifida fusca* Cel9A is a substitution to G, A, N, or S. In certain embodiments the substitution at a position corresponding position E246 in *Thermobifida fusca* Cel9A is a substitution to G.

[0064] In various embodiments the enzyme is a variant of a glycoside hydrolase from an organism selected from the group consisting of archaea, bacteria, and eukaryota. In
15 certain embodiments the enzyme is a variant of a glycoside hydrolase from a gram negative bacterium. In certain embodiments the enzyme is a variant of a glycoside hydrolase from a gram positive bacterium. In certain embodiments the enzyme is a variant of glycoside hydrolase from a bacterial family selected from the group consisting of *Thermofidia*, *Micromonospora*, *Cellulomonas*, *Listeria*, *Pseudomonas*, *Ruminococcus*, *Saccharophagus*,
20 *Streptomyces*, *Vibrio*, *Xanthomonas*, and *Clostridium*.

[0065] In certain embodiments the variant comprises or consists of a variant of *Thermobifida fusca* Cel9A and/or *Clostridium phytofermentans* Cphy3367. In certain embodiments the variant comprises or consists of *Thermobifida fusca* Cel9A having an amino acid substitution at F64, and/or A226, and/or E246. In certain embodiments the variant
25 comprises or consists of *Thermobifida fusca* Cel9A having an amino acid substitution of F64 to I, V, L, or M, and/or of A226 to V, I, L, M, A, or T, and/or of E246 to G, A, N, or S. In certain embodiments the variant comprises or consists of *Thermobifida fusca* Cel9A having an amino acid substitution of F64I, and/or of A226V, and/or E246G. In certain embodiments the variant comprises or consists of *Thermobifida fusca* Cel9A having an amino acid
30 substitution of F64I, A226V, and E246G.

[0066] In certain embodiments the variant comprises or consists of the amino acid sequence of F1, G03, or G11.

[0067] In certain embodiments variant enzymes comprising conservative substitutions in any of the variant enzymes described above are contemplated. In various embodiments the variant comprises no more than 1, or 2, or 3, or 4, or 5, or 6, or 7, or 8, or 9, or 10, or 11, or 12, or 13, or 14, or 15, or 16, or 17, or 18, or 19, or 20 conservative
5 substitutions.

[0068] In certain embodiments isolated polypeptides comprising a fragment of the above-described amino acid sequences are provided. In various embodiments these fragments preferably include at least 20 contiguous amino acids, more preferably at least 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, or even more contiguous
10 amino acids. In certain embodiments the fragments comprise particular domains of the peptide(s) and/or possess a particular activity (*e.g.*, carbohydrate binding, hydrolase activity, *etc.*). In certain embodiments peptides comprising, consisting essentially of, or consisting of the catalytic domain of the variant cellulolytic enzyme are provided. In this context "consisting essentially of" refers to a protein variant or fragment thereof that preserves the
15 catalytic activity of the full-length variant peptide.

[0069] In certain embodiments, fusions between the above-described polypeptide sequences and heterologous polypeptides are provided. The heterologous sequences can, for example, include sequences designed to facilitate purification, *e.g.* histidine tags, and/or visualization of recombinantly-expressed proteins. Other non-limiting examples of protein
20 fusions include those that permit display of the encoded protein on the surface of a phage or a cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), fusion to signal peptides to direct polypeptide processing and export, fusion to cellulose binding module(s), fusion to dockerin domain(s), fusion to cohesion domain(s), fusion to fibronectin-like domain(s), fusions to the IgG Fc region, and the like. The fusions can be
25 direct or can be by way of intervening peptide linker regions/domains.

Preparation of host cell transformants and expression of variant enzymes.

[0070] In various embodiments polynucleotide fragments encoding one or more variant enzymes described herein and, optionally other polypeptides (*e.g.*, other enzymes) may be incorporated into heterologous nucleic acid constructs or vectors, capable of
30 introduction into, and replication in a host cell (*e.g.*, a filamentous fungal or yeast cell, a bacterial cell, *etc.*). Generally any vector may be used as long as it is replicable and viable in the cells into which it is introduced. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. Cloning and expression

vectors are also described in Sambrook *et al.*, 1989, Ausubel F M *et al.*, 1989, and Strathern *et al.*, The Molecular Biology of the Yeast *Saccharomyces*, 1981, each of which is expressly incorporated by reference herein.

Vectors.

5 [0071] In various embodiments nucleic acids encoding one or more of the variant cellulolytic enzymes (cellulases) described herein are inserted into vectors suitable for expressing the enzyme(s) in a host cell. In such vector(s), the nucleic acid encoding the enzyme(s) is operably linked to one or more promoters and/or other regulatory sequences.

[0072] The term "operably linked" refers herein to a configuration in which a
10 control sequence is appropriately placed at a position relative to the coding sequence of the DNA sequence such that the control sequence influences the expression of a polypeptide. When used herein, the term "coding sequence" is intended to cover a nucleotide sequence that directly specifies the amino acid sequence of its protein product. The boundaries of the coding sequence are generally determined by an open reading frame that usually begins
15 with the ATG start codon. The coding sequence typically includes a DNA, cDNA, and/or recombinant nucleotide sequence.

[0073] As used herein, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and secretion.

20 [0074] The term "expression vector" refers herein to a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of the invention, and which is operably linked to additional segments that provide for its transcription.

[0075] In certain embodiments nucleic acid constructs comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast
25 artificial chromosome (YAC), and the like, into which a nucleic acid sequence(s) of the variant enzyme(s) have been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

30 [0076] Polynucleotides encoding the variant enzyme(s) described herein can be incorporated into any of a variety of expression vectors suitable for expressing a polypeptide. Suitable vectors include chromosomal, nonchromosomal and synthetic DNA

sequences, *e.g.*, derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated virus, retroviruses and many others. Any vector that transduces and/or expresses genetic material
5 into a cell, and, if replication is desired, which is replicable and viable in the relevant host can be used.

[0077] When incorporated into an expression vector, the nucleic acid sequence(s) encoding the desired enzyme(s) are operatively linked to an appropriate transcription control sequence (promoter) to direct mRNA synthesis, *e.g.*, T5 promoter. Examples of
10 such transcription control sequences particularly suited for use in transgenic plants include the cauliflower mosaic virus (CaMV) and figwort mosaic virus (FMV). Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses and which can be used in some embodiments include SV40 promoter, *E. coli* lac or trp promoter, phage lambda P_L promoter, tac promoter, T7 promoter, and the like. Examples of
15 suitable promoters useful for directing the transcription of the nucleotide constructs in a filamentous fungal host cell include promoters such as *cbh1*, *cbh2*, *egl1*, *egl2*, *pepA*, *hfb1*, *hfb2*, *xyn1*, *amy*, and *glaA* (*see, e.g.*, Nunberg *et al.* (1984) *Mol. Cell Biol.*, 4: 2306-2315; Boel *et al.* (1984) *EMBO J.* 3: 1581-1585)). In bacterial host cells, suitable promoters include the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase
20 gene (*dagA*), a gene from a *Bacillus* sp., such as, for example, the *Bacillus subtilis* levansucranse gene (*sacB*), the *Bacillus licheniformis* alpha-amylase gene (*amyl*), the *Bacillus megaterium* *InhA* gene, the *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), the *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus subtilis* *xylA* and *xylB* genes, the xylose promoter (P_{xyl}) from *Bacillus megaterium*, the promoter
25 obtained from the prokaryotic beta-lactamase gene, and so forth.

[0078] In various embodiments an expression vector optionally contains a ribosome binding site for translation initiation, and a transcription terminator, such as PinII. The vector also optionally includes appropriate sequences for amplifying expression, *e.g.*, an enhancer.

30 [0079] In various embodiments the vector or DNA construct may also generally include a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and which directs the encoded polypeptide into the cell's

secretory pathway. Suitable signal peptides include, but are not limited to the *Bacillus megaterium* penicillin G acylase signal peptide sequence.

[0080] Other illustrative signal peptide coding regions for bacterial host cells may be obtained from the genes of *Bacillus* NCIB 11837 maltogenic amylase, *B.*

5 *stearothermophilus* alpha-amylase, *B. licheniformis* subtilisin, *B. licheniformis* beta-lactamase, *B. stearothermophilus* neutral proteases (nprT, nprS, nprM) and *B. subtilis* prsS. Further illustrative signal sequences are described in Simonen and Palva (1993), *Microbiological Reviews* 57: 109-137. Effective signal peptide coding regions for filamentous fungal host cells include but are not limited to the signal peptide coding regions
10 obtained from *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase and *Humicola lanuginosa* lipase. Variants of these signal peptides and other signal peptides are suitable, as well as expression mutants thereof having one or more silent mutations.

15 [0081] In various embodiments the expression vectors optionally contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells. Suitable marker genes include those coding for antibiotic resistance such as, ampicillin, kanamycin, chloramphenicol, or tetracycline resistance. Further examples include the antibiotic spectinomycin or streptomycin (*e.g.*, the *aada* gene), the streptomycin
20 phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance. Additional selectable marker genes include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, and tetracycline or ampicillin resistance in *E. coli*.

25 [0082] The vector may further contain genetic elements that facilitate integration by either homologous or non-homologous recombination. Genetic elements that facilitate integration by homologous recombination have sequence homology to targeted integration sites in the genomic sequence of the desired expression host cell. Genetic elements or
30 techniques which facilitate integration by non-homologous recombination include restriction enzyme-mediated integration (REMI) (see Manivasakam *et al.* (1998) *Mol. Cell Biol.* 18(3): 1736-1745), transposon-mediated integration, and other elements and methods that are well known in the art.

[0083] In certain embodiments the cellulase polynucleotides described herein can also be fused, for example, in-frame to nucleic acids encoding a secretion/localization sequence, to target polypeptide expression to a desired cellular compartment, membrane, or organelle of a cell, or to direct polypeptide secretion to the periplasmic space, to the cell membrane or cell wall, or into the cell culture media. Such sequences are known to those of skill, and include secretion leader peptides, organelle targeting sequences (*e.g.*, nuclear localization sequences, endoplasmic reticulum (ER) retention signals, mitochondrial transit sequences, peroxisomal transit sequences, and chloroplast transit sequences), membrane localization/anchor sequences (*e.g.*, stop transfer sequences, GPI anchor sequences), and the like.

[0084] An illustrative expression vector for the expression of variant cellulases described herein is detailed in Example 1, hereinbelow. Vectors as described herein can be employed to transform an appropriate host to permit the host to express one or more variant cellulolytic enzymes described herein.

Host cell transformants (Expression Hosts)

[0085] In certain embodiments, host cells host cells transformed with nucleic acid molecules or vectors encoding one or more of the cellulolytic enzymes described herein. In some embodiments, these cells carry the nucleic acid sequences on vectors, which may, but need not, be freely replicating vectors. In other embodiments, the nucleic acids have been integrated into the genome of the host cells.

[0086] Suitable host cells include, but are not limited to bacterial cells, algal cells, plant cells, fungal cells, insect cells and mammalian cells. In one illustrative embodiment suitable host cells include *E. coli* (*e.g.*, SHuffle™ competent *E. coli* available from New England BioLabs in Ipswich, MA). As used herein, genetically modified or recombinant host cell includes the progeny of host cell(s) that comprise a polynucleotide that encodes a variant enzyme described herein.

[0087] In some embodiments, the genetically modified or recombinant host cell is a eukaryotic cell. Suitable eukaryotic host cells include, but are not limited to, fungal cells, algal cells, insect cells, and plant cells. Suitable fungal host cells include, but are not limited to, *Ascomycota*, *Basidiomycota*, *Deuteromycota*, *Zygomycota*, *Fungi imperfecti*. Certain preferred fungal host cells include yeast cells and filamentous fungal cells. Suitable filamentous fungi host cells include, for example, any filamentous forms of the subdivision

Eumycotina and Oomycota. (see, e.g., Hawksworth *et al.*, In Ainsworth and Bisby's Dictionary of The Fungi, 8.sup.th edition, 1995, CAB International, University Press, Cambridge, UK, which is incorporated herein by reference). Filamentous fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose and other complex polysaccharides. The filamentous fungi host cells are morphologically distinct from yeast.

[0088] In certain illustrative, but non-limiting embodiments the filamentous fungal host cell may be a cell of a species of, but not limited to *Achlya*, *Acremonium*, *Aspergillus*, *Aureobasidium*, *Bjerkandera*, *Ceriporiopsis*, *Cephalosporium*, *Chrysosporium*, *Cochliobolus*, *Corynascus*, *Cryphonectria*, *Cryptococcus*, *Coprinus*, *Coriolus*, *Diplodia*, *Endothia*, *Fusarium*, *Gibberella*, *Gliocladium*, *Humicola*, *Hypocrea*, *Myceliophthora* (e.g., *Myceliophthora thermophila*), *Mucor*, *Neurospora*, *Penicillium*, *Podospora*, *Phlebia*, *Piromyces*, *Pyricularia*, *Rhizomucor*, *Rhizopus*, *Schizophyllum*, *Scytalidium*, *Sporotrichum*, *Talaromyces*, *Thermoascus*, *Thielavia*, *Trametes*, *Tolypocladium*, *Trichoderma*, *Verticillium*, *Volvariella*, or teleomorphs, or anamorphs, and synonyms or taxonomic equivalents thereof.

[0089] In some embodiments the filamentous fungal host cell is of the *Trichoderma* species, e.g., *T. longibrachiatum*, *T. viride* (e.g., ATCC 32098 and 32086), *Hypocrea jecorina* or *T. reesei*, *T. koningii*, and *T. harzianum*, and the like.

[0090] In some embodiments of the invention, the filamentous fungal host cell is of the *Aspergillus* species, e.g., *A. awamori*, *A. funigatus*, *A. japonicus*, *A. nidulans*, *A. niger*, *A. aculeatus*, *A. foetidus*, *A. oryzae*, *A. sojae*, and *A. kawachi*.

[0091] In some embodiments the filamentous fungal host cell is of the *Chrysosporium* species, e.g., *C. lucknowense*, *C. keratinophilum*, *C. tropicum*, *C. merdarium*, *C. inops*, *C. pannicola*, and *C. zonatum*.

[0092] In some embodiments of the invention, the filamentous fungal host cell is of the *Fusarium* species, e.g., *F. bactridioides*, *F. cerealis*, *F. crookwellense*, *F. culmorum*, *F. graminearum*, *F. graminum*, *F. oxysporum*, *F. roseum*, and *F. venenatum*. In some embodiments of the invention, the filamentous fungal host cell is of the *Neurospora* species, (e.g., *N. crassa*), the *Humicola* species, (e.g., *H. insolens*, *H. grisea*, and *H. lanuginosa*) the *Mucor* species (e.g., *M. miehei* and *M. circinelloides*) the *Rhizopus* species (e.g., *R. oryzae* and *R. niveus*) the *Trichoderma* species (e.g., *T. reesei*), the *Trametes* species (e.g., *T. villosa* and *T. versicolor*), and the like.

[0093] In certain embodiments enzyme(s) described herein are expressed in a yeast host. Suitable yeast host cells include, but are not limited to *Candida*, *Hansenula*, *Saccharomyces*, *Schizosaccharomyces*, *Pichia*, *Kluyveromyces*, and *Yarrowia*. In some embodiments the yeast cell is *Hansenula polymorpha*, *Saccharomyces cerevisiae*,
5 *Saccaromyces carlsbergensis*, *Saccharomyces diastaticus*, *Saccharomyces norbensis*,
Saccharomyces kluyveri, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia finlandica*,
Pichia trehalophila, *Pichia kodamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia quercuum*, *Pichia pijperi*, *Pichia stipitis*, *Pichia methanolica*, *Pichia angusta*, *Kluyveromyces lactis*, *Candida albicans*, or *Yarrowia*
10 *lipolytica*.

[0094] In certain embodiments the host cell is an algal such as, *Chlamydomonas* (e.g., *C. Reinhardtii*) and *Phormidium* (*P. sp.* ATCC29409).

[0095] In other embodiments, the host cell is a prokaryotic cell. Suitable prokaryotic cells include gram positive, gram negative and gram-variable bacterial cells. The host cell
15 may be a species of, but not limited to *Agrobacterium*, *Alicyclobacillus*, *Anabaena*,
Anacystis, *Acinetobacter*, *Acidothermus*, *Arthrobacter*, *Azobacter*, *Bacillus*,
Bifidobacterium, *Brevibacterium*, *Butyrivibrio*, *Buchnera*, *Campestris*, *Campylobacter*,
Clostridium, *Corynebacterium*, *Chromatium*, *Coprococcus*, *Escherichia*, *Enterococcus*,
Enterobacter, *Erwinia*, *Fusobacterium*, *Faecalibacterium*, *Francisella*, *Flavobacterium*,
20 *Geobacillus*, *Haemophilus*, *Helicobacter*, *Klebsiella*, *Lactobacillus*, *Lactococcus*,
Ilyobacter, *Micrococcus*, *Microbacterium*, *Mesorhizobium*, *Methylobacterium*,
Methylobacterium, *Mycobacterium*, *Neisseria*, *Pantoea*, *Pseudomonas*, *Prochlorococcus*,
Rhodobacter, *Rhodopseudomonas*, *Rhodopseudomonas*, *Roseburia*, *Rhodospirillum*,
Rhodococcus, *Scenedesmus*, *Streptomyces*, *Streptococcus*, *Synecoccus*,
25 *Saccharomonospora*, *Staphylococcus*, *Serratia*, *Salmonella*, *Shigella*,
Thermoanaerobacterium, *Tropheryma*, *Tularensis*, *Temecula*, *Thermosynechococcus*,
Thermococcus, *Ureaplasma*, *Xanthomonas*, *Xylella*, *Yersinia* and *Zymomonas*.

[0096] In some embodiments the bacterial host strain is an industrial strain. Numerous bacterial industrial strains are known and suitable in the methods and
30 compositions described herein.

[0097] In some embodiments the bacterial host cell is of the *Agrobacterium* species (e.g., *A. radiobacter*, *A. rhizogenes*, *A. rubi*), the *Arthrobacter* species (e.g., *A. aurescens*, *A. citreus*, *A. globiformis*, *A. hydrocarboglutamicus*, *A. mysorens*, *A. nicotianae*, *A. paraffineus*,

A. protohonniae, *A. roseoparaffinus*, *A. sulfureus*, *A. ureafaciens*), the *Bacillus* species (e.g., *B. thuringiensis*, *B. anthracis*, *B. megaterium*, *B. subtilis*, *B. lentus*, *B. circularis*, *B. pumilus*, *B. lautus*, *B. coagulans*, *B. brevis*, *B. firmus*, *B. alkaophilus*, *B. licheniformis*, *B. clausii*, *B. stearothermophilus*, *B. halodurans* and *B. amyloliquefaciens*). In particular

5 *embodiments*, the host cell will be an industrial *Bacillus* strain including but not limited to *B. subtilis*, *B. pumilus*, *B. licheniformis*, *B. megaterium*, *B. clausii*, *B. stearothermophilus* and *B. amyloliquefaciens* (some preferred embodiments of a *Bacillus* host cell include *B. subtilis*, *B. licheniformis*, *B. megaterium*, *B. stearothermophilus* and *B. amyloliquefaciens*), the *Clostridium* species (e.g., *C. acetobutylicum*, *C. tetani* E88, *C. lituseburense*, *C.*

10 *saccharobutylicum*, *C. perfringens*, *C. beijerinckii*), the *Corynebacterium* species (e.g., *C. glutamicum*, *C. acetoacidophilum*) the *Escherichia* species (e.g., *E. coli*), the *Erwinia* species (e.g., *E. uredovora*, *E. carotovora*, *E. ananas*, *E. herbicola*, *E. punctata*, *E. terreus*), the *Pantoea* species (e.g., *P. citrea*, *P. agglomerans*), the *Pseudomonas* species, (e.g., *P. putida*, *P. aeruginosa*, *P. mevalonii*), the *Streptococcus* species (e.g., *S. equisimiles*, *S.*

15 *pyogenes*, *S. uberis*), the *Streptomyces* species (e.g., *S. ambofaciens*, *S. achromogenes*, *S. avermitilis*, *S. coelicolor*, *S. aureofaciens*, *S. aureus*, *S. fungicidicus*, *S. griseus*, *S. lividans*), the *Zymomonas* species (e.g., *Z. mobilis*, *Z. lipolytica*), and the like.

[0098] In various embodiments strains that may be used in the practice of the invention including both prokaryotic and eukaryotic strains, are readily accessible to the

20 public from a number of culture collections such as American Type Culture Collection (ATCC), Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH (DSMZ), Centraalbureau Voor Schimmelcultures (CBS), and Agricultural Research Service Patent Culture Collection, Northern Regional Research Center (NRRL).

Introduction of the nucleic acid encoding a variant cellulase into a host cell

25 [0099] Introduction of a vector or DNA construct into a host cell can be effected by any of a number of methods known to those of skill in the art. Illustrative methods include, but are not limited to calcium phosphate transfection, DEAE-Dextran mediated transfection, electroporation, or other common techniques (*see, e.g., Davis et al. (1986) Basic Methods in Molecular Biology*).

30 [0100] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, and/or amplifying the cellulase polynucleotide. Culture conditions, such as temperature, pH and the like, are those typically used with the host cell selected for expression, and will be apparent to those

skilled in the art (*see, e.g.*, Sambrook, Ausubel and Berger, as well as, for example, Freshney (1994) *Culture of Animal Cells, a Manual of Basic Technique*, third edition, Wiley-Liss, New York; Payne *et al.* (1992) *Plant Cell and Tissue Culture in Liquid Systems* John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds) (1995) *Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual*, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) *The Handbook of Microbiological Media* (1993) CRC Press, Boca Raton, Fla.).

[0101] The methods of transformation may result in the stable integration of all or part of the transformation vector into the genome of the host cell. However, transformation resulting in the maintenance of a self-replicating extra-chromosomal transformation vector is also contemplated.

Production and Recovery of Cellulase Variants

[0102] In certain embodiments methods of making the variant enzymes described herein are provided. The methods comprising providing a host cell transformed with a nucleic acid construct encoding one or more of the enzymes described herein; culturing the transformed host cell in a culture medium under conditions that cause the polynucleotide to express the encoded variant enzyme(s); and optionally recovering or isolating the expressed variant enzyme, or recovering or isolating the culture medium containing the expressed enzyme variant(s), or recovering or isolating cells displaying the expressed enzyme variant(s). In certain embodiments the methods further provides optionally lysing the transformed host cells after expressing the encoded variant enzyme(s) and optionally recovering or isolating the expressed variant enzyme(s) from the cell lysate.

[0103] Typically, recovery or isolation of the variant enzyme(s) is from the host cell culture medium, the host cell or both, using protein recovery techniques that are well known in the art, including those described herein.

[0104] Following transformation of a suitable host strain and growth (cultivating or culturing) of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract may be retained for further purification. Microbial cells employed in expression of proteins can be disrupted by any

convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, or other methods, which are well known to those skilled in the art.

[0105] Many references are available for the culture and production of many cells, including cells of bacterial, plant, animal (especially mammalian) and archebacterial origin.

5 See *e.g.*, Sambrook, Ausubel, and Berger (all supra), as well as Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley-Liss, New York and the references cited therein; Doyle and Griffiths (1997) Mammalian Cell Culture: Essential Techniques John Wiley and Sons, NY; Humason (1979) Animal Tissue Techniques, fourth edition W.H. Freeman and Company; and Ricciardelli, *et al.*, (1989) *In vitro* Cell Dev. Biol.

10 25:1016-1024, all of which are incorporated herein by reference. For plant cell culture and regeneration, Payne *et al.* (1992) Plant Cell and Tissue Culture in Liquid Systems_ John Wiley & Sons, Inc. New York, N.Y.; Gamborg and Phillips (eds) (1995) Plant Cell,--Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York); Jones, ed. (1984) Plant Gene Transfer and Expression Protocols,

15 Humana Press, Totowa, N.J. and Plant Molecular Biology (1993) R. R. D. Croy, Ed. Bios Scientific Publishers, Oxford, U.K. ISBN 0 12 198370 6, all of which are incorporated herein by reference. Cell culture media in general are set forth in Atlas and Parks (eds.) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, Fla., which is incorporated herein by reference. Additional information for cell culture is found in

20 available commercial literature such as the Life Science Research Cell Culture Catalogue (1998) from Sigma-Aldrich, Inc (St Louis, Mo.) ("Sigma-LSRCCC") and, for example, The Plant Culture Catalogue and supplement (1997) also from Sigma-Aldrich, Inc (St Louis, Mo.) ("Sigma-PCCS"), all of which are incorporated herein by reference.

[0106] In some embodiments, cells expressing the variant enzyme(s) described

25 herein are grown under batch or continuous fermentations conditions. Classical batch fermentation is a closed system, wherein the compositions of the medium is set at the beginning of the fermentation and is not subject to artificial alternations during the fermentation. A variation of the batch system is a fed-batch fermentation which also finds use in the present methods described herein. In this variation, the substrate is added in

30 increments as the fermentation progresses. Fed-batch systems are useful when catabolite repression is likely to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the medium. Batch and fed-batch fermentations are common and well known in the art. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of

conditioned medium is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth. Continuous fermentation systems strive to maintain steady state growth conditions. Methods for modulating nutrients and growth factors for continuous
5 fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

[0107] The resulting variant enzyme(s) can be recovered/isolated and optionally purified by any of a number of methods known in the art. For example, the polypeptide may be isolated from the nutrient medium by conventional procedures including, but not limited
10 to, centrifugation, filtration, extraction, spray-drying, evaporation, chromatography (*e.g.*, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), or precipitation. Protein refolding steps can be used, as desired, in completing the configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed in the final purification steps. In addition to the references noted
15 supra, a variety of purification methods are well known in the art, including, for example, those set forth in Sandana (1997) *Bioseparation of Proteins*, Academic Press, Inc.; Bollag *et al.* (1996) *Protein Methods*, 2nd Edition, Wiley-Liss, NY; Walker (1996) *The Protein Protocols Handbook* Humana Press, NJ; Harris and Angal (1990) *Protein Purification Applications: A Practical Approach*, IRL Press at Oxford, Oxford, England; Harris and
20 Angal *Protein Purification Methods: A Practical Approach*, IRL Press at Oxford, Oxford, England; Scopes (1993) *Protein Purification: Principles and Practice* 3rd Edition, Springer Verlag, NY; Janson and Ryden (1998) *Protein Purification: Principles, High Resolution Methods and Applications*, Second Edition, Wiley-VCH, NY; and Walker (1998) *Protein Protocols on CD-ROM*, Humana Press, NJ

[0108] In certain embodiments cell-free transcription/translation systems can also be
25 employed to produce the variant enzyme(s) described herein. Several such systems are commercially available. A general guide to *in vitro* transcription and translation protocols is found in Tymms (1995) *In vitro Transcription and Translation Protocols: Methods in Molecular Biology*, Volume 37, Garland Publishing, NY, which is incorporated herein by
30 reference.

Enzymes and enzyme systems displayed on surfaces.

[0109] In certain embodiments the variant enzymes described herein or enzyme complexes comprising the variant enzymes are displayed on the surfaces of substrates, particles, and/or microorganisms.

5 **Substrates, microparticles, and nanoparticles.**

[0110] In certain embodiments the variant enzymes described herein are displayed a surface. Illustrative surfaces include, for example, the surface of a particle, microbead or particle (both termed "microparticles" for ease of discussion), nanoparticle, a planar substrate (*e.g.*, a microarray substrate), a wall, channel or chamber in a fluidics or
10 microfluidics device, a surface of a cell culture vessel, and a surface comprising a bioreactor, and the like.

[0111] Such surfaces may comprise any of a number of materials including, but not limited to a ceramic, a plastic, a glass, a metal, a mineral, and the like. Illustrative materials include, but are not limited to cellulose and cellulose variants (methylcellulose,
15 hydroxyethyl cellulose, *etc.*), sepharose, polystyrene, polymethylacrylate, polypropylene, latex, polytetrafluoroethylene, polyacrylonitrile, polycarbonate, silica, quartz, metals, metal oxides or nitrides, semiconductors, nanorods, quantum dots, and the like. In various embodiments the materials and may be fluorescent and/or magnetic and/or paramagnetic.

Microorganism display systems.

20 [0112] In various embodiments the variant enzyme(s) described herein can be displayed on the surface of a microorganism (*e.g.*, a yeast or other fungus, a bacterium, a phage *etc.*). The expression and display/presentation of heterologous proteins on the surface of microorganisms is well known to those of skill in the art. Typically in "cell-based" display systems the polypeptides (*e.g.*, variant enzyme(s) to be displayed are inserted
25 into or replace all or part of a gene encoding a cellular protein that is expressed on the cell surface (package surface protein).

[0113] Numerous cell-based surface display platforms are well known in the art. These include, for example, prokaryotic cells such as *E. coli*, *S. typhimurium*, *P. aeruginosa*, *B. subtilis*, *P. aeruginosa*, *V. cholerae*, *K pneumonia*, *N. gonorrhoeae*, *N. meningitides*, and the like. They also include eukaryotic cells such as yeast cells,
30 filamentous fungal cells, and the like.

[0114] As an illustration of prokaryotic based surface display, Wu *et al.* (2006) *FEMS Microbiol. Lett.* 256: 119-25 describe cell surface display of Chi92 on Escherichia coli using ice nucleation protein. Cho *et al.* (2002) *Appl. Environ. Microbiol.* 68: 2026-2030 describe cell surface display of organophosphorus hydrolase in E. coli, Lee *et al.* (2005) *Appl Environ Microbiol.* 71: 8581-8586 reported cell surface display of lipase in Pseudomonas putida KT2442 using OprF as an anchoring motif, Shimazu *et al.* (2003) *Biotechnol Prog.* 19:1612-1614 describe cell surface display of a protein (organophosphorus hydrolase) in Pseudomonas putida using an ice-nucleation protein anchor. In addition, Desvaux *et al.* (2006) *FEMS Microbiol Lett.* 256: 1-15 reviews cell surface display of proteins in Gram-positive bacteria in general.

[0115] Examples of yeast display systems include, but are not limited to for example, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Hansenula*, or *Pichia pastoris*. Other display systems include, but are not limited to phage display, insect display, and mammalian cell display systems.

[0116] The foregoing display systems are intended to be illustrative and no limiting. Using the teachings provided herein numerous other methods and systems for displaying the variant enzyme(s) described herein on the surface of microorganisms will be available to one of skill in the art.

Cellulosomes and minicellulosomes

[0117] The digestion of cellulose and hemicellulose can be facilitated by the use of several types of enzymes acting cooperatively. Often, at least three categories of enzymes are utilized to convert cellulose into fermentable sugars: endoglucanases that cut the cellulose chains at random; cellobiohydrolases that cleave cellobiosyl units from the cellulose chain ends and beta-glucosidases that convert cellobiose and soluble cellodextrins into glucose. Among these three categories of enzymes involved in the biodegradation of cellulose, cellobiohydrolases are useful for the degradation of native crystalline cellulose. Cellobiohydrolase I, also referred to as a cellulose 1,4-beta-cellobiosidase or an exoglucanase, exo-cellobiohydrolase or 1,4-beta-cellobiohydrolase catalyzes the hydrolysis of 1,4-beta-D-glucosidic linkages in cellulose and cellotetraose, by the release of cellobiose from the non-reducing ends of the chains. Cellobiohydrolase II activity is identical, except that cellobiohydrolase II attacks from the reducing ends of the chains.

[0118] In various embodiments the cellulolytic enzymes (including one or more of the variant enzymes described herein) are organized into a cellulosome or minicellulosome. Cellulosomes are complexes of cellulolytic enzymes natively created by bacteria such as *Clostridium* and *Bacteroides*. They consist of catalytic subunits such as glycoside
5 hydrolases, polysaccharide lyases and carboxyl esterases typically bound together by scaffoldins consisting of cohesins connected to other functional units such as the enzymes and carbohydrate binding modules typically via dockerins. They assist in digestion or degradation of plant cell wall materials, most notably cellulose. Artificial cellulosomes (e.g., minicellulosomes) can be engineered into and expressed in a variety of organisms,
10 including, but not limited to certain bacteria, (e.g., *Clostridia*, *Bacteroides*), yeasts, and fungi.

[0119] The displayed cellulosomes can be simple cellulosome systems containing a single scaffoldin, or complex cellulosome systems exhibit multiple types of interacting scaffoldins. In various embodiments each scaffoldin can contain one, two, three, four, five,
15 six, seven, eight, nine, or 10 or more cohesin domains. The arrangement of the modules on the scaffoldin subunit and the specificity of the cohesin(s) and/or dockerin for their modular counterpart determine the overall architecture of the cellulosome. Several different types of scaffoldins have been described: the primary scaffoldins incorporate the various dockerin-bearing subunits directly into the cellulosome complex, adaptor scaffoldins increase the
20 repertoire or number of components into the complex, and the anchoring scaffoldins attach the complex to the bacterial cell surface.

[0120] In certain embodiments the cellulolytic enzymes comprising the cellulosome or individually displayed on the surface of a microorganism include one or more variant enzymes described herein and, optionally, one or more additional enzymes collected from
25 the group consisting of an exoglucanase, an endoglucanase, a glycosyl hydrolase, a cellulase, a hemicellulase, a xylanase, a cellobiohydrolase, a β -glucosidase, a mannanase, a xylosidase (e.g., a β -xylosidase), an arabinofuranosidase, and/or a glucose oxidase.

[0121] Methods of designing and expressing cellulosomes and minicellulosomes are well known to those of skill in the art. Thus, for example, Mingardon *et al.* (2007) *Applied and Environmental Microbiology*, 73(12): 3822-3832 describe the assembly of
30 minicellulosomes on *Saccharomyces*. Tsai *et al.* (2010) *Appl Environ Microbiol.* 76(22):7514-7520 describing the expression of minicellulosomes using a synthetic yeast consortium comprising four different engineered yeast strains capable of either displaying a

trifunctional scaffoldin, Scaf-ctf (SC), carrying three divergent cohesin domains from *Clostridium thermocellum* (t), *Clostridium cellulolyticum* (c), and *Ruminococcus flavefaciens* (f), or secreting one of the three corresponding dockerin-tagged cellulases (endoglucanase [AT], exoglucanase [EC/CB], or β -glucosidase [BF]). Mingardon *et al.* (2005) *Appl Environ Microbiol.* 71(3):1215-1222 describe heterologous production, assembly, and secretion of a minicellulosome by *Clostridium acetobutylicum*. Hyeon *et al.* (2010) *FEMS Microbiol Lett* 310(1): 39-47 describe engineering *Saccharomyces cerevisiae* for assembly of minicellulosomes by heterologous expression of a recombinant scaffolding protein from *Clostridium cellulovorans* and a chimeric endoglucanase E from *Clostridium thermocellum*. The chimeric endoglucanase E fused with the dockerin domain of endoglucanase B from *C. cellulovorans* was assembled with the recombinant scaffolding protein.

[0122] Due to the production of highly versatile cellulosomes and the anaerobic, thermophilic, ethanogenic nature, of *C. thermocellum*, it, and other thermophilic microorganisms are good cellulosome expression systems for consolidated bioprocessing (CBP). CBP features the production of cellulases and hemicellulases, hydrolysis of cellulose and hemicellulose, and, fermentation of hydrolysis products, all in one step. Using a thermophilic strain such as *C. thermocellum* means less time for cooling and easy removal of ethanol at higher temperatures. It also means no addition of oxygen during the biorefining process and fermentation of glucose to produce ethanol and organic acids (*see, e.g. Demain et al. (2005) Microbiol Mole Biol Rev., 69(1): 124-154*).

[0123] It will be recognized that the cellulosome systems identified above are intended to be illustrative and not limiting. Using the teachings provided herein, numerous cellulosomes comprising one or more of the variant enzymes described herein and expression systems the display of such cellulosomes will be available to one of skill in the art.

Uses of the variant cellulolytic enzymes.

[0124] The variant cellulolytic enzymes and the nucleic acid constructs encoding such enzymes find utility in a wide variety of applications some of which are described below.

[0125] In certain embodiments the variant cellulolytic enzymes described herein find utility in detergent compositions that exhibit enhanced cleaning ability, function as a

softening agent and/or improve the feel of cotton fabrics (*e.g.*, "stone washing" or "biopolishing"), in compositions for degrading wood pulp into sugars (*e.g.*, for bio-ethanol production), and/or in feed compositions.

[0126] For example, the rate of hydrolysis of cellulosic products may be increased
5 by using a transformant expressing one or more copies of the enzymes having greater
cellulolytic activity described herein. This permits degradation of products that contain
cellulose or heteroglycans at a faster rate and to a greater extent. Products made from
cellulose such as paper, cotton, cellulosic diapers and the like can be degraded more
efficiently in a landfill. Thus, the product obtainable from the transformants or the
10 transformants alone may be used in compositions to help degrade by liquefaction a variety
of cellulose products that add to the overcrowded landfills.

[0127] Separate saccharification and fermentation is a process whereby cellulose
present in biomass, *e.g.*, corn stover, is converted to glucose and subsequently yeast strains
convert glucose into ethanol. Simultaneous saccharification and fermentation is a process
15 whereby cellulose present in biomass, *e.g.*, corn stover, is converted to glucose and, at the
same time and in the same reactor, yeast strains convert glucose into ethanol. Thus, in
another approach, the variant cellulase enzymes described herein find utility in the
degradation of biomass to ethanol. Ethanol production from readily available sources of
cellulose provides a stable, renewable fuel source.

20 [0128] Cellulose-based feedstocks are comprised of agricultural wastes, grasses and
woods and other low-value biomass such as municipal waste (*e.g.*, recycled paper, yard
clippings, *etc.*). Ethanol may be produced from the fermentation of any of these cellulosic
feedstocks. However, the cellulose must first be converted to sugars before there can be
conversion to ethanol.

25 [0129] A large variety of feedstocks may be used with the cellulolytic enzyme
variants described herein, and the one selected for use may depend on the region where the
conversion is being done. For example, in the Midwestern United States agricultural wastes
such as wheat straw, corn stover and bagasse may predominate while in California rice
straw may predominate. However, it should be understood that any available cellulosic
30 biomass may be used in any region.

[0130] In certain embodiments a cellulase composition containing one or more
variant cellulases described herein finds utility in ethanol production. Ethanol from this
process can be further used as an octane enhancer or directly as a fuel in lieu of gasoline

which is advantageous because ethanol as a fuel source is more environmentally friendly than petroleum derived products. It is believed that the use of ethanol will improve air quality and possibly reduce local ozone levels and smog. Moreover, utilization of ethanol in lieu of gasoline can be of strategic importance in buffering the impact of sudden shifts in
5 non-renewable energy and petrochemical supplies.

[0131] In various embodiments ethanol can be produced via saccharification and fermentation processes from cellulosic biomass such as trees, herbaceous plants, municipal solid waste and agricultural and forestry residues. However, the ratio of individual cellulase enzymes within a naturally occurring cellulase mixture produced by a microbe may not be
10 the most efficient for rapid conversion of cellulose in biomass to glucose. It is known that endoglucanases act to produce new cellulose chain ends which themselves are substrates for the action of cellobiohydrolases and thereby improve the efficiency of hydrolysis of the entire cellulase system. Therefore, the use of increased or optimized enzymes as described herein can greatly enhance the production of ethanol.

15 [0132] Thus, the enzyme variants described herein find use in the hydrolysis of cellulose to its sugar components. In one embodiment, a variant enzyme is added to the biomass prior to the addition of a fermentative organism. In a second embodiment, a variant enzyme is added to the biomass at the same time as a fermentative organism. Optionally, there may be other cellulase components present in either embodiment.

20 [0133] In certain embodiments the variant cellulase is displayed on a particle (or other surface), or on a microorganism (*e.g.*, on a yeast or other fungus, or bacteria).

[0134] In certain embodiments the variant cellulase is provided as a component of a cellulosome or minicellulosome displayed on a particle or a microorganism (*e.g.*, on a yeast or other fungus, or bacteria).

25 [0135] Where the enzyme(s), cellulosome, and/or minicellulosome is presented on a microorganism, the microorganism can be dormant or inactive and in which case the enzyme, cellulosome, and/or minicellulosome simply acts as an enzyme or enzyme complex facilitating the degradation of cellulosic materials to produce sugars.

[0136] In certain embodiments the microorganism is active and the enzyme,
30 cellulosome, and/or minicellulosome is displayed and contacted to the cellulosic material in a culture system (*e.g.*, in a consolidated bioreactor).

[0137] In another embodiment the cellulosic feedstock can be pretreated. Pretreatment may be by elevated temperature and the addition of either of dilute acid, concentrated acid or dilute alkali solution. In certain embodiments the pretreatment solution can be added for a time sufficient to at least partially hydrolyze the hemicellulose components and then neutralized.

[0138] In a typical biomass conversion process, enzymatic saccharification can produce sugars that are made available for biological or chemical conversions to other intermediates or end-products. Therefore, the sugars generated from biomass find use in a variety of processes in addition to the generation of ethanol. Non-limiting examples of such conversions are fermentation of glucose to ethanol, and other biological conversions of glucose to 2,5-diketo-D-gluconate (*see, e.g.* U.S. Patent No: 6,599,722), lactic acid, succinate, 1,3-propanediol, 2,3-butanediol, the chemical and biological conversions of xylose to xylitol (*see, e.g.*, WO 1998/021339), and the like.

[0139] In various embodiments detergent compositions employ besides the variant cellulase(s) described herein, a surfactant, including anionic, non-ionic and ampholytic surfactants, optionally a hydrolase, building agents, bleaching agents, bluing agents and fluorescent dyes, caking inhibitors, solubilizers, cationic surfactants and the like.

[0140] In various embodiments the cellulase(s) described herein can be added to the detergent composition either in a liquid diluent, in granules, in emulsions, in gels, in pastes, and the like. Such forms are well known to the skilled artisan. When a solid detergent composition is employed, the cellulase composition is preferably formulated as granules. Preferably, the granules can be formulated so as to contain a cellulase protecting agent (*see, e.g.*, U.S. Patent No: 6,162,782).

[0141] In various illustrative, but non-limiting embodiments the cellulase compositions can be employed from about 0.00005 weight percent to about 5 weight percent relative to the total detergent composition. More typically the cellulase compositions are employed from about 0.0002 weight percent to about 2 weight percent relative to the total detergent composition.

[0142] These uses are intended to be illustrative and non-limiting. Using the teachings provided herein other uses of the variant cellulases described herein will be available to one of skill in the art.

EXAMPLES

[0143] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 15 **Expressing and testing activity of Cellulase Variants****Generating Plasmid**

[0144] DNA encoding novel variants was PCR amplified using LIC-tailed gene-specific primers as follows: 0.3 μ M primers (forward 5'- TAC TTC CAA TCC AAT GCA ATG GAA CCG GCA TTC AAC TAC-3' (SEQ ID NO:8); reverse 5'-TTA TCC ACT
10 TCC AAT GTT ATT ATG CCA GAG CGC ACG-3'(SEQ ID NO:9), 1 ng template DNA, 1x KOD Hot Start Master Mix (#71842-3, EMD Chemicals), water to 50 μ L. Reactions were incubated at 95 °C for 2 min to activate the polymerase followed by 30x thermal cycling at 95 °C for 20 s, 60 °C for 10 s, and 70 °C for 30 s. Resulting amplicons were band-purified by agarose gel electrophoresis to remove non-specific amplicons and purified
15 by using the illustra GFX PCR DNA and Gel Band Purification kit (#28-9034-71, GE Healthcare) according to manufacture protocols. 0.2 pmol of purified amplicon was treated at 22 °C for 30 min and 75 °C for 20 min with 1 U of LIC-qualified T4 DNA polymerase (#70099-3, EMD Chemicals) in a 20 μ L reaction containing 2.5 mM dCTP, 5.0 mM DTT, and 1x T4 DNA polymerase buffer to generate sticky overhangs. 0.02 pmol of the treated
20 product was hybridized with 0.014 pmol of similarly prepared LIC vector in a 3.4 μ L reaction at 25 °C for 5 min followed by the addition of 1 μ L 25 mM EDTA and another 25 °C 5 min incubation.

Transforming E. coli

[0145] The hybridization reaction above was transformed into SHuffle™ competent
25 *E. coli* (#C3029H, New England Biolabs) as follows: 4.4 μ L of hybridized insert and vector were added to thawed SHuffle™ cells and incubated on ice for 30 min. The mixture was heat shocked at 42 °C for 30 s followed by a 1 hr outgrowth at 30 °C in 250 μ L SOC media. The mixture was split and plated onto 2 pre-warmed LB plates containing 100 μ g/mL carbenicillin and grown inverted for 16 hr at 37°C.

Generating Cellulase

[0146] Individual colonies were picked into 1 mL Magic Media (#K6803, Life Technologies) and incubated at 37 °C for 6 hrs followed by 25 °C for 18 hrs, both with shaking at 900 rpm. OD600 measurements were made on 50 µL of sample to determine cell growth. Cell pellets were recovered by decanting the supernatant after centrifuging at 3,000 rpm at 4 °C for 10 min. Cell pellets were lysed for 20 min in 200 µL BugBuster™ (#71456-3, EMD Chemicals) stock solution (40 mL BugBuster™, 400 µL 100 mM PMSF, 40 µL 1 mg/mL pepstatin, 40 µL 1 mg/mL leupeptin, 40 mg lysozyme). Lysed cells were centrifuged at 3,000 rpm at 4 °C for 30 min and the cleared lysate supernatant containing expressed cellulase variant was recovered for testing.

Determining Activity of Generated Cellulases

[0147] 5 µL of cleared lysate diluted 1:10 in 50 mM HEPES buffer pH 7.2 was combined with 45 µL of 1.1% carboxymethyl cellulose (CMC) and incubated at 50 °C for 30 min. After incubation, 140 µL of stock *p*-hydroxy-benzoic-acid-hydrazide (PHBAH) solution (0.179% solution of PHBAH in 0.1M NaOH) was combined with 10 µL of the CMC digestion and incubated at 95°C for 5 min. Absorbance at 410 nm from a 100 µL sample was measured to determine the amount of liberated reducing sugars by the cellulase variant. Measurements were calibrated to a glucose standard curve and compared to wild-type activity under identical conditions. Improved variant F1 and wild-type Cel9A and were tested at 24, 48, and 120 hours. Results are shown below in FIG. 3.

Example 2

Biofuel production by using enzyme variant F1

[0148] The following procedure may be used to produce ethanol from biomass. Generally, the procedure comprises simultaneous saccharification and fermentation (SSF) of pretreated lignocellulosic biomass whereby cellulases convert the biomass into accessible sugar and yeast ferment the sugar into ethanol.

[0149] Biomass may be pretreated in various manners. One method is to solubilize the biomass in concentrated phosphoric acid then precipitate the swollen cellulose using cold water. The cellulose is collected and washed with sufficient water to neutralize the pH. Alternatively, the biomass may be pretreated with dilute sulfuric acid. Briefly, milled and washed biomass at 20% total solids concentration is treated for 3–12 min in 0.5–1.41%

(w/w liquid phase) H₂SO₄ at 165–183°C. Following treatment, the biomass is washed with water.

[0150] Simultaneous saccharification and fermentation is conducted in a shaking incubator (150 rpm) at a working volume of 100 mL in 250-mL baffled flasks. The washed
5 pretreated biomass is loaded to a level of 6–7% (w/w) cellulose fraction and combined with 35 mg cellulase enzyme variant F1 per gram of cellulose and beta-glucosidase such as ACCELLERASE™ BG (Danisco A/S, Copenhagen, DK) at 0.05 mL product per gram of cellulose. The medium consists of yeast extract (1% [w/v]), peptone (2% [w/v]), and citrate
10 buffer (0.05 M). The initial pH is adjusted to 5.2 using NaOH, and then the culture is inoculated with the yeast, *Saccharomyces cerevisiae* D₅A, to achieve an initial optical density (at 600 nm) of 0.5. The flask is maintained at 32-38°C for 7 days.

[0151] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of
15 this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

CLAIMS

What is claimed is:

1. A variant cellulolytic enzyme, said variant cellulolytic enzyme comprising a glycoside hydrolase with a substitution at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A.
5
2. A variant cellulolytic enzyme, said variant cellulolytic enzyme comprising a glycoside hydrolase comprising a substitution at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A.
- 10 3. The cellulolytic enzyme of claim 2, wherein said enzyme has at least 70% sequence identity with *Thermobifida fusca* Cel9A.
4. The cellulolytic enzyme of claim 2, wherein said enzyme comprises no more than 5 additional variations at positions other than those corresponding to F64, A226, and/or E246 in *Thermobifida fusca* Cel9A.
- 15 5. The cellulolytic enzyme according to any one of claims 1-4, wherein said variant is a variant of a family 9 glycoside hydrolase.
6. The cellulolytic enzyme according to any one of claims 1-5, wherein said variant is a variant of a theme B subfamily of said family 9 glycoside hydrolase.
- 20 7. The cellulolytic enzyme according to any one of claims 1-6, wherein said variant comprises a substitution at a position corresponding to F64 in *Thermobifida fusca* Cel9A.
8. The cellulolytic enzyme according to any one of claims 1-6, wherein said variant comprises a substitution at a position corresponding to A226 in *Thermobifida fusca* Cel9A.
- 25 9. The cellulolytic enzyme according to any one of claims 1-6, wherein said variant comprises a substitution at a position corresponding to E246 in *Thermobifida fusca* Cel9A.

10. The cellulolytic enzyme according to any one of claims 1-6, wherein said variant comprises a substitution at positions corresponding to F64 and A226 in *Thermobifida fusca* Cel9A.

11. The cellulolytic enzyme according to any one of claims 1-6, wherein
5 said variant comprises a substitution at positions corresponding to F64 and E246 in *Thermobifida fusca* Cel9A.

12. The cellulolytic enzyme according to any one of claims 1-6, wherein said variant comprises a substitution at positions corresponding to A226 and E246 in *Thermobifida fusca* Cel9A.

10 13. The cellulolytic enzyme according to any one of claims 1-6, wherein said variant comprises a substitution at positions corresponding to F64, A226, and E246 in *Thermobifida fusca* Cel9A.

14. The cellulolytic enzyme according to any one of claims 1-13, wherein said substitution at a position corresponding position F64 in *Thermobifida fusca* Cel9A is a
15 substitution to I, V, L, or M.

15. The cellulolytic enzyme of claim 14, herein said substitution at a position corresponding position F64 in *Thermobifida fusca* Cel9A is a substitution to I.

16. The cellulolytic enzyme according to any one of claims 1-15, wherein said substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is
20 a substitution to V, I, L, M, or T.

17. The cellulolytic enzyme of claim 16, wherein said substitution at a position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V, I, L, or M.

18. The cellulolytic enzyme of claim 16, wherein said substitution at a
25 position corresponding position A226 in *Thermobifida fusca* Cel9A is a substitution to V.

19. The cellulolytic enzyme according to any one of claims 1-18, wherein said substitution at a position corresponding position E246 in *Thermobifida fusca* Cel9A is a substitution to G, A, N, or S.

20. The cellulolytic enzyme of claim 19, wherein said substitution at a position corresponding position E246 in *Thermobifida fusca* Cel9A is a substitution to G.

21. The cellulolytic enzyme according to any one of claims 1-20, wherein said enzyme is a variant of a glycoside hydrolase from an organism selected from the group consisting of archaea, bacteria, and eukaryota.

22. The cellulolytic enzyme of claim 21, wherein said enzyme is a variant of a glycoside hydrolase from a gram negative bacterium.

23. The cellulolytic enzyme of claim 21, wherein said enzyme is a variant of a glycoside hydrolase from a gram positive bacterium.

24. The cellulolytic enzyme of claim 21, wherein said enzyme is a variant of glycoside hydrolase from a bacterial family selected from the group consisting of *Thermofidia*, *Micromonospora*, *Cellulomonas*, *Listeria*, *Pseudomonas*, *Ruminococcus*, *Saccharophagus*, *Streptomyces*, *Vibrio*, *Xanthomonas*, and *Clostridium*.

25. The cellulolytic enzyme of claim 21, wherein said enzyme is a variant of *Thermobifida fusca* Cel9A.

26. The cellulolytic enzyme of claim 21, wherein said enzyme is a variant of *Clostridium phytofermentans* Cphy3367.

27. A variant cellulolytic enzyme having at least 70% sequence identity with the G03 (SEQ ID NO:6) enzyme and/or the G11 (SEQ ID NO:7) enzyme wherein said enzyme does not include a wild-type peptide.

28. The cellulolytic enzyme according to any one of claims 1-27, wherein said enzyme is attached to or operably linked to a signal peptide.

29. The cellulolytic enzyme according to any one of claims 1-28, wherein said cellulolytic enzyme is present and/or displayed on the surface of an organism selected from the group consisting of a phage, a fungus, an alga, and a bacterium.

30. The cellulolytic enzyme according to any one of claims 1-28, wherein said cellulolytic enzyme is present and/or displayed on a substrate or the surface of a particle.

31. The cellulolytic enzyme of claim 30, wherein said substrate or particle comprises a material selected from the group consisting of a plastic, a glass, a mineral, a synthetic polymer, a biological polymer, and a metal.

32. The cellulolytic of claim 31, wherein said substrate or particle
5 comprises a biological polymer selected from the group consisting of a carbohydrate, a protein, a nucleic acid, and a polysaccharide.

33. The cellulolytic enzyme of claim 30, wherein said substrate or particle comprises a surface of a microfluidic channel or chamber, a surface of a bioreactor, or a surface of a bioreactor.

10 34. The cellulolytic enzyme according to any one of claims 1-28, wherein said cellulolytic enzyme is a component of a cellulosome or a minicellulosome.

35. The cellulolytic enzyme of claim 34, wherein said cellulosome is on the surface of a yeast, bacteria, or non-yeast fungus.

15 36. The cellulolytic enzyme according to any one of claims 34-35, wherein said cellulosome comprises one or more additional enzymes selected from the group consisting of an endocellulase, an exocellulase, a beta-glucosidase (cellobiase), an oxidative cellulose, a xylanase, a hemicellulase, a lichenase, a chitinase, a xylanase, a cellulose phosphorylase, and a cellulose disrupting protein.

20 37. A nucleic acid that encodes a cellulolytic enzyme according to any one of claims 1-28.

38. The nucleic acid of claim 37, wherein said nucleic acid comprises codons optimized for expression in a host cell.

25 39. The nucleic acid of claim 37, wherein said nucleic acid comprises codons optimized for expression in a host selected from the group consisting of a bacterium, a yeast, a fungus, an alga, an insect, and a mammalian cell.

40. The nucleic acid of claim 37, wherein said nucleic acid comprises codons optimized for expression in *E. coli*.

41. A vector comprising the nucleic acid according to any one of claims 37-40.

42. A host cell comprising the nucleic acid according to any one of claims 37-40 and/or the vector of claim 41.

43. The host cell of claim 42, wherein said host cell is selected from the group consisting of an archiobacterium, a bacterium, a yeast cell, a fungal cell, an algal cell,
5 a plant cell, an insect cell, and a mammalian cell.

44. A method of producing a cellulase, said method comprising:
a) culturing the host cell according to any one of claims 42-43 in a suitable culture medium under suitable conditions to produce cellulase; and
(b) obtaining the produced cellulase.

10 45. A method of producing an enzyme variant, comprising introducing a substitution in the amino acid sequence of an endoglucanase at one or more positions corresponding to one or more of residues F64, A226, and/or E246 in *Thermobifida fusca* Cel9A.

15 46. A method of degrading cellulosic biomass into fermentable sugars, said method comprising:
contacting said cellulosic biomass with a cellulolytic enzyme according to any one of claims 1-36 under conditions in which said enzyme partially or fully degrades cellulose in said cellulosic biomass to form one or more fermentable sugars.

20 47. The method of claim 46, wherein said cellulosic biomass comprises one or more materials selected from the group consisting of an agricultural plant waste (*e.g.*, corn stover, cereal straw, sugarcane bagasse), a plant waste from an industrial processes (*e.g.*, sawdust, paper pulp), an a non-food energy crop (*e.g.*, switchgrass).

25 48. The method of claim 47, wherein said cellulosic biomass comprises one or more materials selected form the group consisting of grasses, rice hulls, bagasse, jute, hemp, flax, bamboo, sisal, abaca, straw, corn cobs, corn stover, alfalfa, hay, coconut hair, seaweed, and algae.

49. A cellulolytic composition comprising a plurality of cellulolytic enzymes at least one of which is an enzyme according to any one of claims 1-26.

30 50. The composition of claim 49, wherein said composition comprise a second enzyme selected from the group consisting of an endocellulase, an exocellulase, a

beta-glucosidase (cellobiase), an oxidative cellulose, a xylanase, a hemicellulase, a lichenase, a chitinase, a xylanase, and a cellulose phosphorylase.

51. A detergent composition comprising a cellulolytic enzyme according to any one of claims 1-26, and a surfactant.

5 52. The detergent composition of claim 51, wherein said detergent comprises a laundry detergent, a dish detergent, or an industrial detergent.

53. The use of a cellulolytic enzyme according to any one of claims 1-26 in the treatment of a cellulose containing textile.

10 54. The use of claim 48, wherein said use comprises stone washing or dyeing of a fabric.

55. The use of a cellulolytic enzyme according to any one of claims 1-26 in the treatment of wood pulp or other lignocellulosic material.

56. The use of a cellulolytic enzyme according to any one of claims 1-26 in the reduction of biomass to glucose.

15

Amino acid sequence of novel *Thermobifida fusca* Cel9A variant ("F1")

EPAFNYAEALQKSMFFYEAQRSGLPENNRVSWRGDSGLNDGADVGLDLTGG
WYDAGDHVKFGIPMAFTATMLAWGAIESPGEYIRSGQMPYLKDNLRWVNDYF
IKAHPSPNVLYVQVGDGDADHKWWGPAEVMPEMERPSFKVDPSCPGSDVAAET
AAAMAASSIVFADDDPAYAATLVQHAKQLYTFADTYRQVYSDCVPAFAFYNS
WSGYQDELVWGAYWLYKVTGDDSYLAKAEYEYDFLSTGQQTDLRSYRWTIAW
DDKSYGTYVLLAKETGKQKYIDDANRWLDYWTVGVNGQRVPYSPGGMAVLDT
WGALRYAANTAFAVALVYAKVIDDPVRKQRYHDFAVRQINYALGDNPRNSSYV
VGFNNPPRNPHHRTAHGSWTDSIASPAENRHVLYGALVGGPGSPNDAYTDD
RQDYVANEVATDYNAGFSSALAMLVEEYGGTPLADFPPTTEEPDGPEIFVEAQ
INTPGTTFTEIKAMIRNQSGWPARMKDKGTFRYWFTLDEGVDPADITVSSAY
NQCATPEDVHHVSGDLYYVEIDCTGEKIFPGGQSEHRREVQFRIAGGPGWDP
SNDWSFQIGNELAPAPYIVLYDDGVPVWGTAPEEGEPEGGGEGPGGGGEEPG
EDVTPPSAPGSPAARDVTSTSAVLTWSASSDTGGSGVAGYDVFLRAGTGQEQ
KVGSTTRTSFTLTGLEPDTTYIAAVVARDNAGNVSQRSTVSFTTLAENGGGP
DASCTVGYSTNDWDSGFTASIRITYHGTAPLSSWELSFTFPAGQQVTHGWNA
TWRQDGAAVTATPMSWNSSLAPGATVEVGFNGSWSGSNTPTDFTLNGEPCA
LA

Fig. 1A

DNA sequence of F1

GAACCGGCATTCAACTACGCGGAAGCACTGCAAAAATCCATGTTCTTCTACGAAGCTCAAC
 GCTCAGGTAAACTGCCGAAAATAACCGTGTTTCGTGGCGCGGTGATAGCGGTCTGAACGA
 TGGTGCAGACGTCCGGCCTGGATCTGACCGGCGGTTGGTATGATGCCGGTGACCATGTGAAA
 TTTGGCaTCCCgATGGCTTTCACCGCGACGATGCTGGCCTGGGGTGCAATTGAATCTCCGG
 AAGGTTATATCCGCAGTGGCCAGATGCCGTACCTGAAAAGATAACCTGCGTTGGGTGAATGA
 CTATTTTATTAAAGCGCATCCGTCTCCGAATGTTCTGTACGTTCAAGTCGGTGACGGCGAT
 GCCGACCACAAATGGTGGGGTCCGGCAGAAGTGATGCCGATGGAACGCCCGAGTTTCAAAG
 TCGATCCGTCTGTGCCCGGCGACGACGTGGCAGCAGAAACCGCAGCTGCCGATGGCCGCAAG
 CTCTATCGTCTTTGCCGATGACGATCCGGCGTATGCCGCAACCCTGGTGCAGCACGCTAAA
 CAACTGTACACCTTCGCGGACACGTATCGTGGTGTCTACTCTGATTGTGTGCCGGCTGGTG
 CGTTTTATAACAGTTGGTCCGGCTACCAGGATGAACTGGTGTGGGGTGCTTATTGGCTGTA
 CAAAGtGACCGGCGACGATAGCTATCTGGCCAAAGCAGAATATGAATACGATTTTCTGAGC
 ACCGgACAGCAAACGGATCTGCGTAGCTACCGCTGGACCATTGCGTGGGACGATAAAAAGCT
 ATGGCACCTATGTGCTGCTGGCCAAAGAAACGGGCAAACAGAAATATATCGACGATGCAAA
 CCGCTGGCTGGATTACTGGACCGTGGGTGTTAATGGCCAACGTGTTCCGTATAGCCGGGC
 GGTATGGCCGTCTTGATACTGGGGTGCCTGCGCTATGCCGCAAATACGGCTTTCGTGG
 CGCTGGTTTACGCCAAAGTTATTGACGATCCGGTCCGTAAACAGCGCTATCATGATTTTGC
 TGTGCGCAAATCAACTACGCGCTGGGTGATAACCCGCGTAATAGTTCCTATGTGGTTGGT
 TTCGGCAAACAATCCGCCGCGTAATCCGCATCACCGTACC GCGCATGGCTCGTGGACGGATA
 GCATTGCCTCTCCGGCAGAAAACCGCCACGTCTGTATGGTGCCTGGTGGGCGGTCCGGG
 CTCCCCGAATGACGCGTATAACCGACGATCGTCAGGATTACGTGGCCAACGAAGTTGCAACG
 GATTATAATGCCGGCTTTTCATCGGCTCTGGCGATGCTGGTTGAAGAATACGGTGGCACCC
 CGCTGGCAGACTTTCGCCGACGGAAGAACCGGATGGTCCGGAAATTTTCGTTGAAGCGCA
 GATCAACACCCCGGGCACCACGTTTACGGAAATTAAGCTATGATCCGTAATCAAAGCGGT
 TGGCCGGCGCGCATGCTGGACAAAGGCACCTTTCGTTATTGGTTCACGCTGGATGAAGGTG
 TTGATCCGGCGGACATTACCGTTAGCTCTGCTTACAACAGTGCGCGACGCCGGAAGATGT
 CCATCACGTGTCCGGTGACCTGTATTACGTGGAAATTGATTGTACCGGCGAAAAAATCTTC
 CCGGGCGGTCAATCAGAACATCGTCGCGAAGTTCAATTTTCGTATCGCCGGCGGTCCGGGTT
 GGGACCCGTCTAACGACTGGAGTTTTCAGGGTATTGGCAATGAACTGGCCCCGGCACCGTA
 TATCGTGTGTACGACGATGGTGTCCCGTGTGGGGCACC GCACC GGAAGAAGGCGAAGAA
 CCGGGCGGTGGCGAAGGTCCGGGTGGCGGTGAAGAACC GGGCGAAGATGTCACCCCGCCGT
 CCGCACCGGGCTCACCGGCAGTTCGTGATGTCACCTCAACGTCGGCCGTTCTaACCTGGTC
 CGCAAGTTCGACACGGGCGGTT CAGGCGTGGCTGGCTATGATGTTTTCTGCGCGGGC
 ACCGGCCAGGAACAAAAGTGGGTCTACCACGCGTACGAGTTTACCCTGACGGGCCTGG
 AACCGGATACCAGTATATTGCTGCGGTCTGGCTCGCGATAACGCGGGTAATGTTAGTCA
 GCGTTCCACCGTCTCATTACACGCTGGCAGAAAACGCGGGTGGCCCGGATGCATCGTGC
 ACCGTTGGTTATAGCACGAATGATTGGGACTCCGGCTTACC GCCTCAATTCGCATCACCT
 ATCATGGCACCGCACCGCTGTATCGTGGAACTGAGTTTTACCTTCCCGGCTGGTCAGCA
 AGTGACCCACGGCTGGAATGCCACGTGGCGTCAGGATGGTGCCGCAGTTACC GCGACGCCG
 ATGTCTTGAACAGCTCTCTGGCTCCGGGTGCAACCGTTGAAGTCGGTTTTTAATGGCAGTT
 GGAGTGGTAGCAACACCCCGCCGACCGATTTACCCTGAATGGCGAACCGTGCCTCTGGC
 A

Fig. 1B

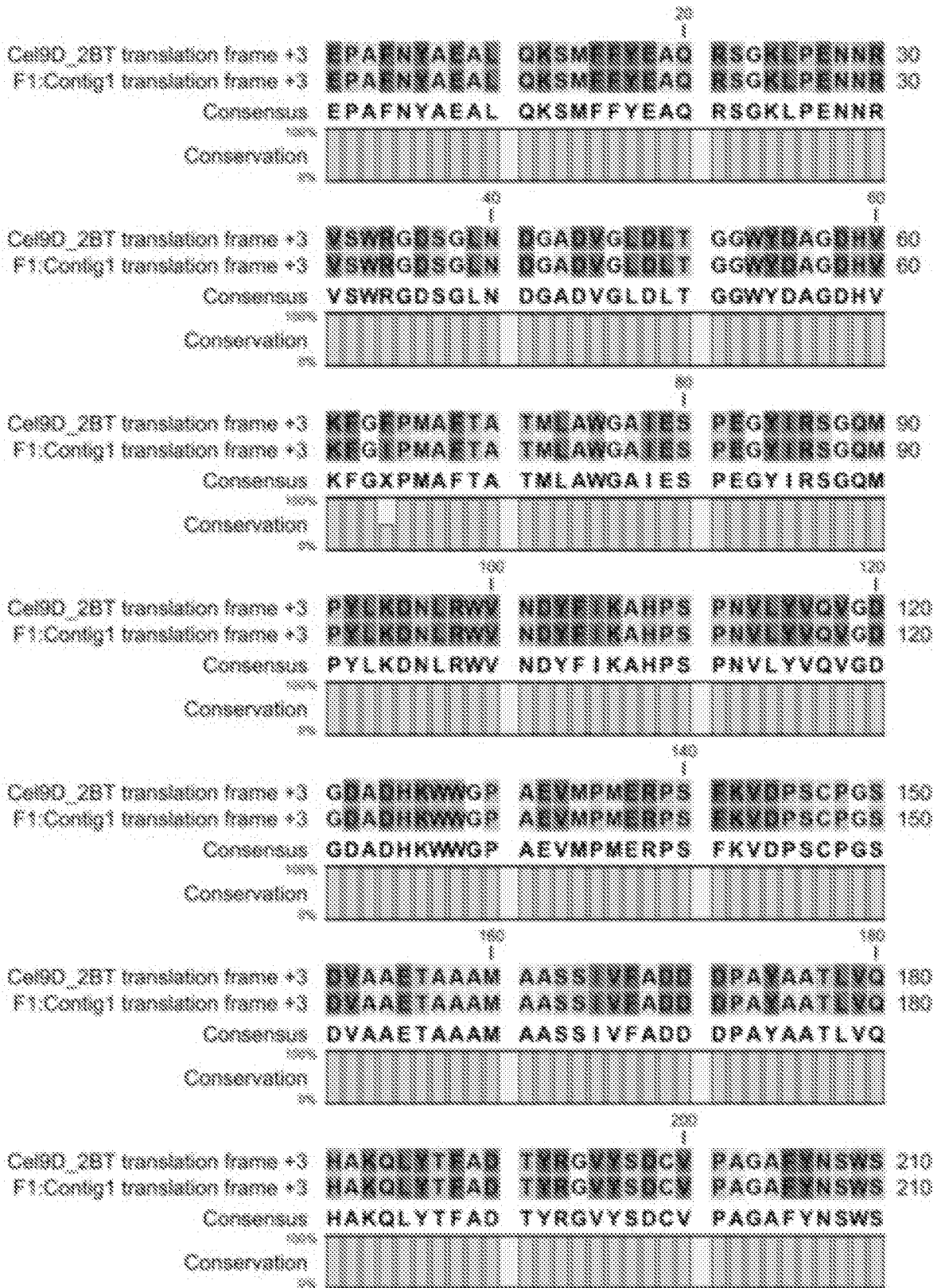


Fig. 2

4/11

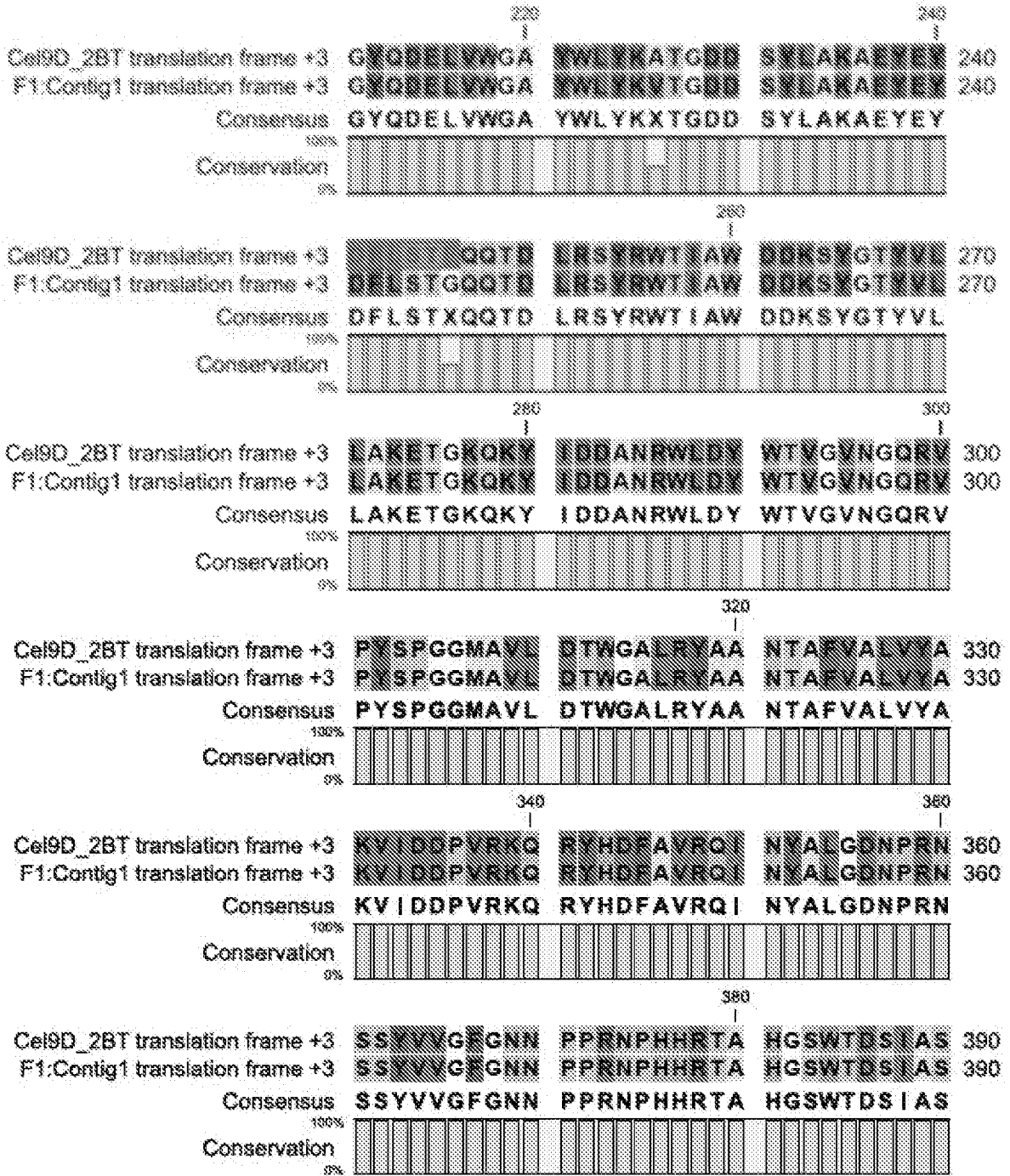


Fig. 2, cont'd.

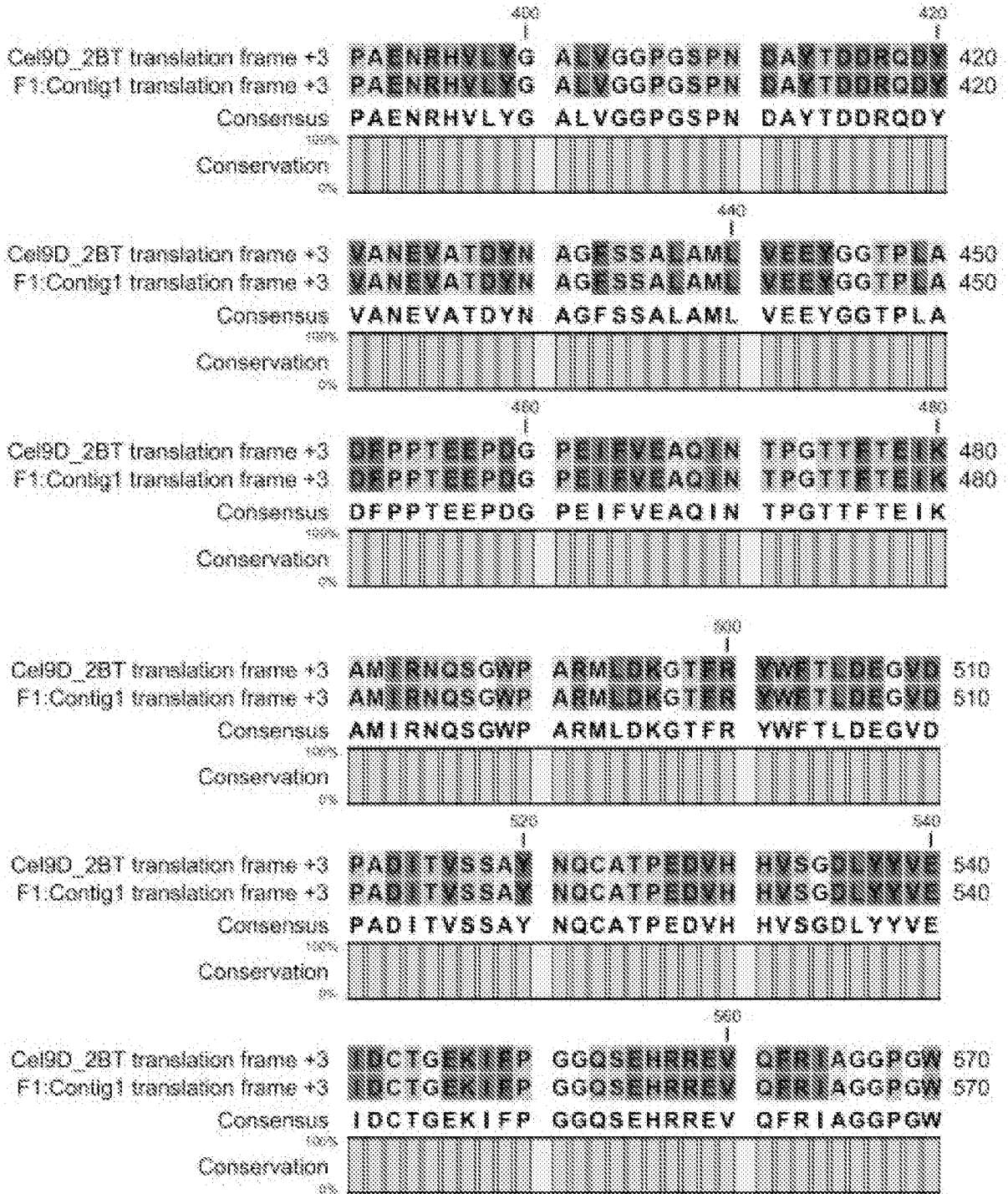


Fig. 2, cont'd.

6/11

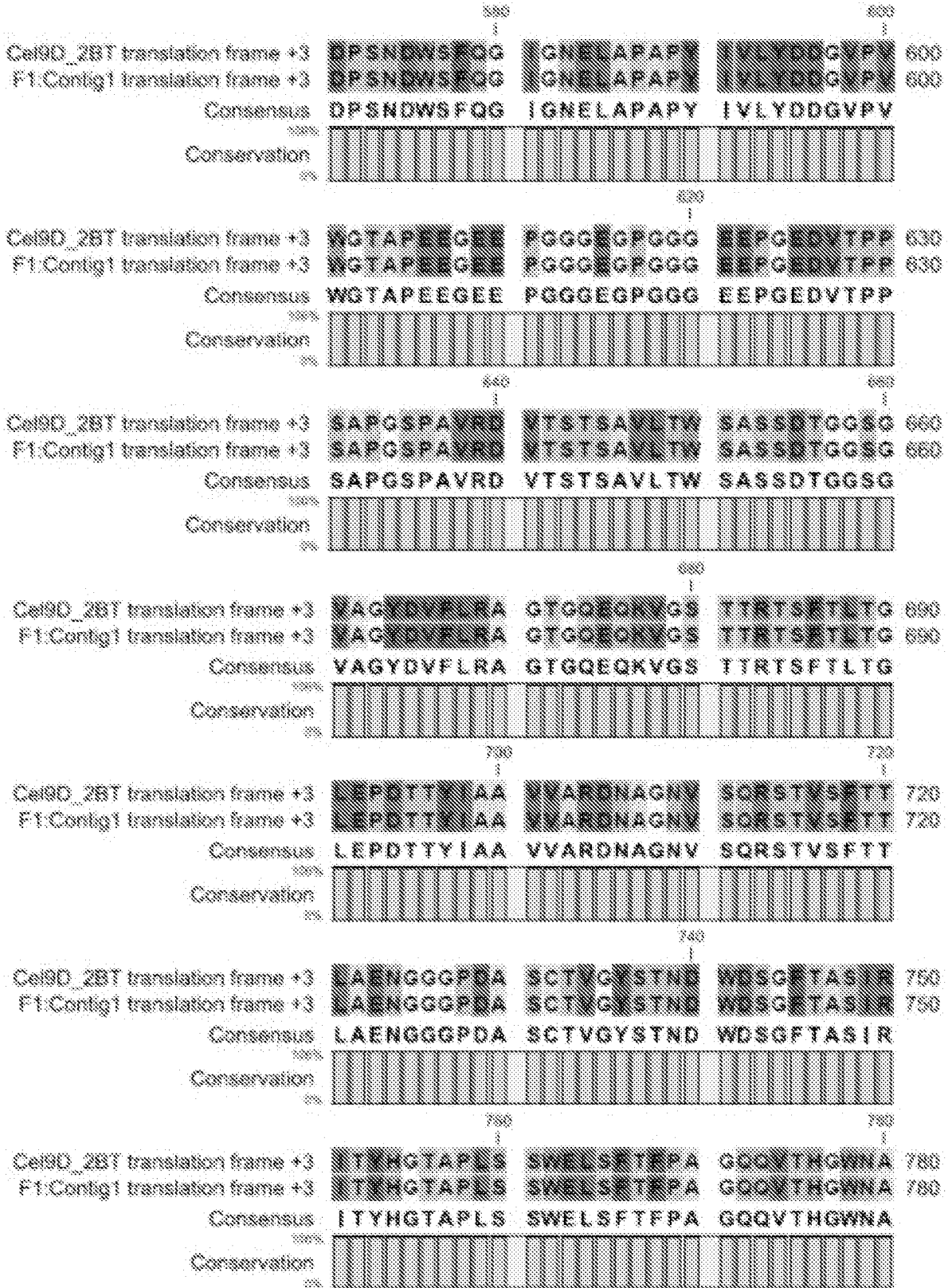


Fig. 2, cont'd.

7/11

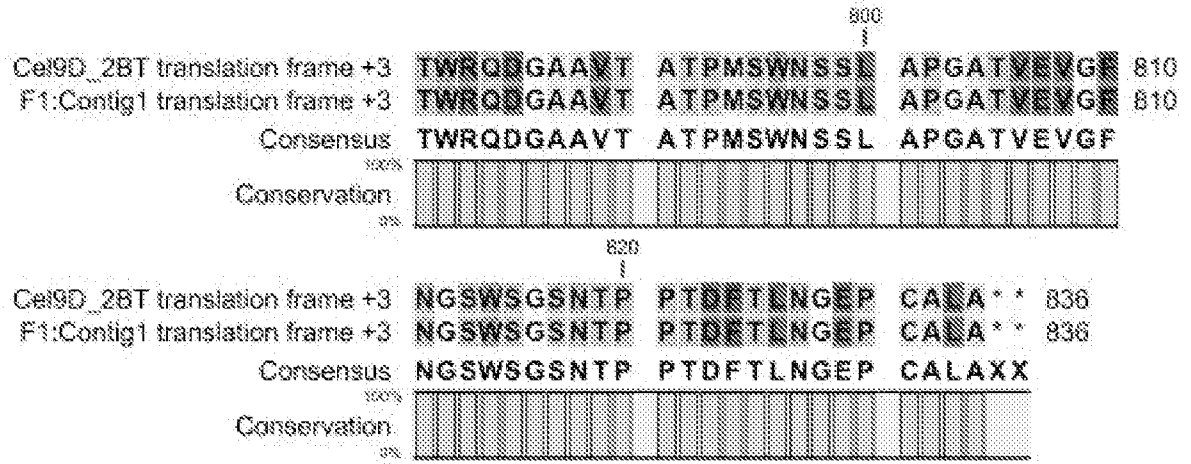


Fig. 2, cont'd.

8/11

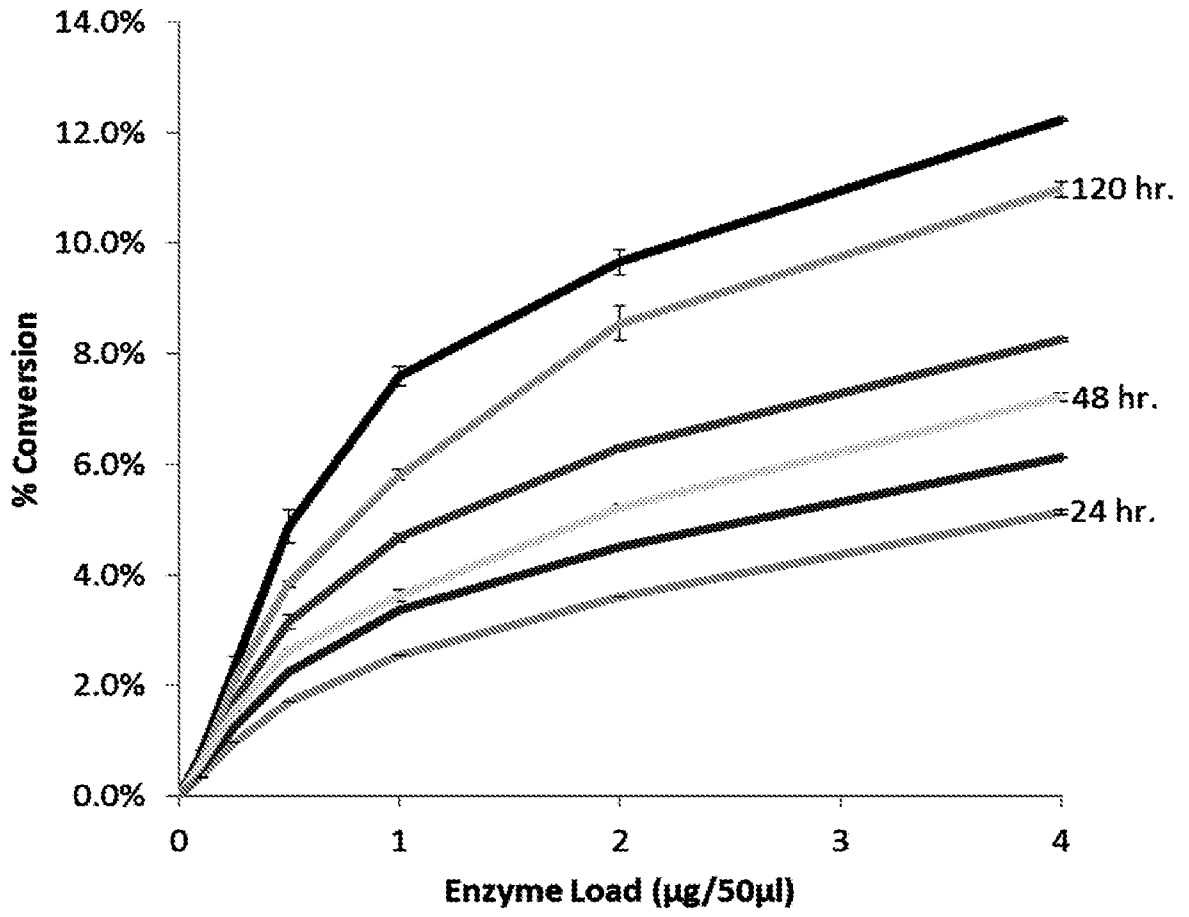


Fig. 3

9/11

Alignment between wild-type *Thermobifida fusca* Cel9A and *Clostridium phytofermentans* Cphy3367 and structures showing 3D similarity

Score = 612 bits (1577), Expect = 4e-179, Method: Compositional matrix adjust.

Identities = 318/618 (52%), Positives = 407/618 (66%), Gaps = 22/618 (3%)

Query = *Thermobifida fusca* Cel9A, Sbjct = *Clostridium phytofermentans* Cphy3367

Query	1	ETNYNYGEALQKSIMFYEFQRSGLPSTIRNNWRGDSGLTDGADVGLDLTGGWYDAGDHV	60
Sbjct	1	E +NY EALQKS+ FYE QRSGLP R +WRGDSGL DGADVGLDLTGGWYDAGDHV	60
Query	61	↓ KFNLPLAYTIVTMLAWAVYEEEEATLSKAGQLSYLLDEIKWSSDYLIKCHPQANVFYYQVGN	120
Sbjct	61	KFGFPMAFTATMLAWGAIESPEGYIRSGQMPYLKDNLRWVNDYFIKAHPSPNVLYVQVGD	120
Query	121	GNTDHSWWGPAEVMQMARPSYKVDLNNPGSTVVGEAAAALAATALIYKTKDPTYSATCLR	180
Sbjct	121	G+ DH WWGPAEVM M RPS+KVD + PGS V E AAA+AA++++ DP Y+AT ++	180
Query	181	HAKELFNFADTTK---SDAGYTAASGFYTSYSGFYDELSSWAATWIIYLASGEATYLDKAES	237
Sbjct	181	HAK+L+ FADT + SD A FY S+SG+ DEL W A W+Y A+G+ +YL KAE	238
Query	238	↓ YVAKWGTEPQSSTLSYKWAQNWDDVHYGAALLLARITNKAIYKNNIEMHLDYWTGYNCS	297
Sbjct	239	TE Q+ SY+W WDD YG +LLA+ T K Y ++ LDYWT G NG	298
Query	298	RITYTPKGLAWLDSWGALRYATTTAFSLASVYADWSGCSAGKVSTYNAFAKQQVDYALGST	357
Sbjct	299	R+ Y+P G+A LD+WGALRYA TAF+A VYA K Y+ FA +Q++YALG	357
Query	358	GR--SFVVGYGVNSPTRPHHRTAHSSWADSQTEPNYHRHTIYGALVGGPGN-NDSYEDNI	414
Sbjct	358	R S+VVG+G N P PHHRTAH SW DS P +RH +YGALVGGPG+ ND+Y D+	417
Query	415	NNYVNNEIACDYNAGFVGALAKVYKTYGGTPIANFKAIETVTNDELFIQAGINASGPSFI	474
Sbjct	418	+YV NE+A DYNAGF ALA + + YGGTP+A+F E E+F++A IN G +F	477
Query	475	EVKALVFNETGWPARTDKLSFKYFIDISEYVAKGYTKNDFTVSTNYNNGATTSALLPWD	534
Sbjct	478	E+KA++ N++GWPART DK +F+Y+ + E G D TVS+ YN AT +	531
Query	535	AANNIYVNVDFSGTKIYPGGQSAYKKEVQFRIAGPQNVNIWDNSNDYSFTQIANVSSGN	594
Sbjct	532	+ ++YV +D +G KI+PGGQS +++EVQFRIAG WD SND+SF I N	584
Query	595	TVKTTYIPLYDNGKLVFG	612
Sbjct	585	YI LYD+G V+G	602
Query	595	LAPAPYIVLYDDGVPVWG	602

Fig. 4

10/11

Clostridium phytofermentans Cphy3367



Thermobifida fusca Cel9A



Fig. 5

Additional examples of novel variants exhibiting increased cellulolytic activity

"G03"

EPAFN^YAEALQKSMFF^YEAQRSGLPENNRVSWRGDSGLNDGADVGLNLTGGWYDAGDHVK
 FGF^PMAFTATMLAWGAI^ESP^EGYIRSGQMPYLKDNLRWVNDYFIKAHSSPNVLYVQVGDGD
 ADHKW^GPAEVM^PMERPSFKVDPSCPGSDVAAETAAMAASSIVFADDDPAYAATLVQHAK
 QLYTFADTYRGVYSDCV^PPASAFYNSWSGYQDELVWGAYWLYKATGDDSYLAKAEYEYDFLS
 TEQQ^TDLRSYRWTIAWDDKSYGTYVLLAKETGKQKYIDDANRWLDYWTVGVNGQ^RVPYSPG
 GMAVLDTWGALRYAANTAFVALVYAKVIDDPVRKQRYHDFAVRQIN^YALGDNPRNSSYVVG
 FGNNPPRNPHHRTAHGSWTDSIASPAENRHVLYGALVGGPGSPNDAYTDDRQDYVANEVAT
 DYNAGFSSALAMLVEEYGGTPLADFPPT^EEPDGPEIFVEAQINTPGTTFTEIKAMIRN^QSG
 WPARMLDKGTFRYWFTLDEGVDPADITVSSAYNQCATPEDVHHVSGDLYYVEIDCTGEKIF
 PGGQSEHRREVQFRIAGGPGWDPSNDWSFQIGNELAPAPYIVLYDDGVPVWGTAPEEGEE
 PGGGEGPGGGE^EPGEDVTPPSAPGSPAVRDVTSTSAVLTWSASSDTGGSGVAGYDVFLRAG
 TGQE^QKVGS^TTRTSFTLLTGLEPDTTYIAAVVARDNAGNVSQRSTV^SF^TTLAENGGGPDASC
 TVGYSTNDWDSGFTASIRITYHGTA^PLSSWELSFTFPAGQ^VTHGWNATWRQDGTAVTATP
 MSWNSSLAPGATVEVGFNGSWSGSNT^PPTDFTLN^GEP^CALA

"G11"

EPAFN^YAEALQKSMFF^YEAQRSGLPENNRVSWRGDSGLNDGADVGLDLTGGWYDAGDHVK
 FGF^PMAFTATMLAWGAI^ESP^EGYIRSGQMPYLKDNLRWVNDYFIKAHSPSNVLYVQVGDGD
 ADHKW^GPAEVM^PMERPSFKVDPSCPGSDVAAETAAMAASSIVFADDDPAYAATLVQHAK
 QLYTFADTYRGVYSDCV^PPAGAFYNSWSGYQDELVWGAYWLYKATGDDSYLAKAEYEYDFLS
 TEQQ^TDLRSYRWTIAWDDKSYGTYVLLAKETGKQKYIDDANRWLDYWTVGVNGQ^RVPYSPG
 GMAVLDTWGALRYAANTAFVALVYAKVIDDPVRKQRYHDFAVRQIN^YALGDNPRNSSYVVG
 FGNNPPRNPHHRTAHGSWTDSIASPAENRHVLYGALVGGPGSPNDAYTDDRQDYVANEVAT
 DYN^SGFSSALAMLVEEYGGTPLADFPPT^EEPDGPEIFVEAQINTPGTTFTEIKAMIRN^QSG
 WPARMLDKGTFRYWFTLDEGVDPADITVSSAYNQCATPEDVHHVSGDLYYVEIDCTGEKIF
 PGGQSEHRREVQFRIAGGPGWDPSNDWSFQIGNELAPAPYIVLYNDGVPVWGTAPEEGEE
 PGGGEGPGGGE^EPGEDVTPPSAPGSPAVRDVTSTSAVLTWSASSDTGGSGVAGYDVFLRAG
 TGQE^QKVGS^TTRTSFTLLTGLEPDTTYIAAVVARDNAGNVSQRSTV^SF^TTLAENGGGPDASC
 TVGYSTNDWDSGFTASIR^FTYHGTA^SLSSWELSFTFPAGQ^VTHGWNATWRQDGA^AVTATP
 MSWNSSLAPGATVEVGFNGSWSGSNT^PPTDFTLN^GEP^CALA

Fig. 6