(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 May 2014 (08.05.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/070837 A1

(51) International Patent Classification:
C12N 9/42 (2006.01)

(21) International Application Number:
PCT/US2013/067413

(22) International Filing Date:
30 October 2013 (30.10.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/720,697 31 October 2012 (31.10.2012) US

(71) Applicant: **DANISCO US INC.** [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).

(72) Inventors: **BOWER, Benjamin, S.**; Danisco US Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US). **FUJDALA, Meredith, K.**; Danisco US Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US).

(74) Agent: **CHEN, Wenfang**; Danisco US Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

— *with international search report (Art. 21(3))*

(54) Title: BETA-GLUCOSIDASE FROM MAGNAPORTHE GRISEA

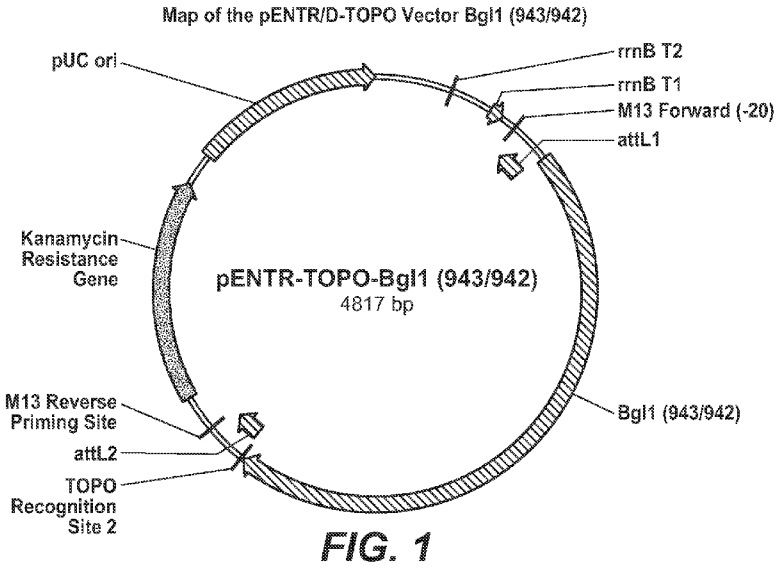

FIG. 1

(57) Abstract: The present compositions and methods relate to a beta-glucosidase from *Magnaporthe grisea*, polynucleotides encoding the beta-glucosidase, and methods of make and/or use thereof. Formulations containing the beta-glucosidase are suitable for use in hydrolyzing lignocellulosic biomass substrates.

WO 2014/070837 A1

BETA-GLUCOSIDASE FROM MAGNAPORTHE GRISEA

PRIORITY

[001] The present application claims priority to U.S. Provisional Application Serial No. 61/720,697, filed on October 31, 2012, which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[002] The present compositions and methods relate to a beta-glucosidase polypeptide obtainable from *Magnaporthe grisea* polynucleotides encoding the beta-glucosidase polypeptide, and methods of making and using thereof. Formulations and compositions comprising the beta-glucosidase polypeptide are useful for degrading or hydrolyzing lignocellulosic biomass.

DESCRIPTION OF THE BACKGROUND

[003] Cellulose and hemicellulose are the most abundant plant materials produced by photosynthesis. They can be degraded and used as an energy source by numerous microorganisms (*e.g.*, bacteria, yeast and fungi) that produce extracellular enzymes capable of hydrolysis of the polymeric substrates to monomeric sugars (Aro *et al.,* (2001) J. Biol. Chem., 276: 24309-24314). As the limits of non-renewable resources approach, the potential of cellulose to become a major renewable energy resource is enormous (Krishna *et al.,* (2001) Bioresource Tech., 77: 193-196). The effective utilization of cellulose through biological processes is one approach to overcoming the shortage of foods, feeds, and fuels (Ohmiya *et al.,* (1997) Biotechnol. Gen. Engineer Rev., 14: 365-414).

[004] Cellulases are enzymes that hydrolyze cellulose (comprising 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 ([beta]-D-glucoside glucohydrolase; EC 3.2.1.21) ("BG") (Knowles *et al.,* (1987) TIBTECH 5: 255-261; and Schulein, (1998) Methods Enzymol., 160: 234-243). Endoglucanases act mainly on the amorphous parts of the cellulose fiber, whereas cellobiohydrolases are also able to degrade crystalline cellulose (Nevalainen and Penttila, (1995) Mycota, 303-319). Thus, the presence of a cellobiohydrolase in a cellulase system is required for efficient solubilization of crystalline cellulose (Suurnakki *et al.,* (2000) Cellulose, 7: 189-209). Beta-glucosidase acts to liberate D-glucose units from cellobiose, cello-oligosaccharides, and other glucosides (Freer, (1993) J. Biol. Chem., 268: 9337-9342).

[005] Cellulases are known to be produced by a large number of bacteria, yeast and fungi. Certain fungi produce a complete cellulase system capable of degrading crystalline forms of cellulose. These fungi can be fermented to produce suites of cellulases or cellulase mixtures. The same fungi and other fungi can also be engineered to produce or overproduce certain cellulases, resulting in mixtures of cellulases that comprise different types or proportions of cellulases. The fungi can also be engineered such that they produce in large quantities *via* fermentation the various cellulases. Filamentous fungi play a special role since many yeast, such as *Saccharomyces cerevisiae,* lack the ability to hydrolyze cellulose in their native state (*see, e.g.,* Wood *et al.,* (1998) Methods in Enzymology, 160: 87-116).

[006] The fungal cellulase classifications of CBH, EG and BG can be further expanded to include multiple components within each classification. For example, multiple CBHs, EGs and BGs have been isolated from a variety of fungal sources including *Trichoderma reesei* (also referred to as *Hypocrea jecorina*), which contains known genes for two CBHs, *i.e.,* CBH I ("CBH1") and CBH II ("CBH2"), at least eight EGs, *i.e.,* EG I, EG II, EG III, EGIV, EGV, EGVI, EGVII and EGVIII, and at least five BGs, *i.e.,* BG1, BG2, BG3, BG4, BG5 and BG7 (Foreman *et al.* (2003), J. Biol. Chem. 278(34):31988-31997) . EGIV, EGVI and EGVIII also have xyloglucanase activity.

[007] In order to efficiently convert crystalline cellulose to glucose the complete cellulase system comprising components from each of the CBH, EG and BG classifications is required, with isolated components less effective in hydrolyzing crystalline cellulose (Filho *et al.,* (1996) Can. J. Microbiol., 42:1-5). Endo-1,4-beta-glucanases (EG) and exo-cellobiohydrolases (CBH) catalyze the hydrolysis of cellulose to cellooligosaccharides (cellobiose as a main product), while beta-glucosidases (BGL) convert the oligosaccharides to glucose. A synergistic relationship has been observed between cellulase components from different classifications. In particular, the EG-type cellulases and CBH-type cellulases synergistically interact to efficiently degrade cellulose. The beta-glucosidases serves the important role of liberating glucose from the cellooligosaccharides such as cellobiose, which is toxic to the microorganisms, such as, for example, yeasts, that are used to ferment the sugars into ethanol; and which is also inhibitory to the activities of endoglucanases and cellobiohydrolases, rendering them ineffective as further hydrolyzing the crystalline cellulose.

[008] In view of the important role played by beta-glucosidases in the degradation or conversion of cellulosic materials, discovery, characterization, preparation, and application of beta-glucosidase homologs with improved efficacy or capability to hydrolyze cellulosic feedstock is desirable and advantageous.

SUMMARY OF THE INVENTION

Beta-glucosidase obtainable from *Magnaporthe grisea* and their use

[009] Enzymatic hydrolysis of cellulose remains one of the main limiting steps of the biological production from lignocellulosic biomass feedstock of a material, which may be cellulosic sugars and/or downstream products. Beta-glucosidases play the important role of catalyzing the last step of that process, releasing glucose from the inhibitory cellobiose, and therefore its activity and efficacy directly contributes to the overall efficacy of enzymatic lignocellulosic biomass conversion, and consequently to the cost in use of the enzyme solution. Accordingly there is great interest in finding, making and using new and more effective beta-glucosidases.

[0010] While a number of beta-glucosidases are known, including the beta-glucosidases Bgl1, Bgl3, Bgl5, Bgl7, etc, from *Trichoderma reesei* or *Hyprocrea jecorina* (Korotkova O.G. *et al.*, (2009) Biochemistry 74:569-577; Chauve, M. et al., (2010) Biotechnol. Biofuels 3:3-3), the beta-glucosidases from *Humicola grisea var. thermoidea* (Nascimento, C.V. *et al.*, (2010) J. Microbiol. 48, 53-62); from *Sporotrichum pulverulentum,* Deshpande V. *et al.*, (1988) Methods Enzymol., 160:415-424); of *Aspergillus oryzae* (Fukuda T. et al, (2007) Appl. Microbiol. Biotechnol. 76:1027-1033, from *Talaromyces thermophilus* CBS 236.58 (Nakkharat P. *et al.*, (2006) J. Biotechnol., 123:304-313), from *Talaromyces emersonii* (Murray P., et al, (2004) Protein Expr. Purif. 38:248-257), so far the *Trichoderma reesei* beta-glucosidase Bgl1 and the *Aspergillus niger* beta-glucosidase SP188 are deemed benchmark beta-glucosidases against which the activities and performance of other beta-glucosidases are evaluated. It has been reported that *Trichoderma reesei* Bgl1 has higher specific activity than *Aspergillus niger* beta-glucosidase SP188, but the former can be poorly secreted, while the latter is more sensitive to glucose inhibition (Chauve, M. *et al.*, (2010) Biotechnol. Biofuels, 3(1):3).

[0011] One aspect of the present compositions and methods is the application or use of a highly active beta-glucosidase isolated from the fungal species *Magnaporthe grisea*, to hydrolyze a lignocellulosic biomass substrate. The genome of *Magnaporthe grisea,* a rice blast fungus, has been annotated in 2005 by Dean *et al.*, in Nature, 434(7036):980-986 (2005). The herein described sequence of SEQ ID NO:2 was published by National Center for Biotechnology Information, U.S. National Library of Medicine (NCBI) with the Accession No. XP_003709907, and designated to be a beta-glucosidase I. The enzyme has not been previously made in recombinant forms, or included in an enzyme composition useful for hydrolyzing a lignocellulosic biomass substrate. Nor has it or a composition comprising such an enzyme been applied to a lignocellulosic biomass substrate in a suitable method of enzymatic hydrolysis of

such a substrate. Furthermore, the beta-glucosidase of *Magnaporthe grisea* has not previously been expressed by an engineered microorganism. Nor has it been co-expressed with one or more cellulase genes and/or one or more hemicellulase genes. Expression in suitable microorganisms, which have, through many years of development, become highly effective and efficient producers of heterologous proteins and enzymes, with the aid of an arsenal of genetic tools, makes it possible to express these useful beta-glucosidases in substantially larger amounts than when they are expressed endogenously in an unengineered microorganism, or when they are expressed in plants. Enzymes classified as beta-glucosidases are diverse not only in their origins but also in their activities on lignocellulosic substrates, although most if not all beta-glucosidases can catalyze cellobiose hydrolysis under suitable conditions. For example, some are active on not only cellobiose but also on longer-chain oligosaccharides, whereas others are more exclusively active only on cellobiose. Even for those beta-glucosidases that have similar substrate preferences, some have enzyme kinetics profiles that make them more catalytically active and efficient, and accordingly more useful in industrial applications where the enzymatically catalyzed hydrolysis cannot afford to take longer than a few days at most. Furthermore, no fermenting or ethanologen microorganism capable of converting cellulosic sugars obtained from enzymatic hydrolysis of lignocellulosic biomass has been engineered to express a beta-glucosidase from *Magnaporthe grisea,* such as a Mg3A polypeptide herein. Expression of beta-glucosidases in ethanologen microorganisms provides an important opportunity to further liberating D-glucose from the remaining cellobiose that are not completely converted by the enzyme saccharification, where the D-glucose thus produced can be immediately consumed or fermented just in time by the ethanologen.

[0012] An aspect of the present composition and methods pertains to beta-glucosidase polypeptides of glycosyl hydrolase family 3 derived from *Magnaporthe grisea,,* referred to herein as "Mg3A" or "Mg3A polypeptides," nucleic acids encoding the same, compositions comprising the same, and methods of producing and applying the beta-glucosidase polypeptides and compositions comprising thereof in hydrolyzing or converting lignocellulosic biomass into soluble, fermentable sugars. Such fermentable sugars can then be converted into cellulosic ethanol, fuels, and other biochemicals and useful products. In certain embodiments, the Mg3A beta-glucosidase polypeptides have higher beta-gluclosidase activity and/or exhibits an increased capacity to hydrolyze a given lignocellulosic biomass substrate as compared to the benchmark *Trichderma reesei* Bgl1, which is a known, high fidelity beta-glucosidase. (Chauve, M. *et al.*, (2010) Biotechnol. Biofuels, 3(1):3).

[0013] In some embodiments, a Mg3A polypeptide is applied together with, or in the presence of, one or more other cellulases in an enzyme composition to hydrolyze or breakdown a suitable biomass substrate. The one or more other cellulases may be, for example, other beta-glucosidases, cellobiohydrolases, and/or endoglucanases. For example, the enzyme composition may comprise a Mg3A polypeptide, a cellobiohydrolase, and an endoglucanase. In some embodiments, the Mg3A polypeptide is applied together with, or in the presence of, one or more hemicellulases in an enzyme composition. The one or more hemicellulases may be, for example, xylanases, beta-xylosidases, and/or L-arabinofuranosidases. In further embodiments, the Mg3A polypeptide is applied together with, or in the presence of, one or more cellulases and one or more hemicellulases in an enzyme composition. For example, the enzyme composition comprises a Mg3A polypeptide, no or one or two other beta-glucosidases, one or more cellobiohydrolases, one or more endoglucanases; optionally no or one or more xylanases, no or one or more beta-xylosidases, and no or one or more L-arabinofuranosidases.

[0014] In certain embodiments, a Mg3A polypeptide, or a composition comprising the Mg3A polypeptide is applied to a lignocellulosic biomass substrate or a partially hydrolyzed lignocellulosic biomass substrate in the presence of an ethanologen microbe, which is capable of metabolizing the soluble fermentable sugars produced by the enzymatic hydrolysis of the lignocellulosic biomass substrate, and converting such sugars into ethanol, biochemicals or other useful materials. Such a process may be a strictly sequential process whereby the hydrolysis step occurs before the fermentation step. Such a process may, alternatively, be a hybrid process, whereby the hydrolysis step starts first but for a period overlaps the fermentation step, which starts later. Such a process may, in a further alternative, be a simultaneous hydrolysis and fermentation process, whereby the enzymatic hydrolysis of the biomass substrate occurs while the sugars produced from the enzymatic hydrolysis are fermented by the ethanologen.

[0015] The Mg3A polypeptide, for example, may be a part of an enzyme composition, contributing to the enzymatic hydrolysis process and to the liberation of D-glucose from oligosaccharides such as cellobiose. In certain embodiments, the Mg3A polypeptide may be genetically engineered to express in an ethanologen, such that the ethanologen microbe expresses and/or secrets such a beta-glucosidase activity. Moreover, the Mg3A polypeptide may be a part of the hydrolysis enzyme composition while at the same time also expressed and/or secreted by the ethanologen, whereby the soluble fermentable sugars produced by the hydrolysis of the lignocellulosic biomass substrate using the hydrolysis enzyme composition is metabolized and/or converted into ethanol by an ethanologen microbe that also expresses and/or secrets the Mg3A polypeptide. The hydrolysis enzyme composition can comprise the Mg3A polypeptide in

addition to one or more other cellulases and/or one or more hemicellulases. The ethanologen can be engineered such that it expresses the Mg3A polypeptide, one or more other cellulases, one or more other hemicellulases, or a combination of these enzymes. One or more of the beta-glucosidases may be in the hydrolysis enzyme composition and expressed and/or secreted by the ethanologen. For example, the hydrolysis of the lignocellulosic biomass substrate may be achieved using an enzyme composition comprising a Mg3A polypeptide, and the sugars produced from the hydrolysis can then be fermented with a microorganism engineered to express and/or secret Mg3A polypeptide. Alternatively, an enzyme composition comprising a first beta-glucosidase participates in the hydrolysis step and a second beta-glucosidase, which is different from the first beta-glucosidase, is expressed and/or secreted by the ethanologen. For example, the hydrolysis of the lignocellulosic biomass substrate may be achieved using a hydrolysis enzyme composition comprising *Trichoderma reesei* Bgl1, and the fermentable sugars produced from hydrolysis are fermented by an ethanologen microorganism expressing and/or secreting a Mg3A polypeptide, or *vice versa.*

[0016] As demonstrated herein, Mg3A polypeptides and compositions comprising Mg3A polypeptides have improved efficacy at conditions under which saccharification and degradation of lignocellulosic biomass take place. The improved efficacy of an enzyme composition comprising a Mg3A polypeptide is shown when its performance of hydrolyzing a given biomass substrate is compared to that of an otherwise comparable enzyme composition comprising Bgl1 of *Trichoderma reesei.*

[0017] In certain embodiments, the improved or increased beta-glucosidase activity is reflected in an improved or increased cellobiase activity of the Mg3A polypeptides, which is measured using cellobiose as substrate, for example, at a temperature of about 30°C to about 65°C (e.g., about 35°C to about 60°C, about 40°C to about 55°C, about 45°C to about 55°C, about 48°C to about 52°C, about 40°C, about 45°C, about 50°C, about 55°C, etc). In some embodiments, the improved beta-glucosidase activity of a Mg3A polypeptide as compared to that of *Trichoderma reesei* Bgl1, is observed when the beta-glucosidase polypeptides are used to hydrolyze a phosphoric acid swollen cellulose (PASC), for example, a thus pretreated Avicel pretreated using an adapted protocol of Walseth, TAPPI 1971, 35:228 and Wood, Biochem. J. 1971, 121:353-362. In some embodiments, the improved beta-glucosidase activity of a Mg3A polypeptide as compared to that of *Trichoderma reesei* Bgl1, is observed when the beta-glucosidase polypeptides are used to hydrolyze a dilute ammonia pretreated corn stover, for example, one described in International Published Patent Applications: WO2006110891,

WO2006110899, WO2006110900, WO2006110901, and WO2006110902; US Patent Nos. 7,998,713, 7,932,063.

[0018] In some embodiments, the improved beta-glucosidase activity is reflected in an increased robustness in the Mg3A polypeptide's capacity to catalyze the conversion or hydrolysis of a biomass substrate in a saccharification reaction conducted at a higher temperature, as compared to that of *Trichoderma reesei* Bgl1. For example, the difference or level of increased hydrolysis observed with Mg3A over what is observed with *Trichoderma reesei* Bgl1, is greater at a higher temperature than at a lower temperature, thus indicating that the Mg3A polypeptides of the disclosure are particularly suitable for biomass hydrolysis at a higher temperature due to its higher thermostability. In some embodiments, at a hydrolysis temperature of about 55°C, the Mg3A polypeptide can achieve at least 2%, for example, at least 5%, at least 7%, at least 10%, or even at least 15% higher level of hydrolysis of the same substrate, in the presence of equal amounts of other cellulases and hemicellases as compared to a *Trichoderma reesei* Bgl1 dosed at the same level. In certain embodiments, that difference between the level of hydrolysis of Mg3A vs. the level of hydrolysis of Bgl1 is higher at 55°C than at 50°C. It is noted that this improved thermostability or high temperature-performance of Mg3A is rather unexpected because the melting temperature (Tm) that is typically used to measure the unfolding or breakdown of polypeptides at higher temperature actually indicated a much higher Tm (e.g., about 10°C higher) for *Trichoderma reesei* Bgl1 than for Mg3A. While not wishing to be bound by theory, this surprising observation may suggest that Tm is not a good indicator or predictor of thermostability in the presence of biomass substrates.

[0019] In some aspects, a Mg3A polypeptide and/or as it is applied in an enzyme composition or in a method to hydrolyze a lignocellulosic biomass substrate is (a) derived from, obtainable from, or produced by *Magnaporthe grisea*; (b) a recombinant polypeptide comprising an amino acid sequence that is at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) identical to the amino acid sequence of SEQ ID NO:2; (c) a recombinant polypeptide comprising an amino acid sequence that is at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) identical to the catalytic domain of SEQ ID NO:2, namely amino acid residues 19-873; (d) a recombinant polypeptide comprising an amino acid sequence that is at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) identical to the mature form of amino acid sequence of SEQ ID NO:3, namely amino acid residues 19-873 of SEQ ID NO:2; or (e) a fragment of (a), (b), (c) or (d) having beta-glucosidase

activity. In certain embodiments, it is provided a variant polypeptide having beta-glucosidase activity, which comprises a substitution, a deletion and/or an insertion of one or more amino acid residues of SEQ ID NO:2.

[0020] In some aspects, a Mg3A polypeptide and/or as it is applied in an enzyme composition or in a method to hydrolyze a lignocellulosic biomass substrate is (a) a polypeptide encoded by a nucleic acid sequence that is at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%) sequence identity to SEQ ID NO:1, or (b) one that hybridizes under medium stringency conditions, high stringency conditions or very high stringency conditions to SEQ ID NO:1 or to a subsequence of SEQ ID NO:1 of at least 100 contiguous nucleotides, or to the complementary sequence thereof, wherein the polypeptide has beta-glucosidase activity. In some embodiments, a Mg3A polypeptide and/or as it is applied in a composition or in a method to hydrolyze a lignocellulosic biomass substrate is one that, due to the degeneracy of the genetic code, does not hybridize under medium stringency conditions, high stringency conditions or very high stringency conditions to SEQ ID NO:1 or to a subsequence of SEQ ID NO:1 of at least 100 contiguous nucleotide, but nevertheless encodes a polypeptide having beta-glucosidase activity and comprising an amino acid sequence that is at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100%) identical to that of SEQ ID NO:2 or to the mature beta-glucosidase sequence of SEQ ID NO:3. The nucleic acid sequences can be synthetic, and is not necessarily derived from *Magnaporthe grisea,* but the nucleic acid sequence encodes a polypeptide having beta-glucosidase activity and comprises an amino acid sequence that is least 75% identical to SEQ ID NO:2 or to SEQ ID NO:3.

[0021] In some preferred embodiments, the Mg3A polypeptide or the composition comprising the Mg3A polypeptide has improved beta-glucosidase activity, as compared to that of the wild type *Trichoderma reesei* Bgl1 (of SEQ ID NO: 4), or the enzyme composition comprising the *Trichoderma reesei* Bgl1. In certain embodiments, the cellulase activity of the Mg3A polypeptide of the compositions and methods herein, as measured using a Chloro-nitro-phenyl-glucoside (CNPG) hydrolysis assay, is at least about 5% higher, or about 10% higher, or about 15% higher, or about 20% higher than that of *Trichoderma reesei* Bgl1. The CNPG assay is described in **Example 2B** herein. In some embodiments, the cellulase activity of the Mg3A polypeptide of the compositions and methods herein, as measured using a CNPG hydrolysis assay, is at least about 2-fold higher, at least about 5-fold higher, at least about 6-fold higher, at least about 7-fold higher, at least about 8-fold higher, at least about 9-fold higher, or even at least about 10-fold higher than that of *Aspergillus niger* B-glu.

[0022] For example, the beta-glucosidase activity of the Mg3A polypeptide of the compositions and methods herein, as measured using a cellobiose hydrolysis assay, is at least about 20% higher (e.g., at least about 20% higher, at least about 30% higher, at least about 40% higher, at least about 50% higher, at least about 60% higher, at least about 70% higher, at least about 80% higher, at least about 85% higher, such as, for example at least about 87% higher) than that of the *Trichoderma reesei* Bgl1. The cellobiose hydrolysis assay is described in **Example 2C** herein. In some embodiments, the beta-glucosidase activity of the Mg3A polypeptide of the compositions and methods herein, as measured using a cellobiose hydrolysis assay, is about 2% less, about 5% less, about 7% less, about 10% less, such as, for example, about 13% less than that of the *Aspergillus niger* B-glu.

[0023] In some embodiments, the Mg3A polypeptides of the compositions and methods herein have substantially increased (e.g., at least about 20% higher, at least about 30% higher, at least about 40% higher, at least about 50% higher, at least about 60% higher, at least about 70% higher, at least about 80% higher, at least about 85% higher, such as, for example at least about 87% higher) cellobiose hydrolysis activity but somewhat less dramatically increased (e.g., about 5% higher, or about 10% higher, or about 15% higher, or about 20% higher) capacity to hydrolyze chloro-nitro-phenyl-glucoside (CNPG) as compared to the *Trichoderma reesei* Bgl1. For example, the Mg3A polypeptides of the compositions and methods herein have substantially increased CNPG hydrolysis activity as compared to the *Aspergillus niger* B-glu, but somewhat similar or slightly less cellobiase or cellobiose hydrolysis activity when compared to the *Aspergillus niger* B-glu.

[0024] In some embodiments, the recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, has an about 10% reduced, about 20% reduced, about 30% reduced, or even about 40% reduced hydrolysis activity ratio over CNPG/cellobiose. In some embodiments, the Mg3A polypeptide, as compared to the *Aspergillus niger* B-glu, has about 2-fold, about 5-fold, about 7-fold, about 10-fold, about 15 fold, or even about 20-fold higher relative hydrolysis activity ratio over CNPG/cellobiose.

[0025] In certain aspects, the Mg3A polypeptides and the compositions comprising the Mg3A polypeptides of the invention have improved performance hydrolyzing lignocellulosic biomass substrates, as compared to that of the wild type *Trichoderma reesei* Bgl1 (of SEQ ID NO:4). In some embodiments, the improved hydrolysis performance of Mg3A polypeptides or compositions comprising Mg3A polypeptides is observable by the production of a greater amount of glucose from a given lignocellulosic biomass substrate, pretreated in a certain way, as

compared to the level of glucose produced by *Trichoderma reesei* Bgl1 or an identical enzyme composition comprising *Trichoderma reesei* Bgl1 from the same biomass pretreated the same way, under the same saccharification conditions. For example, the amount of glucose produced by the Mg3A polypeptides or by the enzyme compositions comprising the Mg3A polypeptides is at least about 5% (e.g., at least about 5%, at least about 10%, at least about 15%, at least about 20%, or at least about 25) greater than the amount of glucose produced by the *Trichoderma reesei* Bgl1 or an otherwise identical enzyme composition comprising the *Trichoderma reesei* Bgl1 (rather than a Mg3A polypeptide), when 0-10 mg (e.g., about 1 mg, about 2 mg, about 3 mg, about 4 mg, about 5 mg, about 6 mg, about 7 mg, about 8 mg, about 9 mg, about 10 mg) of beta-glucosidase (a Mg3A polypeptide or *Trichoderma reesei* Bgl1) is used to hydrolyze 1 g glucan in the biomass substrate.

[0026] In some aspects, the improved hydrolysis performance of Mg3A polypeptides or compositions comprising Mg3A polypeptides is observable by increased % glucan conversion from a given lignocellulosic biomass substrate pretreated in a certain way, as compared to the level of % glucan conversion by *Trichoderma reesei* Bgl1 or an otherwise identical enzyme composition comprising *Trichoderma reesei* Bgl1 from the same biomass pretreated the same way, under the same saccharification conditions. For example, the %glucan conversion by the Mg3A polypeptides or the enzyme compositions comprising the Mg3A polypeptides is at least about 5% (e.g., at least about 5%, at least about 10%, or at least about 15%) higher than the %glucan conversion by *Trichoderma reesei* Bgl1 or an otherwise identical enzyme composition comprising *Trichoderma reesei* Bgl1 (rather than a Mg3A polypeptide), when 0-10 mg (e.g., about 1 mg, about 2 mg, about 3 mg, about 4 mg, about 5 mg, about 6 mg, about 7 mg, about 8 mg, about 9 mg, about 10 mg) of beta-glucosidase (a Mg3A polypeptide or *Trichoderma reesei* Bgl1) is used to hydrolyze 1 g glucan in the biomass substrate.

[0027] In further aspects, the improved hydrolysis performance of Mg3A polypeptides and compositions comprising Mg3A polypeptides is observable by a higher cellobiase activity and/or reduced amount of residual cellobiose in the product mixture, from hydrolyzing a given lignocellulosic biomass substrate pretreated in a certain way, as compared to the residual amount of cellobiose when the same biomass substrate is hydrolyzed by *Trichoderma reesei* Bgl1 or an otherwise identical composition comprising *Trichoderma reesei* Bgl1 under the same saccharification conditions. For example, the amount of residual cellobiose in the product mixture produced from the hydrolysis of a given biomass substrate pretreated a certain way, by the Mg3A polypeptides or the compositions comprising the Mg3A polypeptides is at least about 5% (e.g., at least about 5%, at least about 10%, at least about 15%, or even at least about 20) less

than the amount of residual cellobiose produced in the product mixture produced from hydrolysis of the same biomass substrate pretreated the same way by the *Trichoderma reesei* Bgl1 or by an otherwise identical enzyme composition comprising *Trichoderma reesei* Bgl1 under the same saccharification conditions. This is the case when 0-10 mg beta-glucosidase (e.g., about 1 mg, about 2 mg, about 3 mg, about 4 mg, about 5 mg, about 6 mg, about 7 mg, about 8 mg, about 9 mg, about 10 mg) of beta-glucosidase (e.g., a Mg3A polypeptide or a *Trichoderma reesei* Bgl1) is used to hydrolyze 1 g glucan in the biomass substrate.

[0028] Aspects of the present compositions and methods include a composition comprising a recombinant Mg3A polypeptide as detailed above and a lignocellulosic biomass. Suitable lignocellulosic biomass may be, for example, derived from an agricultural crop, a byproduct of a food or feed production, a lignocellulosic waste product, a plant residue, including, for example, a grass residue, or a waste paper or waste paper product. In certain embodiments, the lignocellulosic biomass has been subject to one or more pretreatment steps in order to render xylan, hemicelluloses, cellulose and/or lignin material more accessible or susceptible to enzymes and thus more amendable to enzymatic hydrolysis. A suitable pretreatment method may be, for example, subjecting biomass material to a catalyst comprising a dilute solution of a strong acid and a metal salt in a reactor. *See, e.g.,* U.S. Patent Nos. 6,660,506, 6,423,145. Alternatively, a suitable pretreatment may be, for example, a multi-stepped process as described in U.S. Patent No. 5,536,325. In certain embodiments, the biomass material may be subject to one or more stages of dilute acid hydrolysis using about 0.4% to about 2% of a strong acid, in accordance with the disclosures of U.S. Patent No. 6,409,841. Further embodiments of pretreatment methods may include those described in, for example, U.S. Patent No. 5,705,369; in Gould, (1984) Biotech. & Bioengr., 26:46-52; in Teixeira et al., (1999) Appl. Biochem & Biotech., 77-79:19-34; in International Published Patent Application WO2004/081185; or in U.S. Patent Publication No. 20070031918, or International Published Patent Application WO06110901.

[0029] The present invention also pertains to isolated polynucleotides encoding polypeptides having beta-glucosidase activity, wherein the isolated polynucleotides are selected from:

(1) a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to SEQ ID NO:2 or to SEQ ID NO:3;

(2) a polynucleotide having at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) identity to SEQ ID NO:1, or hybridizes under

medium stringency conditions, high stringency conditions, or very high stringency conditions to SEQ ID NO:1, or to a complementary sequence thereof.

[0030] Aspects of the present compositions and methods include methods of making or producing a Mg3A polypeptide having beta-glucosidase activity, employing an isolated nucleic acid sequence encoding the recombinant polypeptide comprising an amino acid sequence that is at least 75% identical (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) to that of SEQ ID NO:2, or that of the mature sequence SEQ ID NO:3. In some embodiments, the polypeptide further comprises a native or non-native signal peptide such that the Mg3A polypeptide that is produced is secreted by a host organism, for example, the signal peptide comprises a sequence that is at least 90% identical to SEQ ID NO:13 (the signal sequence of *Trichoderma reesei* Bgl1). In certain embodiments the isolated nucleic acid comprises a sequence that is at least 75% (e.g., at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%) identical to SEQ ID NO:1. In certain embodiments, the isolated nucleic acid further comprises a nucleic acid sequence encoding a signal peptide sequence. In certain embodiments, the signal peptide sequence may be one selected from SEQ ID NOs:13-42. In certain particular embodiments, a nucleic acid sequence encoding the signal peptide sequence of SEQ ID NO:13 is used to express a Mg3A polypeptide in *Trichoderma reesei.*

[0031] Aspects of the present compositions and methods include an expression vector comprising the isolated nucleic acid as described above in operable combination with a regulatory sequence.

[0032] Aspects of the present compositions and methods include a host cell comprising the expression vector. In certain embodiments, the host cell is a bacterial cell or a fungal cell. In certain embodiments, the host cell comprising the expression vector is an ethanologen microbe capable of metabolizing the soluble sugars produced from a hydrolysis of a lignocellulosic biomass, wherein the hydrolysis is the result of a chemical and/or enzymatic process.

[0033] Aspects of the present compositions and methods include a composition comprising the host cell described above and a culture medium. Aspects of the present compositions and methods include a method of producing a Mg3A polypeptide comprising: culturing the host cell described above in a culture medium, under suitable conditions to produce the beta-glucosidase.

[0034] Aspects of the present compositions and methods include a composition comprising a Mg3A polypeptide in the supernatant of a culture medium produced in accordance with the methods for producing the beta-glucosidase as described above.

[0035] In some aspects the present invention is related to nucleic acid constructs, recombinant expression vectors, engineered host cells comprising a polynucleotide encoding a polypeptide having beta-glucosidase activity, as described above and herein. In further aspects, the present invention pertains to methods of preparing or producing the beta-glucosidase polypeptides of the invention or compositions comprising such beta-glucosidase polypeptides using the nucleic acid constructs, recombinant expression vectors, and/or engineered host cells. In particular, the present invention is related, for example, to a nucleic acid constructs comprising a suitable signal peptide operably linked to the mature sequence of the beta-glucosidase that is at least 75% identical to SEQ ID NO:2 or to the mature sequence of SEQ ID NO:3, or is encoded by a polynucleotide that is at least 75% identical to SEQ ID NO:1, an isolated polynucleotide, a nucleic acid construct, a recombinant expression vector, or an engineered host cell comprising such a nucleic acid construct. In some embodiments, the signal peptide and beta-glucosidase sequences are derived from different microorganisms.

[0036] Also provided is an expression vector comprising the isolated nucleic acid in operable combination with a regulatory sequence. Additionally, a host cell is provided comprising the expression vector. In still further embodiments, a composition is provided, which comprises the host cell and a culture medium.

[0037] In some embodiments, the host cell is a bacterial cell or a fungal cell. In certain embodiments, the host cell is an ethanologen microbe, which is capable of metabolizing the soluble sugars produced from hydrolyzing a lignocellulosic biomass substrate, wherein the hydrolyzing can be through a chemical hydrolysis or enzymatic hydrolysis or a combination of these processes, but is also capable of expression of heterologous enzymes. In some embodiments, the host cell is a *Saccharomyces cerevisiae* or a *Zymomonas mobilis* cell, which are not only capable of expressing a heterologous polypeptide such as a Mg3A polypeptide of the invention, but also capable of fermenting sugars into ethanol and/or downstream products. In certain particular embodiments, the *Saccharomyces cerevisiae* cell or *Zymomonas mobilis* cell, which expresses the beta-glucosidase, is capable of fermenting the sugars produced from a lignocellulosic biomass by an enzyme composition comprising one or more beta-glucosidases. The enzyme composition comprising one or more beta-glucosidases may comprise the same beta-glucosidase or may comprise one or more different beta-glucosidases. In certain embodiments, the enzyme composition comprising one or more beta-glucosidases may be an enzyme mixture produced by an engineered host cell, which may be a bacterial or a fungal cell. When a *Saccharomyces cerevisiae* or a *Zymomonas mobilis* cell expressing the Mg3A polypeptide of the present disclosure, the Mg3A polypeptide may be expressed but not secreted.

Accordingly the cellobiose must be introduced or "transported" into such a host cell in order for the beta-glucosidase Mg3A polypeptide to catalyze the liberation of D-glucose. Therefore in certain embodiments, the *Saccharomyces cerevisiae* or a *Zymomonas mobilis* cell are transformed with a cellobiose transporter gene in addition to one that encodes the Mg3A polypeptide. A cellobiose transporter and a beta-glucosidase have been expressed in *Saccharomyces cerevisiae* such that the resulting microbe is capable of fermenting cellobiose, for example, in Ha *et al.*, (2011) PNAS, 108(2):504-509. Another cellobiose transporter has been expressed in a *Pichia* yeast, for example in published U.S. Patent Application No. 20110262983. A cellobiose transporter has been introduced into an *E.coli*, for example, in Sekar *et al.*, (2012) Applied Environmental Microbiology, 78(5):1611-1614.

[0038] In further embodiments, the Mg3A polypeptide is heterologously expressed by a host cell. For example, the Mg3A polypeptide is expressed by an engineered microorganism that is not *Magnaporthe grisea.* In some embodiments, the Mg3A polypeptide is co-expressed with one or more different cellulase genes. In some embodiments, the Mg3A polypeptide is co-expressed with one or more hemicellulase genes.

[0039] In some aspects, compositions comprising the recombinant Mg3A polypeptides of the preceding paragraphs and methods of preparing such compositions are provided. In some embodiments, the composition further comprises one or more other cellulases, whereby the one or more other cellulases are co-expressed by a host cell with the Mg3A polypeptide. For example, the one or more other cellulases can be selected from no or one or more other beta-glucosidases, one or more cellobiohydrolyases, and/or one or more endoglucanases. Such other beta-glucosidases, cellobiohydrolases and/or endoglucanases, if present, can be co-expressed with the Mg3A polypeptide by a single host cell. At least two of the two or more cellulases may be heterologous to each other or derived from different organisms. For example, the composition may comprise two beta-glucosidases, with the first one being a Mg3A polypeptide, and the second beta-glucosidase being not derived from a *Magnaporthe grisea* strain. For example, the composition may comprise at least one cellobiohydrolase, one endoglucanase, or one beta-glucosidase that is not derived from *Magnaporthe grisea.* In some embodiments, one or more of the cellulases are endogenous to the host cell, but are overexpressed or expressed at a level that is different from that would otherwise be naturally-occurring in the host cell. For example, one or more of the cellulases may be a *Trichoderma reesei* CBH1 and/or CBH2, which are native to a *Trichoderma reesei* host cell, but either or both CBH1 and CBH2 are

overexpressed or underexpressed when they are co-expressed in the *Trichoderma reesei* host cell with a Mg3A polypeptide.

[0040] In certain embodiments, the composition comprising the recombinant Mg3A polypeptide may further comprise one or more hemicellulases, whereby the one or more hemicellulases are co-expressed by a host cell with the Mg3A polypeptide. For example, the one or more hemicellulases can be selected from one or more xylanase, one or more beta-xylosidases, and/or one or more L-arabinofuranosidases. Such other xylanases, beta-xylosidases and L-arabinofuranosidases, if present, can be co-expressed with the Mg3A polypeptide by a single host cell. In some embodiments, the composition may comprise at least one beta-xylosidase, xylanase or arabinofuranosidase that is not derived from *Magnaporthe grisea*.

[0041] In further aspects, the composition comprising the recombinant Mg3A polypeptide may further comprise one or more other cellulases and one or more hemicellulases, whereby the one or more cellulases and/or one or more hemicellulases are co-expressed by a host cell with the Mg3A polypeptide. For example, a Mg3A polypeptide may be co-expressed with one or more other beta-glucosidases, one or more cellobiohydrolases, one or more endoglucanases, one or more endo-xylanases, one or more beta-xylosidases, and one or more L-arabinofuranosidases, in addition to other non-cellulase non-hemicellulase enzymes or proteins in the same host cell. Aspects of the present compositions and methods accordingly include a composition comprising the host cell described above co-expressing a number of enzymes in addition to the Mg3A polypeptide and a culture medium. Aspects of the present compositions and methods accordingly include a method of producing a Mg3A -containing enzyme composition comprising: culturing the host cell, which co-expresses a number of enzymes as described above with the Mg3A polypeptide in a culture medium, under suitable conditions to produce the Mg3A and the other enzymes. Also provided are compositions that comprise the Mg3A polypeptide and the other enzymes produced in accordance with the methods herein in supernatant of the culture medium. Such supernatant of the culture medium can be used as is, with minimum or no post-production processing, which may typically include filtration to remove cell debris, cell-kill procedures, and/or ultrafiltration or other steps to enrich or concentrate the enzymes therein. Such supernatants are called “whole broths” or “whole cellulase broths” herein.

[0042] In further aspects, the present invention pertains to a method of applying or using the composition as described above under conditions suitable for degrading or converting a cellulosic material and for producing a substance from a cellulosic material.

[0043] In a further aspect, methods for degrading or converting a cellulosic material into fermentable sugars are provided, comprising: contacting the cellulosic material, preferably having already been subject to one or more pretreatment steps, with the Mg3A polypeptides or the compositions comprising such polypeptides of one of the preceding paragraphs to yield fermentable sugars.

[0044] Accordingly the instant specification is drawn to the following particular aspects:

[0045] In a first aspect, a recombinant polypeptide comprising an amino acid sequence that is at least 75% identical to SEQ ID NO:2 or to the mature sequence SEQ ID NO:3, wherein the polypeptide has beta-glucosidase activity.

[0046] In a second aspect, the recombinant polypeptide of the first aspect, wherein the polypeptide has improved beta-glucosidase activity or the capacity to hydrolyze a lignocellulosic biomass substrate as compared to the *Trichoderma reesei* Bgl1 when the recombinant polypeptide and the *Trichoderma reesei* Bgl1 are used to hydrolyze lignocellulosic biomass substrates.

[0047] In a third aspect, the recombinant polypeptide of the first or the second aspect, as above, wherein the improved beta-glucosidase activity is an increased cellobiase activity or improved capacity to hydrolyze cellobiose, thereby liberating D-glucose.

[0048] In a fourth aspect, the recombinant polypeptide of the first, second, or third aspect, as above, wherein the improved beta-glucosidase activity is an increased yield of glucose from a given lignocellulosic biomass under a given set of saccharification conditions.

[0049] In a fifth aspect, the recombinant polypeptide of any one of the first to fourth aspects above, wherein the lignocellulosic biomass is one that has been subject to a pretreatment prior to saccharification. The pretreatment may suitably be those known in the art that renders the lignocellulosic biomass substrate more amenable to the enzymatic access and hydrolysis, which may include, for example those pretreatment methods described herein.

[0050] In a sixth aspect, the recombinant polypeptide of any one of the first to fifth aspects above, wherein the polypeptide comprises an amino acid sequence that is at least 80% identical to SEQ ID NO:2 or to the mature sequence of SEQ ID NO:3.

[0051] In a seventh aspect, the recombinant polypeptide of any one of the first to sixth aspects above, wherein the polypeptide comprises an amino acid sequence that is at least 90% identical to SEQ ID NO:2 or to the mature sequence of SEQ ID NO:3.

[0052] In an eighth aspect, a composition comprising the recombinant polypeptide of any one of the first to seventh aspects above, further comprising one or more other cellulases.

[0053] In a ninth aspect, the composition of the eighth aspect, wherein the one or more other cellulases are selected from no or one or more other beta-glucosidases, one or more cellobiohydrolases and one or more endoglucanases.

[0054] In a tenth aspect, a composition comprising the recombinant polypeptide of any one of the first to seven aspects above, further comprising one or more hemicellulases.

[0055] In an eleventh aspect, the composition of the eighth or the ninth aspect as above, further comprising one or more hemicellulases.

[0056] In a twelfth aspect, the composition of the tenth or the eleventh aspect as above, wherein the one or more hemicellulases are selected from one or more xylanases, one or more beta-xylosidases, and one or more L-arabinofuranosidases.

[0057] In a thirteenth aspect, a nucleic acid encoding the recombinant polypeptide of any one of the first to seventh aspects.

[0058] In a fourteenth aspect, the nucleic acid of the thirteenth aspect, further comprising a signal sequence.

[0059] In a fifteenth aspect, the nucleic acid of the fourteenth aspect, wherein the signal sequence is selected from the group consisting of SEQ ID NOs: 13-42.

[0060] In a sixteenth aspect, an expression vector comprising the nucleic acid of any one of the thirteenth to fifteenth aspects in operable combination with a regulatory sequence.

[0061] In a seventeenth aspect, a host cell comprising the expression vector of the sixteenth aspect.

[0062] In an eighteenth aspect, the host cell of the seventeenth aspect, wherein the host cell is a bacterial cell or a fungal cell. A number of bacterial cells are known to be suitable host cells as described herein. A number of fungal cells are also suitable. In some embodiments, the bacterial or fungal host cell may be one that is an ethanologen, capable of fermenting or metabolizing certain monomeric sugars into ethanol. For example, the bacterial ethanolgen *Zymomonas mobilis* may be a host cell expressing a beta-glucosidase polypeptide of the present disclosure. For example, a fungal ethanologen *Saccharomyces cerevisiae* yeast may also serve as a host cell to produce a beta-glucosidase polypeptide of the present disclosure.

[0063] In a nineteenth aspect, a composition comprising the host cell of the sisteenth or the seventeenth aspect and a culture medium.

[0064] In a twentieth aspect, a method of producing a beta-glucosidase comprising culturing the host cell of the seventeenth or eighteenth aspect, in a culture medium, under suitable conditions to produce the beta-glucosidase.

[0065] In a 21st aspect, a composition comprising the beta-glucosidase produced in accordance with the method of the 20th aspect above, in supernatant of the culture medium.

[0066] In a 22nd aspect, a method for hydrolyzing a lignocellulosic biomass substrate, comprising contacting the lignocellulosic biomass substrate with the polypeptide of any one of the first to seventh aspects or with the composition of the 21st aspect, to yield a glucose and/or other sugars.

[0067] These and other aspects of Mg3A compositions and methods will be apparent from the following description.

DESCRIPTION OF THE DRAWINGS

[0068] Figure 1 depicts a map of the pENTR/D-TOPO-Bgl1(943/942) vector.

[0069] Figure 2 depicts a map of the pTrex3g 943/942 construct.

[0070] Figures 3A-3F provide 2 sets of experiments using somewhat different Mg3A enzymes preparations to compare hydrolysis performance of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 using a phosphoric acid swollen cellulase (PASC) as substrate, at 50°C, and 1.5 h, wherein the Mg3A and Bgl1 were added to a background whole cellulase produced from an engineered *Trichoderma reesei* strain in accordance with what is described in Published Patent Application WO 2011/038019, and the beta-glucosidase+whole cellulase mixture was mixed with the PASC substrate at various beta-glucosidase doses. **Figure 3A** depicts the measurements and comparison of % glucan conversion at various beta-glucosidase doses in accordance with the conditions described in **Example 4-A** herein. **Figure 3B** depicts the measurements and comparison of total glucose production at various beta-glucosidase doses in accordance with the conditions described in **Example 4-A** herein. **Figure 3C** depicts the measurements and comparison of cellobiose produced from hydrolyzing the PASC at various beta-glucosidase doses, in accordance with the conditions described in **Example 4-A** herein. **Figure 3D** depicts the measurements and comparison of % glucan conversion at various beta-glucosidase doses in accordance with the conditions described in **Example 4-B** herein. **Figure 3E** depicts the measurements and comparison of total glucose production at various beta-

glucosidase doses in accordance with the conditions described in **Example 4-B** herein. **Figure 3F** depicts the measurements and comparison of cellobiose produced from hydrolyzing the PASC at various beta-glucosidase doses, in accordance with the conditions described in **Example 4-B** herein.

[0071] Figures 4A-4B provide comparison of hydrolysis performance of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 using a dilute ammonia pretreated corn stover (DACS) as substrate, at 50°C for 2 days, wherein the Mg3A and Bgl1 were added to a background whole cellulase produced from an engineered *Trichoderma reesei* strain in accordance with what is described in Published Patent Application WO 2011/038019, and the beta-glucosidase+cellulase mixture was mixed with the DACS substrate at various total protein to cellulose doses. **Figure 4A** depicts the measurements and comparison of total glucan conversion at various total protein to cellulose doses. **Figure 4B** depicts the measurements and comparison of total glucose production at various total protein to cellulose doses.

[0072] Figures 5A-5B provide comparison of hydrolysis performance of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 using a dilute ammonia pretreated corn stover (DACS) as substrate, at 50°C for 2 days, wherein various doses of Mg3A and Bgl1 were added to a whole cellulase that is used at a constant13.4 mg protein/g cellulose, produced from an engineered *Trichoderma reesei* strain in accordance with what is described in Published Patent Application WO 2011/038019. **Figure 5A** depicts the measurements and comparison of total glucan conversion at various beta-glucosidase loading (with the whole cellulase background being held at the constant 13.4 mg protein/g cellulose). **Figure 5B** depicts the measurements and comparison of total glucose production at various beta-glucosidase loading (with the whole cellulase background being held at the constant 13.4 mg protein/g cellulose)

[0073] Figures 6A-6C provide comparison of hydrolysis performance of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 using a dilute ammonia pretreated corn stover (DACS) as substrate, at 55°C for 2 days, wherein various dose of Mg3A and Bgl1 were added to a whole cellulase background that contained 10 mg protein/g cellulose, produced from an engineered *Trichoderma reesei* strain in accordance with what is described in Published Patent Application WO 2011/038019. **Figure 6A** depicts the measurements and comparison of total glucan conversion at various beta-glucose loadings (with the whole cellulase background being held at the constant 10 mg protein/g cellulose). **Figure 6B** depicts the measurements and comparison of total glucose production at various beta-glucosidase loadings (with the whole cellulase background being held at the constant 10 mg/g protein/g cellulose. **Figure 6C** depicts the dose

curve of depletion of cellobiose by the presence of Mg3A and Bgl1 at various doses in the saccharification reactions carried out at 55°C.

[0074] **Figure 7** depicts a yeast shuttle vector pSC11 construct comprising a Mg3A gene optimized and synthesized for expression of the Mg3A polypeptide in a *Saccharomyces cerevisiae* ethanologen.

[0075] **Figure 8** depicts a *Zymomonas mobilis* integration vector pZC11 comprising an Mg3A gene optimized and synthesized for expression of the Mg3A polypeptide in a *Zymomonas mobilis* ethanologen.

[0076] **Figures 9A-9E** depict the sequences and sequence identifiers of the present disclosure.

DETAILED DESCRIPTION

I. Overview

[0077] Described herein are compositions and methods relating to a recombinant beta-glucosidase Mg3A belonging to glycosyl hydrolase family 3 from *Magnaporthe grisea.* The present compositions and methods are based, in part, on the observations that recombinant Mg3A polypeptides have higher cellulase activities and are more robust as a component of an enzyme composition when the composition is used to hydrolyze a lignocellulosic biomass material or feedstock than, for example, a known benchmark high fidelity beta-glucosidase Bgl1 of *Trichoderma reesei.* These features of Mg3A polypeptides make them, or variants thereof, suitable for use in numerous processes, including, for example, in the conversion or hydrolysis of a lignocellulosic biomass feedstock.

[0078] Before the present compositions and methods are described in greater detail, it is to be understood that the present compositions and methods are not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present compositions and methods will be limited only by the appended claims.

[0079] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the present compositions and methods. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges and are also

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

[0080] Certain ranges are presented herein with numerical values being preceded by the term "about." The term "about" is used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating unrecited number may be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number. For example, in connection with a numerical value, the term “about” refers to a range of -10% to +10% of the numerical value, unless the term is otherwise specifically defined in context. In another example, the phrase a “pH value of about 6” refers to pH values of from 5.4 to 6.6, unless the pH value is specifically defined otherwise.

[0081] The headings provided herein are not limitations of the various aspects or embodiments of the present compositions and methods which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

[0082] The present document is organized into a number of sections for ease of reading; however, the reader will appreciate that statements made in one section may apply to other sections. In this manner, the headings used for different sections of the disclosure should not be construed as limiting.

[0083] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the present compositions and methods belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present compositions and methods, representative illustrative methods and materials are now described.

[0084] All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present compositions and methods are not entitled to antedate such

publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0085] In accordance with this detailed description, the following abbreviations and definitions apply. Note that the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "an enzyme" includes a plurality of such enzymes, and reference to "the dosage" includes reference to one or more dosages and equivalents thereof known to those skilled in the art, and so forth.

[0086] It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[0087] The term "recombinant," when used in reference to a subject cell, nucleic acid, polypeptides/enzymes or vector, indicates that the subject has been modified from its native state. Thus, for example, recombinant cells express genes that are not found within the native (non-recombinant) form of the cell, or express native genes at different levels or under different conditions than found in nature. Recombinant nucleic acids may differ from a native sequence by one or more nucleotides and/or are operably linked to heterologous sequences, *e.g.*, a heterologous promoter, signal sequences that allow secretion, etc., in an expression vector. Recombinant polypeptides/enzymes may differ from a native sequence by one or more amino acids and/or are fused with heterologous sequences. A vector comprising a nucleic acid encoding a beta-glucosidase is, for example, a recombinant vector.

[0088] It is further noted that the term "consisting essentially of," as used herein refers to a composition wherein the component(s) after the term is in the presence of other known component(s) in a total amount that is less than 30% by weight of the total composition and do not contribute to or interferes with the actions or activities of the component(s).

[0089] It is further noted that the term "comprising," as used herein, means including, but not limited to, the component(s) after the term "comprising." The component(s) after the term "comprising" are required or mandatory, but the composition comprising the component(s) may further include other non-mandatory or optional component(s).

[0090] It is also noted that the term "consisting of," as used herein, means including, and limited to, the component(s) after the term "consisting of." The component(s) after the term "consisting of" are therefore required or mandatory, and no other component(s) are present in the composition.

[0091] As will be apparent to those of skill in the art upon reading this disclosure, each of the individual embodiments described and illustrated herein has discrete components and features which may be readily separated from or combined with the features of any of the other several embodiments without departing from the scope or spirit of the present compositions and methods described herein. Any recited method can be carried out in the order of events recited or in any other order which is logically possible.

II. Definitions

[0092] "Beta-glucosidase" refers to a beta-D-glucoside glucohydrolase of E.C. 3.2.1.21. The term "beta-glucosidase activity" therefore refers the capacity of catalyzing the hydrolysis of beta-D-glucose or cellobiose to release D-glucose. Beta-glucosidase activity may be determined using a cellobiase assay, for example, which measures the capacity of the enzyme to catalyze the hydrolysis of a cellobiose substrate to yield D-glucose, as described in **Example 2C** of the present disclosure.

[0093] As used herein, "Mg3A" or "a Mg3A polypeptide" refers to a beta-glucosidase belonging to glycosyl hydrolase family 3 (e.g., a recombinant beta-glucosidase) derived from *Magnaporthe grisea* (and variants thereof), that has improved performance hydrolyzing a lignocellulosic biomass substrate when compared to a benchmark beta-glucosidase, the wild type *Trichoderma reesei* Bgl1 polypeptide having the amino acid sequence of SEQ ID NO:4. According to aspects of the present compositions and methods, Mg3A polypeptides include those having the amino acid sequence depicted in SEQ. ID NO:2, as well as derivative or variant polypeptides having 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% sequence identity to the amino acid sequence of SEQ. ID NO:2, or to the mature sequence SEQ ID NO:2, or to a fragment of at least 100 residues in length of SEQ. ID NO:2, wherein the Mg3A polypeptides not only have beta-glucosidase activity and capable of catalyzing the conversion of cellobiose into D-glucose, but also have higher beta-glucosidase activity and have higher capacity to catalyze the conversion of cellobiose to D-glucose than *Trichoderma reesei* Bgl1.

[0094] The Mg3A polypeptides to be used in the compositions and methods of the present disclosure would have at least 5%, at least 10%, preferably at least 20%, more preferably at least 30%, and even more preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and preferably at least 70%, more preferably at least 90%, even more preferably at least 100% or more of the beta-glucosidase activity of the polypeptide of the amino

acid sequence of SEQ ID NO:2, or of the polypeptide consisting of residues 19 to 873 of the SEQ ID NO:2; or of the mature sequence SEQ ID NO:3.

[0095] "Family 3 glycosyl hydrolase" or "GH3" refers to polypeptides falling within the definition of glycosyl hydrolase family 3 according to the classification by Henrissat, Biochem. J. 280:309-316 (1991), and by Henrissat & Cairoch, Biochem. J., 316:695-696 (1996).

[0096] Mg3A polypeptides according to the present compositions and methods described herein can be isolated or purified. By purification or isolation is meant that the Mg3A polypeptide is altered from its natural state by virtue of separating the Mg3A from some or all of the naturally occurring constituents with which it is associated in nature. Such isolation or purification may be accomplished by art-recognized separation techniques such as ion exchange chromatography, affinity chromatography, hydrophobic separation, dialysis, protease treatment, ammonium sulphate precipitation or other protein salt precipitation, centrifugation, size exclusion chromatography, filtration, microfiltration, gel electrophoresis or separation on a gradient to remove whole cells, cell debris, impurities, extraneous proteins, or enzymes undesired in the final composition. It is further possible to then add constituents to the Mg3A-containing composition which provide additional benefits, for example, activating agents, anti-inhibition agents, desirable ions, compounds to control pH or other enzymes or chemicals.

[0097] As used herein, "microorganism" refers to a bacterium, a fungus, a virus, a protozoan, and other microbes or microscopic organisms.

[0098] As used herein, a "derivative" or "variant" of a polypeptide means a polypeptide, which is derived from a precursor polypeptide (e.g., the native polypeptide) 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 more amino acids at either or both ends of the polypeptide 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. The preparation of a Mg3A derivative or variant may be achieved in any convenient manner, e.g., by modifying a DNA sequence which encodes the native polypeptides, transformation of that DNA sequence into a suitable host, and expression of the modified DNA sequence to form the derivative/variant Mg3A. Derivatives or variants further include Mg3A polypeptides that are chemically modified, e.g., glycosylation or otherwise changing a characteristic of the Mg3A polypeptide. While derivatives and variants of Mg3A are encompassed by the present compositions and methods, such derivates and variants will display

improved beta-glucosidase activity when compared to that of the wild type *Trichoderma reesei* Bgl1 of SEQ ID NO:4, under the same lignocellulosic biomass substrate hydrolysis conditions.

[0099] In certain aspects, a Mg3A polypeptide of the compositions and methods herein may also encompasses functional fragment of a polypeptide or a polypeptide fragment having beta-glucosidase activity, which is derived from a parent polypeptide, which may be the full length polypeptide comprising or consisting of SEQ ID NO:2, or the mature sequence comprising or consisting SEQ ID NO:3. The functional polypeptide may have been truncated either in the N-terminal region, or the C-terminal region, or in both regions to generate a fragment of the parent polypeptide. For the purpose of the present disclosure, a functional fragment must have at least 20%, more preferably at least 30%, 40%, 50%, or preferably, at least 60%, 70%, 80%, or even more preferably at least 90% of the beta-glucosidase activity of that of the parent polypeptide.

[00100] In certain aspects, a Mg3A derivative/variant will have anywhere from 75% to 99% (or more) amino acid sequence identity to the amino acid sequence of SEQ. ID NO:2, or to the mature sequence SEQ ID NO:3, e.g., 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to the amino acid sequence of SEQ. ID NO:2 or to the mature sequence SEQ ID NO:3. In some embodiments, amino acid substitutions are "conservative amino acid substitutions" using L-amino acids, wherein one amino acid is replaced by another biologically similar amino acid. Conservative amino acid substitutions are those that preserve the general charge, hydrophobicity/hydrophilicity, and/or steric bulk of the amino acid being substituted. Examples of conservative substitutions are those between the following groups: Gly/Ala, Val/Ile/Leu, Lys/Arg, Asn/Gln, Glu/Asp, Ser/Cys/Thr, and Phe/Trp/Tyr. A derivative may, for example, differ by as few as 1 to 10 amino acid residues, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue. In some embodiments, a Mg3A derivative may have an N-terminal and/or C-terminal deletion, where the Mg3A derivative excluding the deleted terminal portion(s) is identical to a contiguous sub-region in SEQ ID NO: 2 or SEQ ID NO:3.

[00101] As used herein, "percent (%) sequence identity" with respect to the amino acid or nucleotide sequences identified herein is defined as the percentage of amino acid residues or nucleotides in a candidate sequence that are identical with the amino acid residues or nucleotides in a Mg3A sequence, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity.

[00102] By "homologue" shall mean an entity having a specified degree of identity with the subject amino acid sequences and the subject nucleotide sequences. A homologous sequence is taken to include an amino acid sequence that is at least 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or even 99% identical to the subject sequence, using conventional sequence alignment tools (*e.g.*, Clustal, BLAST, and the like). Typically, homologues will include the same active site residues as the subject amino acid sequence, unless otherwise specified.

[00103] Methods for performing sequence alignment and determining sequence identity are known to the skilled artisan, may be performed without undue experimentation, and calculations of identity values may be obtained with definiteness. See, for example, Ausubel *et al.*, eds. (1995) *Current Protocols in Molecular Biology*, Chapter 19 (Greene Publishing and Wiley-Interscience, New York); and the ALIGN program (Dayhoff (1978) in *Atlas of Protein Sequence and Structure* 5:Suppl. 3 (National Biomedical Research Foundation, Washington, D.C.). A number of algorithms are available for aligning sequences and determining sequence identity and include, for example, the homology alignment algorithm of Needleman *et al.* (1970) *J. Mol. Biol. 48*:443; the local homology algorithm of Smith *et al.* (1981) *Adv. Appl. Math. 2*:482; the search for similarity method of Pearson *et al.* (1988) *Proc. Natl. Acad. Sci. 85*:2444; the Smith-Waterman algorithm (*Meth. Mol. Biol. 70*:173-187 (1997); and BLASTP, BLASTN, and BLASTX algorithms (see Altschul *et al.* (1990) *J. Mol. Biol. 215*:403-410).

[00104] Computerized programs using these algorithms are also available, and include, but are not limited to: ALIGN or Megalign (DNASTAR) software, or WU-BLAST-2 (Altschul *et al.,* (1996) Meth. Enzym., 266:460-480); or GAP, BESTFIT, BLAST, FASTA, and TFASTA, available in the Genetics Computing Group (GCG) package, Version 8, Madison, Wisconsin, USA; and CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, California. Those skilled in the art can determine appropriate parameters for measuring alignment, including algorithms needed to achieve maximal alignment over the length of the sequences being compared. Preferably, the sequence identity is determined using the default parameters determined by the program. Specifically, sequence identity can determined by using Clustal W (Thompson J.D. *et al.* (1994) Nucleic Acids Res. 22:4673-4680) with default parameters, i.e.:

Gap opening penalty:	10.0
Gap extension penalty:	0.05
Protein weight matrix:	BLOSUM series
DNA weight matrix:	IUB
Delay divergent sequences %:	40
Gap separation distance:	8

DNA transitions weight:	0.50
List hydrophilic residues:	GPSNDQEKR
Use negative matrix:	OFF
Toggle Residue specific penalties:	ON
Toggle hydrophilic penalties:	ON
Toggle end gap separation penalty	OFF

[00105] As used herein, "expression vector" means a DNA construct including a DNA sequence which is operably linked to a suitable control sequence capable of affecting the expression of the DNA in a suitable host. Such control sequences may include a promoter to affect transcription, an optional operator sequence to control transcription, a sequence encoding suitable ribosome-binding sites on the mRNA, and sequences which control termination of transcription and translation. Different cell types may be used with different expression vectors. An exemplary promoter for vectors used in *Bacillus subtilis* is the AprE promoter; an exemplary promoter used in *Streptomyces lividans* is the A4 promoter (from *Aspergillus niger*); an exemplary promoter used in *E. coli* is the Lac promoter, an exemplary promoter used in *Saccharomyces cerevisiae* is *PGK1,* an exemplary promoter used in *Aspergillus niger* is *glaA,* and an exemplary promoter for *Trichoderma reesei* is *cbhI.* 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 present compositions and methods are intended to include other forms of expression vectors which 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 DNA 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 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 λ, e.g., NM989, and other DNA phages, e.g., M13 and filamentous single stranded DNA phages, yeast plasmids such as the 2μ plasmid or derivatives thereof, vectors useful in eukaryotic cells, such as vectors useful in animal cells 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 compositions and methods are known in the art and are described generally in, for example, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor

Press (1989). Often, such expression vectors including the DNA sequences described herein are transformed into a unicellular host by direct insertion into the genome of a particular species through an integration event (see e.g., Bennett & Lasure, More Gene Manipulations in Fungi, Academic Press, San Diego, pp. 70-76 (1991) and articles cited therein describing targeted genomic insertion in fungal hosts).

[00106] As used herein, "host strain" or "host cell" means a suitable host for an expression vector including DNA according to the present compositions and methods. Host cells useful in the present compositions and methods are generally prokaryotic or eukaryotic hosts, including any transformable microorganism in which expression can be achieved. Specifically, host strains may be *Bacillus subtilis, Streptomyces lividans, Escherichia coli, Trichoderma reesei, Saccharomyces cerevisiae* or *Aspergillus niger.* In certain embodiments, the host cell may be an ethanologen microbe, which may be, for example, a yeast such as *Saccharomyces cerevisiae* or a bacterium ethanologen such as a *Zymomonas mobilis.* When a *Saccharomyces cerevisiae* or *Zymomonas mobilis* is used as the host cell, and if the beta-glucosidase gene is not made to secret from host cell but is expressed intracellularly, a cellobiose transporter gene can be introduced into the host cell in order to allow the intracellularly expressed beta-glucosidase to act upon the cellobiose substrate and liberate glucose, which will then be metabolized subsequently or immediately by the microorganisms and converted into ethanol.

[00107] Host cells are transformed or transfected with vectors constructed using recombinant DNA techniques. Such transformed host cells may be capable of one or both of replicating the vectors encoding Mg3A (and its derivatives or variants (mutants)) and expressing the desired peptide product. In certain embodiments according to the present compositions and methods, "host cell" means both the cells and protoplasts created from the cells of *Trichoderma sp.*

[00108] The terms "transformed," "stably transformed," and "transgenic," used with reference to a cell means that the cell contains a non-native (*e.g.*, heterologous) nucleic acid sequence integrated into its genome or carried as an episome that is maintained through multiple generations.

[00109] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", "transformation" or "transduction," as known in the art.

[00110] A "host strain" or "host cell" is an organism into which an expression vector, phage, virus, or other DNA construct, including a polynucleotide encoding a polypeptide of interest (*e.g.*, a beta-glucosidase) has been introduced. Exemplary host strains are microbial cells (*e.g.*,

bacteria, filamentous fungi, and yeast) capable of expressing the polypeptide of interest. The term "host cell" includes protoplasts created from cells.

[00111] The term "heterologous" with reference to a polynucleotide or polypeptide refers to a polynucleotide or polypeptide that does not naturally occur in a host cell.

[00112] The term "endogenous" with reference to a polynucleotide or polypeptide refers to a polynucleotide or polypeptide that occurs naturally in the host cell.

[00113] The term "expression" refers to the process by which a polypeptide is produced based on a nucleic acid sequence. The process includes both transcription and translation.

[00114] Accordingly the process of converting a lignocellulosic biomass substrate to an ethanol can, in some embodiments, comprise two beta-glucosidase activities. For example, a first beta-glucosidase activity may be applied to the lignocellulosic biomass substrate during the saccharification or hydrolysis step, and a second beta-glucosidase activity can be applied as part of the ethanologen microbe in the fermentation step during which the monomeric or fermentable sugars that resulted from the saccharification or hydrolysis step are metabolized. The first and second beta-glucosidase activities may, in some embodiments, result from the presence of the same beta-glucosidase polypeptide. For example, the first beta-glucosidase activity in the saccharification may result from the presence of a Mg3A polypeptide of the invention, whereas the second beta-glucosidase activity in the fermentation stage may result from the expression of a different beta-glucosidase by the ethanologen microbe. In another example, the first and second beta-glucosidase activities may result from the presence of the same polypeptide in the saccharification or hydrolysis step and the fermentation step. For example, the same Mg3A polypeptide of the invention may, in some embodiments, provide the beta-glucosidase activities for both the hydrolysis or saccharification step and the fermentation step.

[00115] In certain other embodiments, the process of converting a lignocellulosic biomass substrate to an ethanol can, comprise two beta-glucosidase activities whereas the saccharification or hydrolysis step and the fermentation step occurs simultaneously, for example, in the same tank. Two or more beta-glucosidase polypeptides may contribute to the beta-glucosidase activities, one of which may be a Mg3A polypeptide of the invention.

[00116] In certain further embodiments, the process of converting a lignocellulosic biomass to an ethanol can comprise a single beta-glucosidase activity whereas either the saccharification or hydrolysis step or the fermentation step, but not both steps involves the participation of a beta-glucosidase. For example, a Mg3A polypeptide of the invention or a composition comprising the Mg3A polypeptide may be used in the saccharification step. In another example,

the enzyme composition that is used to hydrolyze the lignocellulosic biomass substrate does not comprise a beta-glucosidase activity, whereas the ethanologen microbe expresses a beta-glucosidase polypeptide, for example, a Mg3A polypeptide of the invention.

[00117] As used herein, "signal sequence" means a sequence of amino acids bound to the N-terminal portion of a polypeptide which facilitates the secretion of the mature form of the polypeptide outside of the cell. This definition of a signal sequence is a functional one. The mature form of the extracellular polypeptide lacks the signal sequence which is cleaved off during the secretion process. While the native signal sequence of Mg3A may be employed in aspects of the present compositions and methods, other non-native signal sequences may be employed (e.g., SEQ ID NO: 13). The term "mature," when referring to a polypeptide herein, is meant a polypeptide in its final form(s) following translation and any post-translational modifications. For example, the Mg3A polypeptides of the invention has one or more mature forms, at least one of which has the amino acid sequence of SEQ ID NO:3.

[00118] The beta-glucosidase polypeptides of the invention may be referred to as "precursor," "immature," or "full-length," in which case they include a signal sequence, or may be referred to as "mature," in which case they lack a signal sequence. Mature forms of the polypeptides are generally the most useful. Unless otherwise noted, the amino acid residue numbering used herein refers to the mature forms of the respective amylase polypeptides. The beta-glucosidase polypeptides of the invention may also be truncated to remove the N or C-termini, so long as the resulting polypeptides retain beta-glucosidase activity.

[00119] The beta-glucosidase polypeptides of the invention may also be a "chimeric" or "hybrid" polypeptide, in that it includes at least a portion of a first beta-glucosidase polypeptide, and at least a portion of a second beta-glucosidase polypeptide (such chimeric beta-glucosidase polypeptides may, for example, be derived from the first and second beta-glucosidases using known technologies involving the swapping of domains on each of the beta-glucosidases). The present beta-glucosidase polypeptides may further include heterologous signal sequence, an epitope to allow tracking or purification, or the like. When the term "heterologous" is used to refer to a signal sequence used to express a polypeptide of interest, it is meant that the signal sequence is, for example, derived from a different microorganism as the polypeptide of interest. Examples of suitable heterologous signal sequences for expressing the Mg3A polypeptides herein, may be, for example, those from *Trichoderma reesei.*

[00120] As used herein, "functionally attached" or "operably linked" means that a regulatory region or functional domain having a known or desired activity, such as a promoter, terminator, signal sequence or enhancer region, is attached to or linked to a target (e.g., a gene or

polypeptide) in such a manner as to allow the regulatory region or functional domain to control the expression, secretion or function of that target according to its known or desired activity.

[00121] As used herein, the terms "polypeptide" and "enzyme" are used interchangeably to refer to polymers of any length comprising amino acid residues linked by peptide bonds. The conventional one-letter or three-letter codes for amino acid residues are used herein. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified naturally or by intervention; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation or modification, such as conjugation with a labeling component. Also included within the definition are, for example, polypeptides containing one or more analogs of an amino acid (including, for example, unnatural amino acids, etc.), as well as other modifications known in the art.

[00122] As used herein, "wild-type" and "native" genes, enzymes, or strains, are those found in nature.

[00123] The terms "wild-type," "parental," or "reference," with respect to a polypeptide, refer to a naturally-occurring polypeptide that does not include a man-made substitution, insertion, or deletion at one or more amino acid positions. Similarly, the term "wild-type," "parental," or "reference," with respect to a polynucleotide, refers to a naturally-occurring polynucleotide that does not include a man-made nucleoside change. However, a polynucleotide encoding a wild-type, parental, or reference polypeptide is not limited to a naturally-occurring polynucleotide, but rather encompasses any polynucleotide encoding the wild-type, parental, or reference polypeptide.

[00124] As used herein, a "variant polypeptide" refers to a polypeptide that is derived from a parent (or reference) polypeptide by the substitution, addition, or deletion, of one or more amino acids, typically by recombinant DNA techniques. Variant polypeptides may differ from a parent polypeptide by a small number of amino acid residues. They may be defined by their level of primary amino acid sequence homology/identity with a parent polypeptide. Suitably, variant polypeptides have 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 even at least 99% amino acid sequence identity to a parent polypeptide.

[00125] As used herein, a "variant polynucleotide" encodes a variant polypeptide, has a specified degree of homology/identity with a parent polynucleotide, or hybridized under stringent conditions to a parent polynucleotide or the complement thereof. Suitably, a variant

polynucleotide has 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 even at least 99% nucleotide sequence identity to a parent polynucleotide or to a complement of the parent polynucleotide. Methods for determining percent identity are known in the art and described above.

[00126] The term "derived from" encompasses the terms "originated from," "obtained from," "obtainable from," "isolated from," and "created from," and generally indicates that one specified material find its origin in another specified material or has features that can be described with reference to the another specified material.

[00127] As used herein, the term "hybridization conditions" refers to the conditions under which hybridization reactions are conducted. These conditions are typically classified by degree of "stringency" of the conditions under which hybridization is measured. The degree of stringency can be based, for example, on the melting temperature (Tm) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about Tm -5°C (5°C below the Tm of the probe); "high stringency" at about 5-10°C below the Tm; "intermediate stringency" at about 10-20°C below the Tm of the probe; and "low stringency" at about 20-25°C below the Tm. Alternatively, or in addition, hybridization conditions can be based upon the salt or ionic strength conditions of hybridization, and/or upon one or more stringency washes, *e.g.*: 6X SSC = very low stringency; 3X SSC = low to medium stringency; 1X SSC = medium stringency; and 0.5X SSC = high stringency. Functionally, maximum stringency conditions may be used to identify nucleic acid sequences having strict identity or near-strict identity with the hybridization probe; while high stringency conditions are used to identify nucleic acid sequences having about 80% or more sequence identity with the probe. For applications requiring high selectivity, it is typically desirable to use relatively stringent conditions to form the hybrids (*e.g.,* relatively low salt and/or high temperature conditions are used).

[00128] As used herein, the term "hybridization" refers to the process by which a strand of nucleic acid joins with a complementary strand through base pairing, as known in the art. More specifically, "hybridization" refers to the process by which one strand of nucleic acid forms a duplex with, *i.e.*, base pairs with, a complementary strand, as occurs during blot hybridization techniques and PCR techniques. A nucleic acid sequence is considered to be "selectively hybridizable" to a reference nucleic acid sequence if the two sequences specifically hybridize to one another under moderate to high stringency hybridization and wash conditions.

Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex or probe. For example, "maximum stringency" typically occurs at about Tm-5°C (5° below the Tm of the probe); "high stringency" at about 5-10°C below the Tm; "intermediate stringency" at about 10-20°C below the Tm of the probe; and "low stringency" at about 20-25°C below the Tm. Functionally, maximum stringency conditions may be used to identify sequences having strict identity or near-strict identity with the hybridization probe; while intermediate or low stringency hybridization can be used to identify or detect polynucleotide sequence homologs.

[00129] Intermediate and high stringency hybridization conditions are well known in the art. For example, intermediate stringency hybridizations may be carried out with an overnight incubation at 37°C in a solution comprising 20% formamide, 5 x SSC (150mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6), 5 x Denhardt's solution, 10% dextran sulfate and 20 mg/ml denatured sheared salmon sperm DNA, followed by washing the filters in 1x SSC at about 37 - 50°C. High stringency hybridization conditions may be hybridization at 65°C and 0.1X SSC (where 1X SSC = 0.15 M NaCl, 0.015 M Na_3 citrate, pH 7.0). Alternatively, high stringency hybridization conditions can be carried out at about 42°C in 50% formamide, 5X SSC, 5X Denhardt's solution, 0.5% SDS and 100 μg/ml denatured carrier DNA followed by washing two times in 2X SSC and 0.5% SDS at room temperature and two additional times in 0.1X SSC and 0.5% SDS at 42°C. And very high stringent hybridization conditions may be hybridization at 68°C and 0.1X SSC. Those of skill in the art know how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like.

[00130] A nucleic acid encoding a variant beta-glucosidase may have a T_m reduced by 1°C – 3°C or more compared to a duplex formed between the nucleotide of SEQ ID NO: 1 and its identical complement.

[00131] The phrase "substantially similar" or "substantially identical," in the context of at least two nucleic acids or polypeptides, means that a polynucleotide or polypeptide comprises a sequence that has at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, or even at least about 99% identical to a parent or reference sequence, or does not include amino acid substitutions, insertions, deletions, or modifications made only to circumvent the present description without adding functionality.

[00132] As used herein, an "expression vector" refers to a DNA construct containing a DNA sequence that encodes a specified polypeptide and is operably linked to a suitable control sequence capable of effecting the expression of the polypeptides in a suitable host. Such control sequences may include a promoter to effect transcription, an optional operator sequence to control such transcription, a sequence encoding suitable mRNA ribosome binding sites and/or sequences that control termination of transcription and translation. The vector may be a plasmid, a phage particle, or a potential genomic insert. Once transformed into a suitable host, the vector may replicate and function independently of the host genome, or may, in some instances, integrate into the host genome.

[00133] The term "recombinant," refers to genetic material (*i.e.*, nucleic acids, the polypeptides they encode, and vectors and cells comprising such polynucleotides) that has been modified to alter its sequence or expression characteristics, such as by mutating the coding sequence to produce an altered polypeptide, fusing the coding sequence to that of another gene, placing a gene under the control of a different promoter, expressing a gene in a heterologous organism, expressing a gene at a decreased or elevated levels, expressing a gene conditionally or constitutively in a manner different from its natural expression profile, and the like. Generally recombinant nucleic acids, polypeptides, and cells based thereon, have been manipulated by man such that they are not identical to related nucleic acids, polypeptides, and cells found in nature.

[00134] A "signal sequence" refers to a sequence of amino acids bound to the N-terminal portion of a polypeptide, and which facilitates the secretion of the mature form of the polypeptide from the cell. The mature form of the extracellular polypeptide lacks the signal sequence which is cleaved off during the secretion process.

[00135] The term "selective marker" or "selectable marker," refers to a gene capable of expression in a host cell that allows for ease of selection of those hosts containing an introduced nucleic acid or vector. Examples of selectable markers include but are not limited to antimicrobial substances (*e.g.*, hygromycin, bleomycin, or chloramphenicol) and/or genes that confer a metabolic advantage, such as a nutritional advantage, on the host cell.

[00136] The term "regulatory element," refers to a genetic element that controls some aspect of the expression of nucleic acid sequences. For example, a promoter is a regulatory element which facilitates the initiation of transcription of an operably linked coding region. Additional regulatory elements include splicing signals, polyadenylation signals and termination signals.

[00137] As used herein, "host cells" are generally cells of prokaryotic or eukaryotic hosts that are transformed or transfected with vectors constructed using recombinant DNA techniques

known in the art. Transformed host cells are capable of either replicating vectors encoding the polypeptide variants or expressing the desired polypeptide variant. In the case of vectors, which encode the pre- or pro-form of the polypeptide variant, such variants, when expressed, are typically secreted from the host cell into the host cell medium.

[00138] The term "introduced," in the context of inserting a nucleic acid sequence into a cell, means transformation, transduction, or transfection. Means of transformation include protoplast transformation, calcium chloride precipitation, electroporation, naked DNA, and the like as known in the art. (*See*, Chang and Cohen *Mol. Gen. Genet.* 168:111-115, 1979; Smith *et al.* (1986) *Appl. Env. Microbiol.* 51:634; and the review article by Ferrari *et al.*, in Harwood, *Bacillus*, Plenum Publishing Corporation, pp. 57-72, 1989).

[00139] "Fused" polypeptide sequences are connected, *i.e.*, operably linked, via a peptide bond between two subject polypeptide sequences.

[00140] The term "filamentous fungi" refers to all filamentous forms of the subdivision Eumycotina, particularly Pezizomycotina species.

[00141] An "ethanologenic microorganism" refers to a microorganism with the ability to convert a sugar or oligosaccharide to ethanol.

[00142] Other technical and scientific terms have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains (*See, e.g.*, Singleton and Sainsbury, *Dictionary of Microbiology and Molecular Biology*, 2d Ed., John Wiley and Sons, NY 1994; and Hale and Marham, *The Harper Collins Dictionary of Biology*, Harper Perennial, NY 1991).

III. Beta-glucosidase Polypeptides, Polynucleotides, Vectors, and Host Cells

A. Mg3A Polypeptides

[00143] In one aspect, the present compositions and methods provide a recombinant Mg3A beta-glucosidase polypeptide, fragments thereof, or variants thereof having beta-glucosidase activity. An example of a recombinant beta-glucosidase polypeptide was isolated from *Magnaporthe grisea*. The mature Mg3A polypeptide has the amino acid sequence set forth as SEQ ID NO:3. Similar, substantially similar Mg3A polypeptides may occur in nature, *e.g.*, in other strains or isolates of *Magnaporthe grisea*. These and other recombinant Mg3A polypeptides are encompassed by the present compositions and methods.

[00144] In some embodiments, the recombinant Mg3A polypeptide is a variant Mg3A polypeptide having a specified degree of amino acid sequence identity to the exemplified Mg3A

polypeptide, *e.g.*, at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or even at least 99% sequence identity to the amino acid sequence of SEQ ID NO:2 or to the mature sequence SEQ ID NO:3. Sequence identity can be determined by amino acid sequence alignment, *e.g.*, using a program such as BLAST, ALIGN, or CLUSTAL, as described herein.

[00145] In certain embodiments, the recombinant Mg3A polypeptides are produced recombinantly, in a microorganism, for example, in a bacterial or fungal host organism, while in others the Mg3A polypeptides are produced synthetically, or are purified from a native source (*e.g., Magnaporthe grisea*).

[00146] In certain embodiments, the recombinant Mg3A polypeptide includes substitutions that do not substantially affect the structure and/or function of the polypeptide. Examples of these substitutions are conservative mutations, as summarized in Table I.

Table I. Amino Acid Substitutions

Original Residue	Code	Acceptable Substitutions
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, beta-Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-I-thioazolidine-4- carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys

Original Residue	Code	Acceptable Substitutions
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

[00147] Substitutions involving naturally occurring amino acids are generally made by mutating a nucleic acid encoding a recombinant Mg3A polypeptide, and then expressing the variant polypeptide in an organism. Substitutions involving non-naturally occurring amino acids or chemical modifications to amino acids are generally made by chemically modifying a Mg3A polypeptide after it has been synthesized by an organism.

[00148] In some embodiments, variant recombinant Mg3A polypeptides are substantially identical to SEQ ID NO:2 or SEQ ID NO:3, meaning that they do not include amino acid substitutions, insertions, or deletions that do not significantly affect the structure, function, or expression of the polypeptide. Such variant recombinant Mg3A polypeptides will include those designed to circumvent the present description. In some embodiments, variants recombinant Mg3A polypeptides, compositions and methods comprising these variants are not substantially identical to SEQ ID NO:2 or SEQ ID NO:3, but rather include amino acid substitutions, insertions, or deletions that affect, in certain circumstances, substantially, the structure, function, or expression of the polypeptide herein such that improved characteristics, including, e.g., improved specific activity to hydrolyze a lignocellulosic substrate, improved expression in a desirable host organism, improved thermostability, pH stability, etc, as compared to that of a polypeptide of SEQ ID NO:2 or SEQ ID NO:3 can be achieved.

[00149] In some embodiments, the recombinant Mg3A polypeptide (including a variant thereof) has beta-glucosidase activity. Beta-glucosidase activity can be determined and measured using the assays described herein, for example, those described in Example 2, or by other assays known in the art.

[00150] Recombinant Mg3A polypeptides include fragments of "full-length" Mg3A polypeptides that retain beta-glucosidase activity. Preferably those functional fragments (i.e., fragments that retain beta-glucosidase activity) are at least 100 amino acid residues in length (e.g., at least 100 amino acid residues, at least 120 amino acid residues, at least 140 amino acid residues, at least 160 amino acid residues, at least 180 amino acid residues, at least 200 amino acid residues, at least 220 amino acid residues, at least 240 amino acid residues, at least 260

amino acid residues, at least 280 amino acid residues, at least 300 amino acid residues, at least 320 amino acid residues, or at least 350 amino acid residues in length or longer). Such fragments suitably retain the active site of the full-length precursor polypeptides or full length mature polypeptides but may have deletions of non-critical amino acid residues. The activity of fragments can be readily determined using the assays described herein, for example those described in **Example 2**, or by other assays known in the art.

[00151] In some embodiments, the Mg3A amino acid sequences and derivatives are produced as an N- and/or C-terminal fusion protein, for example, to aid in extraction, detection and/or purification and/or to add functional properties to the Mg3A polypeptides. Examples of fusion protein partners include, but are not limited to, glutathione-S-transferase (GST), 6XHis, GAL4 (DNA binding and/or transcriptional activation domains), FLAG-, MYC-tags or other tags known to those skilled in the art. In some embodiments, a proteolytic cleavage site is provided between the fusion protein partner and the polypeptide sequence of interest to allow removal of fusion sequences. Suitably, the fusion protein does not hinder the activity of the recombinant Mg3A polypeptide. In some embodiments, the recombinant Mg3A polypeptide is fused to a functional domain including a leader peptide, propeptide, binding domain and/or catalytic domain. Fusion proteins are optionally linked to the recombinant Mg3A polypeptide through a linker sequence that joins the Mg3A polypeptide and the fusion domain without significantly affecting the properties of either component. The linker optionally contributes functionally to the intended application.

[00152] The present disclosure provides host cells that are engineered to express one or more Mg3A polypeptides of the disclosure. Suitable host cells include cells of any microorganism (*e.g.*, cells of a bacterium, a protist, an alga, a fungus (*e.g.*, a yeast or filamentous fungus), or other microbe), and are preferably cells of a bacterium, a yeast, or a filamentous fungus.

[00153] Suitable host cells of the bacterial genera include, but are not limited to, cells of *Escherichia*, *Bacillus*, *Lactobacillus*, *Pseudomonas,* and *Streptomyces*. Suitable cells of bacterial species include, but are not limited to, cells of *Escherichia coli*, *Bacillus subtilis*, *Bacillus licheniformis*, *Lactobacillus brevis*, *Pseudomonas aeruginosa,* and *Streptomyces lividans*.

[00154] Suitable host cells of the genera of yeast include, but are not limited to, cells of *Saccharomyces*, *Schizosaccharomyces*, *Candida*, *Hansenula*, *Pichia*, *Kluyveromyces*, and *Phaffia*. Suitable cells of yeast species include, but are not limited to, cells of *Saccharomyces*

cerevisiae, *Schizosaccharomyces pombe*, *Candida albicans*, *Hansenula polymorpha*, *Pichia pastoris*, *P. canadensis*, *Kluyveromyces marxianus,* and *Phaffia rhodozyma*.

[00155] Suitable host cells of filamentous fungi include all filamentous forms of the subdivision *Eumycotina*. Suitable cells of filamentous fungal genera include, but are not limited to, cells of *Acremonium, Aspergillus, Aureobasidium, Bjerkandera, Ceriporiopsis, Chrysoporium, Coprinus, Coriolus, Corynascus, Chaertomium, Cryptococcus, Filobasidium, Fusarium, Gibberella, Humicola, Magnaporthe, Mucor, Myceliophthora, Mucor, Neocallimastix, Neurospora, Paecilomyces, Penicillium, Phanerochaete, Phlebia, Piromyces, Pleurotus,Scytaldium, Schizophyllum, Sporotrichum, Talaromyces, Thermoascus, Thielavia, Tolypocladium, Trametes,* and *Trichoderma.*

[00156] Suitable cells of filamentous fungal species include, but are not limited to, cells of *Aspergillus awamori, Aspergillus fumigatus, Aspergillus foetidus, Aspergillus japonicus, Aspergillus nidulans, Aspergillus niger, Aspergillus oryzae, Chrysosporium lucknowense, Fusarium bactridioides, Fusarium cerealis, Fusarium crookwellense, Fusarium culmorum, Fusarium graminearum, Fusarium graminum, Fusarium heterosporum, Fusarium negundi, Fusarium oxysporum, Fusarium reticulatum, Fusarium roseum, Fusarium sambucinum, Fusarium sarcochroum, Fusarium sporotrichioides, Fusarium sulphureum, Fusarium torulosum, Fusarium trichothecioides, Fusarium venenatum, Bjerkandera adusta, Ceriporiopsis aneirina, Ceriporiopsis aneirina, Ceriporiopsis caregiea, Ceriporiopsis gilvescens, Ceriporiopsis pannocinta, Ceriporiopsis rivulosa, Ceriporiopsis subrufa, Ceriporiopsis subvermispora, Coprinus cinereus, Coriolus hirsutus, Humicola insolens, Humicola lanuginosa, Mucor miehei, Myceliophthora thermophila, Neurospora crassa, Neurospora intermedia, Penicillium purpurogenum, Penicillium canescens, Penicillium solitum, Penicillium funiculosum Phanerochaete chrysosporium, Phlebia radiate, Pleurotus eryngii, Talaromyces flavus, Thielavia terrestris, Trametes villosa, Trametes versicolor, Trichoderma harzianum, Trichoderma koningii, Trichoderma longibrachiatum, Trichoderma reesei,* and *Trichoderma viride.*

[00157] Methods of transforming nucleic acids into these organisms are known in the art. For example, a suitable procedure for transforming *Aspergillus* host cells is described in EP 238 023.

[00158] In some embodiments, the recombinant Mg3A polypeptide is fused to a signal peptide to, for example, facilitate extracellular secretion of the recombinant Mg3A polypeptide. For example, in certain embodiments, the signal peptide is encoded by a sequence selected from SEQ ID NOs:13-42. In particular embodiments, the recombinant Mg3A polypeptide is

expressed in a heterologous organism as a secreted polypeptide. The compositions and methods herein thus encompass methods for expressing a Mg3A polypeptide as a secreted polypeptide in a heterologous organism. In some embodiments the recombinant Mg3A polypeptide is expressed in a heterologous organism intracellularly, for example, when the heterologous organism is an ethanologen microbe such as a *Saccharomyces cerevisiae* or a *Zymomonas mobilis*. In those cases, a cellibiose transporter gene can be introduced into the organism using genetic engineering tools, in order for the Mg3A polypeptide to act on the cellobiose substrate inside the organism to convert cellobiose into D-glucose, which is then metabolized or converted by the organism into ethanol.

[00159] The disclosure also provides expression cassettes and/or vectors comprising the above-described nucleic acids. Suitably, the nucleic acid encoding a Mg3A polypeptide of the disclosure is operably linked to a promoter. Promoters are well known in the art. Any promoter that functions in the host cell can be used for expression of a beta-glucosidase and/or any of the other nucleic acids of the present disclosure. Initiation control regions or promoters, which are useful to drive expression of a beta-glucosidase nucleic acids and/or any of the other nucleic acids of the present disclosure in various host cells are numerous and familiar to those skilled in the art (*see*, for example, WO 2004/033646 and references cited therein). Virtually any promoter capable of driving these nucleic acids can be used.

[00160] Specifically, where recombinant expression in a filamentous fungal host is desired, the promoter can be a filamentous fungal promoter. The nucleic acids can be, for example, under the control of heterologous promoters. The nucleic acids can also be expressed under the control of constitutive or inducible promoters. Examples of promoters that can be used include, but are not limited to, a cellulase promoter, a xylanase promoter, the 1818 promoter (previously identified as a highly expressed protein by EST mapping *Trichoderma*). For example, the promoter can suitably be a cellobiohydrolase, endoglucanase, or beta-glucosidase promoter. A particulary suitable promoter can be, for example, a *T. reesei* cellobiohydrolase, endoglucanase, or beta-glucosidase promoter. For example, the promoter is a cellobiohydrolase I (*cbh*1) promoter. Non-limiting examples of promoters include a *cbh1, cbh2, egl1, egl2, egl3, egl4, egl5, pki1, gpd1, xyn1*, or *xyn2* promoter. Additional non-limiting examples of promoters include a *T. reesei cbh1, cbh2, egl1, egl2, egl3, egl4, egl5, pki1, gpd1, xyn1*, or *xyn2* promoter.

[00161] The nucleic acid sequence encoding a Mg3A polypeptide herein can be included in a vector. In some aspects, the vector contains the nucleic acid sequence encoding the Mg3A polypeptide under the control of an expression control sequence. In some aspects, the expression

control sequence is a native expression control sequence. In some aspects, the expression control sequence is a non-native expression control sequence. In some aspects, the vector contains a selective marker or selectable marker. In some aspects, the nucleic acid sequence encoding the Mg3A polypeptide is integrated into a chromosome of a host cell without a selectable marker.

[00162] Suitable vectors are those which are compatible with the host cell employed. Suitable vectors can be derived, for example, from a bacterium, a virus (such as bacteriophage T7 or a M-13 derived phage), a cosmid, a yeast, or a plant. Suitable vectors can be maintained in low, medium, or high copy number in the host cell. Protocols for obtaining and using such vectors are known to those in the art (*see*, for example, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor, 1989).

[00163] In some aspects, the expression vector also includes a termination sequence. Termination control regions may also be derived from various genes native to the host cell. In some aspects, the termination sequence and the promoter sequence are derived from the same source.

[00164] An nucleic acid sequence encoding a Mg3A polypeptide can be incorporated into a vector, such as an expression vector, using standard techniques (Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 1982).

[00165] In some aspects, it may be desirable to over-express a Mg3A polypeptide and/or one or more of any other nucleic acid described in the present disclosure at levels far higher than currently found in naturally-occurring cells. In some embodiments, it may be desirable to under-express (*e.g.*, mutate, inactivate, or delete) an endogenous beta-glucosidase and/or one or more of any other nucleic acid described in the present disclosure at levels far below that those currently found in naturally-occurring cells.

B. *Mg3a* Polynucleotides

[00166] Another aspect of the compositions and methods described herein is a polynucleotide or a nucleic acid sequence that encodes a recombinant Mg3A polypeptide (including variants and fragments thereof) having beta-glucosidase activity. In some embodiments the polynucleotide is provided in the context of an expression vector for directing the expression of a Mg3A polypeptide in a heterologous organism, such as one identified herein. The polynucleotide that encodes a recombinant Mg3A polypeptide may be operably-linked to regulatory elements (*e.g.*, a promoter, terminator, enhancer, and the like) to assist in expressing the encoded polypeptides.

[00167] An example of a polynucleotide sequence encoding a recombinant Mg3A polypeptide has the nucleotide sequence of SEQ ID NO: 1. Similar, including substantially identical, polynucleotides encoding recombinant Mg3A polypeptides and variants may occur in nature, *e.g.*, in other strains or isolates of *Magnaporthe grisea,* or *Magnaporthe sp.*. In view of the degeneracy of the genetic code, it will be appreciated that polynucleotides having different nucleotide sequences may encode the same Mg3A polypeptides, variants, or fragments.

[00168] In some embodiments, polynucleotides encoding recombinant Mg3A polypeptides have a specified degree of amino acid sequence identity to the exemplified polynucleotide encoding a Mg3A polypeptide, *e.g.*, at least 85%, at least 86%, at least 87%, at least 88%, at least 89%, 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 even at least 99% sequence identity to the amino acid sequence of SEQ ID NO: 2. Homology can be determined by amino acid sequence alignment, *e.g.*, using a program such as BLAST, ALIGN, or CLUSTAL, as described herein.

[00169] In some embodiments, the polynucleotide that encodes a recombinant Mg3A polypeptide is fused in frame behind (*i.e.*, downstream of) a coding sequence for a signal peptide for directing the extracellular secretion of a recombinant Mg3A polypeptide. As described herein, the term "heterologous" when used to refer to a signal sequence used to express a polypeptide of interest, it is meant that the signal sequence and the polypeptide of interest are from different organisms. Heterologous signal sequences include, for example, those from other fungal cellulase genes, such as, e.g., the signal sequence of *Trichoderma reesei* Bgl1, of SEQ ID NO:13. Expression vectors may be provided in a heterologous host cell suitable for expressing a recombinant Mg3A polypeptide, or suitable for propagating the expression vector prior to introducing it into a suitable host cell.

[00170] In some embodiments, polynucleotides encoding recombinant Mg3A polypeptides hybridize to the polynucleotide of SEQ ID NO: 1 (or to the complement thereof) under specified hybridization conditions. Examples of conditions are intermediate stringency, high stringency and extremely high stringency conditions, which are described herein.

[00171] Mg3A polynucleotides may be naturally occurring or synthetic (*i.e.*, man-made), and may be codon-optimized for expression in a different host, mutated to introduce cloning sites, or otherwise altered to add functionality.

C. Mg3A Vectors and Host Cells

[00172] In order to produce a disclosed recombinant Mg3A polypeptide, the DNA encoding the polypeptide can be chemically synthesized from published sequences or can be obtained

directly from host cells harboring the gene (*e.g.*, by cDNA library screening or PCR amplification). In some embodiments, the Mg3A polynucleotide is included in an expression cassette and/or cloned into a suitable expression vector by standard molecular cloning techniques. Such expression cassettes or vectors contain sequences that assist initiation and termination of transcription (*e.g.*, promoters and terminators), and typically can also contain one or more selectable markers.

[00173] The expression cassette or vector is introduced into a suitable expression host cell, which then expresses the corresponding Mg3A polynucleotide. Suitable expression hosts may be bacterial or fungal microbes. Bacterial expression host may be, for example, *Escherichia (e.g., Escherichia coli), Pseudomonas (e.g., P. fluorescens or P. stutzerei), Proteus (e.g., Proteus mirabilis), Ralstonia (e.g., Ralstonia eutropha), Streptomyces, Staphylococcus (e.g., S. carnosus), Lactococcus (e.g., L. lactis),* or *Bacillus (e.g., Bacillus subtilis, Bacillus megaterium, Bacillus licheniformis,* etc.). Fungal expression hosts may be, for example, yeasts, which can also serve as ethanologens. Yeast expression hosts may be, for example, *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Yarrowia lipolytica, Hansenula polymorpha, Kluyveromyces lactis* or *Pichia pastoris*. Fungal expression hosts may also be, for example, filamentous fungal hosts including *Aspergillus niger, Chrysosporium lucknowense, Myceliophthora thermophila, Aspergillus (e.g., A. oryzae, A. niger, A. nidulans,* etc.*)* or *Trichoderma reesei*. Also suited are mammalian expression hosts such as mouse (*e.g.*, NS0), Chinese Hamster Ovary (CHO) or Baby Hamster Kidney (BHK) cell lines. Other eukaryotic hosts such as insect cells or viral expression systems (*e.g.*, bacteriophages such as M13, T7 phage or Lambda, or viruses such as Baculovirus) are also suitable for producing the Mg3A polypeptide.

[00174] Promoters and/or signal sequences associated with secreted proteins in a particular host of interest are candidates for use in the heterologous production and secretion of Mg3A polypeptides in that host or in other hosts. As an example, in filamentous fungal systems, the promoters that drive the genes for cellobiohydrolase I (cbh1), glucoamylase A (glaA), TAKA-amylase (amyA), xylanase (exlA), the gpd-promoter cbh1, cbhll, endoglucanase genes eg1-eg5, Cel61B, Cel74A, gpd promoter, Pgk1, pki1, EF-1alpha, tef1, cDNA1 and hex1 are suitable and can be derived from a number of different organisms (*e.g., A. niger, T. reesei, A. oryzae, A. awamori, A. nidulans*).

[00175] In some embodiments, the Mg3A polynucleotide is recombinantly associated with a polynucleotide encoding a suitable homologous or heterologous signal sequence that leads to

secretion of the recombinant Mg3A polypeptide into the extracellular (or periplasmic) space, thereby allowing direct detection of enzyme activity in the cell supernatant (or periplasmic space or lysate). Suitable signal sequences for *Escherichia coli*, other Gram negative bacteria and other organisms known in the art include those that drive expression of the HlyA, DsbA, Pbp, PhoA, PelB, OmpA, OmpT or M13 phage Gill genes. For *Bacillus subtilis*, Gram-positive organisms and other organisms known in the art, suitable signal sequences further include those that drive expression of the AprE, NprB, Mpr, AmyA, AmyE, Blac, SacB, and for *S. cerevisiae* or other yeast, including the killer toxin, Bar1, Suc2, Mating factor alpha, Inu1A or Ggplp signal sequence. Signal sequences can be cleaved by a number of signal peptidases, thus removing them from the rest of the expressed protein. Fungal expression signal sequences may be one that is selected from, for example, SEQ ID NOs: 13-37, herein. Yeast expression signal sequences may be one that is selected from, for example, SEQ ID NOs:38-40. Signal sequences that might be suitable for use to express Mg3A polypeptides of the invention in *Zymomonas mobilis* may include, for example, one selected from SEQ ID NOs:41-42. (Linger J.G. *et al.*, (2010) *Appl. Environ. Microbiol.* 76(19):6360-6369).

[00176] In some embodiments, the recombinant Mg3A polypeptide is expressed alone or as a fusion with other peptides, tags or proteins located at the N- or C-terminus (*e.g.*, 6XHis, HA or FLAG tags). Suitable fusions include tags, peptides or proteins that facilitate affinity purification or detection (*e.g.*, 6XHis, HA, chitin binding protein, thioredoxin or FLAG tags), as well as those that facilitate expression, secretion or processing of the target beta-glucosidases Suitable processing sites include enterokinase, STE13, Kex2 or other protease cleavage sites for cleavage in vivo or in vitro.

[00177] Mg3A polynucleotides are introduced into expression host cells by a number of transformation methods including, but not limited to, electroporation, lipid-assisted transformation or transfection ("lipofection"), chemically mediated transfection (*e.g.*, CaCl and/or CaP), lithium acetate-mediated transformation (*e.g.*, of host-cell protoplasts), biolistic "gene gun" transformation, PEG-mediated transformation (*e.g.*, of host-cell protoplasts), protoplast fusion (*e.g.*, using bacterial or eukaryotic protoplasts), liposome-mediated transformation, *Agrobacterium tumefaciens*, adenovirus or other viral or phage transformation or transduction.

D. Cell culture media

[00178] Generally, the microorganism is cultivated in a cell culture medium suitable for production of the Mg3A polypeptides described herein. The cultivation takes place in a suitable

nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures and variations known in the art. Suitable culture media, temperature ranges and other conditions for growth and cellulase production are known in the art. As a non-limiting example, a typical temperature range for the production of cellulases by *Trichoderma reesei* is 24°C to 37°C, for example, between 25°C and 30°C.

1. Cell culture conditions

[00179] Materials and methods suitable for the maintenance and growth of fungal cultures are well known in the art. In some aspects, the cells are cultured in a culture medium under conditions permitting the expression of one or more beta-glucosidase polypeptides encoded by a nucleic acid inserted into the host cells. Standard cell culture conditions can be used to culture the cells. In some aspects, cells are grown and maintained at an appropriate temperature, gas mixture, and pH. In some aspects, cells are grown at in an appropriate cell medium.

IV. Activities of Mg3A

[00180] The recombinant Mg3A polypeptides disclosed herein have beta-glucosidase activity or a capacity to hydrolyze cellobiose and liberating D-glucose therefrom. The Mg3A polypeptides herein have higher beta-glucosidase activity and improved or increased capacity to liberate D-glucose from cellobiose than the benchmark high fidelity beta-glucosidase Bgl1 of *Trichoderma reesei*, under the same saccharification conditions. In some embodiments, the Mg3A polypeptides herein may have higher beta-glucosidase activity and/or improved or increased capacity to liberate D-glucose from cellobiose than another benchmark beta-glucosidase B-glu of *Aspergillus niger*.

[00181] As shown in **Examples 3**, the recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, has at least about 5%, at least about 10%, preferably at least about 15%, more preferably at least about 20% higher activity hydrolyzing a chloro-nitro-phenyl-glucoside (CNPG) substrate. In some embodiments, the recombinant Mg3A polypeptide, as compared to the *Aspergillus niger* B-glu, has at least about 2-fold, at least about 5-fold, at least about 7-fold, preferably at least about 9-fold, more preferably at least about 10-fold higher activity hydrolyzing the CNPG substrate.

[00182] The recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, has dramatically improved or increased, for example, at least about 30% higher, more preferably at least about 40% higher, preferably at least about 50% higher, more preferably at least 55% higher, preferably at least 60%, even more preferably at least about 65% higher, preferably at least about 70% higher, more preferably at least about 75% higher, and most preferably at least

about 80% higher cellobiase activity, which measures the enzymes' capability to catalyze the hydrolysis of cellobiose, liberating D-glucose. In some embodiments, the recombinant Mg3A polypeptide, as compared to the *Aspergillus niger* B-glu, has about 1/2, about 1/3, about 1/4, about 1/5, or even about 1/6 of the capacity to catalyze the hydrolysis of cellobiose, liberating D-glucose.

[00183] In some embodiments, the recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, has an about 10% reduced, about 20% reduced, about 30% reduced, or even about 40% reduced hydrolysis activity ratio over CNPG/cellobiose. In some embodiments, the Mg3A polypeptide, as compared to the *Aspergillus niger* B-glu, has about 2-fold, about 5-fold, about 7-fold, about 10-fold, about 15 fold, or even about 20-fold higher relative hydrolysis activity ratio over CNPG/cellobiose.

[00184] As shown in **Example 4,** the recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, produced more glucose but equal or less amount of total sugars from a phosphoric acid swollen cellulose substrate.

[00185] As shown in **Example 5,** the recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, also produced more glucose but equal or less amount of total sugars from a dilute ammonia pretreated corn stover substrate.

[00186] As shown in **Example 6,** the recombinant Mg3A polypeptide, as compared to the *Trichoderma reesei* Bgl1, is more effective at glucan conversion and produces more glucose from a dilute ammonia pretreated corn stover substrate at a higher hydrolysis reaction temperature, 55°C. At a temperature above 50°C, it has been observed that *Trichoderma reesei* Bgl1 imparts poor beta-glucosidase performance, which has been attributed to the loss of enzymatic activity due to poor thermostability. In contrast, although the measured melting temperature (Tm) of *Trichoderma reesei* Bgl1 is about 10°C higher than that of Mg3A, it was surprisingly observed that the Mg3A activity and performance remains intact at a saccharification temperature of as high as 55°C. This thermostability and high-temperature performance benefit of Mg3A makes it a particularly suitable beta-glucosidase for biomass conversion.

V. Compositions Comprising a Recombinant Beta-Glucosidase Mg3A Polypeptide

[00187] The present disclosure provides engineered enzyme compositions (e.g., cellulase compositions) or fermentation broths enriched with a recombinant Mg3A polypeptides. In some aspects, the composition is a cellulase composition. The cellulase composition can be, *e.g.,* a filamentous fungal cellulase composition, such as a *Trichoderma* cellulase composition. In

some aspects, the composition is a cell comprising one or more nucleic acids encoding one or more cellulase polypeptides. In some aspects, the composition is a fermentation broth comprising cellulase activity, wherein the broth is capable of converting greater than about 50% by weight of the cellulose present in a biomass sample into sugars. The term "fermentation broth" and "whole broth" as used herein refers to an enzyme preparation produced by fermentation of an engineered microorganism that undergoes no or minimal recovery and/or purification subsequent to fermentation. The fermentation broth can be a fermentation broth of a filamentous fungus, for example, a *Trichoderma, Humicola, Fusarium, Aspergillus, Neurospora, Penicillium, Cephalosporium, Achlya, Podospora, Endothia, Mucor, Cochliobolus, Pyricularia, Myceliophthora* or *Chrysosporium* fermentation broth. In particular, the fermentation broth can be, for example, one of *Trichoderma spp.* such as a *Trichoderma reesei,* or *Penicillium spp.,* such as a *Penicillium funiculosum.* The fermentation broth can also suitably be a cell-free fermentation broth. In one aspect, any of the cellulase, cell, or fermentation broth compositions of the present invention can further comprise one or more hemicellulases.

[00188] In some aspects, the whole broth composition is expressed in *T. reesei* or an engineered strain thereof. In some aspects the whole broth is expressed in an integrated strain of *T. reesei* wherein a number of cellulases including a Mg3A polypeptide has been integrated into the genome of the *T. reesei* host cell. In some aspects, one or more components of the polypeptides expressed in the integrated *T. reesei* strain have been deleted.

[00189] In some aspects, the whole broth composition is expressed in *A. niger* or an engineered strain thereof.

[00190] Alternatively, the recombinant Mg3A polypeptides can be expressed intracellularly. Optionally, after intracellular expression of the enzyme variants, or secretion into the periplasmic space using signal sequences such as those mentioned above, a permeabilization or lysis step can be used to release the recombinant Mg3A polypeptide into the supernatant. The disruption of the membrane barrier is effected by the use of mechanical means such as ultrasonic waves, pressure treatment (French press), cavitation, or by the use of membrane-digesting enzymes such as lysozyme or enzyme mixtures. A variation of this embodiment includes the expression of a recombinant Mg3A polypeptide in an ethanologen microbe intracellularly. For example, a cellobiose transporter can be introduced through genetic engineering into the same ethanologen microbe such that cellobiose resulting from the hydrolysis of a lignocellulosic biomass can be transported into the ethanologen organism, and can therein be hydrolyzed and turned into D-glucose, which can in turn be metabolized by the ethanologen.

[00191] In some aspects, the polynucleotides encoding the recombinant Mg3A polypeptide are expressed using a suitable cell-free expression system. In cell-free systems, the polynucleotide of interest is typically transcribed with the assistance of a promoter, but ligation to form a circular expression vector is optional. In some embodiments, RNA is exogenously added or generated without transcription and translated in cell-free systems.

VI. Uses of Mg3A Polypeptides to Hydrolyze a Lignocellulosic Biomass Substrate

[00192] In some aspects, provided herein are methods for converting lignocelluloses biomass to sugars, the method comprising contacting the biomass substrate with a composition disclosed herein comprising a Mg3A polypeptide in an amount effective to convert the biomass substrate to fermentable sugars. In some aspects, the method further comprises pretreating the biomass with acid and/or base and/or mechanical or other physical means In some aspects the acid comprises phosphoric acid. In some aspects, the base comprises sodium hydroxide or ammonia. In some aspects, the mechanical means may include, for example, pulling, pressing, crushing, grinding, and other means of physically breaking down the lignocellulosic biomass into smaller physical forms. Other physical means may also include, for example, using steam or other pressurized fume or vapor to "loosen" the lignocellulosic biomass in order to increase accessibility by the enzymes to the cellulose and hemicellulose. In certain embodiments, the method of pretreatment may also involve enzymes that are capable of breaking down the lignin of the lignocellulosic biomass substrate, such that the accessibility of the enzymes of the biomass hydrolyzing enzyme composition to the cellulose and the hemicelluloses of the biomass is increased.

[00193] **Biomass:** The disclosure provides methods and processes for biomass saccharification, using the enzyme compositions of the disclosure, comprising a Mg3A polypeptide. The term "biomass," as used herein, refers to any composition comprising cellulose and/or hemicellulose (optionally also lignin in lignocellulosic biomass materials). As used herein, biomass includes, without limitation, seeds, grains, tubers, plant waste (such as, for example, empty fruit bunches of the palm trees, or palm fiber wastes) or byproducts of food processing or industrial processing (*e.g.*, stalks), corn (including, *e.g.,* cobs, stover, and the like), grasses (including, *e.g.*, Indian grass, such as *Sorghastrum nutans*; or, switchgrass, *e.g.*, *Panicum* species, such as *Panicum virgatum*), perennial canes (*e.g.,* giant reeds), wood (including, *e.g.,* wood chips, processing waste), paper, pulp, and recycled paper (including, *e.g.*, newspaper, printer paper, and the like). Other biomass materials include, without limitation, potatoes, soybean (*e.g.,* rapeseed), barley, rye, oats, wheat, beets, and sugar cane bagasse.

[00194] The disclosure therefore provides methods of saccharification comprising contacting a composition comprising a biomass material, for example, a material comprising xylan, hemicellulose, cellulose, and/or a fermentable sugar, with a Mg3A polypeptide of the disclosure, or a Mg3A polypeptide encoded by a nucleic acid or polynucleotide of the disclosure, or any one of the cellulase or non-naturally occurring hemicellulase compositions comprising a Mg3A polypeptide, or products of manufacture of the disclosure.

[00195] The saccharified biomass (*e.g.,* lignocellulosic material processed by enzymes of the disclosure) can be made into a number of bio-based products, *via* processes such as, *e.g.,* microbial fermentation and/or chemical synthesis. As used herein, "microbial fermentation" refers to a process of growing and harvesting fermenting microorganisms under suitable conditions. The fermenting microorganism can be any microorganism suitable for use in a desired fermentation process for the production of bio-based products. Suitable fermenting microorganisms include, without limitation, filamentous fungi, yeast, and bacteria. The saccharified biomass can, for example, be made it into a fuel (*e.g.*, a biofuel such as a bioethanol, biobutanol, biomethanol, a biopropanol, a biodiesel, a jet fuel, or the like) *via* fermentation and/or chemical synthesis. The saccharified biomass can, for example, also be made into a commodity chemical (*e.g.*, ascorbic acid, isoprene, 1,3-propanediol), lipids, amino acids, polypeptides, and enzymes, *via* fermentation and/or chemical synthesis.

[00196] **Pretreatment:** Prior to saccharification or enzymatic hydrolysis and/or fermentation of the fermentable sugars resulting from the saccharifiction, biomass (*e.g.*, lignocellulosic material) is preferably subject to one or more pretreatment step(s) in order to render xylan, hemicellulose, cellulose and/or lignin material more accessible or susceptible to the enzymes in the enzymatic composition (for example, the enzymatic composition of the present invention comprising a Mg3A polypeptide) and thus more amenable to hydrolysis by the enzyme(s) and/or the enzyme compositions.

[00197] In some aspects, a suitable pretreatment method may involve subjecting biomass material to a catalyst comprising a dilute solution of a strong acid and a metal salt in a reactor. The biomass material can, *e.g.,* be a raw material or a dried material. This pretreatment can lower the activation energy, or the temperature, of cellulose hydrolysis, ultimately allowing higher yields of fermentable sugars. See, *e.g.*, U.S. Patent Nos. 6,660,506; 6,423,145.

[00198] In some aspects, a suitable pretreatment method may involve subjecting the biomass material to a first hydrolysis step in an aqueous medium at a temperature and a pressure chosen to effectuate primarily depolymerization of hemicellulose without achieving significant

depolymerization of cellulose into glucose. This step yields a slurry in which the liquid aqueous phase contains dissolved monosaccharides resulting from depolymerization of hemicellulose, and a solid phase containing cellulose and lignin. The slurry is then subject to a second hydrolysis step under conditions that allow a major portion of the cellulose to be depolymerized, yielding a liquid aqueous phase containing dissolved/soluble depolymerization products of cellulose. See, *e.g.*, U.S. Patent No. 5,536,325.

[00199] In further aspects, a suitable pretreatment method may involve processing a biomass material by one or more stages of dilute acid hydrolysis using about 0.4% to about 2% of a strong acid; followed by treating the unreacted solid lignocellulosic component of the acid hydrolyzed material with alkaline delignification. See, *e.g.*, U.S. Patent No. 6,409,841.

[00200] In yet further aspects, a suitable pretreatment method may involve pre-hydrolyzing biomass (*e.g.*, lignocellulosic materials) in a pre-hydrolysis reactor; adding an acidic liquid to the solid lignocellulosic material to make a mixture; heating the mixture to reaction temperature; maintaining reaction temperature for a period of time sufficient to fractionate the lignocellulosic material into a solubilized portion containing at least about 20% of the lignin from the lignocellulosic material, and a solid fraction containing cellulose; separating the solubilized portion from the solid fraction, and removing the solubilized portion while at or near reaction temperature; and recovering the solubilized portion. The cellulose in the solid fraction is rendered more amenable to enzymatic digestion. See, *e.g.*, U.S. Patent No. 5,705,369. In a variation of this aspect, the pre-hydrolyzing can alternatively or further involves pre-hydrolysis using enzymes that are, for example, capable of breaking down the lignin of the lignocellulosic biomass material.

[00201] In yet further aspects, suitable pretreatments may involve the use of hydrogen peroxide H_2O_2. See Gould, 1984, Biotech, and Bioengr. 26:46-52.

[00202] In other aspects, pretreatment can also comprise contacting a biomass material with stoichiometric amounts of sodium hydroxide and ammonium hydroxide at a very low concentration. See Teixeira *et al.*, (1999) Appl. Biochem.and Biotech. 77-79:19-34.

[00203] In some embodiments, pretreatment can comprise contacting a lignocellulose with a chemical (*e.g.*, a base, such as sodium carbonate or potassium hydroxide) at a pH of about 9 to about 14 at moderate temperature, pressure, and pH. *See,* Published International Application WO2004/081185. Ammonia is used, for example, in a preferred pretreatment method. Such a pretreatment method comprises subjecting a biomass material to low ammonia concentration

under conditions of high solids. See, *e.g.*, U.S. Patent Publication No. 20070031918 and Published international Application WO 06110901.

A. The Saccharification Process

[00204] In some aspects, provided herein is a saccharification process comprising treating biomass with an enzyme composition comprising a polypeptide, wherein the polypeptide has beta-glucosidase activity and wherein the process results in at least about 50 wt.% (e.g., at least about 55 wt.%, 60 wt.%, 65 wt.%, 70 wt.%, 75 wt.%, or 80 wt.%) conversion of biomass to fermentable sugars. In some aspects, the biomass comprises lignin. In some aspects the biomass comprises cellulose. In some aspects the biomass comprises hemicellulose. In some aspects, the biomass comprising cellulose further comprises one or more of xylan, galactan, or arabinan. In some aspects, the biomass may be, without limitation, seeds, grains, tubers, plant waste (e.g., empty fruit bunch from palm trees, or palm fiber waste) or byproducts of food processing or industrial processing (*e.g.*, stalks), corn (including, *e.g.,* cobs, stover, and the like), grasses (including, *e.g.*, Indian grass, such as *Sorghastrum nutans*; or, switchgrass, *e.g.*, *Panicum* species, such as *Panicum virgatum*), perennial canes (*e.g.,* giant reeds), wood (including, *e.g.,* wood chips, processing waste), paper, pulp, and recycled paper (including, *e.g.*, newspaper, printer paper, and the like), potatoes, soybean (*e.g.,* rapeseed), barley, rye, oats, wheat, beets, and sugar cane bagasse. In some aspects, the material comprising biomass is subject to one or more pretreatment methods/steps prior to treatment with the polypeptide. In some aspects, the saccharification or enzymatic hydrolysis further comprises treating the biomass with an enzyme composition comprising a Mg3A polypeptide of the invention. The enzyme composition may, for example, comprise one or more other cellulases, in addition to the Mg3A polypeptide. Alternatively, the enzyme composition may comprise one or more other hemicellulases. In certain embodiments, the enzyme composition comprises a Mg3A polypeptide of the invention, one or more other cellulases, one or more hemicellulases. In some embodiments, the enzyme composition is a whole broth composition.

[00205] In some aspects, provided is a saccharification process comprising treating a lignocellulosic biomass material with a composition comprising a polypeptide, wherein the polypeptide has at least about 75% (e.g., at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%) sequence identity to SEQ ID NO:2, and wherein the process results in at least about 50% (e.g., at least about 55%, 60%, 65%, 70%, 75%, 80%, 85%, or 90%) by weight conversion of biomass to fermentable sugars. In some aspects, lignocellulosic

biomass material has been subject to one or more pretreatment methods/steps as described herein.

[00206] Other aspects and embodiments of the present compositions and methods will be apparent from the foregoing description and following examples.

EXAMPLES

[00207] The following examples are provided to demonstrate and illustrate certain preferred embodiments and aspects of the present disclosure and should not be construed as limiting.

EXAMPLE 1

1-A. Cloning & Expression of Gene Expression of Mg3A and benchmark *T. reesei* Bgl1

1-A-a. Construction of the *T. reesei bgl1* expression vector

[00208] The N-terminal portion of the native *T. reesei* β-glucosidase gene *bgl1* was codon optimized (DNA 2.0, Menlo Park, CA). This synthesized portion comprised the first 447 bases of the coding region of this enzyme. This fragment was then amplified by PCR using primers SK943 and SK941 (below). The remaining region of the native *bgl1* gene was PCR amplified from a genomic DNA sample extracted from *T. reesei* strain RL-P37 (Sheir-Neiss, G *et al.* (1984) Appl. Microbiol. Biotechnol. 20:46-53), using the primers SK940 and SK942 (below). These two PCR fragments of the *bgl1* gene were fused together in a fusion PCR reaction, using primers SK943 and SK942:

Forward Primer SK943: (5'– CACCATGAGATATAGAACAGCTGCCGCT-3') (SEQ ID NO:5)

Reverse Primer SK941: (5'-CGACCGCCCTGCGGAGTCTTGCCCAGTGGTCCCGCGACAG-3') (SEQ ID NO: 6)

Forward Primer (SK940): (5'-CTGTCGCGGGACCACTGGGCAAGACTCCGCAGGGCGGTCG-3') (SEQ ID NO:7)

Reverse Primer (SK942): (5'– CCTACGCTACCGACAGAGTG-3') (SEQ ID NO:8)

[00209] The resulting fusion PCR fragments were cloned into the Gateway ® Entry vector pENTR™/D-TOPO® (Figure 1), and transformed into *E. coli One Shot® TOP10* Chemically Competent cells (Invitrogen) resulting in the intermediate vector, pENTR TOPO-Bgl1(943/942) (**Figure 1**). The nucleotide sequence of the inserted DNA was determined. The pENTR-943/942 vector with the correct *bgl1* sequence was recombined with pTrex3g using a LR clonase® reaction (*see,* protocols outlined by Invitrogen). The LR clonase reaction mixture was

transformed into *E. coli One Shot®TOP10* Chemically Competent cells (Invitrogen), resulting in the expression vector, pTrex3g 943/942 (**Figure 2**). The vector also contained the *Aspergillus nidulans amdS* gene, encoding acetamidase, as a selectable marker for transformation of *T. reesei*. The expression cassette was PCR amplified with primers SK745 and SK771 (below) to generate the product for transformation.

Forward Primer SK771: (5' – GTCTAGACTGGAAACGCAAC -3') (SEQ ID NO:9)

Reverse Primer SK745: (5' – GAGTTGTGAAGTCGGTAATCC -3') (SEQ ID NO:10)

1-A-b. Construction of the *mg3a* expression vector

[00210] The open reading frame of the beta-glucosidase gene was amplified by PCR using genomic DNA extracted from *Magnaporthe grisea* as the template. The open reading frame was amplified with the native signal sequence. The PCR thermocycler used was DNA Engine Tetrad 2 Peltier Thermal Cycler (BioRad Laboratories). The DNA polymerase used was PfuUltra II Fusion HS DNA Polymerase (Stratagene) or a similar quality proofreading DNA polymerase. The primers used to amplify the open reading frame were as follows:

Mg3A-F: 5'-CAC CAT GCG TTT CTC CGG GAT CGT -3' (SEQ ID NO:11)
Mg3A-R: 5'-TCA GTT CAG GTC AGC ACT CAG ATG GAG C-3' (SEQ ID NO:12)

[00211] The Mg3A-F forward primer included four additional nucleotides (sequences – CACC) at the 5'-end to facilitate directional cloning into pENTR/D-TOPO. The PCR product of the open reading frame was purified using a Qiaquick PCR Purification Kit (Qiagen, Valencia, CA). The purified PCR product was cloned into the pENTR/D-TOPO vector (Invitrogen), transformed into TOP10 chemically competent *E. coli* cells (Invitrogen, Carlsbad, CA) and plated on LA plates with 50 ppm kanamycin. Plasmid DNA was obtained from the *E. coli* transformants using a QIAspin plasmid preparation kit (Qiagen).

[00212] Sequence data for the DNA inserted in the pENTR/D-TOPO vector was obtained using M13 forward and reverse primers. A pENTR/D-TOPO vector with the correct DNA sequence of the open reading frame was recombined with the pTrex3gM destination vector **(Figure 2)** using LR clonase reaction mixture (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

[00213] The product of the LR clonase reaction was subsequently transformed into TOP10 chemically competent *E. coli* cells which were then plated on LA containing 50 ppm carbenicillin. The resulting pExpression construct was pTrex3gM containing the Mg3A open reading frame and the *Aspergillus tubingensis* acetamidase selection marker (*amdS*). DNA of

the pExpression construct was isolated using a Qiagen miniprep kit and used for transformation of *Trichoderma reesei*.

[00214] Either the pExpression plasmid or a PCR product of the expression cassette was transformed into a *T. reesei* six-fold-delete strain (*see, e.g.,* the description in Published International Patent Application Publication No. WO 2010/141779) using the PEG-mediated protoplast method with slight modifications as described below. For protoplast preparation, spores were grown for 16-24 h at 24°C in *Trichoderma* Minimal Medium MM, which contained 20 g/L glucose, 15 g/L KH_2PO_4, pH 4.5, 5 g/L $(NH_4)_2SO_4$, 0.6 g/L $MgSO_4x7H_2O$, 0.6 g/L $CaCl_2$ x $2H_2O$, 1 mL of 1000 X *T. reesei* Trace elements solution (which contained 5 g/L $FeSO_4$ x $7H_2O$, 1.4 g/L $ZnSO_4$ x $7H_2O$, 1.6 g/L $MnSO_4$ x H_2O, 3.7 g/L $CoCl_2$ x $6H_2O$) with shaking at 150 rpm. Germinating spores were harvested by centrifugation and treated with 50 mg/mL of Glucanex G200 (Novozymes AG) solution to lyse the fungal cell walls. Further preparation of the protoplasts was performed in accordance with a method described by Penttilä *et al.* (1987) Gene 61: 155-164. The transformation mixtures, which contained about 1 µg of DNA and 1-5x 10^7 protoplasts in a total volume of 200 µL, were each treated with 2 mL of 25% PEG solution, diluted with 2 volumes of 1.2 M sorbitol/10 mM Tris, pH7.5, 10 mM $CaCl_2$, mixed with 3% selective top agarose MM containing 5 mM uridine and 20 mM acetamide. The resulting mixtures were poured onto 2% selective agarose plate containing uridine and acetamide. Plates were incubated further for 7-10 d at 28°C before single transformants were picked onto fresh MM plates containing uridine and acetamide. Spores from independent clones were used to inoculate a fermentation medium in shake flasks.

[00215] The fermentation media was 36 mL of defined broth containing glucose/sophorose and 2 g/L uridine, such as Glycine Minimal media (6.0 g/L glycine; 4.7 g/L $(NH_4)_2SO_4$; 5.0 g/L KH_2PO_4; 1.0 g/L $MgSO_4•7H_2O$; 33.0 g/L PIPPS; pH 5.5) with post sterile addition of ~2% glucose/sophorose mixture as the carbon source, 10 ml/L of 100g/L of $CaCl_2$, 2.5 ml/L of *T. reesei* trace elements (400X) : 175g/L Citric acid anhydrous; 200g/L $FeSO_4•7H_2O$; 16g/L $ZnSO_4•7H_2O$; 3.2 g/L $CuSO_4•5H_2O$; 1.4 g/L $MnSO_4•H_2O$; 0.8 g/L H_3BO_3, in 250 ml Thomson Ultra Yield Flasks (Thomson Instrument Co., Oceanside, CA).

1-A-c. Construction of a yeast shuttle vector pSC11

[00216] A yeast shuttle vector can be constructed in accordance with the vector map of **Figure 7.** This vector can be used to express a Mg3A polypeptide in *Saccharomyces cerevisiae* intracellularly. A cellobiose transporter can be introduced into the *Saccharomyces cerevisiae* in

the same shuttle vector or in a separate vector using known methods, such as, for example, those described by Ha *et al.*, (2011), PNAS, 108(2): 504-509.

[00217] Transformation of expression cassettes can be performed using the yeast EZ-Transformation kit. Transformants can be selected using YSC medium, which contains 20 g/L cellobiose. The successful introduction of the expression cassettes into yeast can be confirmed by colony PCR with specific primers.

[00218] Yeast strains can be cultivated in accordance with known methods and protocols. For example, they can be cultivated at 30°C in a YP medium (10 g/L yeast extract, 20 g/L Bacto peptone) with 20 g/L glucose. To select transformants using an amino acid auxotrophic marker, yeast synthetic complete (YSC) medium may be used, which contains 6.7 g/L yeast nitrogen base plus 20 g/L glucose, 20 g/L agar, and CSM-Leu-Trp-Ura to supply nucleotides and amino acids.

1-A-d. Construction of a *Zymomonas mobilis* integration vector pZC11.

[00219] A *Zymomonas mobilis* integration vector pZC11 can be constructed in accordance with the vector map of **Figure 8.** This vector can be used to express a Mg3A polypeptide in *Zymomonas mobilis* intracellularly. A cellobiose transporter can be introduced into the *Zymomonas mobilis* in the same integration vector or in a separate vector using known methods of introducing those transporters into a bacterial cell, such as, for example, those described by Sekar *et al.*, (2012) Applied Environmental Microbiology, 78(5):1611-1614..

[00220] Successful introduction of the integration vector as well as the cellobiose transporter gene can be confirmed using various known approaches, for example by PCR using confirmatory primers specifically designed for this purpose.

[00221] *Zymomonas mobilis* strains can be cultivated and fermented according to known methods, such as, for example, those described in U.S. Patent No. 7,741,119.

1-B. Purification of *T. reesei* Bgl1 & Mg3A

[00222] *T. reesei* Bgl1 was over-expressed in, and purified from, the fermentation broth of a six-fold-deleted *Trichoderma reesei* host strain (*see, e.g.,* the description in Published International Patent Application Publication No. WO 2010/141779). A concentrated broth was loaded onto a G25 SEC column (GE Healthcase Bio-Sciences) and was buffer-exchanged against 50 mM sodium acetate, pH 5.0. The buffer exchanged Bgl1 was then loaded onto a 25 mL column packed with amino benzyl-S-glucopyranosyl sepharose affinity matrix. After extensive washing with 250 mM sodium chloride in 50 mM sodium acetate, pH 5.0, the bound

fraction was eluted with 100 mM glucose in 50 mM sodium acetate and 250 mM sodium chloride, pH 5.0. The eluted fractions that tested positive for chloro-nitro-phenyl glucoside (CNPG) activity were pooled and concentrated. A single band corresponding to the MW of the *T. reesei* Bgl1 on SDS-PAGE and confirmed by mass spectrometry verified the purity of the eluted Bgl1. The final stock concentration was determined to be 2.2 mg/mL by absorbance at 280 nm.

[00223] A Mg3A expressed by *Tricoderma reesei* as described above can be purified from a concentrated fermentation broth by first diluting 100 mg into a 50 mM MES buffer, pH 6.0. The Mg3A can be enriched by loading 2 mg protein per mL resin onto a SP Sepharose ion exchange resin (GE Healthcare) charged at pH 6. The Mg3A can then be eluted in the flow-through. The enriched Mg3A can then be concentrated using a 10,000 MW cut-off membrane (Vivaspin, GE Healthcare) to a volume 5 times lower from the original volume. The other background components can be removed from Mg3A by adding 40% ammonium sulfate (w/v) in batch mode. Pure Mg3A is then recovered on the supernatant after centrifugation. Mg3A is then simultaneously dialyzed and concentrated in 50 mM MES buffer, pH 6.0, using a 10,000MW cut off membrane (Vivaspin, GE Healthcare). The activity and purity of Mg3A can be assessed by the chloro-nitro-phenyl glucoside assay and SDS-PAGE, respectively. The supernatant is then dialyzed extensively against 50 mM MES, 100 mM NaCl buffer, pH 6.0 using a 7,000 MW cut-off membrane dialysis cassette (PIERCE). The activity of the final Mg3A batch can be determined by chloro-nitro-phenyl glucoside assay. The concentration can be determined by the bicinchoninic acid assay (PIERCE) and by the absorbance assay at a wavelength of 280 nm using a molar extinction coefficient calculated by GPMAW v 7.0.

Example 2: Various assays

2-A. Protein concentration measurement by UPLC

[00224] An Agilent HPLC 1290 Infinity system was used for protein quantitation with a Waters ACQUITY UPLC BEH C4 Column (1.7 µm, 1 x 50 mm). A six minute program with an initial gradient from 5% to 33% acetonitrile (Sigma-Aldrich) in 0.5 min, followed by a gradient from 33% to 48% in 4.5 min, and then a step gradient to 90% acetronitrile was used. A protein standard curve based on the purified *Trichoderma reesei* Bgl1 was used to quantify the Mg3A polypeptides.

2-B. Chloro-nitro-phenyl-glucoside (CNPG) Hydrolysis Assay

[00225] Two hundred (200) µL of a 50 mM sodium acetate buffer, pH 5 was added to individual wells of a microtiter plate. Five (5) µL of enzyme, diluted in 50 mM sodium acetate

buffer, pH 5, was also added to individual wells. The plate was covered and allowed to equilibrate at 37°C for 15 min in an Eppendorf Thermomixer. Twenty (20) µL of 2 mM 2-Chloro-4-nitrophenyl-beta-D-Glucopyranoside (CNPG, Rose Scientific Ltd., Edmonton, CA) prepared in Millipore water was added to individual wells and the plate was quickly transferred to a spectrophotometer (SpectraMax 250, Molecular Devices). A kinetic read was performed at OD 405 nm for 15 min and the data recorded as Vmax. The extinction coefficient for CNP was used to convert Vmax from units of OD/sec to µM CNP/sec. Specific activity (µM CNP/sec/mg Protein) was determined by dividing µM CNP/sec by the mg of enzyme used in the assay. Standard error for the CNPG assay was determined to be 3%.

2-C. Cellobiose Hydrolysis Assay

[00226] Cellobiase activity was determined at 50°C using the method of Ghose, T.K. *Pure & Applied Chemistry*, 1987, *59* (2), 257-268. Cellobiose units (derived as described in Ghose) are defined as 0.0926 divided by the amount of enzyme required to release 0.1 mg glucose under the assay conditions. Standard error for the cellobiase assay was determined to be 10%.

2-D. Preparation of Phosphoric Acid Swollen Cellulose (PASC)

[00227] Phosphoric acid swollen cellulose (PASC) was prepared from Avicel using an adapted protocol of Walseth, TAPPI 1971, 35:228 and Wood, Biochem. J. 1971, 121:353-362. In short, Avicel PH-101 was solubilized in concentrated phosphoric acid then precipitated using cold deionized water. The cellulose was collected and washed with more water to neutralize the pH. It was diluted to 1% solids in 50 mM sodium acetate pH5.

Example 3. Improved Hydrolysis Performance of Mg3A over the benchmark *Trichoderma reesei* Bgl1 or over the benchmark *Aspergillus niger* B-glu, as seen in CNPG and cellobiase assays.

3-A. CNPG and cellobiase activity of beta-glucosidases produced in shake flask

[00228] The concentration of Mg3A in the crude shake flask broth was measured by UPLC (described herein) and determined to be 0.041 g/L. Two cellobiohydrolases were included in the following experiments as controls for beta-glucosidase activity in the expression strain background and were below the detection limit of the assays. Purified *Trichoderma reesei* Bgl1 was used from a stock of 2.2 mg/mL (A280 measurement). Purified *A. niger* beta-glucosidase B-glu was obtained from Megazyme International, without BSA (Megazyme International Ireland Ltd., Wicklow, Ireland, Lot No. 031809).

[00229] The activity of each enzyme on the model substrates chloro-nitro-phenyl-glucoside (CNPG) and cellobiose were measured. The assays were each carried out at the temperature in the standard protocol; CNPG at 37°C and cellobiose at 50°C.

Table 3-1.

		Ratio to *T. reesei* Bgl1	
Enzyme	Purified	CNPG	Cellobiose
T. reesei Bgl1	Y	1	1
A.niger B-glu	Y	0.13	12
Mg3A	N	1.19	1.87

[00230] Mg3A had 20% higher CNPG activity than that of *Trichoderma reesei* Bgl1, and over 80% higher cellobiose hydrolysis activity (or cellobiase activity). Mg3A had about 9-fold higher activity on CNPG than that of *Aspergillus niger* B-glu, but about 6-fold less activity on cellobiose than that of *Aspergillus niger* B-glu. The cellobiohydrolases had no activity on cellobiose (no glucose was observed for any wells).

Table 3-2.

Enzyme	Purified	CNPG/Cellobiose
T. reesei Bgl1	Y	62
A.niger B-glu	Y	1.6
Mg3A	N	39

[00231] To compare the activity of each molecule independent of protein determination, the ratio of CNPG to cellobiase activity was calculated. The ratio of hydrolysis activities on CNPG/cellobiose for Mg3A is somewhat less than a half of the ratio of hydrolysis activities on CNPG/cellobiose for *Trichoderma reesei* Bgl1. The ratio of hydrolysis activities on CNPG/cellobiose for Mg3A is over 20-fold higher than the ratio of hydrolysis activities on CNPG/cellobiose for *Aspergillus niger* B-glu.

Example 4: Improved Hydrolysis Performance of Mg3A polypeptides on PASC substrates.

4-A. Dose curves depicting the measurements and comparison of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 hydrolysis of PASC, in a background whole cellulase composition produced by a strain described in Published International Patent Application No. WO 2011/038019.

[00232] The beta-glucosidases were added from 0-10 mg protein/g cellulose to a constant loading of 10 mg protein/g glucan whole cellulase background produced by a strain described in Published International Patent Application No. WO 2011/038019, which expresses Fv43D, Fv3A, Fv51A, AfuXyn2, EG4, and etc, at a total protein concentration of 88.8 g/L). The mixtures were used to hydrolyze phosphoric acid swollen cellulose (PASC). Each sample dose was assayed in quadruplicate.

[00233] All enzyme dilutions were made into 50 mM sodium acetate buffer, pH 5.0. One hundred and fifty (150) µL of cold 0.6% PASC was added to 30 µL of enzyme solution in microtiter plates (NUNC flat bottom PS, cat. no. 269787). The enzyme mixture therefore contained 10 mg protein/g glucan of the whole cellulase plus 0-10 mg of Mg3A or Bgl1/g glucan. The plates were covered with aluminum plate seals and incubated for 1.5 h at 50°C, 200 rpm in an Innova incubator/shaker. The reaction was quenched with 100 µL of 100 mM Glycine, pH 10, filtered (Millipore vacuum filter plate cat. no. MAHVN45) and the soluble sugars were measured on an Agilent 5042-1385 HPLC with an Aminex HPX-87P column.

[00234] Percent glucan conversion was determined as (mg glucose + mg cellobiose + mg cellotriose) / mg cellulose in the reaction.

[00235] The results are indicated in **Figures 3A-3C**. The horizontal lines at about 46% and 62.5% conversion represent 10 mg/g and 20 mg/g loading of just the whole cellulase background activities.

[00236] Mg3A produced more glucose than the same dose of *T. reesei* Bgl1. This is consistent with Mg3A having higher cellobiase activity than *T. reesei* Bgl1. When total sugars: adding glucose, cellobiose and cellotriose concentrations together, was measured to determine total % glucan conversion, Mg3A outperformed *T. reesei* Bgl1.

4-B. A second set of dose curves depicting the measurements and comparison of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 hydrolysis of PASC, in a background whole cellulase composition produced by a strain described in International Published Patent Application No. WO 2011/038019, under different conditions as 4-A.

[00237] Shake flask produced Mg3A culture broth was concentrated >20 fold using 10,000 molecular weight cut off PES spin concentrators. Protein concentration was determined by UPLC, compared to a *Trichoderma reesei* Bgl1 standard curve. PASC hydrolysis experiments were repeated with higher protein loadings than in 4-A, above. The specific cellobiase activity of the concentrated samples was equal to the original shake flask supernatant. The PASC assays

were conducted as described above (50°C for 1.5 h). The reaction was then quenched with 100 µL of 100 mM glycine buffer, at pH 10. Each dose was assayed in quadruplicate.

[00238] The concentration of Mg3A was 1.55 g/L, as measured using UPLC. The concentration of the benchmark *Trichoderma reesei* Bgl1 was 2.2 mg/L.

[00239] Results are shown in **Figures 3D-3F**

[00240] The results confirmed that Mg3A produced more glucose from hydrolyzing the same PASC than the same dose of *Trichoderma reesei* Bgl1 did. Mg3A has a higher cellobiase activity than Bgl1, which yielded from the same PASC substrate under the same hydrolysis conditions, a lower cellobiose concentration. When the total sugars was measured: adding the total concentration of glucose and cellobiose, in order to measure total % glucan conversion, Mg3A substantially outperformed *T. reesei* Bgl1.

Example 5: Improved Hydrolysis Performance of Mg3A polypeptides on Dilute Ammonia Pretreated Corn Stover substrates.

5-A. Dose curves depicting the measurements and comparison of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 hydrolysis of DACS, in a background whole cellulase composition produced by a strain described in Published International Patent Application No. WO 2011/038019.

[00241] Shake flask produced Mg3A culture broth was concentrated >20 fold using 10,000 molecular weight cut off PES spin concentrators. Protein concentration was determined by UPLC, compared to a *Trichoderma reesei* Bgl1 standard curve. The concentrated shake flask supernatant was used in saccharification assays using dilute ammonia pretreated corn stover (DACS) as substrates. Each enzyme mixture sample was blended with 10% Mg3A or Bgl1, with a whole cellulase produced by an engineered *Trichoderma reesei* strain as described in International Published Patent Application No. WO 2011/038019. Dose response curves for hydrolyzing the DACS substrate was generated by adding 3-53 mg total protein/g glucan enzyme mixture to the substrate.

[00242] *Trichoderma reesei* Bgl1 was added to the hydrolysis assay from a purified stock of 2.2 g/L total protein. Mg3A was added from the 1.55 mg/mL concentrated sample.

[00243] Dilute ammonia pre-treated corn stover (DACS) was slurried in 20 mM Sodium Acetate, pH 5 for a final 7% glucan (21.5% solids) content. If needed, the slurry was adjusted to pH 5 and the slurry was transferred into 96-well microtiter plates.

[00244] All enzymes were loaded based on mg protein/g glucan in the substrate. All enzyme dilutions were made into 50 mM Sodium Acetate buffer, pH 5.0. Thirty (30) µL of enzyme solution and 45 µg DACS substrate were added per well in 96-well microtiter plates. Each sample dose was tested in quadruplicate. The plates were covered with an aluminum seal and incubated for 2 days at 50°C, 200 rpm in the Innova incubator/shaker. The reaction was quenched with 100 µL of 100 mM Glycine, pH 10, filtered, and the soluble sugars were measured by HPLC.

[00245] Percent glucan conversion is defined as (mg glucose + mg cellobiose + mg cellotriose) / mg cellulose in the DACS substrate.

[00246] The dose response curves are in **Figures 4A-4B.**

[00247] Mixing Mg3A and the whole cellulase background composition provided a higher level of glucan conversion than when the benchmark *Trichoderma reesei* Bgl1 is mixed with the same whole cellulase background composition, or than when the glucan conversion is level measured with the cellulase background composition alone. The improvement was particularly prominent when the beta-glucosidases were dosed at below 13 mg/g.

5-B: Dose curves depicting the measurements and comparison of Mg3A vs. the benchmark *Trichoderma reesei* Bgl1 hydrolysis of DACS, wherein the Mg3A and Bgl1 were added at increasing doses to a whole cellulase composition at 13.4 mg/g produced by a strain described in International Published Patent Application No. WO 2011/038019.

[00248] In this experiment, the beta-glucosidases were added in increasing dose to a constant loading of 13.4 mg protein/g of glucan of a whole cellulase produced by an engineered *Trichoderma reesei* strain in accordance with International Published Patent Application No. WO 2011/038019. The mixtures were used to hydrolyze DACS (4% glucan) for 2 days at 50°C. To prepare the mixture, *Trichoderma reesei* Bgl1 was added to the mixture from a purified stock of 2.2 g/L total protein, and the whole cellulase was added to the mixture from an 88.8 g/L total protein stock. Mg3A was added from the 1.55 mg/mL concentrate.

[00249] Dilute ammonia pretreated corn stover in microtiter plates was prepared as described above. All enzymes were loaded based on mg protein/g glucan in the substrate. All enzyme dilutions were made into 50 mM Sodium Acetate buffer, pH 5.0. Thirty (30) µL of enzyme solution was added to 45 µg substrate per well in microtiter plates. The plates were covered with foil seals and incubated for 2 days at 50°C, 200 rpm in an Innova incubator/shaker. The reaction was quenched with 100 µL of 100 mM Glycine, pH 10, filtered and the soluble sugars

measured by HPLC (Agilent 100 series equipped with a de-ashing column (Biorad 125-0118) and carbohydrate column (Aminex HPX-87P). The mobile phase was water with a flow rate of 0.6 mL/min and 20 min run time. A glucose standard curve was generated and used for quantitation.

[00250] Percent glucan conversion is defined as (mg glucose + mg cellobiose + mg cellotriose) / mg cellulose in the substrate.

[00251] Results are shown in **Figures 5A-5B.**

[00252] Mg3A out performed *Trichoderma reesei* Bgl1 at all doses.

Example 6: Improved Thermostability and High-temperature Hydrolysis Performance of Mg3A polypeptides on Dilute Ammonia Pretreated Corn Stover substrates.

[00253] In this experiment, the beta-glucosidases were added in increasing dose to a constant loading of 10 mg protein/g of glucan of a whole cellulase produced by an engineered *Trichoderma reesei* strain in accordance with International Published Patent Application No. WO 2011/038019. The mixtures were used to hydrolyze DACS (4% glucan) for 2 days at 55°C. To prepare the mixture, *Trichoderma reesei* Bgl1 was added to the mixture from a purified stock of 2.2 g/L total protein, and the whole cellulase was added to the mixture from an 88.8 g/L total protein stock. Mg3A was added from a 1.55 mg/mL concentrate prepared in accordance with the descriptions above.

[00254] Dilute ammonia pretreated corn stover in microtiter plates was prepared as described above. All enzymes were loaded based on mg protein/g glucan in the substrate. All enzyme dilutions were made into 50 mM Sodium Acetate buffer, pH 5.3. Thirty (30) µL of enzyme solution was added to 75 µg substrate per well in microtiter plates. The plates were covered with foil seals and incubated for 2 days at 55°C, 200 rpm in an Innova incubator/shaker. The reaction was quenched with 100 µL of 100 mM Glycine, pH 10, filtered and the soluble sugars measured by HPLC (Agilent 100 series equipped with a de-ashing column (Biorad 125-0118) and carbohydrate column (Aminex HPX-87P). The mobile phase was water with a flow rate of 0.6 mL/min and 20 min run time. A glucose standard curve was generated and used for quantitation.

[00255] Percent glucan conversion is defined as (mg glucose + mg cellobiose + mg cellotriose) / mg cellulose in the substrate.

[00256] Results are shown in **Figures 6A-5C**.

[00257] Mg3A out performed *Trichoderma reesei* Bgl1 at all doses for glucan conversion, glucose production. Both Mg3A and Bgl1 effectively deplete the cellobiose in the saccharification mixtures.

CLAIMS

We claim:

1. A recombinant polypeptide comprising an amino acid sequence that is at least 75% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3, wherein the polypeptide has beta-glucosidase activity.

2. The recombinant polypeptide of claim 1, wherein the polypeptide has improved beta-glucosidase activity as compared to *Trichoderma reesei* Bgl1 when the recombinant polypeptide and the *Trichoderma reesei* Bgl1 are used to hydrolyze lignocellulosic biomass substrates.

3. The recombinant polypeptide of claim 1 or 2, wherein the improved beta-glucosidase activity is an increased cellobiase activity.

4. The recombinant polypeptide of any one of claims 1-3, wherein the improved beta-glucosidase activity is an increased yield of glucose from a lignocellulosic biomass under the same saccharification conditions.

5. The recombinant polypeptide of claim 4, wherein the lignocellulosic biomass is subject to a pretreatment prior to saccharification.

6. The recombinant polypeptide of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence that is at least 80% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3.

7. The recombinant polypeptide of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence that is at least 90% identical to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:3.

8. A composition comprising the recombinant polypeptide of any one of claims 1-7, further comprising one or more other cellulases.

9. The composition of claim 8 wherein the one or more other cellulases are selected from no or one or more other beta-glucosidases, one or more cellobiohydrolases, and one or more endoglucanases.

10. A composition comprising the recombinant polypeptides of any one of claims 1-9, further comprising one or more hemicellulases.

11. The composition of claim 8 or 9, further comprising one or more hemicellulases.

12. The composition of claim 10 or 11, wherein the one or more hemicellulases are selected from one or more xylanases, one or more beta-xylosidases, and one or more L-arabinofuranosidases.

13. A nucleic acid encoding the recombinant polypeptide of any one of claims 1-7.

14. The nucleic acid of claim 13, wherein the polypeptide further comprises a signal peptide sequence.

15. The nucleic acid of claim 14, wherein the signal peptide sequence is selected from the group consisting of SEQ ID NOs:13-42.

16. An expression vector comprising the nucleic acid of any one of claims 13-15 in operable combination with a regulatory sequence.

17. A host cell comprising the expression vector of claim 16.

18. The host cell of claim 17, wherein the host cell is a bacterial cell or a fungal cell.

19. A composition comprising the host cell of claim 17 or 18 and a culture medium.

20. A method of producing a beta-glucosidase, comprising: culturing the host cell of claim 17 or 18 in a culture medium, under suitable conditions to produce the beta-glucosidase.

21. A composition comprising the beta-glucosidase produced in accordance with the method of claim 20 in supernatant of the culture medium.

22. A method for hydrolyzing a lignocellulosic biomass substrate, comprising: contacting the lignocellulosic biomass substrate with the polypeptide of any one of claims 1-7, or the composition of claim 21, to yield a glucose or other sugars.

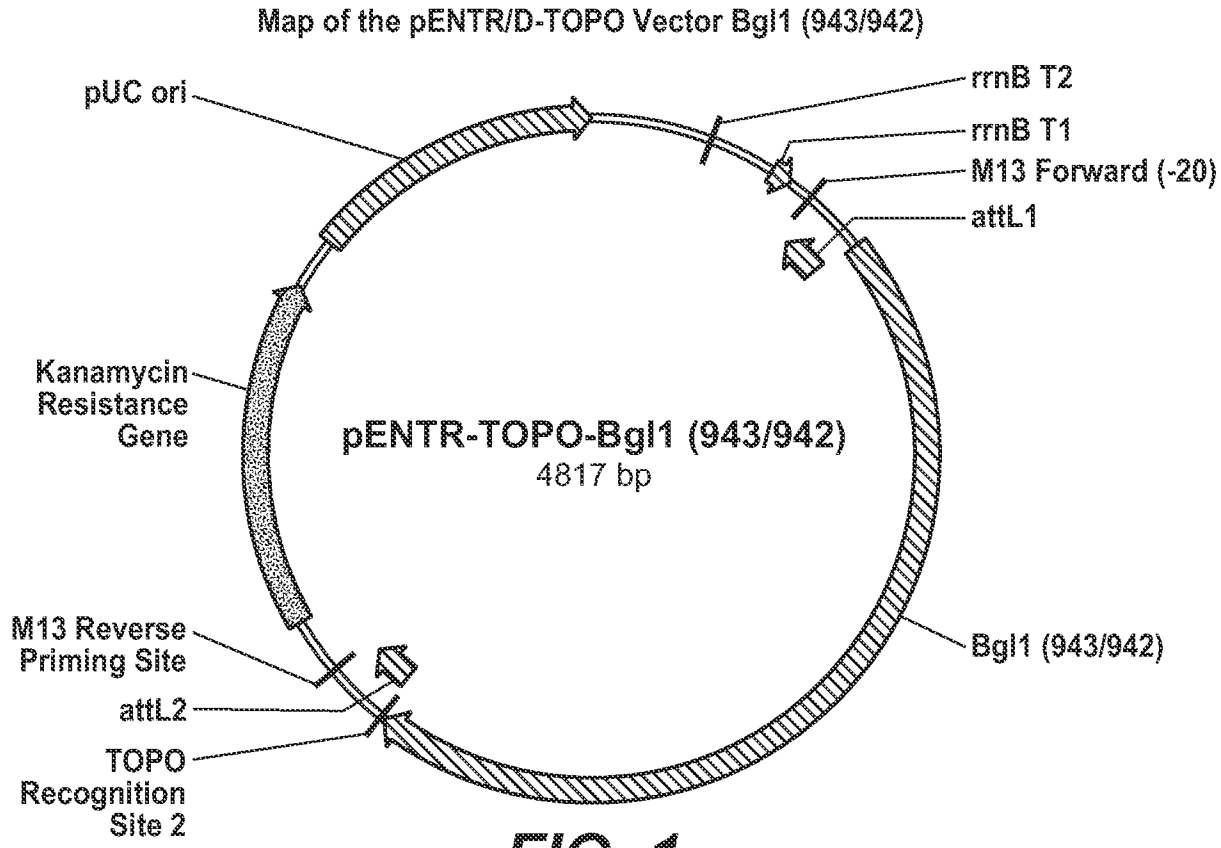
Map of the pENTR/D-TOPO Vector Bgl1 (943/942)
pUC ori
rrnB T2
rrnB T1
M13 Forward (-20)
attL1
Kanamycin Resistance Gene
pENTR-TOPO-Bgl1 (943/942)
4817 bp
M13 Reverse Priming Site
attL2
TOPO Recognition Site 2
Bgl1 (943/942)

FIG. 1

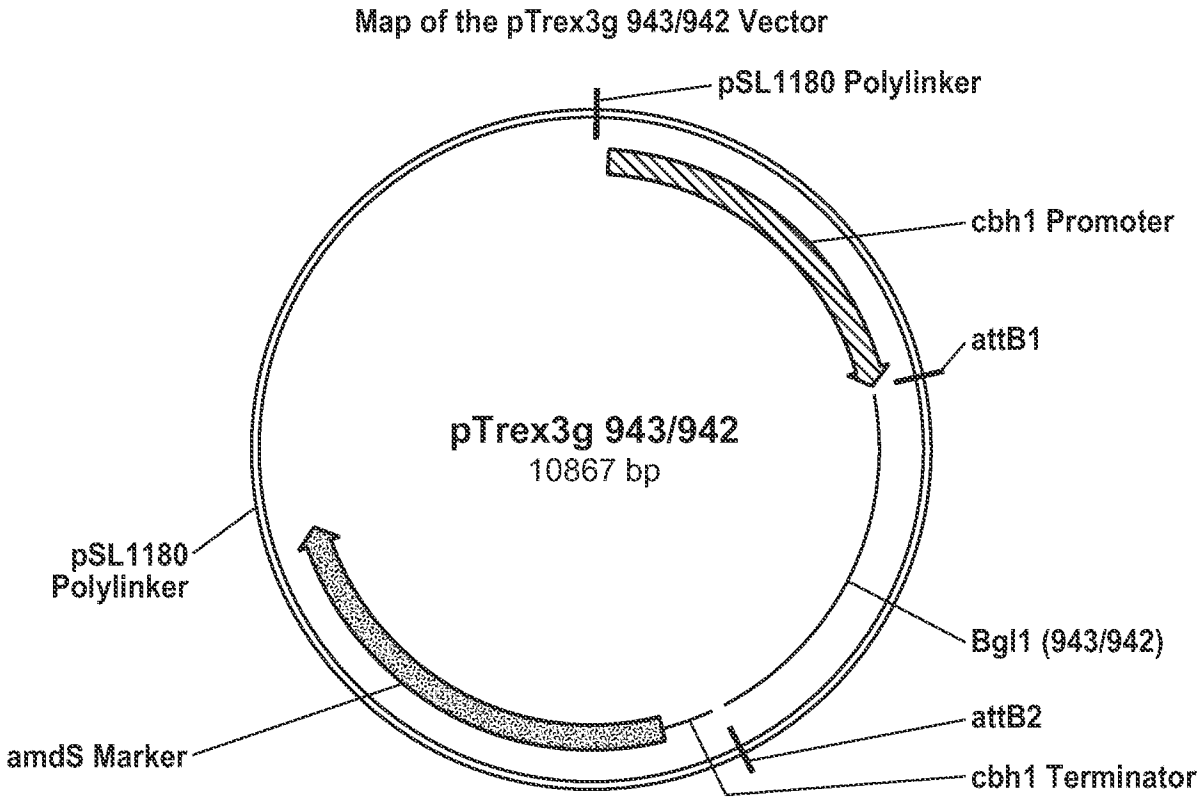
Map of the pTrex3g 943/942 Vector
pSL1180 Polylinker
cbh1 Promoter
attB1
pTrex3g 943/942
10867 bp
pSL1180 Polylinker
Bgl1 (943/942)
attB2
amdS Marker
cbh1 Terminator

FIG. 2

FIG. 3: Dose curves measuring and comparing hydrolysis of PASC (phosphoric acid swollen cellulose)

3A: Dose curves depicting the measurements and a comparison of total % glucan conversion from a given PASC substrate, by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising same concentrations of *T. reesei* Bgl1 vs. Mg3A. Measurements were conducted in accordance with the conditions of **Example 4-A.**

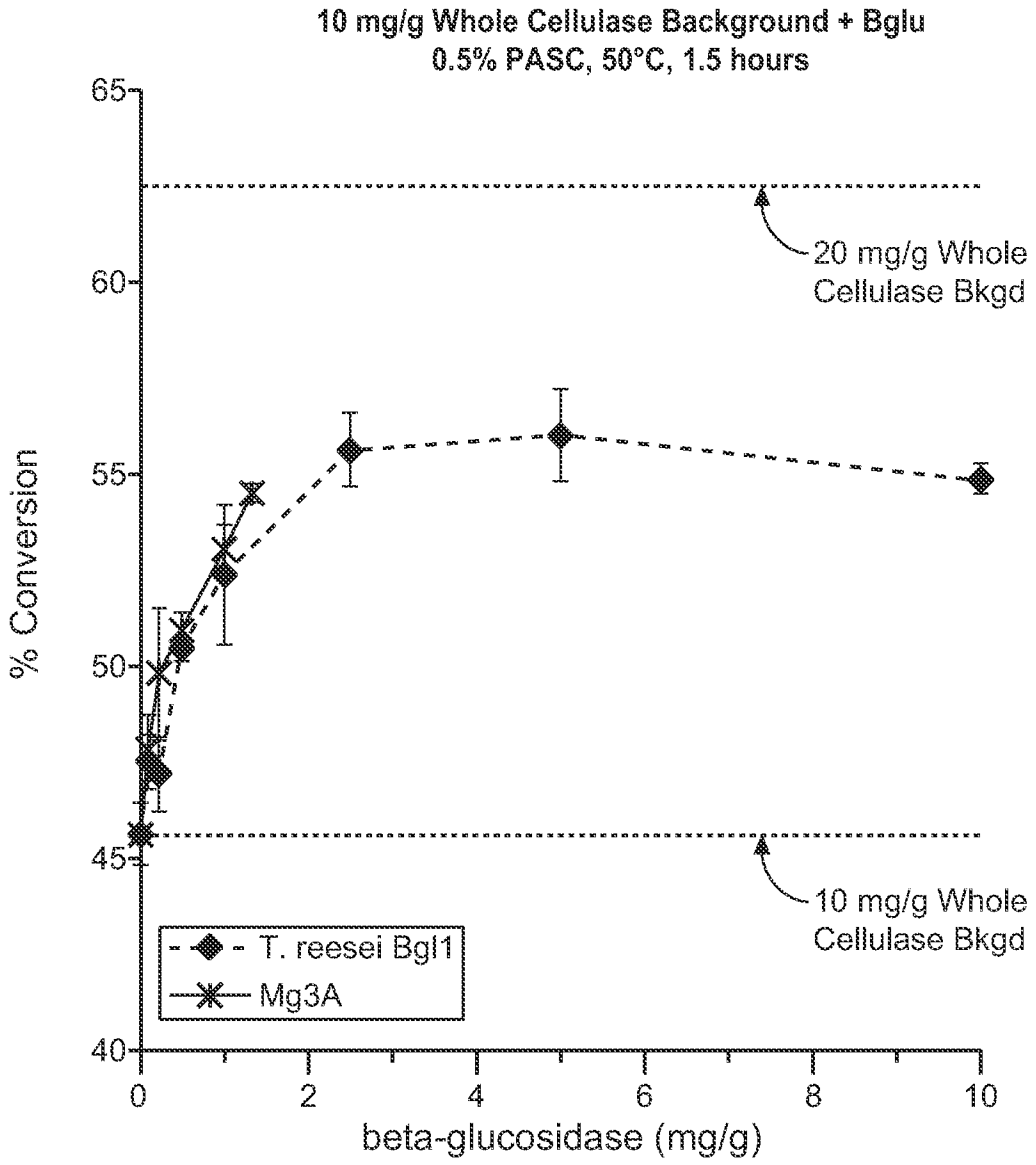

FIG. 3A

3B: Dose curves depicting the measurements and a comparison of glucose yields from hydrolysis of a given PASC substrate by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising the same concentrations of *T. reesei* Bgl1 vs. Mg3A. Measurements were conducted in accordance with the conditions of **Example 4-A.**

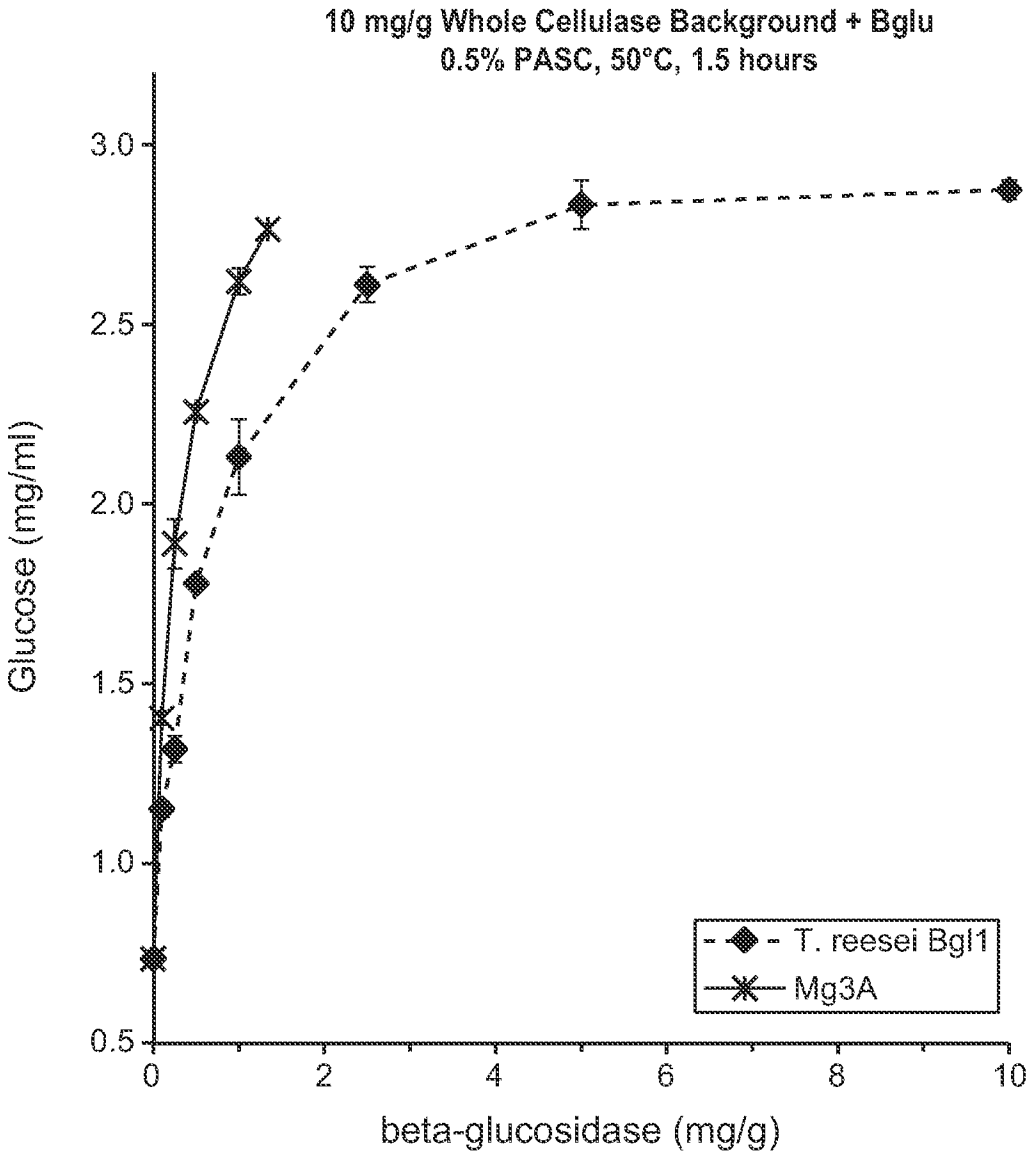

FIG. 3B

3C: Dose curves depicting the measurements and a comparison of the cellobiose yields from hydrolysis of a given PASC substrate by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising the same concentrations of *T. reesei* Bgl1 vs. Mg3A. Measurements were conducted in accordance with the conditions of **Example 4-A.**

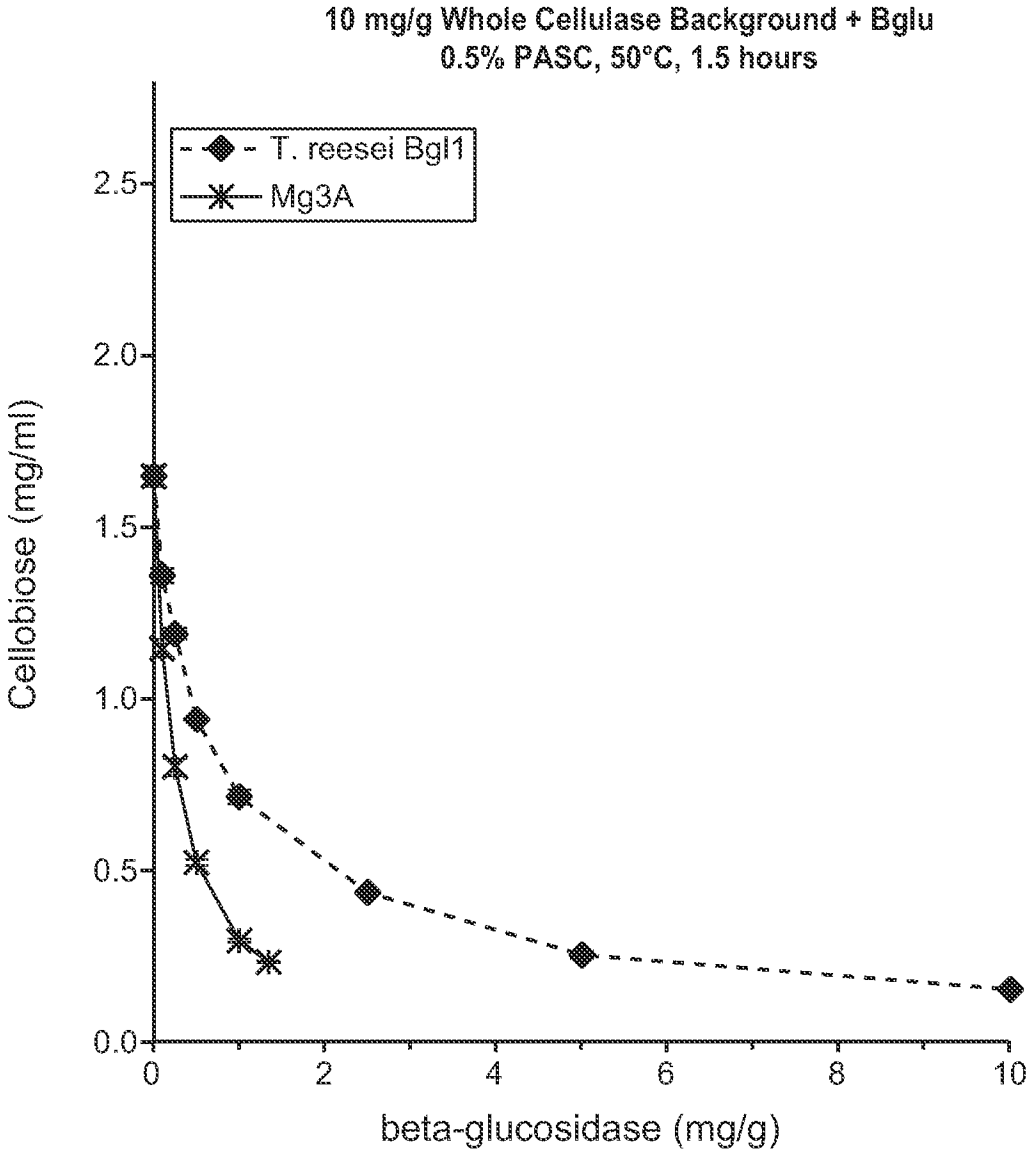

FIG. 3C

3D: Dose curves depicting the measurements and a comparison of total % glucan conversion from a given PASC substrate, by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising same concentrations of *T. reesei* Bgl1 vs. Mg3A. Measurements were conducted in accordance with the conditions of **Example 4-B.**

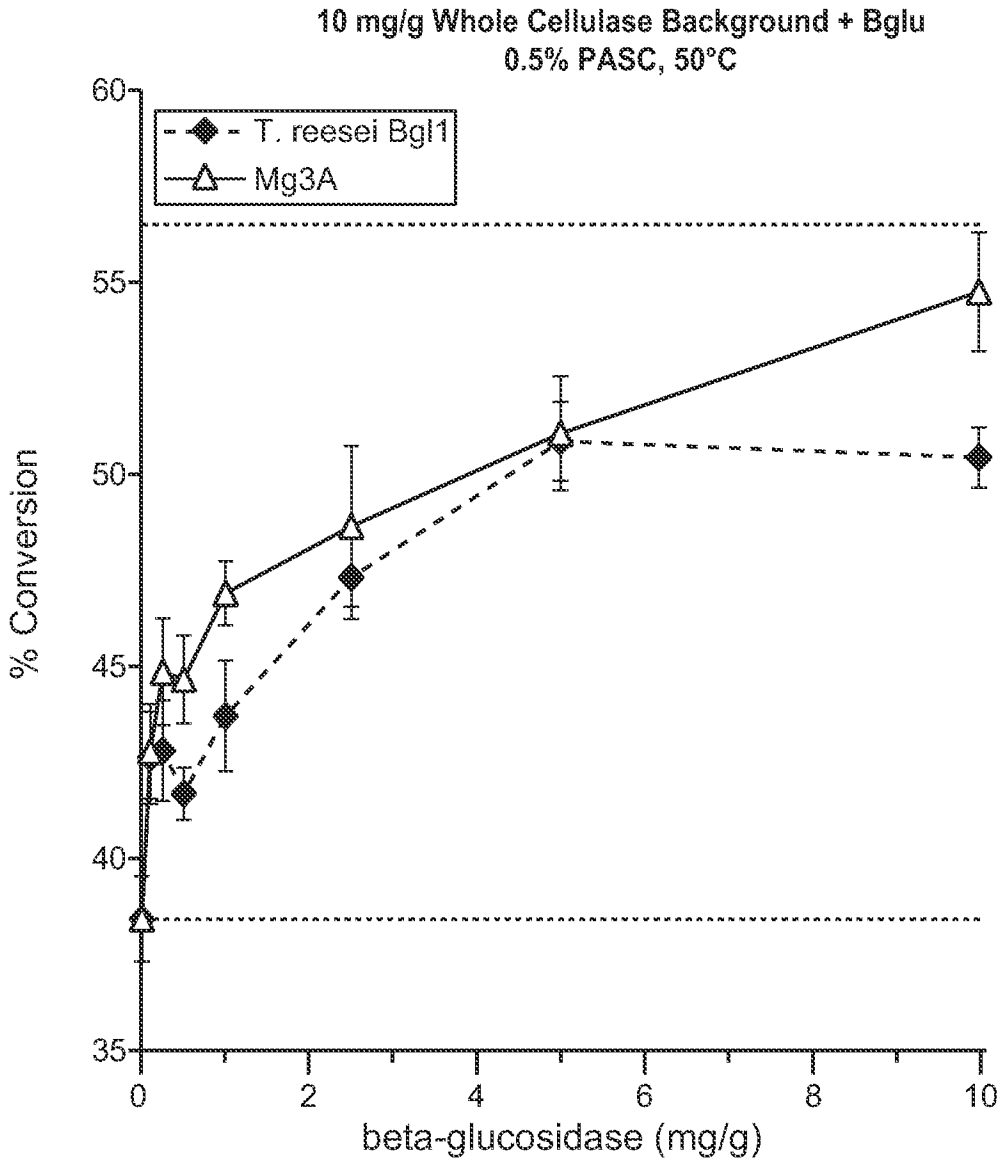

FIG. 3D

3E: Dose curves depicting the measurements and a comparison of glucose yields from hydrolysis of a given PASC substrate by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising the same concentrations of *T. reesei* Bgl1 vs. Mg3A. Measurements were conducted in accordance with the conditions of **Example 4-B.**

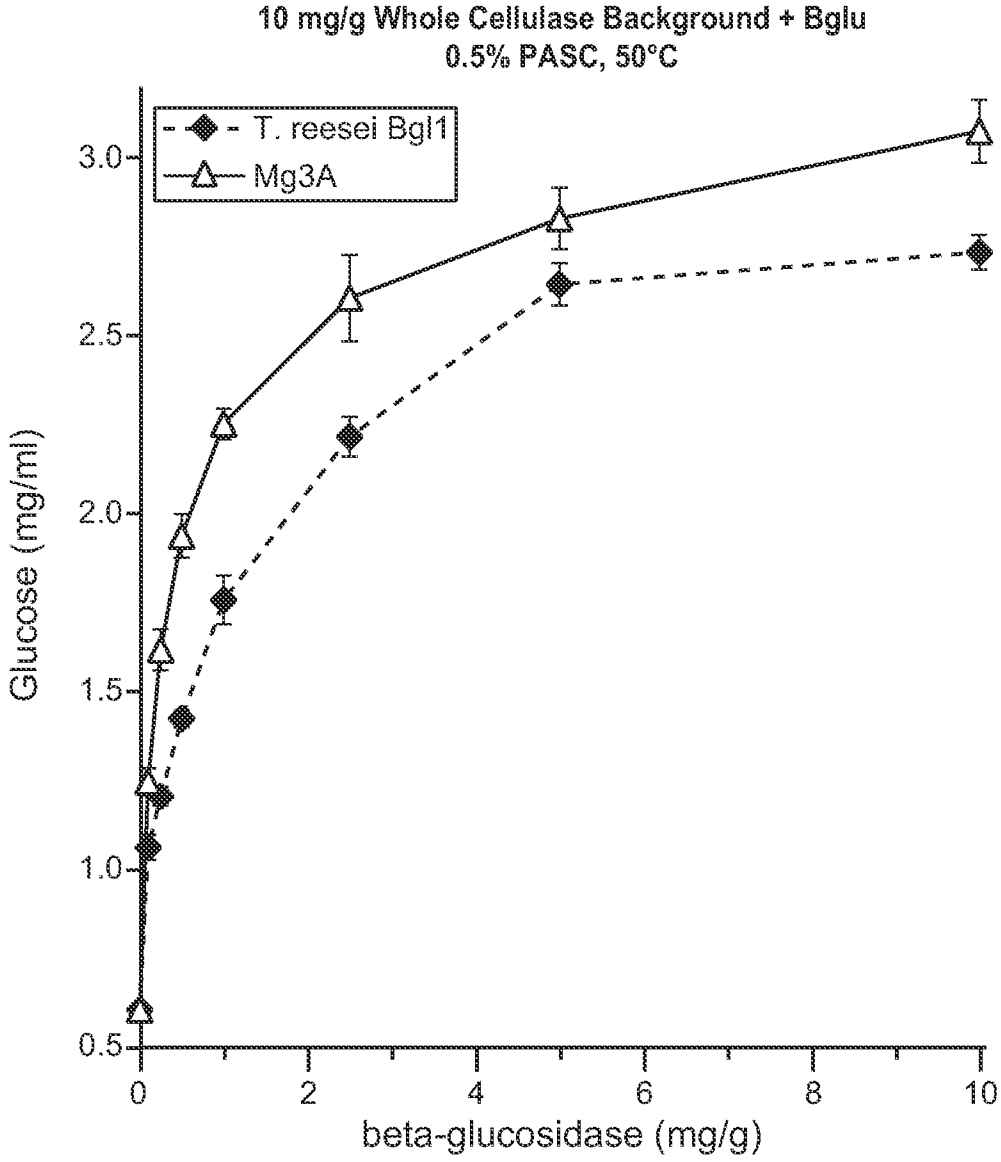

FIG. 3E

3F: Dose curves depicting the measurements and a comparison of the cellobiose yields from hydrolysis of a given PASC substrate by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising the same concentrations of *T. reesei* Bgl1 vs. Mg3A. Measurements were conducted in accordance with the conditions of **Example 4-B.**

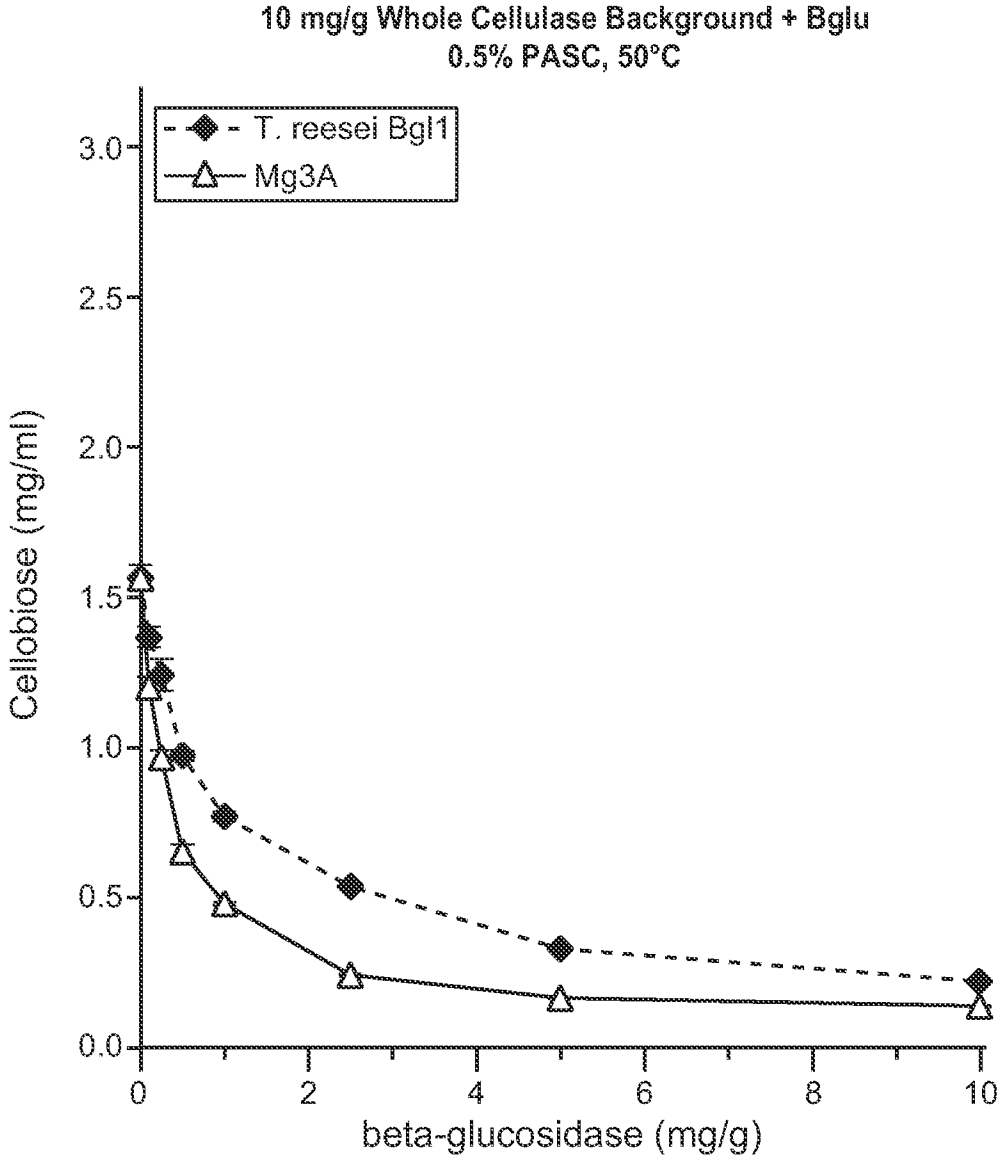

FIG. 3F

FIG. 4: Dose curves measuring and comparing hydrolysis of dilute ammonia pretreated corn stover (DACS)

4A: Dose curves depicting the measurements and a comparison of total glucan conversion from a given DACS substrate, by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising same concentrations of *T. reesei* Bgl1 vs. Mg3A.

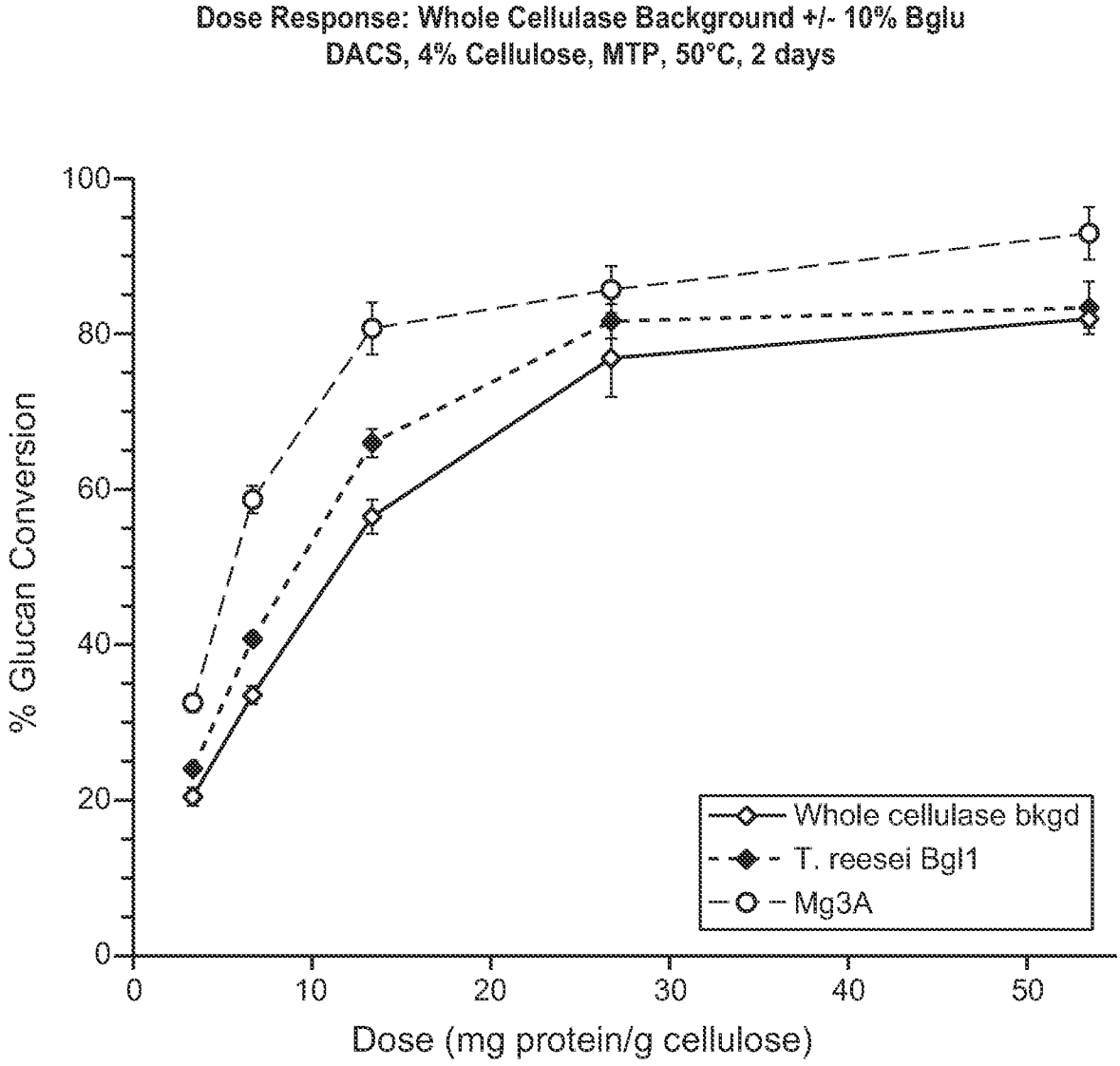

FIG. 4A

4B: Dose curves depicting the measurements and a comparison of glucose yields from hydrolysis of a given DACS substrate by the whole cellulase composition (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019) comprising the same concentrations of *T. reesei* Bgl1 vs. Mg3A.

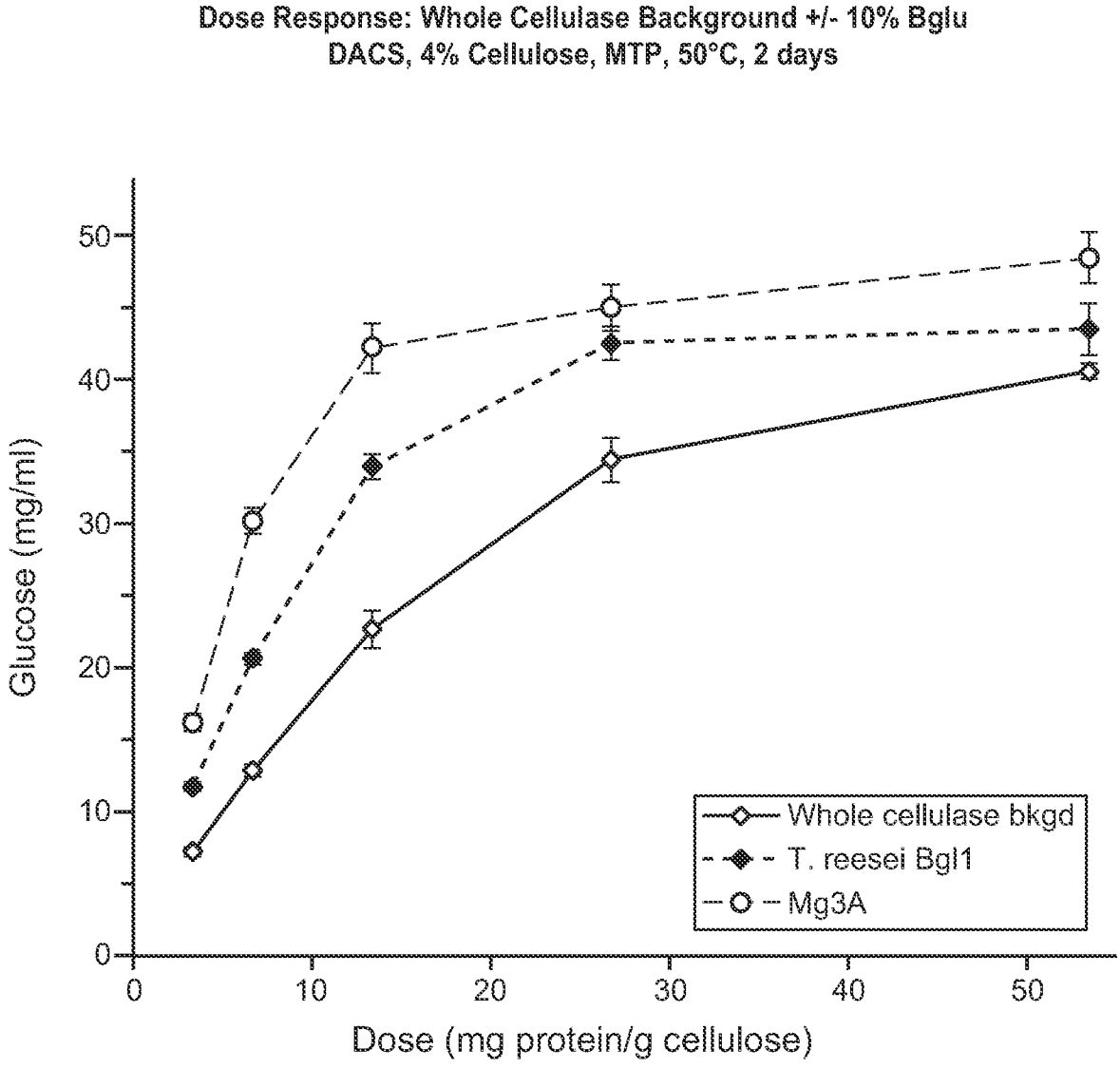

FIG. 4B

FIG. 5: Dose curves depicting the measurements and comparison of hydrolysis performance on DACS substrates, adding different doses of Mg3A and *T. reesei* Bgl1 to a whole cellulase at 13.4 mg/g total protein.

5A: Dose curve depicting the measurements and a comparison of total glucan conversion from a given DACS substrate, by a mixture of Mg3A vs. *T. reesei* Bgl1, added at increasing doses to a 13.4 mg/g whole cellulase (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019).

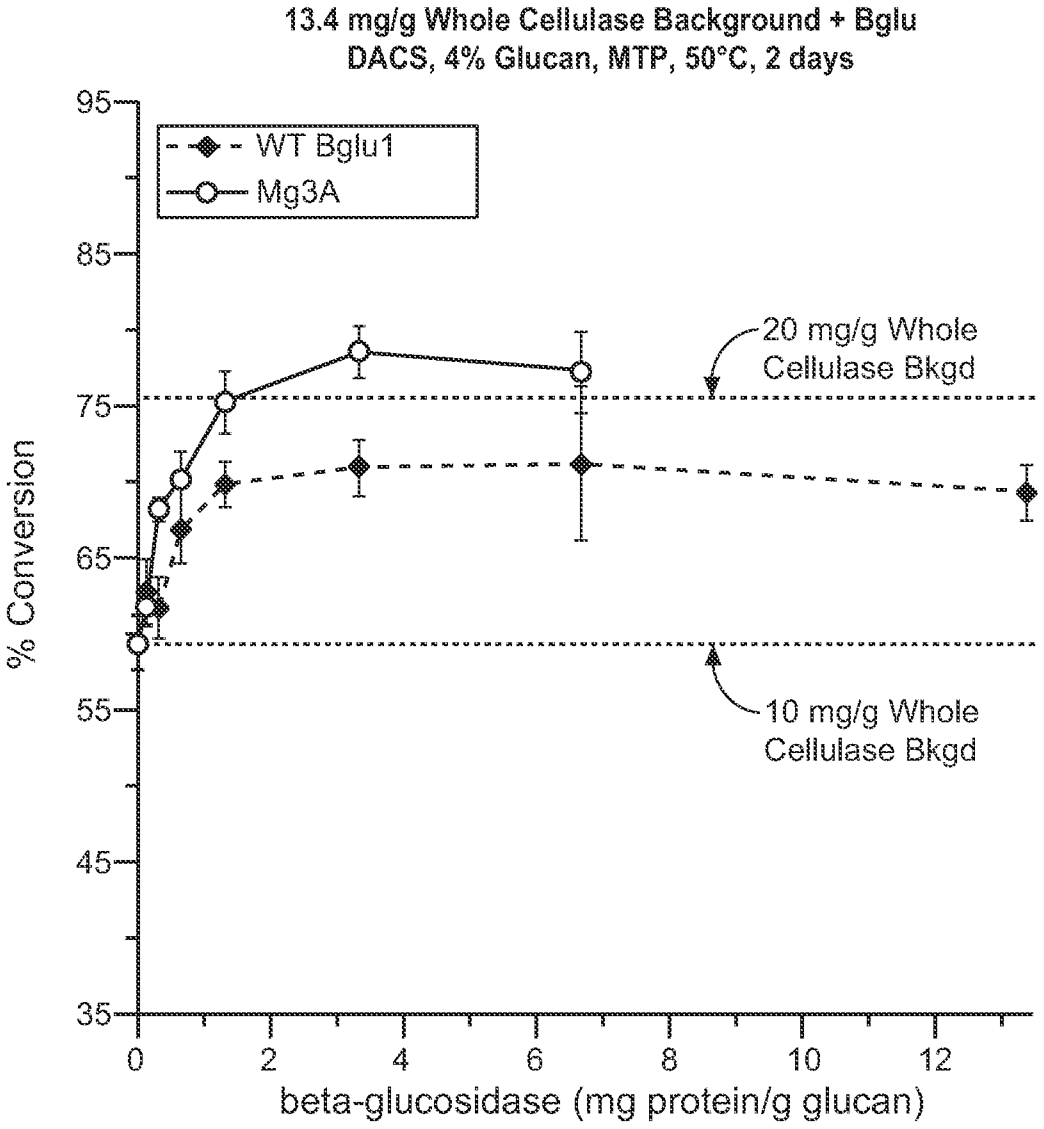

FIG. 5A

5B: Dose curve depicting the measurements and a comparison of glucose yields from a given DACS substrate, by a mixture of Mg3A vs. *T. reesei* Bgl1, added at increasing doses to a 13.4 mg/g whole cellulase (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019).

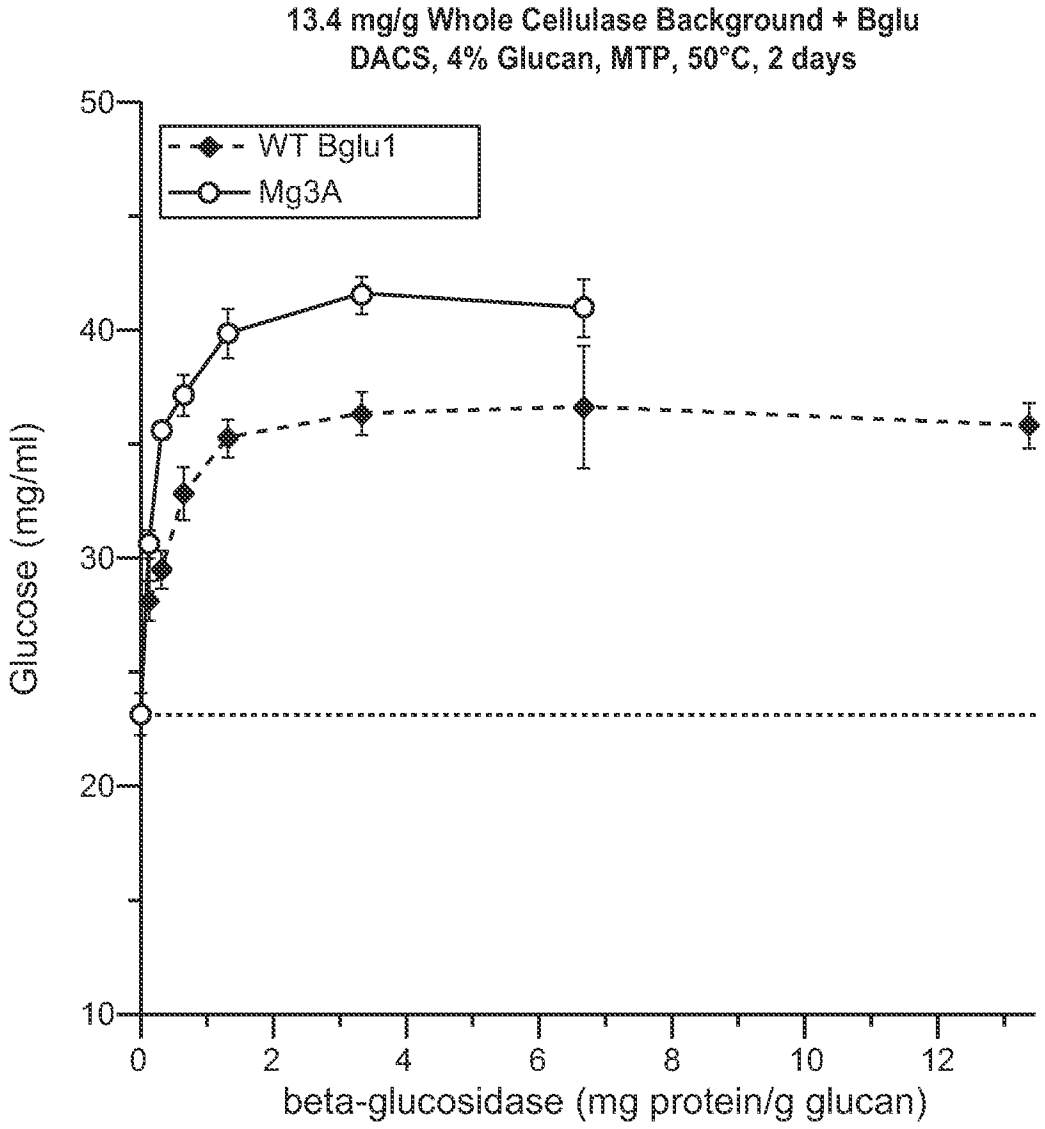

FIG. 5B

FIG 6: Dose curves depicting the measurements and comparison of hydrolysis performance on DACS substrates at 55°C, adding different doses of Mg3A and *T. reesei* Bgl1 to a whole cellulase at 10.0 mg/g total protein.

6A: Dose curve depicting the measurements and a comparison of total glucan conversion from a given DACS substrate, by a mixture of Mg3A vs. *T. reesei* Bgl1, added at increasing doses to a 10 mg/g whole cellulase (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019).

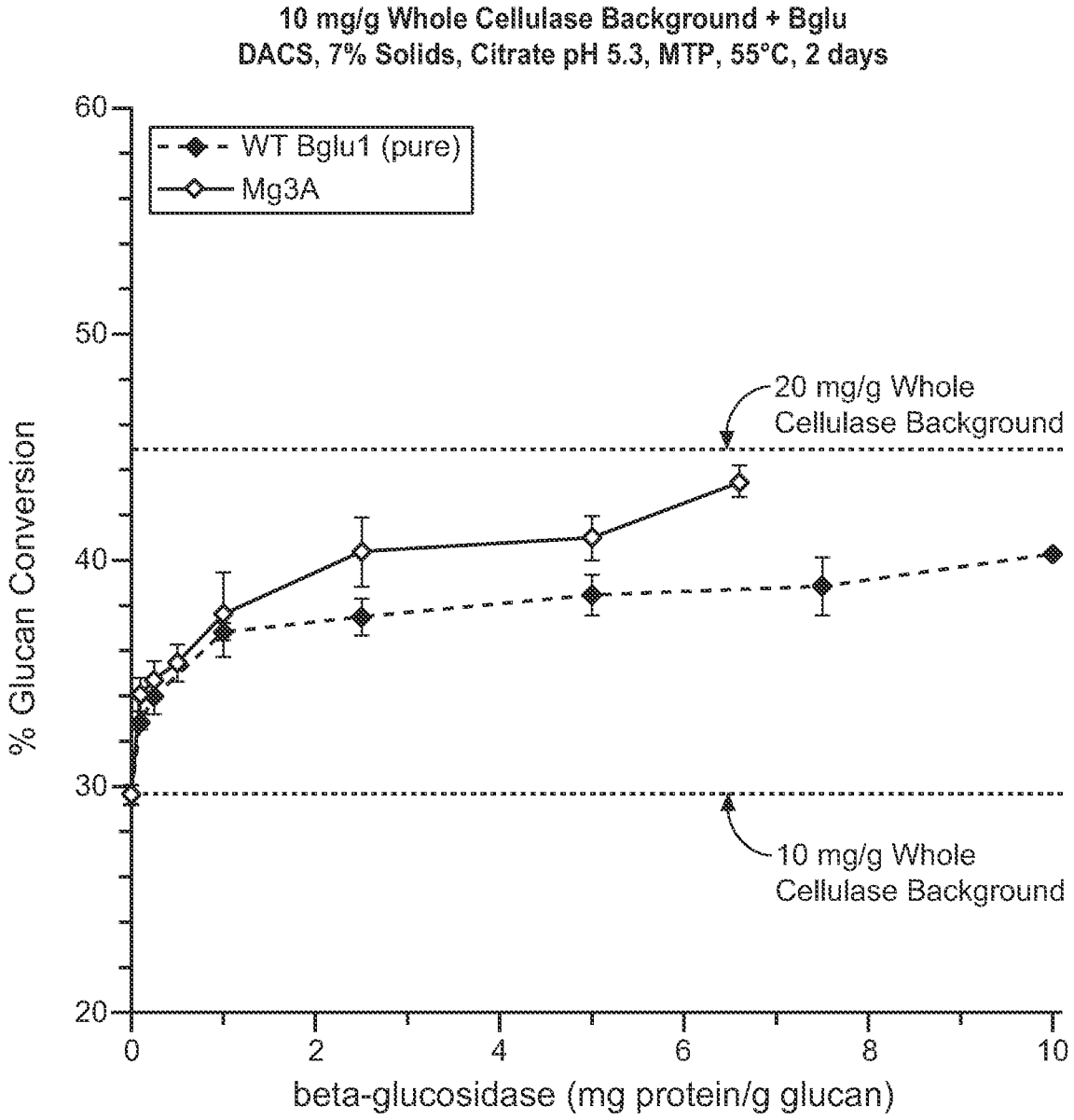

FIG. 6A

6B: Dose curve depicting the measurements and a comparison of glucose yields from a given DACS substrate, by a mixture of Mg3A vs. *T. reesei* Bgl1, added at increasing doses to a 10mg/g whole cellulase (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019).

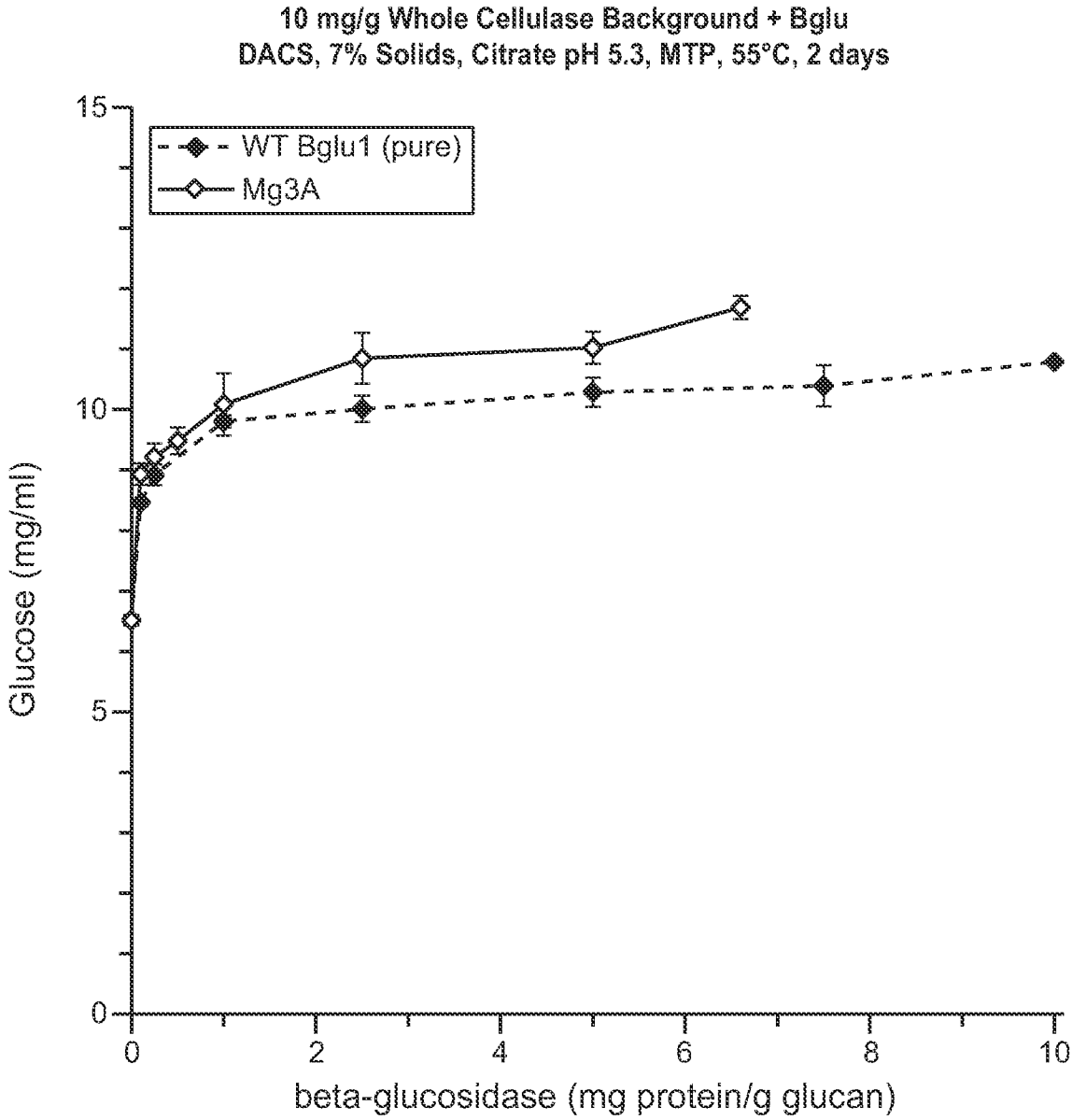

FIG. 6B

6C: Dose curve depicting the measurements and a comparison of cellobiose depletion from a given saccharification reaction on DACS substrate, by a mixture of Mg3A vs. *T. reesei* Bgl1, added at increasing doses to a 10 mg/g whole cellulase (produced from an engineered *Trichoderma reesei* strain in accordance with what is described in International Published Patent Application WO 2011/038019).

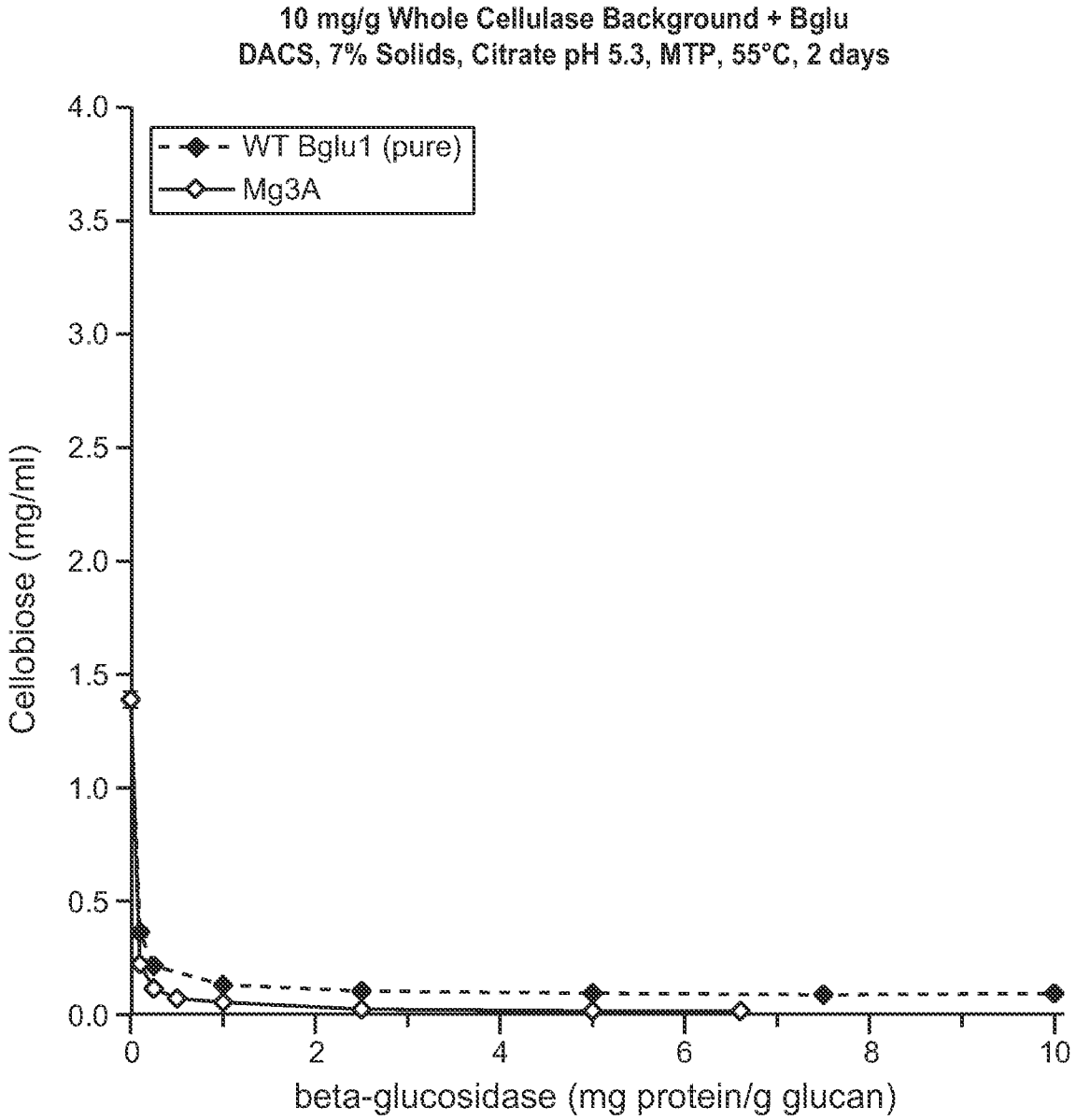

FIG. 6C

A Yeast Shuttle Vector pSC11 Construct for Expressing Mg3A in a *Saccharomyces cerevisiae* Ethanologen.

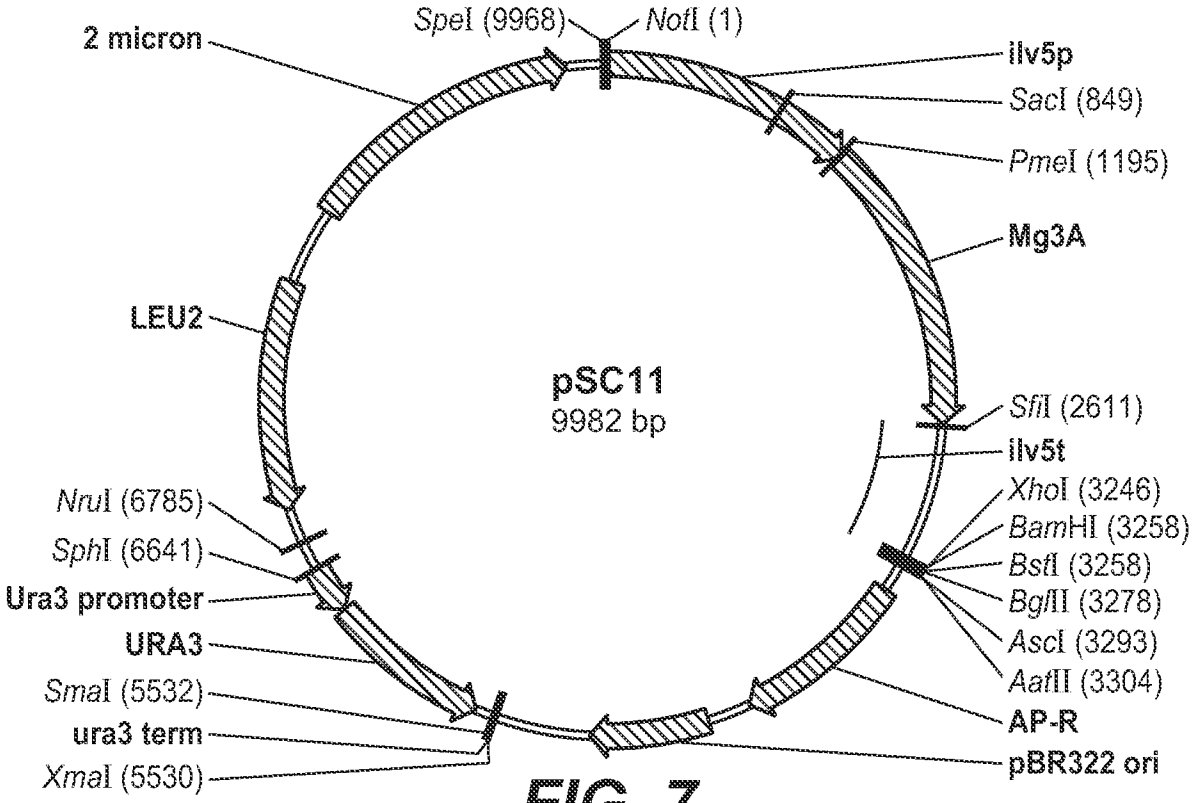

FIG. 7

A *Zymomonas mobilis* Integration Vector pZC11 for Expressing Mg3A in a *Zymomonas mobilis* Ethanologen.

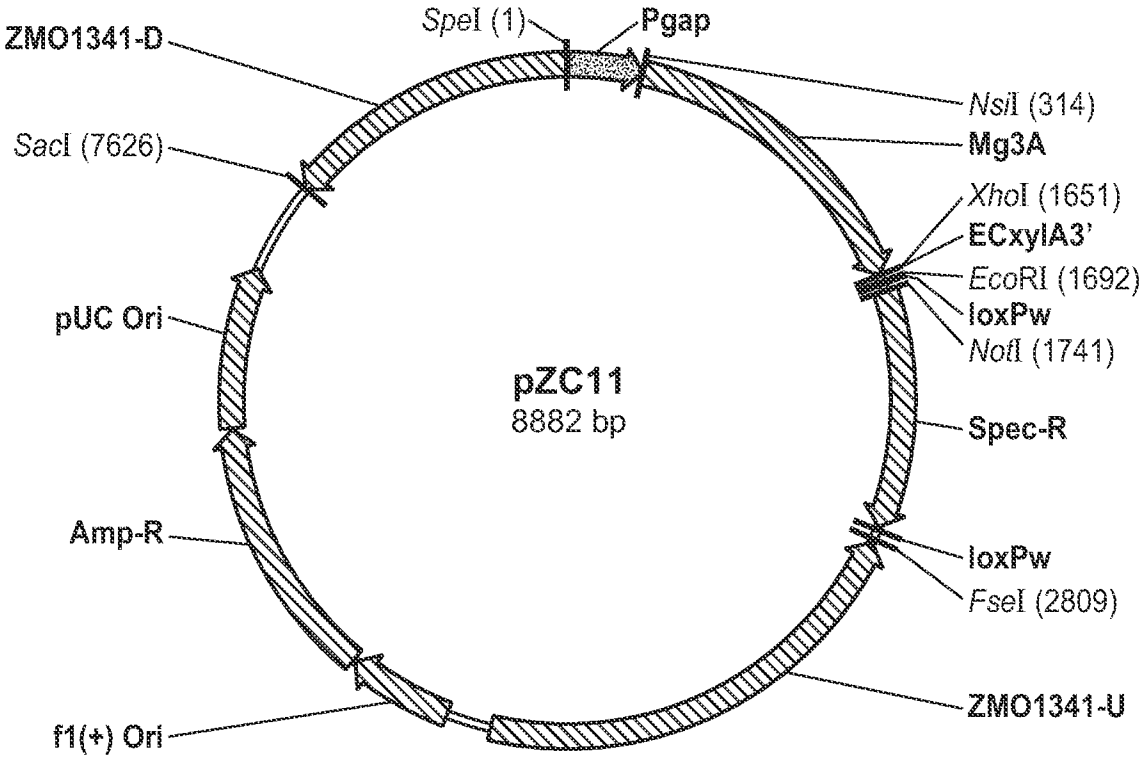

FIG. 8

SEQ ID NO:1: Nucleic acid sequence encoding Mg3A

atgcgtttctccgggatcgtcgccacgcttgtggcgggtgccggcgtgtctgcccaccctggtgacta
ctcaaaggtatgataaaaaataaaaataaaaaggctctcaagcctgacgactcgggaacctttttactt
tgtcccttgtttttgcgcgccgcatgctcagttttgttggcactccttgtaggagaggctattgcctt
ttggtcaggttgtggtagggccttgttttgcctattctgtgaccctctccacacaattctgaagcaat
tggcttaaatacgagggcaaagttccttctgtcgcgttgctgttgcaagaaatccctcgggatcgtca
tgtttctggtttcccatcatggcttccatcaaatacaatagtgctaactatacatagcttgagcgtcg
ggccgtggccacatcagaaccccactaccctcagccgtggatgaacccagatgcagacgggtggcagg
aggcctatgtcaaagcaaaggactttgtctcgcagatgaccctgcttgagaaggtcaacttgacaacc
ggtgtagggtaggtcaacaaacccccttttttttcgtcatgacccccctgaatggcaaccggcaacccc
ctccccccaaacaaaccgtagcgaatattgacacttttttggtactggtgtttgtagctgggcaagcga
tctctgcgttggcaacgttggtgccgtcccccgcctgggtctccgaagcctctgtctgcaggactctc
ccaccggtgtccgtttcgccgattgggtctcggtctttcctgctggcatcaccaccggcgccacgttc
gataaggggctgatgtaccgccgtggacaagctatgggccaagaagccaaggacaagggcatcaacgt
cctgcttggtcccgttgctggaggacttggccgtgtggctgctggtaagcctttgctcgatacactgg
ctgatgacaacatagagagatgtctttttttctcacagctaacccgacatcataggtggacgagcttgg
gagtctttcggagctgacccagtccttaccggatatggtatgattgagacaatcaagggtatccagga
cactggtgtcattgctactgccaagcatttcatcggcaacgagcaaggtaagtctatcggttgacctc
tttggccagtcgagatgcctggtgcaagcttgctgtggtttccaggtccatttactaatacccaatgg
catccaacagagcatttccgtcaggtaggcgaagagcgaggccgtggcgtcaacatcagcgagtcgct
ctcctccaacattgatgacaagactatgcacgagttgtacctctggtaagtccaacactgcattcagg
ggatggattggattagtgcctggctgacctgaccataataggccctttgcggatgcagttcgcgcagg
tgtcggctcggttatgtgctcgtacacacaggtcaacaactcgtacggttgccagaacagcaagcttc
tcaacggtttgctcaaggacgagcttggcttccagggattcgtcatgtctgactggcaggctcagcac
actggagcggccagcgctgcggccggtcttgatatgtccatggtcagtcgaagtctccaagttctatt
ctagttcatgacaggcaacgatgctaacgtttctcgaacaaaagcccggtgacaccgagttcaacaca
ggtcttagtttctggggagccaacttgaccctggctgtcgtcaacggaactgttgccgaatggcgtat
cgacgacatggccatgcgaatcatggctgccttcttcaaggttggcaacacgctggaccagcccgaga
tcaactttagctcttggaccaaggacaccttcggtcccccttcactcgtcatctggaaacagaatccag
cagatcaaccagcacgtcgatgtccgtcgcgaccacggcaacttgattcgcgaggttgccgccaaggg
aaccgtgcttctcaagaacaccaacaacgccctcccgctcaacaagcccaagttcctggccgtcattg
gtgacgatgccggctccaacccgcgcggtcccaacggctgcccggaccgcggctgccttttgggaact
ttgggcatggcctggggttctggcacggctgacttcccctacctcatcacccccgacgcggctctcca
agctcaggccatcgaggacggcacacgctacgagagcatcctgagcaactatgccacggcacagactc
aagctcttgtcagccaaacctacgcaaccgccattgtctttgttgctgcttcttcgggagaaggctgt
gagtctgcctttcttgtttcagctgaacgcaccatgtttgaagcaacacccccttactgacaattctt
ctcttcccaccagacatcgactttgacggaaacaagggagaccgcaacaacctgaccctctggtatga
cggcgactcgctggtcaagaatgtcagctccgtgtgcaacaacaccatcgtcgtcatccactccaccg
gcccgaccatcctgaccgagtggtacgacaaccccaacgtcaccgccatcgtctgggcaggtgttccg
ggacaggagtccggccgtgccatcaccgacgtcctgtacggccgcgtcaaccccgccggccgctcgcc
cttcacctggggcaagacccgcgagagctacggcaccgacgtcatgtacaagcccaacaacggcaacg
aggccccgcagcaggactacaccgagggcgtcttcatcgactaccgccactttgaccagcagaaggac
gagcccgtctacgagttcggccacggcttgagctacaccaccctttgagtactccaacatccgcgtcga
caaggccccggccagcgagtacaagcccaccaccggccagaccatcccggcccccgtctttggcgcca
acgtctccaaggacctgtcccagtacactttcccgtccgacgagttcccgcacatctacctcttcatc
tacccgtacctcaacaccagctcctcgggtgaggaggcctcgcgcgaccccaagtacggcggcacggc

FIG. 9A

cgaggagttcctgccgcccaaggcgctcgatggctcgccccagcccttgccccgcgcgtcgggcaaga
actcgcccggcggcaaccgccagctgtacgacaccctgtacaccgtcacggcaaccatcaccaacacc
ggcaagctcgtgggcgaagaggtgccgcagttgtacgtgtcgcacggcggccccgaggacccgcccgt
cgtcctgcgcggcttcgagcgcatccgtctcgaccccggccagagcgccaccttcaaggtcgacctga
ccaggagggacgtcagcaactgggacgtcaaggtccaggactgggtcatctctgagcaccccaagaag
gtgtttgtcggaagcagcagcaggaagctccatctgagtgctgacctgaactga

SEQ ID NO:2: Polypeptide sequence of Mg3A (underlined residues are the predicted signal sequence)

mrfsgivatlvaqaqvsahpgdysklerravatsephypqpwmnpdadgwqeayvkakdfvsqmtlle
kvnlttgvgwasdlcvgnvgavprlglrslclqdsptgvrfadwvsvfpagittgatfdkglmyrrgq
amgqeakdkginvllgpvagglgrvaaggrawesfgadpvltgygmietikgiqdtgviatakhfign
eqehfrqvgeergrgvniseslssniddktmhelylwpfadavragvgsvmcsytqvnnsygcqnskl
lngllkdelgfqgfvmsdwqaqhtgaasaaagldmsmpgdtefntglsfwganltlavvngtvaewri
ddmamrimaaffkvgntldqpeinfsswtkdtfgplhsssgnriqqinqhvdvrrdhgnlirevaakg
tvllkntnnalplnkpkflavigddagsnprgpngcpdrgcllgtlgmawgsgtadfpylitpdaalq
aqaiedgtryesilsnyataqtqalvsqtyataivfvaassgegyidfdgnkgdrnnltlwydgdslv
knvssvcnntivvihstgptiltewydnpnvtaivwagvpgqesgraitdvlygrvnpagrspftwgk
tresygtdvmykpnngneapqqdytegvfidyrhfdqqkdepvyefghglsyttfeysnirvdkapas
eykpttgqtipapvfganvskdlsqytfpsdefphiylfiypylntsssgeeasrdpkyggtaeeflp
pkaldgspqplprasgknspggnrqlydtlytvtatitntgklvgeevpqlyvshggpedppvvlrgf
erirldpgqsatfkvdltrrdvsnwdvkvqdwvisehpkkvfvgsssrklhlsadln

SEQ ID NO:3: Mature Mg3A polypeptide sequence:

hpgdysklerravatsephypqpwmnpdadgwqeayvkakdfvsqmtllekvnlttgvgwasdlcvgn
vgavprlglrslclqdsptgvrfadwvsvfpagittgatfdkglmyrrgqamgqeakdkginvllgpv
agglgrvaaggrawesfgadpvltgygmietikgiqdtgviatakhfigneqehfrqvgeergrgvni
seslssniddktmhelylwpfadavragvgsvmcsytqvnnsygcqnskllngllkdelgfqgfvmsd
wqaqhtgaasaaagldmsmpgdtefntglsfwganltlavvngtvaewriddmamrimaaffkvgntl
dqpeinfsswtkdtfgplhsssgnriqqinqhvdvrrdhgnlirevaakgtvllkntnnalplnkpkf
lavigddagsnprgpngcpdrgcllgtlgmawgsgtadfpylitpdaalqaqaiedgtryesilsnya
taqtqalvsqtyataivfvaassgegyidfdgnkgdrnnltlwydgdslvknvssvcnntivvihstg
ptiltewydnpnvtaivwagvpgqesgraitdvlygrvnpagrspftwgktresygtdvmykpnngne
apqqdytegvfidyrhfdqqkdepvyefghglsyttfeysnirvdkapaseykpttgqtipapvfgan
vskdlsqytfpsdefphiylfiypylntsssgeeasrdpkyggtaeeflppkaldgspqplprasgkn
spggnrqlydtlytvtatitntgklvgeevpqlyvshggpedppvvlrgferirldpgqsatfkvdlt
rrdvsnwdvkvqdwvisehpkkvfvgsssrklhlsadln

FIG. 9B

SEQ ID NO:4: *T.reesei* Bgl1 polypeptide sequence (underlined are predicted signal sequence residues)

mryrtaaalalatgpfaradshstsgasaeavvppagtpwgtaydkakaalaklnlqdkvgivsgvgwnggpcvg
ntspaskisypslclqdgplgvrystgstaftpgvqaastwdvnlirergqfigeevkasgihvilgpvagplgk
tpqggrnwegfgvdpyltgiamgqtingiqsvgvqatakhyilneqelnretissnpddrtlhelytwpfadavq
anvasvmcsynkvnttwacedqytlqtvlkdqlgfpgyvmtdwnaqhttvqsansgldmsmpgtdfngnnrlwgp
altnavnsnqvptsrvddmvtrilaawyltgqdqaqypsfnisrnvqgnhktnvraiardgivllkndanilplk
kpasiavvgsaaiignharnspscndkgcddgalgmgwgsgavnypyfvapydaintrassqgtqvtlsntdnts
sqasaargkdvaivfitadsgegyitvegnagdrnnldpwhngnalvqavagansnvivvvhsvgaiileqilal
pqvkavvwaglpsqesgnalvdvlwgdvspsgklvytiakspndyntrivsggsdsfseglfidykhfddanitp
ryefgyglsytkfnysrlsvlstaksgpatgavvpggpsdlfqnvatvtvdiansgqvtgaevaqlyitypssap
rtppkqlrgfaklnltpgqsgtatfnirrrdlsywdtasqkwvvpsgsfgisvgassrdirltstlsva

FIG. 9C

SEQ ID NOs: 13-39: signal peptide sequences

SEQ ID NO:13: mryrtaaalalatgpfara
SEQ ID NO:14: mvsftsllaasppsrascrpaaevesvavekr
SEQ ID NO:15: mkanvilcllaplvaa
SEQ ID NO:16: mivgilttlatlatlaas
SEQ ID NO:17: myrklavisaflatara
SEQ ID NO:18: mllnlqvaasalslsllgglaea
SEQ ID NO:19: mklnwvaaalsigaagtds
SEQ ID NO:20: masirsvlvsgllaagvna
SEQ ID NO:21: mwltspllfastllgltgvala
SEQ ID NO:22: mrfswllcpllamgsa
SEQ ID NO:23: mrllsfpshllvafltlkeass
SEQ ID NO:24: mqlkflssalllsltgncaa
SEQ ID NO:25: mkvywlvawatsltpala
SEQ ID NO:26: mvrfssilaaaacfvaves
SEQ ID NO:27: mihlkpalaallalstqcva
SEQ ID NO:28: malqtffllaaamlana
SEQ ID NO:29: mklnkpflaiylafnlaea
SEQ ID NO:30: maplslralsllaltgaaaa
SEQ ID NO:31: mVRPTILLTSLLLAPFAAA
SEQ ID NO:32: mhmhslvaalaagtlpllasa
SEQ ID NO:33: mvhlsslaaalaalplvyg
SEQ ID NO:34: mrfslaattllaglata
SEQ ID NO:35: mvvlsklvssilfaslvsa
SEQ ID NO:36: mvqikaaalamlfashvls
SEQ ID NO:37: mKASSVLLGLAPLAALA
SEQ ID NO:38: MRFPSIFTAVLFAASSALA
SEQ ID NO:39:
MRFPSIFTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYLDLEGDFDVAVLPFSNSTNNGLLFINTTIASIAA
KEEGVSLDKR

FIG. 9D

SEQ ID NOs: 40-42: signal peptide sequences

SEQ ID NO:40:MLLQAFLFLLAGFAAKISAR
SEQ ID NO:41:
ATGATAAAAGTCCCGCGGTTCATCTGTATGATCGCGCTTACATCCAGCGTTCTGGCAAGCGGCCTTTCTCAAAGC
GTTTCAGCTCAT
SEQ ID NO:42:
ATGAAAAGAAAGCTTGGTCGTCGCCAGTTATTAACTGGCTTTGTTGCCCTTGGCGGTATGGCGATTACAGCTGGT
AAGGCGCAGGCTTCT

FIG. 9E

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2013/067413

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/42
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, Sequence Search, WPI Data, BIOSIS, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 15 March 2005 (2005-03-15), "SubName: Full=Beta-glucosidase-like protein;", XP002718760, retrieved from EBI accession no. UNIPROT:Q5EMW3 Database accession no. Q5EMW3 sequence ----- -/--	1-22

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 January 2014	28/01/2014

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Griesinger, Irina

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE UniProt [Online] 14 December 2011 (2011-12-14), "SubName: Full=Beta-glucosidase 1;", XP002718761, retrieved from EBI accession no. UNIPROT:G4MQA2 Database accession no. G4MQA2 sequence -----	1-22
X	WO 2007/071818 A1 (ROAL OY [FI]; VEHMAANPERAE JARI [FI]; ALAPURANEN MARIKA [FI]; PURANEN) 28 June 2007 (2007-06-28) claims 1,10, 18, 26; example 20; table 24; sequences 23,24 -----	1-22
A	OHMIYA KUNIO ET AL: "Structure of cellulases and their applications", BIOTECHNOLOGY AND GENETIC ENGINEERING REVIEWS, INTERCEPT LTD., ANDOVER, GB, vol. 14, 1 April 1997 (1997-04-01), pages 365-414, XP008093172, ISSN: 0264-8725 the whole document -----	1-22

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2013/067413

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007071818 A1	28-06-2007	AR 058721 A1	20-02-2008
		AU 2006326963 A1	28-06-2007
		BR PI0620287 A2	08-11-2011
		CA 2632502 A1	28-06-2007
		CA 2833029 A1	28-06-2007
		CN 101384713 A	11-03-2009
		CN 102174492 A	07-09-2011
		EP 1963498 A1	03-09-2008
		EP 2319920 A1	11-05-2011
		EP 2453013 A1	16-05-2012
		EP 2453014 A1	16-05-2012
		FI 20051318 A	28-12-2006
		TW 200801195 A	01-01-2008
		US 2009042266 A1	12-02-2009
		US 2011045544 A1	24-02-2011
		US 2013224801 A1	29-08-2013
		WO 2007071818 A1	28-06-2007