[0010] In one embodiment, the aglycosylated polypeptide is expressed by a gene that comprises a mutation proximal to or a mutation at one or more glycosylation site, wherein the mutation prevents glycosylation at the glycosylation site. In one embodiment, the mutation is at one or more of the threonine at amino acid position 78, the threonine at amino acid position 241, the serine at amino acid position 343, the serine at amino acid position 450, the threonine at amino acid position 599, the serine at amino acid position 616, the threonine at amino acid position 691, the serine at amino acid position 21, the threonine at amino acid position 24, the serine at amino acid position 25, the serine at amino acid position 28, the threonine at amino acid position 38, the threonine at amino acid position 42, the threonine at amino acid position 303, the serine at amino acid position at 398, the serine at amino acid position 435, the serine at amino acid position 436, the threonine at amino acid position 439, threonine at amino acid position 442, the threonine at amino acid position 446, the serine at amino acid position 451, the serine at amino acid position 619, the serine at amino acid position 622, the threonine at amino acid position 623, the serine at amino acid position 626, or the threonine at amino acid position 630 of SEQ ID NO: 1. A mutation proximal to one or more glycosylation site can prevent glycosylation at that site, e.g., by changing the conformation of the polypeptide or changing the consensus site recognized by the glycosylating enzyme such that glycosylation would not occur at the glycosylation site.

[0011] In one embodiment, the aglycosylated polypeptide hydrolyzes a carbohydrate, such as a dimer, a trimer, a tetramer, a pentamer, a hexamer, a heptamer, an octamer, or an oligomer of glucose; or an oligomer of glucose and xylose, into one or more monosaccharide, e.g., glucose. In one embodiment, the cellobiase activity comprises hydrolysis of a beta 1,4 glycosidic linkage of cellobiose.

[0012] In another aspect, the disclosure features a nucleic acid sequence encoding an aglycosylated polypeptide described herein.

[0013] In one embodiment, the nucleic acid sequence encodes a Cel3A enzyme or functional fragment thereof, wherein the nucleic acid sequence comprises (e.g., consists of) at least 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% identity to SEQ ID NO: 2 or SEQ ID NO: 3. In one embodiment, the nucleic acid sequence encodes a Cel3A enzyme, wherein the nucleic acid sequence comprises (e.g., consists of) SEQ ID NO: 2 or SEQ ID NO: 3.

[0031] In one embodiment, the reaction mixture has a ratio between the aglycosylated polypeptide and the remaining enzymes of at least 1:32, e.g., 1:32 to 1:300. In one embodiment, the reaction mixture has a ratio of the aglycosylated polypeptide to a glycosylated polypeptide is at least 1:32, e.g., 1:32 to 1:300.

[0032] In one aspect, the disclosure features a method of producing a product (e.g., hydrogen, a sugar, an alcohol, etc.) from a biomass (or converting a biomass to a product) comprising contacting a biomass, e.g., a biomass that has been treated with radiation, e.g., radiation from an electron beam, with an aglycosylated polypeptide described herein, under conditions suitable for the production of a sugar product. In one embodiment, the method further comprises contacting the biomass with a microorganism that produces one or more biomass-degrading enzyme or an enzyme mixture comprising biomass-degrading enzymes, e.g., an enzyme mixture described herein.

[0033] In one aspect, the disclosure features a method of producing a product (e.g., hydrogen, sugar, alcohol, etc.) from a biomass (or converting a biomass to a product) comprising contacting a biomass with an enzyme mixture described herein under conditions suitable for the production of the product.

[0034] In one embodiment, the product is a sugar product, e.g., a sugar product described herein. In one embodiment, the sugar product is glucose and/or xylose, or other sugar products, such as fructose, arabinose, galactose, and cellobiose.

[0035] In one embodiment, the method further comprises isolating the sugar product. In one embodiment, the isolating of the sugar product comprises precipitation, crystallization, chromatography, centrifugation, and/or extraction.

[0036] In one embodiment, the enzyme mixture comprises at least two of the enzymes selected from B2AF03, CIP1, CIP2, Cel1a, Cel3a, Cel5a, Cel6a, Cel7a, Cel7b, Cel12a, Cel45a, Cel74a, paMan5a, paMan26a, Swollenin, or any of the enzymes listed in Table 1. In an embodiment, the enzymes listed above are isolated from a cell that expresses the enzyme heterologously or endogenously, e.g., *Trichoderma reesei* or *Podospora anserina*.

[0037] In one embodiment, the biomass comprises a starchy material or a starchy material that includes a cellulosic component. In some embodiments, the biomass comprises one or more of an agricultural product or waste, a paper product or waste, a forestry product, or a

general waste, or any combination thereof; wherein: a) an agricultural product or waste comprises sugar cane, jute, hemp, flax, bamboo, sisal, alfalfa, hay, arracacha, buckwheat, banana, barley, cassava, kudzu, oca, sago, sorghum, potato, sweet potato, taro, yams, beans, favas, lentils, peas, grasses, switchgrass, miscanthus, cord grass, reed canary grass, grain residues, canola straw, wheat straw, barley straw, oat straw, rice straw, corn cobs, corn stover, corn fiber, coconut hair, beet pulp, bagasse, soybean stover, grain residues, rice hulls, oat hulls, wheat chaff, barley hulls, or beeswing, or a combination thereof; b) a paper product or waste comprises paper, pigmented papers, loaded papers, coated papers, filled papers, magazines, printed matter, printer paper, polycoated paper, cardstock, cardboard, paperboard, or paper pulp, or a combination thereof; c) a forestry product comprises aspen wood, particle board, wood chips, or sawdust, or a combination thereof; and d) a general waste comprises manure, sewage, or offal, or a combination thereof. In one embodiment, the method further comprises a step of treating the biomass prior to introducing the microorganism or the enzyme mixture to reduce the recalcitrance of the biomass, e.g., by treating the biomass with bombardment with electrons, sonication, oxidation, pyrolysis, steam explosion, chemical treatment, mechanical treatment, and/or freeze grinding.

[0038] In one embodiment, the microorganism that produces a biomass-degrading enzyme is from species in the genera selected from *Bacillus, Coprinus, Myceliophthora, Cephalosporium, Scytalidium, Penicillium, Aspergillus, Pseudomonas, Humicola, Fusarium, Thielavia, Acremonium, Chrysosporium* or *Trichoderma*. In one embodiment, the microorganism that produces a biomass-degrading enzyme is selected from *Aspergillus*, *Humicola insolens* (*Scytalidium thermophilum*), *Coprinus cinereus*, *Fusarium oxysporum*, *Myceliophthora thermophila*, *Meripilus giganteus*, *Thielavia terrestris*, *Acremonium persicinum*, *Acremonium acremonium*, *Acremonium brachypenium*, *Acremonium dichromosporum*, *Acremonium obclavatum*, *Acremonium pinkertoniae*, *Acremonium roseogriseum*, *Acremonium incoloratum*, *Acremonium furatum*, *Chrysosporium lucknowense*, *Trichoderma viride*, *Trichoderma reesei*, or *Trichoderma koningii*.

[0039] In one embodiment, the microorganism has been induced to produce biomass-degrading enzymes by combining the microorganism with an induction biomass sample under conditions suitable for increasing production of biomass-degrading enzymes compared to an uninduced microorganism. In one embodiment, the induction biomass sample comprises paper,

the functional fragment is 700 to 744 amino acids, 650 to 699 amino acids, 600 to 649 amino acids, 550 to 599 amino acids, 500 to 549 amino acids, 450 to 499 amino acids, 400 to 449 amino acids, 350 to 399 amino acids, 300 to 349 amino acids, 250 to 299 amino acids, 200 to 249 amino acids, 150 to 199 amino acids, 100 to 149 amino acids, or 50 to 99 amino acids. With regard to the ranges of amino acid length described above, the lowest and highest values of amino acid length are included within each disclosed range. In an embodiment, the functional fragment has at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 99% of the cellobiase activity as wild-type Cel3a or the polypeptide comprising SEQ ID NO: 1 produced in *E.coli*.

[0080] Assays for detecting cellobiase activity are known in the art. For example, detection of the amount of glucose released from cellobiose can be determined by incubating purified cellobiase with substrate, e.g., cellobiose, D-(+)-cellobiose, and detecting the resultant amount of free glucose after completion of the reaction. The amount of free glucose can be determined using a variety of methods known in the art. For example, dilutions of purified cellobiase are prepared in a buffer containing 50mM sodium citrate monobasic, pH 5.0 NaOH. The cellobiose substrate is added to the purified cellobiase in an amount such that the final concentration of cellobiose in the reaction mixture is 30mM. The reaction mixture is incubated under conditions suitable for the reaction to occur, e.g., in a shaker (700 rpm) at 48°C for 30 minutes. To stop the reaction, the reaction mixture is heated for 5 minutes at 100°C. The reaction mixture is filtered through a 0.45μm filter and the filtrate is analyzed to quantify the amount of glucose and/or cellobiase. A YSI instrument that measures analytes such as glucose can be used to determine the concentration of glucose produced from the reaction. Alternatively, UPLC (Ultra Performance Liquid Chromatography) can be used to determine the concentration of glucose and cellobiose from the reaction. This assay can be formatted in a single reaction or in multiple reaction formats, e.g., 96 well format. In some embodiments, the multiple reaction format may be preferred to generate an activity curve representing cellobiase activity with respect to different concentrations of the purified cellobiase. The concentration of the purified cellobiase can be determined using a standard Bradford assay. Dilutions of the purified cellobiase assay are prepared, e.g., 2-fold dilutions, and are aliquoted into a 96 well plate, e.g., 12 wells of 2-fold dilutions. Cellobiose substrate is added as previously described,

[0085] In embodiments, the functional variant contains one or more mutations wherein one or more glycosylation sites present in the Cel3a polypeptide expressed by a nucleic acid sequence described herein that has been mutated such that a glycan can no longer be attached or linked to the glycosylation site. In another embodiment, the functional variant contains one or more mutations proximal to one or more glycosylation sites present in the Cel3a polypeptide expressed by a nucleic acid sequence described herein that has been mutated such that a glycan can no longer be attached or linked to the glycosylation site. For example, the mutation proximal to a glycosylation site mutates the consensus motif recognized by the glycosylating enzyme, or changes the conformation of the polypeptide such that the polypeptide cannot be glycosylated, e.g., the glycosylation site is hidden or steric hindrance due to the new conformation prevents the glycosylating enzymes from accessing the glycosylation site. A mutation proximal to a glycosylation site in the Cel3a polypeptide is directly adjacent to, or at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 15, at least 20, at least 30 or at least 40 amino acids from the glycosylation site that, as a result of the proximal mutation, will not be glycosylated.

[0086] In an embodiment, one or more of the following glycosylation sites of Cel3a, or SEQ ID NO: 1, are mutated: the threonine at amino acid position 78, the threonine at amino acid position 241, the serine at amino acid position 343, the serine at amino acid position 450, the threonine at amino acid position 599, the serine at amino acid position 616, the threonine at amino acid position 691, the serine at amino acid position 21, the threonine at amino acid position 24, the serine at amino acid position 25, the serine at amino acid position 28, the threonine at amino acid position 38, the threonine at amino acid position 42, the threonine at amino acid position 303, the serine at amino acid position at 398, the serine at amino acid position 435, the serine at amino acid position 436, the threonine at amino acid position 439, threonine at amino acid position 442, the threonine at amino acid position 446, the serine at amino acid position 451, the serine at amino acid position 619, the serine at amino acid position 622, the threonine at amino acid position 623, the serine at amino acid position 626, or the threonine at amino acid position 630, or any combination thereof. In embodiments, the glycosylation site is mutated from a serine or threonine to an alanine. For example, the aglycosylated polypeptide described herein has one or more of the following mutations: T78A, T241A, S343A, S450A, T599A, S616A, T691A, S21A, T24A, S25A, S28A, T38A, T42A,

T303A, T398A, S435A, S436A, T439A, T442A, T446A, S451A, S619A, S622A, T623A, S626A, or T630A, or any combination thereof. Alternatively, one or more amino acids proximal to the glycosylation sites described above is mutated.

[0087] Assays to detect whether a polypeptide is modified by a glycan (e.g., whether the polypeptide is glycosylated or aglycosylated) are known in the art. The polypeptide can be purified or isolated and can be stained for detection and quantification of glycan moieties, or the polypeptide can be analyzed by mass spectrometry, and compared to a corresponding reference polypeptide. The reference polypeptide has the same primary sequence as the test polypeptide (of which the glycosylation state is to be determined), but is either glycosylated or aglycosylated.

[0088] The aglycosylated polypeptides described herein have increased cellobiase activity compared to a corresponding glycosylated polypeptide, e.g., glycosylated Cel3a polypeptide. For example, the aglycosylated polypeptide having cellobiase activity has at least 1%, 2%, 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90% , 100%, or 200% cellobiase activity compared to the glycosylated polypeptide.

***Nucleic Acids***

[0089] The present invention also provides a nucleic acid sequence encoding a polypeptide having cellobiase activity of the present invention. In an embodiment, the nucleic acid sequence encodes a Cel3a enzyme or a functional fragment thereof with the amino acid sequence described herein.

[0090] In an embodiment, the nucleic acid sequence that encodes Cel3a is provided below:

ATGCGTTACCGAACAGCAGCTGCGCTGGCACTTGCCACTGGGCCCTTTGCTAGGGCAGACAGTCACTCAACATCGG
GGGCCTCGGCTGAGGCAGTTGTACCTCCTGCAGGGACTCCATGGGGAACCGCGTACGACAAGGCGAAGGCCGCATT
GGCAAAGCTCAATCTCCAAGATAAGGTCGGCATCGTGAGCGGTGTCGGCTGGAACGGCGGTCCTTGCGTTGGAAAC
ACATCTCCGGCCTCCAAGATCAGCTATCCATCGCTATGCCTTCAAGACGGACCCCTCGGTGTTCGATACTCGACAG
GCAGCACAGCCTTTACGCCGGGCGTTCAAGCGGCCTCGACGTGGGATGTCAATTTGATCCGCGAACGTGGACAGTT
CATCGGTGAGGAGGTGAAGGCCTCGGGGATTCATGTCATACTTGGTCCTGTGGCTGGGCCGCTGGGAAAGACTCCG
CAGGGCGGTCGCAACTGGGAGGGCTTCGGTGTCGATCCATATCTCACGGGCATTGCCATGGGTCAAACCATCAACG
GCATCCAGTCGGTAGGCGTGCAGGCGACAGCGAAGCACTATATCCTCAACGAGCAGGAGCTCAATCGAGAAACCAT
TTCGAGCAACCCAGATGACCGAACTCTCCATGAGCTGTATACTTGGCCATTTGCCGACGCGGTTCAGGCCAATGTC
GCTTCTGTCATGTGCTCGTACAACAAGGTCAATACCACCTGGGCCTGCGAGGATCAGTACACGCTGCAGACTGTGC

GTGATCGCAATAATCTGGACCCGTGGCATAACGGCAACGCCCTGGTTCAGGCAGTGGCAGGCGCAAATAGCAACGT
GATCGTTGTGGTGCATAGCGTGGGTGCCATCATTCTGGAGCAGATCCTGGCCCTGCCGCAAGTTAAGGCAGTTGTG
TGGGCAGGTCTGCCGAGCCAAGAAAGTGGCAATGCCCTGGTGGACGTTCTGTGGGGCGATGTTAGTCCGAGCGGCA
AGCTGGTGTATACAATCGCCAAGAGCCCGAACGACTATAACACCCGCATCGTTAGCGGCGGCAGTGATAGCTTCAG
CGAGGGCCTGTTTATCGACTACAAGCATTTCGATGATGCCAATATTACCCCGCGCTACGAATTTGGTTATGGCCTG
AGCTATACCAAGTTCAACTACAGCCGCCTGAGCGTTTTAAGTACCGCCAAGAGTGGTCCGGCAACAGGTGCCGTGG
TTCCTGGTGGTCCGAGTGATCTGTTTCAGAATGTGGCCACCGTGACCGTGGATATCGCCAACAGTGGTCAGGTTAC
CGGCGCCGAAGTGGCACAGCTGTACATCACCTATCCGAGCAGTGCACCGCGCACCCCGCCGAAACAGCTGCGTGGC
TTCGCCAAATTAAACCTGACCCCGGGCCAGAGCGGTACAGCAACCTTCAATATTCGCCGCCGTGATCTGAGCTATT
GGGACACCGCCAGCCAAAAATGGGTGGTGCCGAGCGGCAGCTTTGGCATTAGTGTGGGTGCAAGTAGCCGCGACAT
TCGCTTAACAAGCACCCTGAGTGTTGCC
(SEQ ID NO: 3)

[0093] In an embodiment, the nucleic acid sequence encoding a Cel3a enzyme or functional variant thereof comprises at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO: 2. In an embodiment, the nucleic acid sequence encoding a Cel3a enzyme or functional variant thereof comprises at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to SEQ ID NO:2 or SEQ ID NO: 3.

[0094] Provided herein is a nucleic acid sequence encoding an aglycosylated polypeptide, e.g., Cel3a polypeptide, as described above, in which one or more glycosylation sites present in the polypeptide has been mutated such that a glycan can no longer be attached or linked to the glycosylation site. In another embodiment, the nucleic acid sequence described herein encoding an aglycosylated polypeptide, e.g., a Cel3a polypeptide, as described above, in which one or more mutations proximal to one or more glycosylation sites present in the polypeptide has been mutated such that a glycan can no longer be attached or linked to the glycosylation site, as previously described. In an embodiment, the nucleic acid sequence encodes a polypeptide comprising one or more mutations at one or more of the following glycosylation sites of Cel3a, or SEQ ID NO: 1: the threonine at amino acid position 78, the threonine at amino acid position 241, the serine at amino acid position 343, the serine at amino acid position 450, the threonine at amino acid position 599, the serine at amino acid position 616, the threonine at amino acid position 691, the serine at amino acid position 21, the threonine at amino acid position 24, the serine at amino acid position 25, the serine at amino acid position 28, the threonine at amino acid position 38, the threonine at amino acid position 42, the

threonine at amino acid position 303, the serine at amino acid position at 398, the serine at amino acid position 435, the serine at amino acid position 436, the threonine at amino acid position 439, threonine at amino acid position 442, the threonine at amino acid position 446, the serine at amino acid position 451, the serine at amino acid position 619, the serine at amino acid position 622, the threonine at amino acid position 623, the serine at amino acid position 626, or the threonine at amino acid position 630, or any combination thereof. In embodiments, the glycosylation site is mutated from a serine or threonine to an alanine. For example, the nucleic acid sequence of the invention encodes an aglycosylated polypeptide comprising one or more of the following mutations: T78A, T241A, S343A, S450A, T599A, S616A, T691A, S21A, T24A, S25A, S28A, T38A, T42A, T303A, T398A, S435A, S436A, T439A, T442A, T446A, S451A, S619A, S622A, T623A, S626A, or T630A, or any combination thereof. The ordinarily skilled artisan could readily modify the nucleic acid sequence of wild-type Cel3a (SEQ ID NO: 2) to encode a polypeptide with one or more glycosylation site mutation by using methods known in the art, such as site-directed mutagenesis.

[0095] The techniques used to isolate or clone a nucleic acid sequence encoding a polypeptide are known in the art and include isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the nucleic acid sequences of the present invention from such genomic DNA can be effected, e.g., by using the well known polymerase chain reaction (PCR) or antibody screening of expression libraries to detect cloned DNA fragments with shared structural features. See, e.g., Innis et al., 1990, PCR: A Guide to Methods and Application, Academic Press, New York. Other amplification procedures such as ligase chain reaction (LCR), ligated activated transcription (LAT) and nucleotide sequence-based amplification (NASBA) may be used. The nucleic acid sequence may be cloned from a strain of *Trichoderma reesei*, e.g., wild-type *T. reesei,* or *T. reesei RUTC30*, or another or related organism and thus, for example, may be an allelic or species variant of the polypeptide encoding region of the nucleic acid sequence.

[0096] The nucleic acid sequence may be obtained by standard cloning procedures used in genetic engineering to relocate the nucleic acid sequence from its natural location to a different site where it will be reproduced. The cloning procedures may involve excision and isolation of a desired fragment comprising the nucleotide sequence encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a host

al., Molecular Cloning: A Laboratory Manual, 1989. In some preferred embodiments, the promoter is inducible, where the addition of a molecule stimulates the transcription and expression of the downstream reading frame.

[00101] Examples of suitable promoters for directing transcription of the nucleic acid constructs of the present invention in a eukaryotic host cell, e.g., in a fungal or yeast cell are promoters obtained from the genes of *Trichoderma Reesei*, methanol-inducible alcohol oxidase (AOX promoter), *Aspergillus nidulans* tryptophan biosynthesis (trpC promoter), *Aspergillus niger var. awamori* flucoamylase (*glaA*), *Saccharomyces cerevisiae* galactokinase (GAL1), or *Kluyveromyces lactis Plac4-PBI promoter.*

[00102] A control sequence present in the expression vector of the present invention may also be a signal sequence that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway, e.g., a secretion signal sequence. The signal sequence may be an endogenous signal sequence, e.g., where the signal sequence is present at the N-terminus of the wild-type polypeptide when endogenously expressed by the organism from which the polypeptide of interest originates from. The signal sequence may be a foreign, or heterologous, signal peptide, in which the signal sequence is from a different organism or a different polypeptide than that of the polypeptide of interest being expressed. Any signal sequence which directs the expressed polypeptide into the secretory pathway of a host cell may be used in the present invention. Typically, signal sequences are composed of between 6 and 136 basic and/or hydrophobic amino acids.

[00103] Examples of signal sequences suitable for the present invention include the signal sequence from *Saccharomyces cerevisiae* alpha-factor.

[00104] Fusion tags may also be used in the expression vector of the present invention to facilitate the detection and purification of the expressed polypeptide. Examples of suitable fusion tags include His-tag (e.g., 3xHis, 6x His (SEQ ID NO: 6), or 8xHis (SEQ ID NO: 7)), GST-tag, HSV-tag, S-tag, T7 tag. Other suitable fusion tags include myc tag, hemagglutinin (HA) tag, and fluorescent protein tags (e.g., green fluorescent protein). The fusion tag is typically operably linked to the N or C terminus of the polypeptide to be expressed. In some embodiments, there may be a linker region between the fusion tag sequence and the N-terminus

[00111] An expression vector comprising a nucleic acid sequence of the present invention is introduced into a host cell so that the vector is maintained (e.g., by chromosomal integration or as a self-replicating extra-chromosomal vector) such that the polypeptide is expressed.

[00112] The host cell may be a prokaryote or a eukaryote. The host cell may be a bacteria, such as an *E. coli* strain, e.g., K12 strains NovaBlue, NovaBlue T1R, JM109, and DH5α. Preferably, the bacteria cell has the capability to fold, or partially fold, exogenously expressed proteins, such as *E. coli* Origami strains, e.g., Origami B, Origami B (DE3), Origami 2, and Origami 2(DE3) strains. In some embodiments, it may be preferred to use a host cell that is deficient for glycosylation, or has an impaired glycosylation pathway such that proteins expressed by the host cell are not significantly glycosylated.

[00113] The host cell may be a yeast or a filamentous fungus, particularly those classified as *Ascomycota*. Genera of yeasts useful as host microbes for the expression of modified TrCel3A beta-glucosidases of the present invention include *Saccharomyces, Pichia, Hansenula, Kluyveromyces, Yarrowia*, and *Arxula*. Genera of fungi useful as microbes for the expression of the polypeptides of the present invention include *Trichoderma, Hypocrea, Aspergillus, Fusarium, Humicola, Neurospora, Chrysosporium, Myceliophthora, Thielavia, Sporotrichum* and *Penicillium*. For example, the host cell may be *Pichia pastoris*. For example, the host cell may be an industrial strain of *Trichoderma reesei,*or a mutant thereof, e.g., *T. reesei RUTC30*. Typically, the host cell is one which does not express a parental cellobiase or Cel3a.

[00114] The selection of the particular host cell, e.g., bacterial cell or a fungal cell, depends on the expression vector (e.g., the control sequences) and/or the method utilized for producing an aglycosylated polypeptide of the invention, as described in further detail below.

[00115] The expression vector of the invention may be introduced into the host cell by any number of methods known by one skilled in the art of microbial transformation, including but not limited to, transformation, treatment of cells with $CaCl_2$, electroporation, biolistic bombardment, lipofection, and PEG-mediated fusion of protoplasts (e.g. White et al., WO 2005/093072, which is incorporated herein by reference). After selecting the recombinant host cells containing the expression vector (e.g., by selection utilizing the selectable marker of the expression vector), the recombinant host cells may be cultured under conditions that induce the expression of the polypeptide having cellobiase activity of the invention.

combination of glycosylation inhibitors is used. Preferably, the concentration of glycosylation inhibitors used in this embodiment is sufficient to inhibit glycosylation of the polypeptide, but do not cause cytotoxicity or inhibition of protein expression of the host cell.

***Methods of Producing Products Using Aglycosylated Polypeptides***

[00126] The present invention provides methods and compositions for converting or processing a biomass into products, using an aglycosylated polypeptide having cellobiase activity, as described herein. Methods for converting a biomass to products, such as sugar products, are known in the art, for example, as described in US Patent Application 2014/0011258, the contents of which are incorporated by reference in its entirety. Briefly, a biomass is optimally pretreated, e.g., to reduce the recalcitrance, and saccharified by a saccharification process that involves incubating the treated biomass with biomass-degrading, or cellulolytic, enzymes to produce sugars (e.g., glucose and/or xylose). The sugar products can then be further processed to produce a final product, e.g., by fermentation or distillation. Final products include alcohols (e.g., ethanol, isobutanol, or n-butanol), sugar alcohols (e.g., erythritol, xylitol, or sorbitol), or organic acids (e.g., lactic acid, pyruvic acid, succinic acid).

[00127] Using the processes described herein, the biomass material can be converted to one or more products, such as energy, fuels, foods and materials. Specific examples of products include, but are not limited to, hydrogen, sugars (e.g., glucose, xylose, arabinose, mannose, galactose, fructose, cellobiose, disaccharides, oligosaccharides and polysaccharides), alcohols (e.g., monohydric alcohols or dihydric alcohols, such as ethanol, n-propanol, isobutanol, sec-butanol, tert-butanol or n-butanol), hydrated or hydrous alcohols (e.g., containing greater than 10%, 20%, 30% or even greater than 40% water), biodiesel, organic acids, hydrocarbons (e.g., methane, ethane, propane, isobutene, pentane, n-hexane, biodiesel, bio-gasoline and mixtures thereof), co-products (e.g., proteins, such as cellulolytic proteins (enzymes) or single cell proteins), and mixtures of any of these in any combination or relative concentration, and optionally in combination with any additives (e.g., fuel additives). Other examples include carboxylic acids, salts of a carboxylic acid, a mixture of carboxylic acids and salts of carboxylic acids and esters of carboxylic acids (e.g., methyl, ethyl and n-propyl esters), ketones

glycosylated polypeptide and the aglycosylated polypeptide have cellobiase activity. The aglycosylated polypeptide having cellobiase activity is any of the aglycosylated polypeptides described herein, e.g., produced using the methods described herein. The glycosylated polypeptide comprising at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to SEQ ID NO: 1 can be isolated or obtained from a microorganism that endogenously expresses the polypeptide.

[00148] In an embodiment, the glycosylated polypeptide and the aglycosylated polypeptide are both the Cel3A enzyme from wild-type *T. reesei*, e.g., comprising SEQ ID NO: 1.

[00149] In embodiments, the enzyme mixture further comprises at least one additional enzyme derived from a microorganism, wherein the additional enzyme has biomass or cellulose-based material-degrading activity, e.g., the additional enzyme is a cellulolytic enzyme, e.g., a cellulase. For example, the additional enzyme is a ligninase, an endoglucanase, a cellobiohydrolase, a xylanase, and a cellobiase. In an embodiment, the mixture further comprises one or more ligninase, one or more endogluconase, one or more cellobiohydrolase, and one or more xylanase. In embodiments, the additional biomass-degrading enzyme is glycosylated. In embodiments, the enzyme mixture further comprises at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 additional biomass-degrading enzymes described herein. For example, the enzyme mixture further comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15 or all of the enzymes listed in Table 1.

[00150] For example, the enzyme mixture further comprises a mixture of additional biomass-degrading enzymes produced by a microorganism, e.g., a fungal cell, such as wild-type *T. reesei*, or a mutant thereof, e.g., *T. Reesei RUTC30*. In an embodiment, the additional biomass-degrading enzymes are isolated from the microorganisms. In an embodiment, the mixture comprises one or more of the following biomass-degrading enzymes: B2AF03, CIP1, CIP2, Cel1a, Cel3a, Cel5a, Cel6a, Cel7a, Cel7b, Cel12a, Cel45a, Cel74a, paMan5a, paMan26a, or Swollenin, or any combination thereof. The additional biomass-degrading enzymes, e.g., listed above, can be endogenously expressed and isolated from the microorganism, e.g., fungal cell, from which the enzyme originates from (listed below in Table 1). Alternatively, the additional biomass-degrading enzymes, e.g., listed above, can be heterologously expressed

(including, but not limited to, e.g., a slime mold), or an algae. When the organisms are compatible, mixtures of organisms can be utilized.

[00199] Suitable fermenting microorganisms have the ability to convert carbohydrates, such as glucose, fructose, xylose, arabinose, mannose, galactose, oligosaccharides or polysaccharides into fermentation products. Fermenting microorganisms include strains of the genus *Saccharomyces* spp. (including, but not limited to, *S. cerevisiae* (baker's yeast), *S. distaticus, S. uvarum*), the genus *Kluyveromyces,* (including, but not limited to, *K. marxianus, K. fragilis*), the genus *Candida* (including, but not limited to, *C. pseudotropicalis,* and *C. brassicae*), *Pichia stipitis* (a relative of *Candida shehatae*), the genus *Clavispora* (including, but not limited to, *C. lusitaniae* and *C. opuntiae*), the genus *Pachysolen* (including, but not limited to, *P. tannophilus*), the genus *Bretannomyces* (including, but not limited to, e.g., *B. clausenii* (Philippidis, G. P., 1996, Cellulose bioconversion technology, in Handbook on Bioethanol: Production and Utilization, Wyman, C. E., ed., Taylor & Francis, Washington, D.C., 179-212) ). Other suitable microorganisms include, for example, *Zymomonas mobilis, Clostridium* spp. (including, but not limited to, *C. thermocellum* (Philippidis, 1996, supra), *C. saccharobutylacetonicum, C. saccharobutylicum, C. Puniceum, C. beijerncki,* and *C. acetobutylicum*), *Moniliella pollinis, Moniliella megachiliensis, Lactobacillus* spp. *Yarrowia lipolytica, Aureobasidium* sp., *Trichosporonoides* sp., *Trigonopsis variabilis, Trichosporon* sp., *Moniliellaacetoabutans* sp., *Typhula variabilis, Candida magnoliae, Ustilaginomycetes* sp., *Pseudozyma tsukubaensis,* yeast species of genera *Zygosaccharomyces, Debaryomyces, Hansenula* and *Pichia,* and fungi of the dematioid genus *Torula.*

[00200] For instance, *Clostridium* spp. can be used to produce ethanol, butanol, butyric acid, acetic acid, and acetone. *Lactobacillus* spp. can be used to produce lactic acid.

[00201] Many such microbial strains are publicly available, either commercially or through depositories such as the ATCC (American Type Culture Collection, Manassas, Va., USA), the NRRL (Agricultural Research Service Culture Collection, Peoria, Ill., USA), or the DSMZ (Deutsche Sammlung von Mikroorganismen and Zellkulturen GmbH, Braunschweig, Germany), to name a few.

[00202] Commercially available yeasts include, for example, Red Star®/Lesaffre Ethanol Red (available from Red Star/Lesaffre, USA), FALI® (available from Fleischmann's Yeast, a

Scientific, Catalog No. BP912). Colonies carrying the recombinant DNA were picked from plates for the inoculation of 2 ml starter cultures, and grown overnight at 37°C, then subsequently used to inoculate 100ml of LB media containing the appropriate antibiotics. Cultures were grown at 37°C until OD600 reached 0.8.

[00211] To induce protein expression, 500μM IPTG (Isopropyl-b-D-thiogalactopyranoside; Fisher Scientific, Catalog No. BP1755) was added. The expression culture was further grown for another 4 hours at 37°C. The cells were harvested by centrifugation at 4000g at room temperature (RT) for 20 minutes using the Sorvall St16 rotor TX400; and the pellet was stored at -80°C.

[00212] For protein extraction, the cell pellet was thawed on ice and resuspended in 2ml of native lysis buffer containing 50mM Tris-HCl pH 7.5, 0.1% Triton X-100, 5mM imidazole (Fisher Scientific, Catalog No. O3196), and 1mg/ml lysozyme (Fisher Scientific, Catalog No. BP535), then incubated on ice for 20 minutes. To digest the DNA in the sample, 2μl of lysonase (EMD Millipore, Catalog No. 71230) was added and incubated for another 10 minutes. The sample was then spun in a microcentrifuge at maximum speed for 10 minutes at 4°C. The clarified lysate was transferred to a fresh tube containing 100μl of pre-equilibrated Profinity (Biorad) Ni-charged IMAC resin slurry (Biorad, Catalog No. 732-4614). The native binding buffer consisted of 50mM Tris HCl pH 7.5, 150mM NaCl, 0.1% Triton X-100, and 5μM imidazole. The protein was batch-bound for 1 hour at RT, and then washed with native buffer containing 25μM imidazole. The protein was eluted in 300ml of native buffer containing 200μM imidazole. Purified Cel3a protein from the expression and purification process described above is shown in Figure 1.

**Example 3: Expression of Cel3a-C'His in Bioreactors**

[00213] In this example, the expression and purification of Cel3a is scaled up to 3 liter and 5 liter bioreactors. The skilled artisan could readily use the process described herein to scale the production further, for example, to 3L and 7L bioreactors.

[00214] For each bioreactor run, the pET Duet-1 Cel3a-C'His construct is transformed into *E. Coli* Origami B (DE3) host cell line, and streaked on plates containing LB medium and 100μg/ml ampicillin (Fisher Scientific, Catalog No. BP1760), 15 μg/ml kanamysin (Fisher

**What is claimed is:**

1. An aglycosylated polypeptide having cellobiase activity comprising at least 90% identity to SEQ ID NO: 1, or a functional fragment thereof.

2. The aglycosylated polypeptide of claim 1, wherein the polypeptide comprises a Cel3A enzyme from wild-type *T. reesei*, or a functional variant or fragment thereof.

3. The aglycosylated polypeptide of claim 2, wherein the Cel3A enzyme comprises the amino acid sequence SEQ ID NO: 1.

4. The aglycosylated polypeptide of any one of claims 1-3, wherein the polypeptide is encoded by a nucleic acid sequence comprising SEQ ID NO: 3 or a nucleic acid sequence at least 80% identical to SEQ ID NO: 3.

5. The aglycosylated polypeptide of any one of claims 1-4, wherein the polypeptide comprises a mutation proximal to or at one or more glycosylation sites, wherein the mutation prevents glycosylation at the one or more glycosylation sites, wherein optionally, the mutation is at one or more of the threonine at amino acid position 78, the threonine at amino acid position 241, the serine at amino acid position 343, the serine at amino acid position 450, the threonine at amino acid position 599, the serine at amino acid position 616, the threonine at amino acid position 691, the serine at amino acid position 21, the threonine at amino acid position 24, the serine at amino acid position 25, the serine at amino acid position 28, the threonine at amino acid position 38, the threonine at amino acid position 42, the threonine at amino acid position 303, the serine at amino acid position at 398, the at serine amino acid position 435, the serine at amino acid position 436, the threonine at amino acid position 439, threonine at amino acid position 442, the threonine at amino acid position 446, the serine at amino acid position 451, the serine at amino acid position 619, the serine at amino acid position 622, the threonine at amino acid position 623, the serine at amino acid position 626, or the threonine at amino acid position 630 of SEQ ID NO: 1.

6. The aglycosylated polypeptide of any one of the preceding claims, wherein the aglycosylated polypeptide:
i) has increased cellobiase activity as compared to glycosylated Cel3A enzyme from wild-type *T. reesei*;
ii) has increased substrate recognition, a more active substrate recognition site, or reduced steric hindrance as compared to glycosylated Cel3A enzyme from wild-type *T. reesei;* and/or
iii) hydrolyzes a carbohydrate such as cellobiose into one or more monosaccharides, e.g., glucose.

7. The aglycosylated polypeptide of any one of the preceding claims, wherein the cellobiase activity comprises hydrolysis of a beta 1,4 glycosidic linkage of cellobiose.

8. A nucleic acid sequence encoding the polypeptide of claims 1-7.

9. A nucleic acid sequence encoding a Cel3A enzyme or functional variant thereof, wherein the nucleic acid sequence comprises at least 80% identity to SEQ ID NO: 3.

10. A nucleic acid molecule comprising a nucleic acid sequence of any of claims 8 and 9.

11. The nucleic acid molecule of claim 10, further comprising a promoter, e.g., a promoter for prokaryotic cell expression, e.g., bacterial cell expression, e.g., expression in *E. coli,* wherein:
i) the promoter sequence is a constitutive promoter sequence, inducible promoter sequence, or a repressible promoter sequence; or
ii) the promoter is a T7 promoter.

12. The nucleic acid molecule of any one of claims 10-11, further comprising a nucleic acid sequence encoding:
i) a tag, e.g., a tag for detection and/or purification and/or for linkage to another molecule, e.g., a His tag; or
ii) one or more signal sequences, e.g., a secretion signals sequence.

13. An expression vector comprising the nucleic acid sequence of claims 8-12, optionally wherein, the vector comprises a nucleic acid sequence encoding a selection marker, optionally wherein the selection marker comprises a kanamycin or an ampicillin marker.

14. A cell comprising the expression vector of any one of claims 10-13, optionally wherein, the cell is a prokaryotic cell or bacterial cell.

15. A cell expressing a polypeptide of any one of claims 1-7, wherein, the cell is a prokaryotic cell or bacterial cell, and optionally wherein:
i) the cell is impaired for glycosylation; or
ii) the cell is an *E. coli* cell, wherein the *E. coli* cell is an origami *E. coli* cell.

16. A method for producing an aglycosylated polypeptide having cellobiase activity comprising at least 90% identity to SEQ ID NO: 1, or a functional fragment thereof, comprising culturing a cell comprising the nucleic acid of any one of claims 8-12 under conditions suitable for the expression of the polypeptide, wherein the cell does not glycosylate the polypeptide, e.g., a bacterial cell, e.g., an *E. coli* cell, e.g., an origami *E. coli* cell.

17. A method for producing an aglycosylated polypeptide having cellobiase activity comprising at least 90% identity to SEQ ID NO:1, or a functional fragment thereof comprising treating the polypeptide with a deglycosylating enzyme, optionally wherein, the deglycosylating enzyme is selected from PGNase and EndoH,

18. A method for producing an aglycosylated polypeptide having cellobiase activity comprising at least 90% identity to SEQ ID NO:1, or a functional fragment thereof, comprising culturing a cell comprising the nucleic acid of any one of claims 8-12, wherein the nucleic acid sequence has one or more mutations which prevent glycosylation of the encoded polypeptide.

19. A method for culturing a cell expressing an aglycosylated polypeptide having cellobioase activity comprising at least 90% identity to SEQ ID NO:1 or a functional fragment thereof, in the presence of a glycosylation inhibitor, e.g., tunicamycin.

20. An enzyme mixture comprising a glycosylated polypeptide comprising an amino acid sequence with at least 90% identity to SEQ ID NO: 1, or a functional fragment thereof, and an aglycosylated polypeptide comprising an amino acid sequence with at least 90% identity to SEQ ID NO: 1, or a functional fragment thereof, wherein both of the glycosylated polypeptide and the aglycosylated polypeptide have cellobiase activity.

21. The enzyme mixture of claim 20, wherein the aglycosylated polypeptide is the aglycosylated polypeptide of any one of claims 1-7.

22. The enzyme mixture of claim 20 or 21, wherein the glycosylated polypeptide and the aglycosylated polypeptide both comprise Cel3A enzyme from wild-type *T. reesei.*

23. The enzyme mixture of any one of claims 20-22, further comprising at least one additional enzyme derived from a microorganism, wherein the additional enzyme has a biomass-degrading activity of a cellulose based material, and optionally wherein, the additional enzyme is selected from a ligninase, an endoglucanase, a cellobiohydrolase, xylanase, and a cellobiase.

24. The enzyme mixture of claim 20 or 21, wherein the mixture further comprises one or more ligninase, one or more endoglucanase, one or more cellobiohydrolase, one or more xylanase.

25. The enzyme mixture of any one of claims 20-24, wherein the ratio between the aglycosylated polypeptide to the remaining enzymes or to the glycosylated polypeptide in the mixture is at least 1:32.

26. A method of producing a product from a biomass or converting a biomass to a product comprising contacting a biomass, with an aglycosylated polypeptide having cellobiase activity

comprising at least 90% identity to SEQ ID NO:1, or a functional fragment thereof, and a microorganism that produces one or more biomass-degrading enzyme or an enzyme mixture comprising biomass-degrading enzymes under conditions suitable for the production of the product.

27. A method of producing a product (e.g., a hydrogen, a sugar, an alcohol) from a biomass comprising contacting a biomass with an enzyme mixture of any one of claims 20-25 under conditions suitable for the production of the product.

28. The method of claim 26 or 27, wherein said product is a sugar product.

29. The method of any one of claims 26-28, further comprising isolating the sugar product, wherein the isolating of the sugar product comprises precipitation, crystallization, chromatography, centrifugation, or extraction.

30. The method of any one of claims 28 or 29, wherein the sugar product is glucose and/or xylose.

31. The method of any one of claims 26-30, wherein the enzyme mixture comprises at least two of the enzymes selected from the group consisting of B2AF03, CIP1, CIP2, Cel1a, Cel3a, Cel5a, Cel6a, Cel7a, Cel7b, Cel12a, Cel45a, Cel74a, paMan5a, paMan26a, Swollenin, and the enzymes listed in Table 1.

32. The method of any one of claims 26-31, wherein the biomass comprises one or more of an agricultural product or waste, a paper product or waste, a forestry product, or a general waste, or any combination thereof; wherein:

a) an agricultural product or waste comprises sugar cane, jute, hemp, flax, bamboo, sisal, alfalfa, hay, arracacha, buckwheat, banana, barley, cassava, kudzu, oca, sago, sorghum, potato, sweet potato, taro, yams, beans, favas, lentils, peas, grasses, switchgrass, miscanthus, cord grass, reed canary grass, grain residues, canola straw,

wheat straw, barley straw, oat straw, rice straw, corn cobs, corn stover, corn fiber, coconut hair, beet pulp, bagasse, soybean stover, grain residues, rice hulls, oat hulls, wheat chaff, barley hulls, or beeswing, or a combination thereof;

b) a paper product or waste comprises paper, pigmented papers, loaded papers, coated papers, filled papers, magazines, printed matter, printer paper, polycoated paper, cardstock, cardboard, paperboard, or paper pulp, or a combination thereof;

c) a forestry product comprises aspen wood, particle board, wood chips, or sawdust, or a combination thereof; and

d) a general waste comprises manure, sewage, or offal, or a combination thereof.

33. The method of any one of claims 26-32, further comprising a step of treating the biomass prior to introducing the microorganism or the enzyme mixture to reduce the recalcitrance of the biomass, wherein the treating comprises bombardment with electrons, sonication, oxidation, pyrolysis, steam explosion, chemical treatment, mechanical treatment, or freeze grinding.

34. The method of any one of claims 26-33 wherein the microorganism that produces a biomass-degrading enzyme is from species in the genera selected from *Bacillus, Coprinus, Myceliophthora, Cephalosporium, Scytalidium, Penicillium, Aspergillus, Pseudomonas, Humicola, Fusarium, Thielavia, Acremonium, Chrysosporium* or *Trichoderma.*

35. The method of any one of claims 26-34, wherein the microorganism that produces a biomass-degrading enzyme is selected from *Aspergillus*, *Humicola insolens (Scytalidium thermophilum) Coprinus cinereus, Fusarium oxysporum, Myceliophthora thermophila, Meripilus giganteus, Thielavia terrestris, Acremonium persicinum, Acremonium acremonium, Acremonium brachypenium, Acremonium dichromosporum, Acremonium obclavatum, Acremonium pinkertoniae, Acremonium roseogriseum, Acremonium incoloratum, Acremonium furatum, Chrysosporium lucknowense, Trichoderma viride, Trichoderma reesei,* or *Trichoderma koningii.*

36. The method of any one of claims 26-35, wherein the microorganism has been induced to produce a biomass-degrading enzyme by combining the microorganism with an induction

biomass sample under conditions suitable for increasing production of a biomass-degrading enzyme compared to an uninduced microorganism.

37. The method of claim 36, wherein said induction biomass sample comprises paper, paper products, paper waste, paper pulp, pigmented papers, loaded papers, coated papers, filled papers, magazines, printed matter, printer paper, polycoated paper, card stock, cardboard, paperboard, cotton, wood, particle board, forestry wastes, sawdust, aspen wood, wood chips, grasses, switchgrass, miscanthus, cord grass, reed canary grass, grain residues, rice hulls, oat hulls, wheat chaff, barley hulls, agricultural waste, silage, canola straw, wheat straw, barley straw, oat straw, rice straw, jute, hemp, flax, bamboo, sisal, abaca, corn cobs, corn stover, soybean stover, corn fiber, alfalfa, hay, coconut hair, sugar processing residues, bagasse, beet pulp, agave bagasse, algae, seaweed, manure, sewage, offal, agricultural or industrial waste, arracacha, buckwheat, banana, barley, cassava, kudzu, oca, sago, sorghum, potato, sweet potato, taro, yams, beans, favas, lentils, peas, or any combination thereof.