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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 March 2019 (21.03.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/055373 A1

(51) International Patent Classification:
C12N 9/10 (2006.01) *C12P 19/04* (2006.01)

(21) International Application Number:
PCT/US2018/050345

(22) International Filing Date:
11 September 2018 (11.09.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/557,834 13 September 2017 (13.09.2017) US

(71) Applicant: **E. I. DU PONT DE NEMOURS AND COMPANY** [US/US]; Chestnut Run Plaza, 974 Centre Road, P. O. Box 2915, Wilmington, Delaware 19805 (US).

(72) Inventors: **LI, Yougen**; 4 Kentsdale Drive, Pennington, New Jersey 08534 (US). **SEMKE, Ellen D.**; 413 Douglas D Alley Drive, Newark, Delaware 19713 (US). **PAYNE, Mark S.**; 4617 Old Linden Hill Road, Wilmington, Delaware 19808 (US). **PARKER, Jared B.**; 111 Milhollan Drive, Elkton, Maryland 21921 (US). **HENNESSEY, Susan Marie**; 32 Truman Lane, Avondale, Pennsylvania 19311 (US). **KRALJ, Slavko**; Dupont Nutrition Biosciences APS, Langebrogade 1, DK-1411 Copenhagen K (DK). **ALKAN, Veli**; 925 Page Mill Road, Palo Alto, California 94304 (US). **BOTT, Richard**; 925 Page Mill Road, Palo Alto, California 94304 (US).

(74) Agent: **CHESIRE, Dennis R.**; E. I. du Pont de Nemours and Company, Legal Patent Records Center, Chestnut Run Plaza 721/2340, 974 Centre Road, PO Box 2915 Wilmington, Delaware 19805 (US).

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

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

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: ENGINEERED GLUCOSYLTRANSFERASES

(57) Abstract: Disclosed herein are glucosyltransferases with modified amino acid sequences. Such engineered enzymes synthesize alpha-glucan products having increased molecular weight. Further disclosed are reactions and methods in which engineered glucosyltransferases are used to produce alpha-glucan.

WO 2019/055373 A1

TITLE

ENGINEERED GLUCOSYLTRANSFERASES

This application claims the benefit of U.S. Provisional Application No. 62/557,834 (filed September 13, 2017), which is incorporated herein by reference in its entirety.

FIELD

The present disclosure is in the field of enzyme catalysis. For example, the disclosure pertains to glucosyltransferase enzymes with modified amino acid sequences. Such modified enzymes synthesize products with increased molecular weight.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

The official copy of the sequence listing is submitted electronically via EFS-Web as an ASCII formatted sequence listing with a file named 20180911_CL6159WOPCT_SequenceListing_ST25 created on September 11, 2018, and having a size of about 315 kilobytes and is filed concurrently with the specification. The sequence listing contained in this ASCII-formatted document is part of the specification and is herein incorporated by reference in its entirety.

BACKGROUND

Driven by a desire to use polysaccharides in various applications, researchers have explored for polysaccharides that are biodegradable and that can be made economically from renewably sourced feedstocks. One such polysaccharide is alpha-1,3-glucan, an insoluble glucan polymer characterized by having alpha-1,3-glycosidic linkages. This polymer has been prepared, for example, using a glucosyltransferase enzyme isolated from *Streptococcus salivarius* (Simpson et al., *Microbiology* 141:1451-1460, 1995). Also for example, U.S. Patent No. 7000000 disclosed the preparation of a spun fiber from enzymatically produced alpha-1,3-glucan. Various other glucan materials have also been studied for developing new or enhanced applications. For example, U.S. Patent Appl. Publ. No. 2015/0232819 discloses enzymatic synthesis of several insoluble glucans having mixed alpha-1,3 and -1,6 linkages.

While these and other advances have been made in producing glucan polymers using glucosyltransferase enzymes, less attention appears to have been drawn to enhancing the molecular weight of insoluble glucan products synthesized by such enzymes. Addressing this technological gap, disclosed herein are glucosyltransferases

with modified amino acid sequences that produce higher molecular weight insoluble glucan products.

SUMMARY

In one embodiment, the present disclosure concerns a non-native glucosyltransferase comprising at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, wherein the non-native glucosyltransferase synthesizes insoluble alpha-glucan comprising 1,3-linkages, and the molecular weight of the insoluble alpha-glucan is higher than the molecular weight of insoluble alpha-glucan synthesized by a second glucosyltransferase that only differs from the non-native glucosyltransferase at the substitution position(s).

In another embodiment, the present disclosure concerns a polynucleotide comprising a nucleotide sequence encoding a non-native glucosyltransferase as presently disclosed, optionally wherein one or more regulatory sequences are operably linked to the nucleotide sequence, and preferably wherein the one or more regulatory sequences include a promoter sequence.

In another embodiment, the present disclosure concerns a reaction composition comprising water, sucrose, and a non-native glucosyltransferase as presently disclosed.

In another embodiment, the present disclosure concerns a method of producing insoluble alpha-glucan comprising: (a) contacting at least water, sucrose, and a non-native glucosyltransferase enzyme as presently disclosed, whereby insoluble alpha-glucan is produced; and (b) optionally, isolating the insoluble alpha-glucan produced in step (a).

In another embodiment, the present disclosure concerns a method of preparing a polynucleotide sequence encoding a non-native glucosyltransferase, the method comprising: (a) identifying a polynucleotide sequence encoding a parent glucosyltransferase that (i) comprises an amino acid sequence that is at least about 40% identical to SEQ ID NO:4 or positions 55-960 of SEQ ID NO:4, and (ii) synthesizes insoluble alpha-glucan comprising 1,3-linkages; and (b) modifying the polynucleotide sequence identified in step (a) to substitute at least one amino acid of the parent glucosyltransferase at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661,

Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, thereby providing a polynucleotide sequence encoding a non-native glucosyltransferase that synthesizes insoluble alpha-glucan with a molecular weight that is higher than the molecular weight of insoluble alpha-glucan synthesized by the parent glucosyltransferase.

BRIEF DESCRIPTION OF THE SEQUENCES

Table 1. Summary of Nucleic Acid and Protein SEQ ID Numbers[b]

Description	Nucleic acid SEQ ID NO.	Protein SEQ ID NO.
GTF 0874, *Streptococcus sobrinus*. The first 156 amino acids of the protein are deleted compared to GENBANK Identification No. 450874; a start methionine is included.	1 [a]	2 (1435 aa)
GTF 6855, *Streptococcus salivarius* SK126. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 228476855 (Acc. No. ZP_04061500.1); a start methionine is included.	3 [a]	4 (1341 aa)
GTF 2379, *Streptococcus salivarius.* The first 203 amino acids of the protein are deleted compared to GENBANK Identification No. 662379; a start methionine is included.	5 [a]	6 (1247 aa)
GTF 7527 or GTFJ, *Streptococcus salivarius.* The first 42 amino acids of the protein are deleted compared to GENBANK Identification No. 47527; a start methionine is included.	7 [a]	8 (1477 aa)
GTF 1724, *Streptococcus downei.* The first 162 amino acids of the protein are deleted compared to GENBANK Identification No. 121724; a start methionine is included.	9 [a]	10 (1436 aa)
GTF 0544, *Streptococcus mutans.* The first 164 amino acids of the protein are deleted compared to GENBANK Identification No. 290580544; a start methionine is included.	11 [a]	12 (1313 aa)
GTF 5926, *Streptococcus dentirousetti.* The first 144 amino acids of the protein are deleted compared to GENBANK Identification No. 167735926; a start methionine is included.	13 [a]	14 (1323 aa)
GTF 4297, *Streptococcus oralis.* The first 228 amino acids of the protein are deleted compared to GENBANK Identification No. 7684297; a start methionine is included.	15 [a]	16 (1348 aa)
GTF 5618, *Streptococcus sanguinis.* The first 223 amino acids of the protein are deleted compared to GENBANK Identification No. 328945618; a start methionine is included.	17 [a]	18 (1348 aa)
GTF 2765, unknown *Streptococcus* sp. C150. The first 193 amino acids of the protein are deleted compared to	19 [a]	20 (1340 aa)

GENBANK Identification No. 322372765; a start methionine is included.		
GTF 0427, *Streptococcus sobrinus.* The first 156 amino acids of the protein are deleted compared to GENBANK Identification No. 940427; a start methionine is included.	25 [a]	26 (1435 aa)
GTF 2919, *Streptococcus salivarius* PS4. The first 92 amino acids of the protein are deleted compared to GENBANK Identification No. 383282919; a start methionine is included.	27 [a]	28 (1340 aa)
GTF 2678, *Streptococcus salivarius* K12. The first 188 amino acids of the protein are deleted compared to GENBANK Identification No. 400182678; a start methionine is included.	29 [a]	30 (1341 aa)
GTF 3929, *Streptococcus salivarius* JIM8777. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 387783929; a start methionine is included.	33 [a]	34 (1341 aa)
GTF 3298, *Streptococcus* sp. C150. The first 209 amino acids of the protein are deleted compared to GENBANK Identification No. 322373298; a start methionine is included.		59 (1242 aa)
Wild type GTFJ, *Streptococcus salivarius.* GENBANK Identification No. 47527.		60 (1518 aa)
Wild type GTF corresponding to GTF 2678, *Streptococcus salivarius* K12.		61 (1528 aa)
Wild type GTF corresponding to GTF 6855, *Streptococcus salivarius* SK126.		62 (1518 aa)
Wild type GTF corresponding to GTF 2919, *Streptococcus salivarius* PS4.		63 (1431 aa)
Wild type GTF corresponding to GTF 2765, unknown *Streptococcus* sp. C150.		64 (1532 aa)
Shorter version of GTF 7527, *Streptococcus salivarius*, (also referred to as "7527-NT" herein. The first 178 amino acids of the protein are deleted compared to GENBANK Identification No. 47527; a start methionine is included.		65 (1341 aa)
Terminator sequence added to pHY300PLK to derive the pHYT vector.	67	

[a] This DNA coding sequence is codon-optimized for expression in *E. coli*, and is merely disclosed as an example of a suitable coding sequence.

[b] SEQ ID NOs:21-24, 31, 32, 35-58 and 66 are intentionally not included in this table and merely serve as placeholders.

DETAILED DESCRIPTION

The disclosures of all cited patent and non-patent literature are incorporated herein by reference in their entirety.

Unless otherwise disclosed, the terms "a" and "an" as used herein are intended to encompass one or more (i.e., at least one) of a referenced feature.

Where present, all ranges are inclusive and combinable, except as otherwise noted. For example, when a range of "1 to 5" is recited, the recited range should be construed as including ranges "1 to 4", "1 to 3", "1-2", "1-2 & 4-5", "1-3 & 5", and the like.

The terms "alpha-glucan", "alpha-glucan polymer" and the like are used interchangeably herein. An alpha-glucan is a polymer comprising glucose monomeric units linked together by alpha-glycosidic linkages. In typical embodiments, an alpha-glucan herein comprises at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% alpha-glycosidic linkages. Examples of alpha-glucan polymers herein include alpha-1,3-glucan.

The terms "poly alpha-1,3-glucan", "alpha-1,3-glucan", "alpha-1,3-glucan polymer" and the like are used interchangeably herein. Alpha-1,3-glucan is a polymer comprising glucose monomeric units linked together by glycosidic linkages, typically wherein at least about 50% of the glycosidic linkages are alpha-1,3. Alpha-1,3-glucan in certain embodiments comprises at least 90% or 95% alpha-1,3 glycosidic linkages. Most or all of the other linkages in alpha-1,3-glucan herein typically are alpha-1,6, though some linkages may also be alpha-1,2 and/or alpha-1,4.

The terms "glycosidic linkage", "glycosidic bond", "linkage" and the like are used interchangeably herein and refer to the covalent bond that joins a carbohydrate (sugar) molecule to another group such as another carbohydrate. The term "alpha-1,3-glycosidic linkage" as used herein refers to the type of covalent bond that joins alpha-D-glucose molecules to each other through carbons 1 and 3 on adjacent alpha-D-glucose rings. The term "alpha-1,6-glycosidic linkage" as used herein refers to the covalent bond that joins alpha-D-glucose molecules to each other through carbons 1 and 6 on adjacent alpha-D-glucose rings. The glycosidic linkages of a glucan polymer herein can also be referred to as "glucosidic linkages". Herein, "alpha-D-glucose" will be referred to as "glucose".

The glycosidic linkage profile of an alpha-glucan herein can be determined using any method known in the art. For example, a linkage profile can be determined using methods using nuclear magnetic resonance (NMR) spectroscopy (e.g., ^{13}C NMR or ^{1}H NMR). These and other methods that can be used are disclosed in, for example, Food

Carbohydrates: Chemistry, Physical Properties, and Applications (S. W. Cui, Ed., Chapter 3, S. W. Cui, Structural Analysis of Polysaccharides, Taylor & Francis Group LLC, Boca Raton, FL, 2005), which is incorporated herein by reference.

The "molecular weight" of large alpha-glucan polymers herein can be represented as weight-average molecular weight (Mw) or number-average molecular weight (Mn), the units of which are in Daltons or grams/mole. Alternatively, the molecular weight of large alpha-glucan polymers can be represented as DP_w (weight average degree of polymerization) or DP_n (number average degree of polymerization). The molecular weight of smaller alpha-glucan polymers such as oligosaccharides typically can be provided as "DP" (degree of polymerization), which simply refers to the number of glucoses comprised within the alpha-glucan. Various means are known in the art for calculating these various molecular weight measurements such as with high-pressure liquid chromatography (HPLC), size exclusion chromatography (SEC), or gel permeation chromatography (GPC).

The term "sucrose" herein refers to a non-reducing disaccharide composed of an alpha-D-glucose molecule and a beta-D-fructose molecule linked by an alpha-1,2-glycosidic bond. Sucrose is known commonly as table sugar.

The terms "leucrose" and "D-glucopyranosyl-alpha(1-5)-D-fructopyranose" are used interchangeably herein and refer to a disaccharide containing an alpha-1,5 glucosyl-fructose linkage.

The terms "glucosyltransferase", "glucosyltransferase enzyme", "GTF", "glucansucrase" and the like are used interchangeably herein. The activity of a glucosyltransferase herein catalyzes the reaction of the substrate sucrose to make the products alpha-glucan and fructose. Other products (by-products) of a GTF reaction can include glucose, various soluble gluco-oligosaccharides, and leucrose. Wild type forms of glucosyltransferase enzymes generally contain (in the N-terminal to C-terminal direction) a signal peptide (which is typically removed by cleavage processes), a variable domain, a catalytic domain, and a glucan-binding domain. A glucosyltransferase herein is classified under the glycoside hydrolase family 70 (GH70) according to the CAZy (Carbohydrate-Active EnZymes) database (Cantarel et al., *Nucleic Acids Res.* 37:D233-238, 2009).

The term "glucosyltransferase catalytic domain" herein refers to the domain of a glucosyltransferase enzyme that provides alpha-glucan-synthesizing activity to a

glucosyltransferase enzyme. A glucosyltransferase catalytic domain typically does not require the presence of any other domains to have this activity.

The terms “enzymatic reaction”, “glucosyltransferase reaction”, “glucan synthesis reaction”, “reaction composition”, “reaction formulation” and the like are used interchangeably herein and generally refer to a reaction that initially comprises water, sucrose, at least one active glucosyltransferase enzyme, and optionally other components. Components that can be further present in a glucosyltransferase reaction typically after it has commenced include fructose, glucose, leucrose, soluble gluco-oligosaccharides (e.g., DP2-DP7) (such may be considered as products or by-products, depending on the glucosyltransferase used), and/or insoluble alpha-glucan product(s) of DP8 or higher (e.g., DP100 and higher). It would be understood that certain glucan products, such as alpha-1,3-glucan with a degree of polymerization (DP) of at least 8 or 9, are water-insoluble and thus not dissolved in a glucan synthesis reaction, but rather may be present out of solution (e.g., by virtue of having precipitated from the reaction). It is in a glucan synthesis reaction where the step of contacting water, sucrose and a glucosyltransferase enzyme is performed. The term “under suitable reaction conditions” as used herein refers to reaction conditions that support conversion of sucrose to alpha-glucan product(s) via glucosyltransferase enzyme activity.

The terms “percent by volume”, “volume percent”, “vol %”, “v/v %” and the like are used interchangeably herein. The percent by volume of a solute in a solution can be determined using the formula: [(volume of solute)/(volume of solution)] x 100%.

The terms “percent by weight”, “weight percentage (wt%)”, “weight-weight percentage (% w/w)” and the like are used interchangeably herein. Percent by weight refers to the percentage of a material on a mass basis as it is comprised in a composition, mixture, or solution.

The terms “aqueous conditions”, “aqueous reaction conditions”, “aqueous setting”, “aqueous system” and the like are used interchangeably herein. Aqueous conditions herein refer to a solution or mixture in which the solvent is at least about 60 wt% water, for example. A glucosyltransferase reaction herein is performed under aqueous conditions.

The terms “soluble”, “aqueous-soluble”, “water-soluble” and the like as used herein characterize a glucan that has the capability of dissolving in water and/or an aqueous solution herein. Examples of soluble glucans herein are certain

oligosaccharides, such as alpha-1,3-glucan with a DP less than 8. In contrast, a glucan that is "insoluble", "aqueous-insoluble", "water-insoluble" (and like terms) does not dissolve (or does not appreciably dissolve) in water and/or an aqueous solution herein. Optionally, the conditions for determining solubility include a water/solution temperature range of about 1 to 85 °C (e.g., 20-25 °C) and/or a pH range of about 4-9 (e.g., 6-8).

The terms "polynucleotide", "polynucleotide sequence", "nucleic acid molecule" and the like are used interchangeably herein. These terms encompass nucleotide sequences and the like. A polynucleotide may be a polymer of DNA or RNA that is single- or double-stranded, that optionally contains synthetic, non-natural or altered nucleotide bases. A polynucleotide may be comprised of one or more segments of cDNA, genomic DNA, synthetic DNA, or mixtures thereof.

The term "gene" as used herein refers to a DNA polynucleotide sequence that expresses an RNA (RNA is transcribed from the DNA polynucleotide sequence) from a coding region, which RNA can be a messenger RNA (encoding a protein) or a non-protein-coding RNA. A gene may refer to the coding region alone, or may include regulatory sequences upstream and/or downstream to the coding region (e.g., promoters, 5'-untranslated regions, 3'-transcription terminator regions). A coding region encoding a protein can alternatively be referred to herein as an "open reading frame" (ORF). A gene that is "native" or "endogenous" refers to a gene as found in nature with its own regulatory sequences; such a gene is located in its natural location in the genome of a host cell. A "chimeric" gene refers to any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature (i.e., the regulatory and coding regions are heterologous with each other). Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A "foreign" or "heterologous" gene can refer to a gene that is introduced into the host organism by gene transfer. Foreign/heterologous genes can comprise native genes inserted into a non-native organism, native genes introduced into a new location within the native host, or chimeric genes. Polynucleotide sequences in certain embodiments disclosed herein are heterologous. A "transgene" is a gene that has been introduced into the genome by a gene delivery procedure (e.g., transformation). A "codon-

optimized" open reading frame has its frequency of codon usage designed to mimic the frequency of preferred codon usage of the host cell.

As used herein, the term "polypeptide" is defined as a chain of amino acid residues, usually having a defined sequence. As used herein the term polypeptide is interchangeable with the terms "peptides" and "proteins". Typical amino acids contained in polypeptides herein include (respective three- and one-letter codes shown parenthetically): alanine (Ala, A), arginine (Arg, R), asparagine (Asn, N), aspartic acid (Asp, D), cysteine (Cys, C), glutamic acid (Glu, E), glutamine (Gln, Q), glycine (Gly, G), histidine (His, H), isoleucine (Ile, I), leucine (Leu, L), lysine (Lys, K), methionine (Met, M), phenylalanine (Phe, F), proline (Pro, P), serine (Ser, S), threonine (Thr, T), tryptophan (Trp, W), tyrosine (Tyr, Y), valine (Val, V).

The term "heterologous" means not naturally found in the location of interest. For example, a heterologous gene can be one that is not naturally found in a host organism, but that is introduced into the host organism by gene transfer. As another example, a nucleic acid molecule that is present in a chimeric gene can be characterized as being heterologous, as such a nucleic acid molecule is not naturally associated with the other segments of the chimeric gene (e.g., a promoter can be heterologous to a coding sequence).

A "non-native" amino acid sequence or polynucleotide sequence comprised in a cell or organism herein does not occur in a native (natural) counterpart of such cell or organism. Such an amino acid sequence or polynucleotide sequence can also be referred to as being heterologous to the cell or organism.

"Regulatory sequences" as used herein refer to nucleotide sequences located upstream of a gene's transcription start site (e.g., promoter), 5' untranslated regions, introns, and 3' non-coding regions, and which may influence the transcription, processing or stability, and/or translation of an RNA transcribed from the gene. Regulatory sequences herein may include promoters, enhancers, silencers, 5' untranslated leader sequences, introns, polyadenylation recognition sequences, RNA processing sites, effector binding sites, stem-loop structures, and other elements involved in regulation of gene expression. One or more regulatory elements herein may be heterologous to a coding region herein.

A "promoter" as used herein refers to a DNA sequence capable of controlling the transcription of RNA from a gene. In general, a promoter sequence is upstream of the

transcription start site of a gene. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. Promoters that cause a gene to be expressed in a cell at most times under all circumstances are commonly referred to as "constitutive promoters". A promoter may alternatively be inducible. One or more promoters herein may be heterologous to a coding region herein.

A "strong promoter" as used herein refers to a promoter that can direct a relatively large number of productive initiations per unit time, and/or is a promoter driving a higher level of gene transcription than the average transcription level of the genes in a cell.

The terms "3' non-coding sequence", "transcription terminator", "terminator" and the like as used herein refer to DNA sequences located downstream of a coding sequence. This includes polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression.

As used herein, a first nucleic acid sequence is "hybridizable" to a second nucleic acid sequence when a single-stranded form of the first nucleic acid sequence can anneal to the second nucleic acid sequence under suitable annealing conditions (e.g., temperature, solution ionic strength). Hybridization and washing conditions are well known and exemplified in Sambrook J, Fritsch EF and Maniatis T, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989), which is incorporated herein by reference, particularly Chapter 11 and Table 11.1.

The term "DNA manipulation technique" refers to any technique in which the sequence of a DNA polynucleotide sequence is modified. Although the DNA polynucleotide sequence being modified can be used as a substrate itself for modification, it does not have to be physically in hand for certain techniques (e.g., a sequence stored in a computer can be used as the basis for the manipulation technique). A DNA manipulation technique can be used to delete and/or mutate one or more DNA sequences in a longer sequence. Examples of a DNA manipulation technique include recombinant DNA techniques (restriction and ligation, molecular cloning), polymerase chain reaction (PCR), and synthetic DNA methods (e.g., oligonucleotide synthesis and ligation). Regarding synthetic DNA techniques, a DNA manipulation technique can entail observing a DNA polynucleotide *in silico*, determining

desired modifications (e.g., one or more deletions) of the DNA polynucleotide, and synthesizing a DNA polynucleotide that contains the desired modifications.

The term "*in silico*" herein means in or on an information storage and/or processing device such as a computer; done or produced using computer software or simulation, i.e., virtual reality.

The terms "upstream" and "downstream" as used herein with respect to polynucleotides refer to "5' of" and "3' of", respectively.

The term "expression" as used herein refers to (i) transcription of RNA (e.g., mRNA or a non-protein-coding RNA) from a coding region, and/or (ii) translation of a polypeptide from mRNA. Expression of a coding region of a polynucleotide sequence can be up-regulated or down-regulated in certain embodiments.

The term "operably linked" as used herein refers to the association of two or more nucleic acid sequences such that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence. That is, the coding sequence is under the transcriptional control of the promoter. A coding sequence can be operably linked to one (e.g., promoter) or more (e.g., promoter and terminator) regulatory sequences, for example.

The term "recombinant" when used herein to characterize a DNA sequence such as a plasmid, vector, or construct refers to an artificial combination of two otherwise separated segments of sequence, e.g., by chemical synthesis and/or by manipulation of isolated segments of nucleic acids by genetic engineering techniques.

The term "transformation" as used herein refers to the transfer of a nucleic acid molecule into a host organism or host cell by any method. A nucleic acid molecule that has been transformed into an organism/cell may be one that replicates autonomously in the organism/cell, or that integrates into the genome of the organism/cell, or that exists transiently in the cell without replicating or integrating. Non-limiting examples of nucleic acid molecules suitable for transformation are disclosed herein, such as plasmids and linear DNA molecules. Host organisms/cells herein containing a transforming nucleic acid sequence can be referred to as "transgenic", "recombinant", "transformed", "engineered", as a "transformant", and/or as being "modified for exogenous gene expression", for example.

The terms "sequence identity", "identity" and the like as used herein with respect to polynucleotide or polypeptide sequences refer to the nucleic acid residues or amino acid residues in two sequences that are the same when aligned for maximum correspondence over a specified comparison window. Thus, "percentage of sequence identity", "percent identity" and the like refer to the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the results by 100 to yield the percentage of sequence identity. It would be understood that, when calculating sequence identity between a DNA sequence and an RNA sequence, T residues of the DNA sequence align with, and can be considered "identical" with, U residues of the RNA sequence. For purposes of determining "percent complementarity" of first and second polynucleotides, one can obtain this by determining (i) the percent identity between the first polynucleotide and the complement sequence of the second polynucleotide (or vice versa), for example, and/or (ii) the percentage of bases between the first and second polynucleotides that would create canonical Watson and Crick base pairs.

Percent identity can be readily determined by any known method, including but not limited to those described in: 1) Computational Molecular Biology (Lesk, A.M., Ed.) Oxford University: NY (1988); 2) Biocomputing: Informatics and Genome Projects (Smith, D.W., Ed.) Academic: NY (1993); 3) Computer Analysis of Sequence Data, Part I (Griffin, A.M., and Griffin, H.G., Eds.) Humana: NJ (1994); 4) Sequence Analysis in Molecular Biology (von Heinje, G., Ed.) Academic (1987); and 5) Sequence Analysis Primer (Gribskov, M. and Devereux, J., Eds.) Stockton: NY (1991), all of which are incorporated herein by reference.

Preferred methods for determining percent identity are designed to give the best match between the sequences tested. Methods of determining identity and similarity are codified in publicly available computer programs, for example. Sequence alignments and percent identity calculations can be performed using the MEGALIGN program of the

LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI), for example. Multiple alignment of sequences can be performed, for example, using the Clustal method of alignment which encompasses several varieties of the algorithm including the Clustal V method of alignment (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.,* 8:189-191 (1992)) and found in the MEGALIGN v8.0 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). For multiple alignments, the default values can correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10. Default parameters for pairwise alignments and calculation of percent identity of protein sequences using the Clustal method can be KTUPLE=1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. For nucleic acids, these parameters can be KTUPLE=2, GAP PENALTY=5, WINDOW=4 and DIAGONALS SAVED=4. Additionally, the Clustal W method of alignment can be used (described by Higgins and Sharp, *CABIOS*. 5:151-153 (1989); Higgins, D.G. et al., *Comput. Appl. Biosci.* 8:189-191(1992); Thompson, J.D. et al, *Nucleic Acids Research*, 22 (22): 4673-4680, 1994) and found in the MEGALIGN v8.0 program of the LASERGENE bioinformatics computing suite (DNASTAR Inc.). Default parameters for multiple alignment (protein/nucleic acid) can be: GAP PENALTY=10/15, GAP LENGTH PENALTY=0.2/6.66, Delay Divergent Seqs(%)=30/30, DNA Transition Weight=0.5, Protein Weight Matrix=Gonnet Series, DNA Weight Matrix=IUB.

Various polypeptide amino acid sequences and polynucleotide sequences are disclosed herein as features of certain embodiments. Variants of these sequences that are at least about 70-85%, 85-90%, or 90%-95% identical to the sequences disclosed herein can be used or referenced. Alternatively, a variant amino acid sequence or polynucleotide sequence can have at least 70%, 71%, 72%, 73%, 74%, 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% identity with a sequence disclosed herein. The variant amino acid sequence or polynucleotide sequence has the same function/activity of the disclosed sequence, or at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% of the function/activity of the disclosed sequence. Any polypeptide amino acid sequence disclosed herein not beginning with a methionine can typically further comprise at least a start-methionine at the N-terminus of the amino acid sequence. In contrast, any

polypeptide amino acid sequence disclosed herein beginning with a methionine can optionally lack such a methionine residue.

The terms "aligns with", "corresponds with", and the like can be used interchangeably herein. Some embodiments herein relate to a glucosyltransferase comprising at least one amino acid substitution at a position corresponding with at least one particular amino acid residue of SEQ ID NO:62. An amino acid position of a glucosyltransferase or subsequence thereof (e.g., catalytic domain or catalytic domain plus glucan-binding domains) (can refer to such an amino acid position or sequence as a "query" position or sequence) can be characterized to correspond with a particular amino acid residue of SEQ ID NO:62 (can refer to such an amino acid position or sequence as a "subject" position or sequence) if (1) the query sequence can be aligned with the subject sequence (e.g., where an alignment indicates that the query sequence and the subject sequence [or a subsequence of the subject sequence] are at least about 30%, 40%, 50%, 60%, 70%, 80%, or 90% identical), and (2) if the query amino acid position directly aligns with (directly lines up against) the subject amino acid position in the alignment of (1). In general, one can align a query amino acid sequence with a subject sequence (SEQ ID NO:62 or a subsequence of SEQ ID NO:62) using any alignment algorithm, tool and/or software described disclosed herein (e.g., BLASTP, ClustalW, ClustalV, Clustal-Omega, EMBOSS) to determine percent identity. Just for further example, one can align a query sequence with a subject sequence herein using the Needleman-Wunsch algorithm (Needleman and Wunsch, *J. Mol. Biol.* 48:443-453, 1970) as implemented in the Needle program of the European Molecular Biology Open Software Suite (EMBOSS [e.g., version 5.0.0 or later], Rice et al., *Trends Genet.* 16:276-277, 2000). The parameters of such an EMBOSS alignment can comprise, for example: gap open penalty of 10, gap extension penalty of 0.5, EBLOSUM62 (EMBOSS version of BLOSUM62) substitution matrix.

The numbering of particular amino acid residues of SEQ ID NO:62 herein (e.g., Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, Ile-1453) is with respect to the full-length amino acid sequence of SEQ ID NO:62. The first amino acid (i.e., position 1, Met-1) of SEQ ID NO:62 is at the start of the signal peptide. Unless otherwise disclosed, substitutions herein are with respect to the full-length amino acid sequence of SEQ ID NO:62.

A "non-native glucosyltransferase" herein ("mutant", "variant", "modified" and like terms can likewise be used to describe such a glucosyltransferase) has at least one amino acid substitution at a position corresponding with a particular amino acid residue of SEQ ID NO:62. Such at least one amino acid substitution typically is in place of the amino acid residue(s) that normally (natively) occurs at the same position in the native counterpart (parent) of the non-native glucosyltransferase (i.e., although SEQ ID NO:62 is used as a reference for position, an amino acid substitution herein is with respect to the native counterpart of a non-native glucosyltransferase) (considered another way, when aligning the sequence of a non-native glucosyltransferase with SEQ ID NO:62, determining whether a substitution exists at a particular position does not depend in-and-of-itself on the respective amino acid residue in SEQ ID NO:62, but rather depends on what amino acid exists at the subject position within the native counterpart of the non-native glucosyltransferase). The amino acid normally occurring at the relevant site in the native counterpart glucosyltransferase often (but not always) is the same as (or conserved with) the particular amino acid residue of SEQ ID NO:62 for which the alignment is made. A non-native glucosyltransferase optionally can have other amino acid changes (mutations, deletions, and/or insertions) relative to its native counterpart sequence.

It may be instructive to illustrate a substitution/alignment herein. SEQ ID NO:12 (GTF 0544) is a truncated form of a *Streptococcus sobrinus* glucosyltransferase. It is noted that Leu-193 of SEQ ID NO:12 corresponds with Leu-373 of SEQ ID NO:62 (alignment not shown). If SEQ ID NO:12 is mutated at position 193 to substitute the Leu residue with a different residue (e.g., Gln), then it can be stated that the position 193-mutated version of SEQ ID NO:12 represents a non-native glucosyltransferase having an amino acid substitution at a position corresponding with Leu-373 of SEQ ID NO:62, for example.

The term "isolated" means a substance (or process) in a form or environment that does not occur in nature. Non-limiting examples of isolated substances include (1) any non-naturally occurring substance (e.g., a non-native glucosyltransferase herein), (2) any substance including, but not limited to, any host cell, enzyme, variant, nucleic acid, protein, peptide, cofactor, or carbohydrate/saccharide that is at least partially removed from one or more or all of the naturally occurring constituents with which it is associated in nature; (3) any substance modified by the hand of man relative to that substance

found in nature (e.g., a non-native glucosyltransferase herein); or (4) any substance modified by increasing the amount of the substance relative to other components with which it is naturally associated. It is believed that the embodiments (e.g., enzymes and reaction compositions) disclosed herein are synthetic/man-made, and/or have properties that are not naturally occurring.

The term "increased" as used herein can refer to a quantity or activity that is at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 11%, 12%, 13%, 14%, 15%, 16%, 17%, 18%, 19%, 20%, 50%, 100%, or 200% more than the quantity or activity for which the increased quantity or activity is being compared. The terms "increased", "elevated", "enhanced", "greater than", "improved" and the like are used interchangeably herein. These terms can be used to characterize the "over-expression" or "up-regulation" of a polynucleotide encoding a protein, for example.

While advances have been made in producing insoluble glucan polymers using glucosyltransferase enzymes, less attention appears to have been drawn to enhancing the molecular weight of insoluble glucan products synthesized by such enzymes. Addressing this technological gap, disclosed herein are glucosyltransferases with modified amino acid sequences that produce higher molecular weight insoluble glucan products.

Certain embodiments of the present disclosure concern a non-native glucosyltransferase comprising at least one amino acid substitution at a position corresponding with amino acid residue(s) Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, wherein the non-native glucosyltransferase synthesizes insoluble alpha-glucan comprising 1,3-linkages, and the molecular weight of this alpha-glucan is higher than the molecular weight of insoluble alpha-glucan synthesized by a second glucosyltransferase that only differs from the non-native glucosyltransferase at the substitution position(s). Thus, in general, mutant glucosyltransferase enzymes are disclosed herein that can synthesize higher molecular weight insoluble alpha-glucan having alpha-1,3 linkages.

A non-native glucosyltransferase herein synthesizes insoluble alpha-glucan comprising 1,3-linkages. In some aspects, at least about 30%, 31%, 32%, 33%, 34%,

35%, 36%, 37%, 38%, 39%, 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 69%, 70%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 99.5%, or 100% of the glycosidic linkages of such an alpha-glucan can be alpha-1,3 linkages. The linkage profile of an alpha-glucan can optionally be characterized as having a range between any two of these values. The other linkages in any of these aspects having 30%-99% alpha-1,3 linkages can be alpha-1,6, and/or not include any alpha-1,4 or alpha-1,2 linkages, for example.

Alpha-glucan in some aspects can have, for example, less than 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, or 0.5% of alpha-1,2 or alpha-1,4 glycosidic linkages. In another embodiment, an alpha-glucan only has alpha-1,3 and optionally alpha-1,6 linkages (i.e., no alpha-1,2 or alpha-1,4 linkages).

Alpha-glucan in some aspects can be linear/unbranched (no branch points). Alternatively, there can be branches in an alpha-glucan herein. For example, an alpha-glucan can have less than about 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% branch points as a percent of the linkages in the polymer.

In certain aspects, an alpha-glucan can have a molecular weight in DP_w or DP_n of at least about 100. For example, the DP_w or DP_n can be about, or at least about,100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, or 1200. The molecular weight of an alpha-glucan can optionally be expressed as a range between any two of these values (e.g., 100-1200, 400-1200, 700-1200, 100-1000, 400-1000, 700-1000). Molecular weight herein can be measured following any suitable method, including those methods disclosed in the present Examples (below) or as disclosed in U.S. Pat. Appl. Publ. Nos. 2017/0002335, 2015/0064748, or 2015/0232819, for example.

In certain aspects, the polydispersity index (PDI, which equals DP_w/DP_n) of alpha-1,3-glucan produced by a non-native glucosyltransferase can be the same as, or similar to, the PDI of alpha-1,3-glucan produced by a second glucosyltransferase (e.g., parent glucosyltransferase) that only differs from the non-native glucosyltransferase at the substitution position(s). The PDI can be at or below about 2.5, 2.4, 2.3, 2.2, 2.1, or 2.0, or range from about 2.0-2.5, 2.0-2.4, 2.0-2.3, 2.0-2.2, 2.1-2.5, 2.1-2.4, 2.1-2.3, or 2.1-2.2, for example. Amino acid substitutions providing a non-native glucosyltransferase

herein can optionally be characterized to allow enzymatic synthesis of alpha-1,3-glucan with preserved PDI, but with increased molecular weight.

An alpha-glucan produced by a non-native glucosyltransferase herein is water-insoluble. Alpha-1,3-glucan is generally insoluble at a DP_w of 8 or 9 and above in neutral (e.g., pH 6-8) aqueous conditions.

Any of the foregoing linkage profiles and/or molecular weight profiles, for example, can be combined herein to appropriately characterize an alpha-glucan product of a non-native glucosyltransferase of the present disclosure. In some aspects, the linkage and/or molecular weight profile of an alpha-glucan product can be as disclosed in any of the following publications, all of which are incorporated herein by reference: U.S. Patent Nos. 7000000 and 8871474, U.S. Patent Appl. Publ. No. 2015/0232819.

A non-native glucosyltransferase, for example, can comprise the amino acid sequence of any glucosyltransferase disclosed in the following publications that is capable of producing insoluble alpha-glucan as presently disclosed, but with the exception that the non-native glucosyltransferase comprises at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62: U.S. Patent Nos. 7000000 and 8871474; and U.S. Patent Appl. Publ. Nos. 2015/0232819 and 2017/0002335, all of which are incorporated herein by reference. In some aspects, such a non-native glucosyltransferase (i) has at least one of the foregoing substitutions, and (ii) comprises an amino acid sequence that is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to the amino acid sequence of the respective counterpart/parent glucosyltransferase not having the at least one substitution.

In some aspects, a non-native glucosyltransferase (i) comprises at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, and (ii) comprises or consists of an amino acid sequence that is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 26, 28, 30, 34, or 59. Certain information

regarding insoluble alpha-glucan products of glucosyltransferases with most of these amino acid sequences is provided in Table 2.

Table 2. GTF Enzymes and Related Alpha-Glucan Products[a]

GTF ID	SEQ ID NO.	Reducing Sugars	Insoluble Product	Linkages		DP_n
				% alpha-1,3	% alpha-1,6	
0874	2	yes	yes	100	0	60
6855	4	yes	yes	100	0	440
2379	6	yes	yes	37	63	310
7527	8	yes	yes	100	0	440
1724	10	yes	yes	100	0	250
0544	12	yes	yes	62	36	980
5926	14	yes	yes	100	0	260
4297	16	yes	yes	31	67	800
5618	18	yes	yes	34	66	1020
2765	20	yes	yes	100	0	280
0427	26	yes	yes	100	0	120
2919	28	yes	yes	100	0	250
2678	30	yes	yes	100	0	390
3929	34	yes	yes	100	0	280

[a] GTF reactions and product analyses were performed as follows. Reactions were prepared comprising sucrose (50 g/L), potassium phosphate buffer (pH 6.5, 20 mM) and a GTF enzyme (2.5% bacterial cell extract by volume; extracts prepared according to U.S. Appl. Publ. No. 2017/0002335, in a manner similar to procedure disclosed in U.S. Patent No. 8871474). After 24-30 hours at 22-25 °C, insoluble product was harvested by centrifugation, washed three times with water, washed once with ethanol, and dried at 50 °C for 24-30 hours. Approximate linkages and DP_n are shown for each insoluble product. Linkages and DP_n were determined by ^{13}C NMR and SEC, respectively.

In some aspects, a non-native glucosyltransferase (i) comprises at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, and (ii) comprises or consists of a glucosyltransferase catalytic domain that is at least about 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to amino acid residues 55-960 of SEQ ID NO:4, residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20. Such a non-native glucosyltransferase, for instance, is believed to be able to produce alpha-glucan that is water-insoluble and comprise at least about 50% (e.g., ≥90% or ≥95%) alpha-1,3 linkages, and optionally further have a

DP_w of at least 100. It is noted that a glucosyltransferase with amino acid positions 54-957 of SEQ ID NO:65 can produce alpha-1,3-glucan with 100% alpha-1,3 linkages and a DP_w of at least 400 (data not shown, refer to Table 6 of U.S. Pat. Appl. Publ. No. 2017/0002335, which is incorporated herein by reference), for example. It is further noted that SEQ ID NOs:65 (GTF 7527), 30 (GTF 2678), 4 (GTF 6855), 28 (GTF 2919), and 20 (GTF 2765) each represent a glucosyltransferase that, compared to its respective wild type counterpart, lacks the signal peptide domain and all or a substantial portion of the variable domain. Thus, each of these glucosyltransferase enzymes has a catalytic domain followed by a glucan-binding domain. The approximate location of catalytic domain sequences in these enzymes is as follows: 7527 (residues 54-957 of SEQ ID NO:65), 2678 (residues 55-960 of SEQ ID NO:30), 6855 (residues 55-960 of SEQ ID NO:4), 2919 (residues 55-960 of SEQ ID NO:28), 2765 (residues 55-960 of SEQ ID NO:20). The amino acid sequences of the catalytic domains (approx.) of GTFs 2678, 6855, 2919 and 2765 have about 94.9%, 99.0%, 95.5% and 96.4% identity, respectively, with the approximate catalytic domain sequence of GTF 7527 (i.e., amino acids 54-957 of SEQ ID NO:65). Each of these particular glucosyltransferases (GTFs 2678, 6855, 2919 and 2765) can produce alpha-1,3-glucan with 100% alpha-1,3 linkages and a DP_w of at least 400 (data not shown, refer to Table 4 of U.S. Pat. Appl. Publ. No. 2017/0002335). Based on this activity, and the relatedness (high percent identity) of the foregoing catalytic domains, it is contemplated that a non-native glucosyltransferase herein having one of the foregoing catalytic domains further with at least one amino acid substitution as presently disclosed can produce insoluble alpha-glucan comprising at least about 50% (e.g., ≥ 90% or ≥95%) alpha-1,3 linkages and a DP_w of at least 100.

In some aspects, a non-native glucosyltransferase (i) comprises at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, and (ii) comprises or consists of an amino acid sequence that is at least about 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 69%, 70%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to SEQ ID NO:62 or a subsequence thereof such as SEQ ID

NO:4 (without start methionine thereof) or positions 55-960 of SEQ ID NO:4 (approximate catalytic domain).

Although it is believed that a non-native glucosyltransferase in certain aspects need only have a catalytic domain, the non-native glucosyltransferase can be comprised within a larger amino acid sequence. For example, a catalytic domain may be linked at its C-terminus to a glucan-binding domain, and/or linked at its N-terminus to a variable domain and/or signal peptide.

Although amino acid substitutions in a non-native glucosyltransferase are generally disclosed herein with respect to the corresponding positions in SEQ ID NO:62, such substitutions can alternatively be stated simply with respect to its position number in the sequence of the non-native glucosyltransferase itself, as convenience may dictate.

Still further examples of non-native glucosyltransferases can be any as disclosed herein and that include 1-300 (or any integer there between [e.g., 10, 15, 20, 25, 30, 35, 40, 45, or 50]) residues on the N-terminus and/or C-terminus. Such additional residues may be from a corresponding wild type sequence from which the glucosyltransferase enzyme is derived, or may be a heterologous sequence such as an epitope tag (at either N- or C-terminus) or a heterologous signal peptide (at N-terminus), for example. A non-native glucosyltransferase herein typically lacks an N-terminal signal peptide; such an enzyme can optionally be characterized as being mature if its signal peptide was removed during a secretion process.

A non-native glucosyltransferase herein can be derived from any microbial source, for example, such as bacteria. Examples of bacterial glucosyltransferases are those derived from a *Streptococcus* species, *Leuconostoc* species, or *Lactobacillus* species. Examples of *Streptococcus* species include *S. salivarius*, *S. sobrinus*, *S. dentirousetti*, *S. downei*, *S. mutans*, *S. oralis*, *S. gallolyticus* and *S. sanguinis*. Examples of *Leuconostoc* species include *L. mesenteroides*, *L. amelibiosum*, *L. argentinum*, *L. carnosum*, *L. citreum*, *L. cremoris*, *L. dextranicum* and *L. fructosum*. Examples of *Lactobacillus* species include *L. acidophilus*, *L. delbrueckii*, *L. helveticus*, *L. salivarius*, *L. casei*, *L. curvatus*, *L. plantarum*, *L. sakei*, *L. brevis*, *L. buchneri*, *L. fermentum* and *L. reuteri*.

A non-native glucosyltransferase herein can be prepared by fermentation of an appropriately engineered microbial strain, for example. Recombinant enzyme

production by fermentation is well known in the art using microbial species such as *E. coli*, *Bacillus* strains (e.g., *B. subtilis*), *Ralstonia eutropha*, *Pseudomonas fluorescens*, *Saccharomyces cerevisiae*, *Pichia pastoris*, *Hansenula polymorpha*, and species of *Aspergillus* (e.g., *A. awamori*) and *Trichoderma* (e.g., *T. reesei*) (e.g., see Adrio and Demain, *Biomolecules* 4:117-139, 2014, which is incorporated herein by reference). A nucleotide sequence encoding a non-native glucosyltransferase amino acid sequence is typically linked to a heterologous promoter sequence to create an expression cassette for the enzyme, and/or is codon-optimized accordingly. Such an expression cassette may be incorporated in a suitable plasmid or integrated into the microbial host chromosome, using methods well known in the art. The expression cassette may include a transcriptional terminator nucleotide sequence following the amino acid coding sequence. The expression cassette may also include, between the promoter sequence and glucosyltransferase amino acid coding sequence, a nucleotide sequence encoding a signal peptide (e.g., heterologous signal peptide) that is designed for direct secretion of the glucosyltransferase enzyme. At the end of fermentation, cells may be ruptured accordingly (generally when a signal peptide for secretion is not employed) and the glucosyltransferase enzyme can be isolated using methods such as precipitation, filtration, and/or concentration. Alternatively, a lysate or extract comprising a glucosyltransferase can be used without further isolation. If the glucosyltransferase was secreted (i.e., it is present in the fermentation broth), it can optionally be used as isolated from, or as comprised in, the fermentation broth. The activity of a glucosyltransferase enzyme can be confirmed by biochemical assay, such as measuring its conversion of sucrose to glucan polymer.

A non-native glucosyltransferase herein can comprise at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Asn-531 of SEQ ID NO:62 can be with a Gly, Leu, or Met residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Arg-534 of SEQ ID NO:62 can be with a Lys, Gly, Ile, Leu, or Met residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue

Thr-563 of SEQ ID NO:62 can be with an Ala residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Glu-567 of SEQ ID NO:62 can be with a Gln residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Val-586 of SEQ ID NO:62 can be with a Thr residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Gln-588 of SEQ ID NO:62 can be with a Leu residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Ile-591 of SEQ ID NO:62 can be with a Val, Lys, or Arg residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Lys-593 of SEQ ID NO:62 can be with a Met residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Ile-608 of SEQ ID NO:62 can be with a Tyr residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Ala-610 of SEQ ID NO:62 can be with a Cys or Thr residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Leu-661 of SEQ ID NO:62 can be with a Pro residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Arg-722 of SEQ ID NO:62 can be with a His or Asn residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Thr-728 of SEQ ID NO:62 can be with a Ser residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Met-732 of SEQ ID NO:62 can be with a Leu residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Arg-741 of SEQ ID NO:62 can be with a Ser, Ala, Pro, Gln, or Thr residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Asn-743 of SEQ ID NO:62 can be with a Ser, Thr, or Asp residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Ala-777 of SEQ ID NO:62 can be with an Asn residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Tyr-848 of SEQ ID NO:62 can be with a Glu residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Ile-1453 of SEQ ID NO:62 can be with a Gly or Met residue.

A non-native glucosyltransferase in certain embodiments can comprise, in addition to any of the above-listed substitution(s), or alternatively to the above-listed substitution(s), at least one amino acid substitution at a position corresponding with

amino acid residue Phe-424, Asn-475, Trp-511, Arg-609, Asn-614, or Asn-1214 of SEQ ID NO:62. Such a non-native glucosyltransferase can be based on any of the above-disclosed glucosyltransferase/catalytic domain amino acid sequences (and percent identities thereto), for example. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Phe-424 of SEQ ID NO:62 can be with an Ala, Val, Leu, Glu, or Gln residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Asn-475 of SEQ ID NO:62 can be with a Gln or Ser residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Trp-511 of SEQ ID NO:62 can be with a Tyr residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Arg-609 of SEQ ID NO:62 can be with a His residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Asn-614 of SEQ ID NO:62 can be with a Phe residue. In some aspects, the amino acid substitution at a position corresponding with amino acid residue Asn-1214 of SEQ ID NO:62 can be with a Leu or Ile residue.

A non-native glucosyltransferase herein can comprise one, two, three, four, five, six, seven, eight, nine, or more of the presently disclosed amino acid substitutions, for instance. A non-native glucosyltransferase in some aspects can comprise at least one amino acid substitution at a position corresponding with amino acid residue Gln-588, Arg-741, or Arg-722 of SEQ ID NO:62. In some aspects, a non-native glucosyltransferase can comprise substitutions at one of these sites (e.g., Q588), two of these sites (e.g., Q588 and R741), or all three of these sites (Q588, R741 and Arg-722). Such an amino acid substitution(s) can be any of those as disclosed above, for example (e.g., Q588L, R741S, R722H). In some aspects, a non-native glucosyltransferase can comprise substitutions at one, two or all three of these sites, and have a total number of the presently disclosed substitutions as disclosed above.

Suitable substitution sites, and examples of particular substitutions at these sites, can include those as listed in Table 3 in Example 1 (below) that are associated with an increase in the molecular weight (DP_w) of alpha-1,3-glucan product by at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, or 60%, for example. In some alternative aspects, a substitution site, and example(s) of particular substitution(s) at such site, can include any of those as listed in Table 3 in Example 1 (below) that are

associated with a feature that endows a benefit to a glucosyltransferase that synthesizes insoluble alpha-1,3-glucan. The foregoing substitutions as listed in Table 3 are as they correspond with the listed residue position number in SEQ ID NO:62. In some aspects, one or more substitutions are conserved or non-conserved substitutions; such conservation (or not) can be, for instance, with respect to the amino acid that occurs in the native glucosyltransferase from which the non-native glucosyltransferase is derived.

A non-native glucosyltransferase herein can synthesize insoluble alpha-glucan comprising 1,3-linkages with a molecular weight higher than the molecular weight of insoluble alpha-glucan comprising 1,3-linkages synthesized by a second glucosyltransferase (or, simply, "another" glucosyltransferase) (e.g., parent glucosyltransferase) that only differs from the non-native glucosyltransferase at the substitution position(s). A second glucosyltransferase herein, for example, can be comprised of all of, or mostly, native amino acid sequence. Thus, while a second glucosyltransferase herein can be a native glucosyltransferase in some aspects, it can be a prior-modified glucosyltransferase in other aspects (e.g., a glucosyltransferase with one or more other amino acid substitutions differing from the substitution[s] of the present disclosure). In some embodiments, a second glucosyltransferase to which a non-native glucosyltransferase is compared has a native amino acid residue(s) at the substitution position(s). Determining whether an amino acid residue is native can be done by comparing the second glucosyltransferase amino acid sequence to the native/wild type glucosyltransferase amino acid sequence from which the second glucosyltransferase is derived.

In some aspects, a non-native glucosyltransferase herein can synthesize insoluble alpha-glucan with a molecular weight that is at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, or 60% higher than the molecular weight of insoluble alpha-glucan synthesized by a second glucosyltransferase. Such a determination can be made with respect to any glucan synthesis reaction/process as disclosed herein (e.g., taking into account initial sucrose conc., temperature, pH, and/or reaction time), and using any suitable measurement technique (e.g., SEC). Typically, a comparison between non-native and second glucosyltransferases herein can be made under identical or similar reaction conditions. The molecular weight of insoluble alpha-glucan can be expressed as DP_w, for example.

Some embodiments disclosed herein concern a polynucleotide comprising a nucleotide sequence that encodes a non-native glucosyltransferase as presently disclosed (e.g., a non-native glucosyltransferase comprising at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62). Optionally, one or more regulatory sequences are operably linked to the nucleotide sequence, and preferably a promoter sequence is included as a regulatory sequence.

A polynucleotide comprising a nucleotide sequence encoding a non-native glucosyltransferase herein can be a vector or construct useful for transferring a nucleotide sequence into a cell, for example. Examples of a suitable vector/construct can be selected from a plasmid, yeast artificial chromosome (YAC), cosmid, phagemid, bacterial artificial chromosome (BAC), virus, or linear DNA (e.g., linear PCR product). A polynucleotide sequence in some aspects can be capable of existing transiently (i.e., not integrated into the genome) or stably (i.e., integrated into the genome) in a cell. A polynucleotide sequence in some aspects can comprise, or lack, one or more suitable marker sequences (e.g., selection or phenotype marker).

A polynucleotide sequence in certain embodiments can comprise one or more regulatory sequences operably linked to the nucleotide sequence encoding a non-native glucosyltransferase. For example, a nucleotide sequence encoding a non-native glucosyltransferase may be in operable linkage with a promoter sequence (e.g., a heterologous promoter). A promoter sequence can be suitable for expression in a cell (e.g., bacterial cell such as *E. coli* or *Bacillus*; eukaryotic cell such as a fungus, yeast, insect, or mammalian cell) or in an *in vitro* protein expression system, for example. Examples of other suitable regulatory sequences include transcription terminator sequences.

Some aspects herein are drawn to a cell comprising a polynucleotide sequence as presently disclosed; such a cell can be any type disclosed herein (e.g., bacterial cell such as *E. coli* or *Bacillus*; eukaryotic cell such as a fungus, yeast, insect, or mammalian cell). A cell can optionally express a non-native glucosyltransferase encoded by the polynucleotide sequence. In some aspects, the polynucleotide sequence exists

transiently (i.e., not integrated into the genome) or stably (i.e., integrated into the genome) in the cell.

Some embodiments disclosed herein concern reaction compositions comprising water, sucrose, and one or more non-native glucosyltransferases herein (e.g., a non-native glucosyltransferase comprising at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62). Such a reaction composition produces, at least, alpha-glucan comprising 1,3-linkages as disclosed.

The temperature of a reaction composition herein can be controlled, if desired, and can be about 5-50 °C, 20-40 °C, 30-40 °C, 20-30 °C, 20-25 °C, 20 °C, 25 °C, 30 °C, 35 °C, or 40 °C, for example.

The initial concentration of sucrose in a reaction composition herein can be about 20-400 g/L, 75-175 g/L, or 50-150 g/L, for example. In some aspects, the initial sucrose concentration is at least about 50, 75, 100, 150 or 200 g/L, or is about 50-600 g/L, 100-500 g/L, 50-100 g/L, 100-200 g/L, 150-450 g/L, 200-450 g/L, or 250-600 g/L. "Initial concentration of sucrose" refers to the sucrose concentration in a reaction composition just after all the reaction components have been added/combined (e.g., at least water, sucrose, non-native glucosyltransferase enzyme).

The pH of a reaction composition in certain embodiments can be about 4.0-9.0, 4.0-8.5, 4.0-8.0, 5.0-8.0, 5.5-7.5, or 5.5-6.5. In some aspects, the pH can be about 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, or 8.0. The pH can be adjusted or controlled by the addition or incorporation of a suitable buffer, including but not limited to: phosphate, tris, citrate, or a combination thereof. The buffer concentration in a reaction composition herein can be about 0.1-300 mM, 0.1-100 mM, 10-100 mM, 10 mM, 20 mM, or 50 mM, for example.

A reaction composition can be contained within any vessel (e.g., an inert vessel/container) suitable for applying one or more of the reaction conditions disclosed herein. An inert vessel in some aspects can be of stainless steel, plastic, or glass (or comprise two or more of these components) and be of a size suitable to contain a particular reaction. For example, the volume/capacity of an inert vessel (and/or the volume of a reaction composition herein), can be about, or at least about, 1, 10, 50, 100,

500, 1000, 2500, 5000, 10000, 12500, 15000, or 20000 liters. An inert vessel can optionally be equipped with a stirring device.

A reaction composition herein can contain one, two, or more glucosyltransferase enzymes, for example, just as long that at least one of the enzymes is a non-native glucosyltransferase as presently disclosed. In some embodiments, only one or two glucosyltransferase enzymes is/are comprised in a reaction composition. A glucosyltransferase reaction herein can be, and typically is, cell-free (e.g., no whole cells present).

Any of the features disclosed herein (e.g., above and in the below Examples) regarding a reaction composition can characterize appropriate aspects of a glucan production method herein, and *vice versa*.

The present disclosure also concerns a method for producing insoluble alpha-glucan, the method comprising: (a) contacting at least water, sucrose, and at least one non-native glucosyltransferase as disclosed herein that produces insoluble alpha-glucan, whereby insoluble alpha-glucan is produced; and b) optionally, isolating the insoluble alpha-glucan produced in step (a). Conducting such a method, which can optionally be characterized as a glucan synthesis method, is typically also performed when conducting a reaction composition herein.

A glucan synthesis method as presently disclosed comprises contacting at least water, sucrose, and a non-native glucosyltransferase herein that produces insoluble alpha-glucan. These and optionally other reagents can be added altogether or in any order as discussed below. This step can optionally be characterized as providing a reaction composition comprising water, sucrose and a non-native glucosyltransferase enzyme that synthesizes insoluble alpha-glucan. The contacting step herein can be performed in any number of ways. For example, the desired amount of sucrose can first be dissolved in water (optionally, other components may also be added at this stage of preparation, such as buffer components), followed by addition of glucosyltransferase enzyme. The solution may be kept still, or agitated via stirring or orbital shaking, for example. A glucan synthesis method can be performed by batch, fed-batch, continuous mode, or by any variation of these modes.

Completion of a reaction in certain embodiments can be determined visually (e.g., no more accumulation of insoluble glucan), and/or by measuring the amount of sucrose

left in the solution (residual sucrose), where a percent sucrose consumption of at least about 90%, 95%, or 99% can indicate reaction completion. A reaction of the disclosed process can be conducted for about 1 hour to about 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 36, 48, 60, 72, 96, 120, 144, or 168 hours, for example.

The molecular weight of insoluble alpha-glucan produced in some aspects of a glucan synthesis method herein can be at least about 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, or 60% higher than the molecular weight of insoluble alpha-glucan synthesized by a second glucosyltransferase. Such molecular weight enhancement in some aspects is achieved in a reaction conducted for about 16-24 hours (e.g., ~20 hours).

Insoluble alpha-glucan produced in a method herein can optionally be isolated. In certain embodiments, isolating insoluble alpha-glucan can include at least conducting a step of centrifugation and/or filtration. Isolation can optionally further comprise washing alpha-glucan one, two, or more times with water or other aqueous liquid, and/or drying the alpha-glucan product.

An isolated alpha-glucan product herein, as provided in a dry form, can comprise no more than 2.0, 1.5, 1.0, 0.5, 0.25, 0.10, 0.05, or 0.01 wt% water, for example. In some aspects, an alpha-glucan product is provided in an amount of at least 1 gram (e.g., at least about 2.5, 5, 10, 25, 50, 100, 250, 500, 750, 1000, 2500, 5000, 7500, 10000, 25000, 50000, or 100000 g); such an amount can be a dry amount, for example.

Any of the disclosed conditions for synthesizing an alpha-glucan, such as the foregoing or those described in the below Examples, can be applied to practicing a reaction composition as presently disclosed (and *vice versa*), and/or used to characterize features/activity of a non-native glucosyltransferase, accordingly.

The present disclosure also concerns a method of preparing a polynucleotide sequence encoding a non-native glucosyltransferase herein. This method comprises:

(a) identifying a polynucleotide sequence encoding a parent glucosyltransferase that (i) comprises an amino acid sequence that is at least about 40% identical to SEQ ID NO:4 or positions 55-960 of SEQ ID NO:4, and (ii) synthesizes insoluble alpha-glucan comprising 1,3-linkages; and

(b) modifying the polynucleotide sequence identified in step (a) to substitute at least one amino acid of the parent glucosyltransferase at a position corresponding with

amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, thereby providing a polynucleotide sequence encoding a non-native glucosyltransferase that synthesizes insoluble alpha-glucan with a molecular weight that is higher than the molecular weight of insoluble alpha-glucan synthesized by the parent glucosyltransferase.

Such a method can optionally further comprise using a polynucleotide prepared in this manner in a method of expressing the non-native glucosyltransferase encoded by the polynucleotide. Such an expression method can follow any heterologous protein expression method as known in the art, for example. The present method of preparing a polynucleotide can optionally alternatively be characterized as a method of increasing the product molecular weight of a glucosyltransferase.

Identification step (a) herein can, in some instances, comprise identifying an amino acid sequence of a parent glucosyltransferase enzyme. A polynucleotide sequence could be determined from this amino acid sequence according to the genetic code (codons), such as the genetic code used in the species from which the parent glucosyltransferase was identified.

Identifying a polynucleotide encoding a parent glucosyltransferase herein can be performed (a) *in silico*, (b) with a method comprising a nucleic acid hybridization step, (c) with a method comprising a protein sequencing step, and/or (d) with a method comprising a protein binding step, for example.

Regarding *in silico* detection, the amino acid sequences of candidate parent glucosyltransferase enzymes (and/or nucleotide sequences encoding such glucosyltransferase enzymes) stored in a computer or database (e.g., public databases such as GENBANK, EMBL, REFSEQ, GENEPEPT, SWISS-PROT, PIR, PDB) can be reviewed *in silico* to identify a glucosyltransferase enzyme comprising an amino acid sequence with a percent sequence identity as described above for a parent glucosyltransferase. Such review could comprise using any means known in the art such as through use of an alignment algorithm or software as described above (e.g., BLASTN, BLASTP, ClustalW, ClustalV, Clustal-Omega, EMBOSS).

Identifying a parent glucosyltransferase as disclosed above can optionally be performed via a method comprising a nucleic acid hybridization step. Such a method can comprise using DNA hybridization (e.g., Southern blot, dot blot), RNA hybridization

(e.g., northern blot), or any other method that has a nucleic acid hybridization step (e.g., DNA sequencing, PCR, RT-PCR, all of which may comprise hybridization of an oligonucleotide), for example. A polynucleotide sequence encoding SEQ ID NO:4 or a subsequence thereof (e.g., positions 55-960 of SEQ ID NO:4) can be used as a probe, for example, in such a hybridization. Conditions and parameters for carrying out hybridization methods in general are well known and disclosed, for example, in Sambrook J, Fritsch EF and Maniatis T, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1989); Silhavy TJ, Bennan ML and Enquist LW, Experiments with Gene Fusions, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY (1984); Ausubel FM et al., Current Protocols in Molecular Biology, published by Greene Publishing Assoc. and Wiley-Interscience, Hoboken, NJ (1987); and Innis MA, Gelfand DH, Sninsky JJ and White TJ (Editors), PCR Protocols: A Guide to Methods and Applications, Academic Press, Inc., San Diego, CA (1990).

Identifying a parent glucosyltransferase as disclosed above can optionally be performed via a method comprising a protein sequencing step. Such a protein sequencing step can comprise one or more procedures such as N-terminal amino acid analysis, C-terminal amino acid analysis, Edman degradation, or mass spectrometry, for example.

Identifying a parent glucosyltransferase as disclosed above can optionally be performed via a method comprising a protein binding step. Such a protein binding step can be performed using an antibody that binds to a motif or epitope within SEQ ID NO:4 (e.g., within positions 55-960 of SEQ ID NO:4), for example.

A polynucleotide identified in step (a) (i.e., before its modification in step [b]) can, in some aspects, encode a glucosyltransferase comprising an amino acid sequence that is identical to, or at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to, the amino acid sequence of any glucosyltransferase disclosed in Table 1. An alpha-glucan as produced by such a glucosyltransferase can be as disclosed herein, for example.

A method of preparing a polynucleotide sequence encoding a non-native glucosyltransferase herein comprises step (b) of modifying the polynucleotide sequence (encoding a parent glucosyltransferase) identified in step (a). Such modification substitutes at least one amino acid of the parent glucosyltransferase at a position

corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62. The non-native glucosyltransferase (encoded by the modified polynucleotide sequence) resulting from such one or more substitutions can be optionally be characterized as a "child glucosyltransferase" herein.

A parent glucosyltransferase enzyme herein can comprise an amino acid sequence that is at least about 40%, 41%, 42%, 43%, 44%, 45%, 46%, 47%, 48%, 49%, 50%, 51%, 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 61%, 62%, 63%, 64%, 65%, 66%, 67%, 69%, 70%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to SEQ ID NO:4 (optionally without start methionine thereof) or positions 55-960 of SEQ ID NO:4 (approximate catalytic domain), for example. It is noted simply for reference purposes that SEQ ID NO:4 without its start methionine is a subsequence of SEQ ID NO:62.

A suitable modification of a polynucleotide in step (b) can be made following any DNA manipulation technique known in the art. Modifying step (b) can optionally be performed *in silico*, followed by synthesis of the polynucleotide sequence encoding a non-native glucosyltransferase. For example, a polynucleotide sequence identified in step (a) can be manipulated *in silico* using a suitable sequence manipulation program/software (e.g., VECTOR NTI, Life Technologies, Carlsbad, CA; DNAStrider; DNASTAR, Madison, WI). Following such virtual manipulation, the modified polynucleotide sequence can be artificially synthesized by any suitable technique (e.g., annealing-based connection of oligonucleotides, or any technique disclosed in Hughes et al., *Methods Enzymol.* 498:277-309, which is incorporated herein by reference). It should be appreciated that the foregoing methodology is not believed to necessarily rely on having a pre-existing polynucleotide (encoding a parent glucosyltransferase) in hand.

Modifying step (b) can optionally be performed using a physical copy of a polynucleotide sequence identified in step (a) encoding a parent glucosyltransferase. As an example, such a polynucleotide can serve as a template for amplification using primers designed in a manner such that the amplified product encodes a non-native glucosyltransferase herein (e.g., refer to Innis et al., *ibid.*).

The amino acid substitutions in this method can be any of those combinations of substitutions as disclosed herein. Essentially any non-native glucosyltransferase as presently disclosed can be encoded by a polynucleotide as prepared by this method, for instance.

Non-limiting examples of compositions and methods disclosed herein include:

1. A non-native glucosyltransferase comprising at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, wherein the non-native glucosyltransferase synthesizes insoluble alpha-glucan comprising 1,3-linkages, and the molecular weight of the insoluble alpha-glucan is higher than the molecular weight of insoluble alpha-glucan synthesized by a second glucosyltransferase that only differs from the non-native glucosyltransferase at the substitution position(s).

2. The non-native glucosyltransferase of embodiment 1, wherein: (i) the amino acid substitution at the position corresponding with amino acid residue Asn-531 is with a Gly, Leu, or Met residue; (ii) the amino acid substitution at the position corresponding with amino acid residue Arg-534 is with a Lys, Gly, Ile, Leu, or Met residue; (iii) the amino acid substitution at the position corresponding with amino acid residue Thr-563 is with an Ala residue; (iv) the amino acid substitution at the position corresponding with amino acid residue Glu-567 is with a Gln residue; (v) the amino acid substitution at the position corresponding with amino acid residue Val-586 is with a Thr residue; (vi) the amino acid substitution at the position corresponding with amino acid residue Gln-588 is with a Leu residue; (vii) the amino acid substitution at the position corresponding with amino acid residue Ile-591 is with a Val, Lys, or Arg residue; (viii) the amino acid substitution at the position corresponding with amino acid residue Lys-593 is with a Met residue; (ix) the amino acid substitution at the position corresponding with amino acid residue Ile-608 is with a Tyr residue; (x) the amino acid substitution at the position corresponding with amino acid residue Ala-610 is with a Cys or Thr residue; (xi) the amino acid substitution at the position corresponding with amino acid residue Leu-661 is with a Pro residue; (xii) the amino acid substitution at the position corresponding with amino acid residue Arg-722 is with a His or Asn residue; (xiii) the amino acid substitution at the position corresponding with amino acid residue Thr-728 is with a Ser residue; (xiv) the amino

acid substitution at the position corresponding with amino acid residue Met-732 is with a Leu residue; (xv) the amino acid substitution at the position corresponding with amino acid residue Arg-741 is with a Ser, Ala, Pro, Gln, or Thr residue; (xvi) the amino acid substitution at the position corresponding with amino acid residue Asn-743 is with a Ser, Thr, or Asp residue; (xvii) the amino acid substitution at the position corresponding with amino acid residue Ala-777 is with an Asn residue; (xviii) the amino acid substitution at the position corresponding with amino acid residue Tyr-848 is with a Glu residue; and/or (xix) the amino acid substitution at the position corresponding with amino acid residue Ile-1453 is with a Gly or Met residue.

3. The non-native glucosyltransferase of embodiment 1 or 2, comprising at least one amino acid substitution at a position corresponding with amino acid residue Gln-588, Arg-741, or Arg-722 of SEQ ID NO:62; optionally wherein: (i) the amino acid substitution at the position corresponding with amino acid residue Gln-588 is with a Leu residue; (ii) the amino acid substitution at the position corresponding with amino acid residue Arg-741 is with a Ser residue; and/or (iii) the amino acid substitution at the position corresponding with amino acid residue Arg-722 is with a His residue.

4. The non-native glucosyltransferase of embodiment 3, comprising two or more amino acid substitutions at positions corresponding with amino acid residues Gln-588, Arg-741, or Arg-722 of SEQ ID NO:62.

5. The non-native glucosyltransferase of embodiment 1, 2, 3, or 4, wherein the insoluble alpha-glucan produced by the non-native glucosyltransferase comprises at least about 50% alpha-1,3 linkages, and optionally wherein it has a weight average degree of polymerization (DP_w) of at least 100.

6. The non-native glucosyltransferase of embodiment 1, 2, 3, 4, or 5, wherein the insoluble alpha-glucan produced by the non-native glucosyltransferase has a DP_w of at least 650.

7. The non-native glucosyltransferase of embodiment 5 or 6, comprising a catalytic domain that is at least about 90% identical to residues 55-960 of SEQ ID NO:4, residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20.

8. The non-native glucosyltransferase of embodiment 5, 6, or 7, comprising an amino acid sequence that is at least about 90% identical to SEQ ID NO:4, SEQ ID NO:65, SEQ ID NO:30, SEQ ID NO:28, or SEQ ID NO:20.

9. The non-native glucosyltransferase of embodiment 1, 2, 3, 4, 5, 6, 7, or 8, wherein the insoluble alpha-glucan produced by the non-native glucosyltransferase comprises at least about 90% (or at least 95%) alpha-1,3 linkages.

10. The non-native glucosyltransferase of embodiment 1, 2, 3, 4, 5, 6, 7, 8, or 9, wherein the molecular weight of the insoluble alpha-glucan produced by the non-native glucosyltransferase is at least about 10% higher than the molecular weight of insoluble alpha-glucan synthesized by the second glucosyltransferase.

11. A polynucleotide comprising a nucleotide sequence encoding a non-native glucosyltransferase according to any one of embodiments 1-10, optionally wherein one or more regulatory sequences are operably linked to the nucleotide sequence, and preferably wherein the one or more regulatory sequences include a promoter sequence.

12. A reaction composition comprising water, sucrose, and a non-native glucosyltransferase according to any one of embodiments 1-10.

13. A method of producing insoluble alpha-glucan comprising: (a) contacting at least water, sucrose, and a non-native glucosyltransferase enzyme according to any one of embodiments 1-10, whereby insoluble alpha-glucan is produced; and (b) optionally, isolating the insoluble alpha-glucan produced in step (a).

14. A method of preparing a polynucleotide sequence encoding a non-native glucosyltransferase (e.g., of any one of embodiments 1-10), the method comprising: (a) identifying a polynucleotide sequence encoding a parent glucosyltransferase that (i) comprises an amino acid sequence that is at least about 40% identical to SEQ ID NO:4 or positions 55-960 of SEQ ID NO:4, and (ii) synthesizes insoluble alpha-glucan comprising 1,3-linkages; and (b) modifying the polynucleotide sequence identified in step (a) to substitute at least one amino acid of the parent glucosyltransferase at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62, thereby providing a polynucleotide sequence encoding a non-native glucosyltransferase that synthesizes insoluble alpha-glucan with a molecular weight that is higher than the molecular weight of insoluble alpha-glucan synthesized by the parent glucosyltransferase.

15. The method of embodiment 14, wherein the identifying step is performed: (a) *in silico*, (b) with a method comprising a nucleic acid hybridization step, (c) with a method

comprising a protein sequencing step, and/or (d) with a method comprising a protein binding step; and/or wherein the modifying step is performed: (e) *in silico*, followed by synthesis of the polynucleotide sequence encoding the non-native glucosyltransferase enzyme, or (f) using a physical copy of the polynucleotide sequence encoding the parent glucosyltransferase.

EXAMPLES

The present disclosure is further exemplified in the following Examples. It should be understood that these Examples, while indicating certain preferred aspects herein, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of the disclosed embodiments, and without departing from the spirit and scope thereof, can make various changes and modifications to adapt the disclosed embodiments to various uses and conditions.

EXAMPLE 1

Analysis of Amino Acid Sites Affecting Glucosyltransferase Alpha-Glucan Product Molecular Weight

This Example describes screening for glucosyltransferase variants that produce alpha-glucan with increased molecular weight. This screening was performed using a site evaluation library (SEL).

The amino acid sequence of the glucosyltransferase used to prepare amino acid substitutions in this Example was SEQ ID NO:4 (GTF 6855), which essentially is an N-terminally truncated (signal peptide and variable region removed) version of the full-length wild type glucosyltransferase (represented by SEQ ID NO:62) from *Streptococcus salivarius* SK126 (see Table 1). Substitutions made in SEQ ID NO:4 can be characterized as substituting for native amino acid residues, as each amino acid residue/position of SEQ ID NO:4 (apart from the Met-1 residue of SEQ ID NO:4) corresponds accordingly with an amino acid residue/position within SEQ ID NO:62. In reactions comprising at least sucrose and water, the glucosyltransferase of SEQ ID NO:4 typically produces alpha-glucan having about 100% alpha-1,3 linkages and a DP_w of 400 or greater (e.g., refer to U.S. Patent Nos. 8871474 and 9169506, and U.S. Pat. Appl. Publ. No. 2017/0002336, which are incorporated herein by reference). This alpha-

glucan product, which is insoluble, can be isolated following enzymatic synthesis via filtration, for example.

To summarize this Example, GTF 6855 variants (each with a single amino acid substitution) from SELs were each bacterially expressed, purified, and normalized to a concentration of 100 ppm. Each enzyme preparation was then screened (in triplicate) using sucrose as substrate in alpha-1,3-glucan synthesis reactions. Alpha-1,3-glucan polymer obtained in each reaction was analyzed for length (DP_w) using size exclusion chromatography (SEC).

Plasmids for individually expressing various single amino acid-substituted variants of GTF 6855 (SEQ ID NO:4) in a *Bacillus subtilis* host were prepared. Such plasmids were prepared as follows. A DNA expression cassette having (operably linked in 5'-to-3' order) the *B. subtilis* aprE promoter, a codon-optimized sequence encoding SEQ ID NO:4 (GTF 6855), and a BPN' terminator was synthesized. This expression cassette was cloned into the pHYT replicating shuttle vector (forming pHYT-GTF6855) and transformed into *B. subtilis* CBS12-1. The pHYT vector was derived from pHY300PLK (Takara) by adding a terminator sequence (SEQ ID NO:67) after the tetracycline resistance gene using the *BstEII* and *EcoRI* sites. The *HindIII* site in pHY300PLK had been removed by cloning a linker sequence (not shown) into the *BamHI* and *HindIII* sites. The pHYT-GTF6855 plasmid was amplified and used for generating SELs. The resulting plasmids encoding single-amino acid substituted GTFs were sequenced to verify each substitution.

To produce GTF 6855 (SEQ ID NO:4) and single amino acid-substituted variants thereof, *B. subtilis* individually transformed with pHYT-GTF6855 or mutated versions thereof were cultivated in Tryptone Soya Broth (Oxoid Ltd., UK) and Grant's II medium. Heart infusion agar plates (Difco Laboratories, MI) were used to select transformants. Plasmid integrity was maintained by the addition of 25 μg/mL tetracycline. Each GTF targeted for expression was detected in the growth medium after incubation for about 6 hours at 37 °C. After centrifugation and filtration, culture supernatants with expressed GTF were obtained. GTF enzyme present in the supernatant was purified to apparent homogeneity by affinity chromatography using washed (2x MILLIQ 1 x 25 mM NaH_2PO_4 pH 5.7 with intermediate centrifugation steps 100 x *g*) SUPERDEX 200 resin (GE Healthcare). Each GTF was eluted with a 15% solution of Dextran T1 (Pharmacosmos) in 25 mM NaH_2PO_4 pH 5.7 by centrifugation 100 x *g*. Each purified GTF was dialyzed

against 25 mM NaH_2PO_4 pH 5.7 buffer (at least 100x) using a Harvard Apparatus 96-well DISPODIALYZER (10000-Dalton MWCO).

After dialysis, GTF enzyme concentration was determined by OD280 using purified GTF 6855 as a standard. Normalization of each purified GTF to 100 ppm was achieved by diluting appropriately with 25 mM NaH_2PO_4 pH 5.7. Protein concentration for each sample was confirmed using an AGILENT 1200 (Agilent Technologies) HPLC equipped with an AGILENT BIO SEC3 guard-column column (3 µm 100Å (4.6x50 mm). Five (5) µL of sample was injected onto the column for each determination. Compounds were eluted with isocratic flow of 25 mM KH_2PO_4 pH 6.8 + 0.1 M NaCl for 1.3 min at 0.5 mL/min flow rate.

Each GTF (GTF 6855 and each variant thereof) was entered into a reaction with sucrose to produce alpha-glucan. Each reaction was performed as follows: 37.5 µL of 100 ppm enzyme sample (ppm based on a BSA calibration curve) was added to 262.5 µL of 86 g/L sucrose (75 g/L final) in 20 mM Na_2HPO_4 /NaH_2PO_4 pH 5.7 and incubated overnight (about 20 hours) at 30 °C. After this incubation, each reaction was quenched by incubation for 1 hour at 80 °C. Appropriate analyses indicated that each variant enzyme listed in Table 3 below was able to perform a glucosyltransferase reaction producing alpha-1,3-glucan (data not shown).

A 100-µL aliquot of each quenched reaction was diluted 20X in DMSO/2% LiCl and filtered via centrifugation through a 0.2-µm PALL GHP membrane (4000*g* x 30 min x 30 °C) in preparation for SEC analysis. Alpha-1,3-glucan polymer size was approximated using a WATERS APC-SEC system equipped with a WATERS ACQUITY APC-XT 450-Å 2.5-µm 4.6x30 mm column. The column was held at 55 °C with a flow-rate of 0.65 mL/min of DMSO/0.25% LiCl mobile-phase. Dextran analytical standards with molecular weights of 80, 165, 325 and 750 kD were used to approximate alpha-1,3-glucan DP_w's of 280, 480, 660 and 880, respectively, via elution peak apexes.

The molecular weight (DP_w) and polydispersity index (PDI) of alpha-1,3-glucan produced in each reaction (~20 hours) as measured via the above methodology are provided in Table 3.

Table 3. Insoluble Alpha-1,3-Glucan Produced by GTF 6855 (SEQ ID NO:4) and Single Amino Acid-Substituted Variants thereof

Plate 1[a]		
	Alpha-1,3-Glucan	
GTF	DP_w	PDI
6855[b]	611	2.2
6855[b]	635	2.1
6855[b]	633	2.1
6855[b]	626	2.1
V186A[c]	589	2.0
V186M	580	2.0
E194C	580	2.3
L434N	613	2.2
A472C	530	1.8
A472S	374	2.2
A510E	654	2.0
A510I	621	2.1
A510V	655	2.1
M529L	558	2.1
R534G	711	2.2
R534I	789	2.3
R534L	763	2.3
R534M	776	2.2
G576H	436	2.3
Q588L	817	2.1
I591K	816	2.0
I591R	832	2.0
Y605W	524	1.9
F607N	561	2.0
F607W	624	2.0
A610C	799	2.4
N613I	555	2.1
N613M	587	2.1
N613T	526	2.0
N613V	578	2.0
K625A	638	2.3
K625M	623	2.2
A510E	622	2.0
S631T	532	2.1
T635H	539	2.1
T635W	528	2.1
I636H	521	2.1

Y848E	843	2.0
D947G	408	2.1
F951Y	325	2.0
E849M	610	2.4
Q1007A	394	1.9
D1003G	486	2.1
A1022M	303	1.8
D1028L	416	1.9
D1028Q	537	2.1
A1057H	624	2.3
N1096A	562	2.2
Y1104M	611	2.2
N1122K	614	2.3
E1132A	589	2.2
E1132H	611	2.2
E1132K	610	2.2
E1132R	622	2.1
V1135K	612	2.3
V1188E	641	2.2
L1212N	630	2.0
E1250R	606	2.0
T1381E	612	2.1
T1431M	625	2.0
A1442R	609	2.0
E1450F	611	2.0
E1450W	618	2.0
Dead[d]		
Blank[e]		
Blank[e]		
Plate 2[a]		
	Alpha-1,3-Glucan	
GTF	DP_w	PDI
6855[b]	622	2.1
6855[b]	628	2.1
6855[b]	634	2.1
6855[b]	619	2.2
I1453M[c]	635	2.1
V1491F	604	2.0
P1499Y	587	2.0
Y219C	591	2.0
E243H	631	2.1
A377I	514	2.1

I411F	586	2.3
I411S	591	2.1
D425Q	681	2.1
L428V	577	2.0
M529N	560	2.1
N531G	977	2.2
G576R	416	1.8
Y580H	554	2.1
K593M	792	1.9
I608Y	708	1.9
N613G	644	2.0
N613L	618	2.0
D617E	419	2.1
E621T	603	2.1
I627W	506	2.2
S631D	521	2.0
S631E	545	2.0
S631R	521	2.1
G633W	493	2.2
F634A	523	2.1
T635E	561	2.2
T635I	648	2.1
T635Y	518	2.1
R722H	793	2.5
T728S	769	2.4
M732L	791	2.4
A777N	755	2.4
A510E	625	2.2
N904E	554	2.5
K930G	637	2.5
K930V	582	2.3
D947F	619	2.6
D947I	610	2.4
D947K	559	2.5
D947N	635	2.6
D947Q	635	2.9
D947S	603	2.7
D947V	621	2.5
D947Y	624	2.6
Q1007S	578	2.6
D1003N	570	2.6
I1026H	621	2.6
D1028A	568	2.1

D1028M	535	2.1
V1037A	591	2.2
K1041A	583	2.1
K1041M	648	2.3
D1080M	554	2.2
F1244P	589	2.2
F1244Q	534	1.9
E1250H	553	2.1
E1250K	591	2.2
T1431Q	663	2.2
E1450D	585	2.2
G1484P	627	2.2
I1453G	881	2.1
W1437N	654	2.3
R722N	766	2.2
Dead[d]		
Blank[e]		
Blank[e]		

[a] Glucan synthesis reactions were run in microtiter plate format (two plates).
[b] GTF 6855, SEQ ID NO:4. Reactions with this GTF were run in quadruplicate per plate.
[c] Each listed GTF with a substitution is a version of GTF 6855 comprising a substitution at a respective position, where the position number is in correspondence with the residue numbering of SEQ ID NO:62. The wild type residue is listed first (before residue position number) and the substituting residue is listed second (after the residue position number) (this "wild type residue-position number-variant residue" annotation format applies throughout the present disclosure).
[d] GTF with destroyed activity was entered into the reaction. Alpha-1,3-glucan was not detected.
[e] No GTF was added to the reaction. Alpha-1,3-glucan was not detected.

Based on the data in Table 3, individual substitutions at amino acid positions Y848, I591, Q588, A610, R534, N531, I1453, K593, M732, T728, R722, A777, and I608, for example, in GTF 6855 (SEQ ID NO:4) each result in an enzyme that produces insoluble alpha-1,3-glucan with a DP_w at least 10% greater than the average DP_w of insoluble alpha-1,3-glucan produced by the parent non-substituted enzyme (GTF 6855, SEQ ID NO:4). Interestingly, these higher DP_w alpha-1,3-glucan products generally had a PDI that was the same as, or similar to, the average PDI (2.1-2.2) of alpha-1,3-glucan

produced by the parent non-substituted enzyme (GTF 6855, SEQ ID NO:4), meaning that the increase in molecular weight likely did not compromise polymer uniformity.

In a similar SEL study (not all data shown), individual substitutions in GTF 6855 (SEQ ID NO:4) at the following amino acid positions appeared to also have an enhancing effect on insoluble alpha-1,3-glucan molecular weight (substituting residue, resulting DP_w): N614 (P, 1054), R609 (H, 1052), F424 (A, 1051; V, 1023; L, 942; E, 937; Q, 725), W511 (Y, 1020), N475 (Q, 975; S, 954), A610 (T, 972; C, 959), Y848 (E, 922), N1214 (L, 918; I, 863), N531 (L, 879; M, 724; G, 714), I591 (K, 840; R, 824), R534 (K, 759), Q588 (L, 734), and R722 (H, 701) (non-substituted GTF 6855 produced insoluble alpha-1,3-glucan with DP_w 487).

One or more substitutions at any of the foregoing sites in this example are expected to allow for production of insoluble alpha-1,3-glucan with a DP_w significantly higher than the DP_w of alpha-1,3-glucan produced by a parent non-substituted glucosyltransferase.

EXAMPLE 2

Generating Glucosyltransferase Variants that Produce Alpha-Glucan Products of Higher Molecular Weight

This Example describes another screening for glucosyltransferase variants that produce alpha-glucan with increased molecular weight.

Saturation mutagenesis was performed on GTF 6855 (SEQ ID NO:4) to provide a multitude of single amino acid-substituted variants of this glucosyltransferase. Each variant was entered into a glucan synthesis reaction with parameters that were the same as, or similar to, the following: vessel, 250-mL bottom-indented shake flask agitated at 120 rpm; initial pH, 5.7; reaction volume, 50 mL; sucrose, 75 g/L; GTF, 1.5 mL lysate of *E. coli* cells heterologously expressing enzyme; KH_2PO_4, 5 mM; temperature, 30 °C; time, about 20 hours. The molecular weight (DP_w) of alpha-1,3-glucan produced in each reaction (as measured via SEC methodology similar to that disclosed in Example 1) is provided in Table 4.

Table 4. Molecular Weight of Insoluble Alpha-1,3-Glucan Produced by GTF 6855 (SEQ ID NO:4) and Single Amino Acid-Substituted Variants thereof

GTF	Alpha-1,3-Glucan DP_w
6855[a]	558

E567Q	1001
I591V	859
L661P	842
N743D	700
N743S	937
N743T	874
R741A	831
R741P	871
R741Q	886
R741S	887
R741T	693
T563A	910
V586T	874

[a] GTF 6855, SEQ ID NO:4. The reaction with this GTF was performed separately under the same conditions as described above.

Based on the data in Table 4, it is apparent, for example, that each listed single amino acid substitution in GTF 6855 (SEQ ID NO:4) results in an enzyme that produces insoluble alpha-1,3-glucan with increased molecular weight (DP_w). Thus, one or more substitutions at the positions indicated in Table 4 are expected to allow for production of insoluble alpha-1,3-glucan with a DP_w significantly higher than the DP_w of alpha-1,3-glucan produced by a parent non-substituted glucosyltransferase.

EXAMPLE 3

Analysis of the Effects of Amino Acid Substitution Combinations on Glucosyltransferase Alpha-Glucan Synthesis Activity

This Example describes the effects of introducing multiple amino acid substitutions to a glucosyltransferase and determining their effect on its alpha-glucan synthesis function. This analysis indicates, for example, that amino acid substitutions identified above to enhance alpha-glucan product molecular weight can be included in substitution combinations that likewise impart this molecular weight enhancement.

Briefly, certain combinations of amino acid substitutions were made to SEQ ID NO:4 (GTF 6855, see Table 1 and Example 1 for description of this glucosyltransferase) by site-directed mutagenesis of appropriate DNA templates contained in a plasmid. The plasmid sequences encoding each modified glucosyltransferase were individually sequenced to confirm the intended codon changes. Each combination of substitutions is listed in Table 5 below; appropriate analyses indicated that each modified enzyme was

able to perform a glucosyltransferase reaction producing alpha-1,3-glucan (data not shown).

Expression plasmids encoding the modified glucosyltransferases were individually used to transform a *B. subtilis* strain containing nine protease deletions (*amyE*::*xylRPxylAcomK-ermC*, *degUHy32*, *oppA*, Δ*spoIIE3501*, Δ*aprE*, Δ*nprE*, Δ*epr*, Δ*ispA*, Δ*bpr*, Δ*vpr*, Δ*wprA*, Δ*mpr-ybfJ*, Δ*nprB*). Transformed cells were spread onto LB plates supplemented with 5 μg/mL chloramphenicol. Colonies growing on these plates were streaked several times onto LB plates with 25 μg/mL chloramphenicol. Each resulting *Bacillus* strain for expressing a particular variant glucosyltransferase was then grown for 6-8 hours in LB medium containing 25 μg/mL chloramphenicol, and then subcultured into Grants II medium at 30 °C for 2-3 days. The cultures were spun at 15000 g for 30 minutes at 4 °C, and the supernatants were filtered through 0.22-μm filters. The filtered supernatants, each of which contained an expressed secreted variant glucosyltransferase, were aliquoted and frozen at -80 °C, and later used (below) for analyzing alpha-1,3-glucan synthesis activity.

The same amount of each variant enzyme, activity-wise, was entered into a glucan synthesis reaction with parameters that were the same as, or similar to, the following: vessel, 500-mL jacketed reactor with Teflon®-pitched blade turbine (45-degree angle) on a glass stir rod and agitated at 50-200 rpm; initial pH, 5.5; reaction volume, 500 mL; sucrose, 108 g/L; KH_2PO_4, 1 mM; temperature, 39 °C; time, about 18-24 hours; filtrate from a previous alpha-1,3-glucan synthesis reaction, 50 vol%. The molecular weight (DP_w) of alpha-1,3-glucan produced in each reaction (as measured via methodology similar to that disclosed in Example 1) is provided in Table 5.

Table 5. Alpha-1,3 Glucan Products of GTF 6855 (SEQ ID NO:4) Variants with Multiple Amino Acid Substitutions

GTF[a]												Alpha-1,3-Glucan[b] DP_w
A510D	Q588L	F607Y	R741S	D948G			R722H	T877K		M1253I	K1277N	844
A510D	Q588L	F607Y	R741S	D948G			R722H	T877K	V1188E	M1253I	Q957P	744
A510D	Q588L	F607Y	R741S	D948G				T877K	V1188E	M1253I	Q957P	812
A510D	Q588L	F607Y	R741S	D948G						M1253I		750
A510D	Q588L	F607W	R741S	D948G								752
	Q588L	F607Y	R741S	D948G								705
A510D	Q588L	F607Y	R741S	D948G	N628D	T635A		T877K		M1253I	F929L	R1172C 855
A510D	Q588L	F607W	R741S	D948G	S631T	S710G	R722H	T877K	V1188E	M1253I		863
A510D	Q588L	F607W	R741S	D948G	S631T	S710G	R722H	T877K	V1188E			812
A510D	Q588L	F607W	R741S	D948G	S631T	S710G		T877K	V1188E	M1253I		727
A510D	Q588L	F607Y	R741S	D948G								697
A510D	Q588L	F607Y	R741S	D948G					V1188E			703
A510D	Q588L	F607W	R741S	D948G	S631T	S710G			V1188E			676
A510D	Q588L	F607W	R741S	D948G		S710G	R722H	T877K		M1253I		963
A510D	Q588L	F607Y	R741S	D948G	S631T		R722H	T877K	V1188E	M1253I		906
A510D	Q588L	F607W	R741S	D948G	S631T			T877K	V1188E	M1253I		781
A510D	Q588L	F607W	R741S	D948G	S631T				V1188E			687
A510D	Q588L	F607Y	R741S	D948G	S631T		R722H	T877K	V1188E	M1253I		961
A510D	Q588L	F607W	R741S	D948G					V1188E	M1253I		781

[a] Each listed GTF is a version of GTF 6855 (SEQ ID NO:4) comprising substitutions at respective positions, where each position number is in correspondence with the residue numbering of SEQ ID NO:62.

[b] Insoluble alpha-1,3-glucan product.

Based on the data in Table 5, it is apparent that introduction of multiple amino acid substitutions to GTF 6855 (SEQ ID NO:4), including substitutions that enhance molecular weight, can be employed in efforts to produce higher molecular weight insoluble alpha-1,3-glucan; for example, compare these DP_w values to those of GTF 6855 (SEQ ID NO:4) without substitutions shown in Table 3.

It is apparent, for example, that a glucosyltransferase with multiple substitutions, including those at positions corresponding to positions Gln-588, Arg-741, and/or Arg-722 of SEQ ID NO:62, can increase the molecular weight of insoluble alpha-glucan produced by the glucosyltransferase.

CLAIMS

What is claimed is:

1. A non-native glucosyltransferase comprising at least one amino acid substitution at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62,

wherein the non-native glucosyltransferase synthesizes insoluble alpha-glucan comprising 1,3-linkages, and the molecular weight of said insoluble alpha-glucan is higher than the molecular weight of insoluble alpha-glucan synthesized by a second glucosyltransferase that only differs from the non-native glucosyltransferase at the substitution position(s).

2. The non-native glucosyltransferase of claim 1, wherein:

(i) the amino acid substitution at the position corresponding with amino acid residue Asn-531 is with a Gly, Leu, or Met residue;

(ii) the amino acid substitution at the position corresponding with amino acid residue Arg-534 is with a Lys, Gly, Ile, Leu, or Met residue;

(iii) the amino acid substitution at the position corresponding with amino acid residue Thr-563 is with an Ala residue;

(iv) the amino acid substitution at the position corresponding with amino acid residue Glu-567 is with a Gln residue;

(v) the amino acid substitution at the position corresponding with amino acid residue Val-586 is with a Thr residue;

(vi) the amino acid substitution at the position corresponding with amino acid residue Gln-588 is with a Leu residue;

(vii) the amino acid substitution at the position corresponding with amino acid residue Ile-591 is with a Val, Lys, or Arg residue;

(viii) the amino acid substitution at the position corresponding with amino acid residue Lys-593 is with a Met residue;

(ix) the amino acid substitution at the position corresponding with amino acid residue Ile-608 is with a Tyr residue;

(x) the amino acid substitution at the position corresponding with amino acid residue Ala-610 is with a Cys or Thr residue;

(xi) the amino acid substitution at the position corresponding with amino acid residue Leu-661 is with a Pro residue;

(xii) the amino acid substitution at the position corresponding with amino acid residue Arg-722 is with a His or Asn residue;

(xiii) the amino acid substitution at the position corresponding with amino acid residue Thr-728 is with a Ser residue;

(xiv) the amino acid substitution at the position corresponding with amino acid residue Met-732 is with a Leu residue;

(xv) the amino acid substitution at the position corresponding with amino acid residue Arg-741 is with a Ser, Ala, Pro, Gln, or Thr residue;

(xvi) the amino acid substitution at the position corresponding with amino acid residue Asn-743 is with a Ser, Thr, or Asp residue;

(xvii) the amino acid substitution at the position corresponding with amino acid residue Ala-777 is with an Asn residue;

(xviii) the amino acid substitution at the position corresponding with amino acid residue Tyr-848 is with a Glu residue; and/or

(xix) the amino acid substitution at the position corresponding with amino acid residue Ile-1453 is with a Gly or Met residue.

3. The non-native glucosyltransferase of claim 1, comprising at least one amino acid substitution at a position corresponding with amino acid residue Gln-588, Arg-741, or Arg-722 of SEQ ID NO:62; optionally wherein:

(i) the amino acid substitution at the position corresponding with amino acid residue Gln-588 is with a Leu residue;

(ii) the amino acid substitution at the position corresponding with amino acid residue Arg-741 is with a Ser residue; and/or

(iii) the amino acid substitution at the position corresponding with amino acid residue Arg-722 is with a His residue.

4. The non-native glucosyltransferase of claim 3, comprising two or more amino acid substitutions at positions corresponding with amino acid residues Gln-588, Arg-741, or Arg-722 of SEQ ID NO:62.

5. The non-native glucosyltransferase of claim 1, wherein the insoluble alpha-glucan comprises at least about 50% alpha-1,3 linkages, and optionally wherein the insoluble alpha-glucan has a weight average degree of polymerization (DP_w) of at least 100.

6. The non-native glucosyltransferase of claim 5, wherein the insoluble alpha-glucan has a DP_w of at least 650.

7. The non-native glucosyltransferase of claim 6, comprising a catalytic domain that is at least about 90% identical to residues 55-960 of SEQ ID NO:4, residues 54-957 of SEQ ID NO:65, residues 55-960 of SEQ ID NO:30, residues 55-960 of SEQ ID NO:28, or residues 55-960 of SEQ ID NO:20.

8. The non-native glucosyltransferase of claim 7, comprising an amino acid sequence that is at least about 90% identical to SEQ ID NO:4, SEQ ID NO:65, SEQ ID NO:30, SEQ ID NO:28, or SEQ ID NO:20.

9. The non-native glucosyltransferase of claim 7, wherein the insoluble alpha-glucan comprises at least about 90% alpha-1,3 linkages.

10. The non-native glucosyltransferase of claim 1, wherein the molecular weight of said insoluble alpha-glucan is at least about 10% higher than the molecular weight of insoluble alpha-glucan synthesized by said second glucosyltransferase.

11. A polynucleotide comprising a nucleotide sequence encoding a non-native glucosyltransferase according to claim 1, optionally wherein one or more regulatory sequences are operably linked to the nucleotide sequence, and preferably wherein said one or more regulatory sequences include a promoter sequence.

12. A reaction composition comprising water, sucrose, and a non-native glucosyltransferase according to claim 1.

13. A method of producing insoluble alpha-glucan comprising:
(a) contacting at least water, sucrose, and a non-native glucosyltransferase enzyme according to claim 1, whereby insoluble alpha-glucan is produced; and
(b) optionally, isolating the insoluble alpha-glucan produced in step (a).

14. A method of preparing a polynucleotide sequence encoding a non-native glucosyltransferase, said method comprising:
(a) identifying a polynucleotide sequence encoding a parent glucosyltransferase that (i) comprises an amino acid sequence that is at least about 40% identical to SEQ ID NO:4 or positions 55-960 of SEQ ID NO:4, and (ii) synthesizes insoluble alpha-glucan comprising 1,3-linkages; and
(b) modifying the polynucleotide sequence identified in step (a) to substitute at least one amino acid of the parent glucosyltransferase at a position corresponding with amino acid residue Asn-531, Arg-534, Thr-563, Glu-567, Val-586, Gln-588, Ile-591, Lys-593, Ile-608, Ala-610, Leu-661, Arg-722, Thr-728, Met-732, Arg-741, Asn-743, Ala-777, Tyr-848, or Ile-1453 of SEQ ID NO:62,
thereby providing a polynucleotide sequence encoding a non-native glucosyltransferase that synthesizes insoluble alpha-glucan with a molecular weight that is higher than the molecular weight of insoluble alpha-glucan synthesized by the parent glucosyltransferase.

15. The method of claim 14, wherein said identifying step is performed:
(a) *in silico*,
(b) with a method comprising a nucleic acid hybridization step,
(c) with a method comprising a protein sequencing step, and/or
(d) with a method comprising a protein binding step;

and/or wherein said modifying step is performed:

(e) *in silico*, followed by synthesis of the polynucleotide sequence encoding the non-native glucosyltransferase enzyme, or

(f) using a physical copy of the polynucleotide sequence encoding the parent glucosyltransferase.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/050345

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. [X] forming part of the international application as filed:

      [X] in the form of an Annex C/ST.25 text file.

      [ ] on paper or in the form of an image file.

   b. [ ] furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. [ ] furnished subsequent to the international filing date for the purposes of international search only:

      [ ] in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      [ ] on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. [ ] In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2018/050345

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-15(partially)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/050345

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N9/10 C12P19/04
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 C12P

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, BIOSIS, EMBL, FSTA, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	DATABASE UniProt [Online] 15 February 2017 (2017-02-15), XP002785922, Database accession no. A0A1F0BMZ6 the whole document -----	1-15
Y	WO 2013/036918 A2 (DU PONT [US]; O'BRIEN JOHN P [US]; PAYNE MARK S [US]) 14 March 2013 (2013-03-14) claims 1-17; sequence 3 page 16, lines 14-14 page 4, lines 18-24 -----	1-15
Y	US 2017/166938 A1 (NAGY KEVIN D [US] ET AL) 15 June 2017 (2017-06-15) claim 1; sequence 6 -----	1-15

☐ Further documents are listed in the continuation of Box C.

☒ 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
24 October 2018	14/01/2019

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	Obel, Nicolai

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2018/050345

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2013036918 A2	14-03-2013	US 2013244287 A1	19-09-2013
		WO 2013036918 A2	14-03-2013
US 2017166938 A1	15-06-2017	AU 2016371897 A1	07-06-2018
		CA 3006987 A1	22-06-2017
		CN 108463557 A	28-08-2018
		EP 3390651 A1	24-10-2018
		US 2017166938 A1	15-06-2017
		WO 2017106262 A1	22-06-2017

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-15(partially)

Directed to a Ans531 mutant

2-19. claims: 1-15(partially)

Each further mutant listed in claim 1 is considered an independent invention.
